Army Science Conference Proceedings

22-25 June 1992

Volume III

Principal Authors N through Z

This document has been approved for public release and sale; its distribution is unlimited.

Assistant Secretary of the Army
(Research, Development and Acquisition)
Department of the Army

C O R R E C T I O N S

DUE TO PRINTING ERRORS THE FOLLOWING CHANGES ARE NECESSARY:

Volume I	**Page ii**	**Refer to Index in Volume II or III, which are printed correctly.**
Volume III	**Pages 3,4**	**Remove these pages and insert attached.**

DEPARTMENT OF THE ARMY
OFFICE OF THE ASSISTANT SECRETARY
WASHINGTON, DC 20310-0103

SARD-TR

30 JUL 1992

MEMORANDUM FOR SEE DISTRIBUTION

SUBJECT: Proceedings of the 1992 Army Science Conference

The 18th Army Science Conference was held at the Hyatt Hotel, Orlando, Florida, 22-25 June 1992. The conference presented a cross section of the many significant scientific and engineering programs carried out by the Department of the Army (DA). Additionally, it provided an opportunity for DA civilian and military scientists and engineers to present the results of their research and development efforts before a distinguished and critical audience.

These Proceedings of the 1992 Army Science Conference are a compilation of all papers presented at the conference and the supplemental papers that were submitted.

Our purpose for soliciting these papers was to:

a. Stimulate the involvement of scientific and engineering talent within the Department of the Army;

b. Demonstrate Army competence in research and development;

c. Provide a forum wherein Army personnel can demonstrate the full scope and depth of their current projects; and

d. Promote the interchange of ideas among members of the Army scientific and engineering community as well as management and technical personnel from industry and other government agencies.

The information contained in these volumes will be of benefit to those who attended the conference and to others interested in Army research and development. It is requested that these Proceedings be placed in technical libraries where they will be available for reference.

George T. Singley III
Deputy Assistant Secretary
For Research and Technology

DISTRIBUTION:
Office of the Under Secretary of Defense for Research and Advanced Technology, Wash, DC 20310-3080
Office of the Assistant Secretary of the Army (RD&A), ATTN: SARD-TR, Wash, DC 20310-0103
HQDA, Assistant Chief of Staff for Intelligence, ATTN: DAMI-ZA, Wash, DC 20310
HQDA, Deputy Chief of Staff for Logistics, ATTN: DALO-ZX, Wash, DC 20310-0500
HQDA, Deputy Chief of Staff for Operations & Plans, ATTN: DAMO-FD, Wash, DC 20310-0400
HQDA, Director of Information Systems for C4, ATTN: SAIS-ZA, Wash, DC 20310-0400
HQDA, Deputy Chief of Staff for Personnel, ATTN: DAPE-ZA, Wash, DC 20310-0300
Office of the Chief of Engineers, ATTN: CERD-M, 20 Massachusetts Avenue, N.W., Wash, DC 20314-1000
Office of the Surgeon General, ATTN: DASG-ZA, 5109 Leesburg Pike, Falls Church, VA 22041-3258
COMMANDERS/DIRECTORS:
USA Concepts Analysis Agency, 8120 Woodmont Avenue, Bethesda, MD 20814-2797
USA Operational Test & Evaluation Agency, 5600 Columbia Pike, Falls Church, VA 22041-5115
USA Information Systems Engineering Command, Ft. Huachuca, AZ 85613-5000
USA Strategic Defense Command, P. O. Box 15280, Arlington, VA 22215-0150
USA Materiel Command, ATTN: AMCLD, 5001 Eisenhower Avenue, Alexandria, VA 22333-0001
USA Materiel Systems Analysis Activity, ATTN: AMXSY-D, APG, MD 21005-5071
USA Armament, Munitions & Chemical Command, ATTN: AMSMC-CG, Rock Island, IL 61299-6000
USA Armament RD&E Center, ATTN: SMCAR-TD, Picatinny Arsenal, NJ 07806-5000
Fire Support Armament Cen, ATTN: SMCAR-FS, PA, NJ 07806-5000
Armament Engineering Dir, ATTN: SMCAR-AE, PA, NJ 07806-5000
Close Combat Armament Cen, ATTN: SMCAR-CC, PA, NJ 07806-5000
Chemical RD&E Center, ATTN: SMCCR-TD, APG, MD 21010-5423
USA Aviation Systems CMD, ATTN: AMSAV-GTD, 4300 Goodfellow Blvd, St. Louis, MO 63120-1798
USARTA, Ames Research Center, ATTN: SAVRT-D, Moffett Field, CA 94035-1099
USARTA, Aeroflightdynamics Dir, ATTN: SAVRT-AF, Moffett Field, CA 94035
USARTA, Aviation Applied Technology Dir, ATTN: SAVRT-TY, Ft Eustis, VA 23604
USARTA, Propulsion Dir, ATTN: SAVRT-PN, Lewis Research Center, 21000 Brook Park Rd, Cleveland, OH 44135
USARTA, Aerostructures Dir, ATTN: SAVRT-SD, Hampton, VA 22665
Avionics R&D Activity, ATTN: SAVAA, Ft Monmouth, NJ 07703-5000

Structures of these compounds can be found in figure 2. The Bachmann HMX Process is a "one pot" nitrolysis of hexamine in the presence of nitric acid, acetic acid, ammonium nitrate, and a large excess of acetic anhydride. This process has been operated at the Holston Army Ammunition Plant in Kingsport, TN by the Holston Defense Corporation, a subsidiary of Eastman Kodak Corporation.

The GARDEC Process uses a reaction sequence which proceeds stepwise through the formation of DAPT, thence to the acetylation of DAPT to form TAT, and subsequently to the nitration of TAT to HMX.

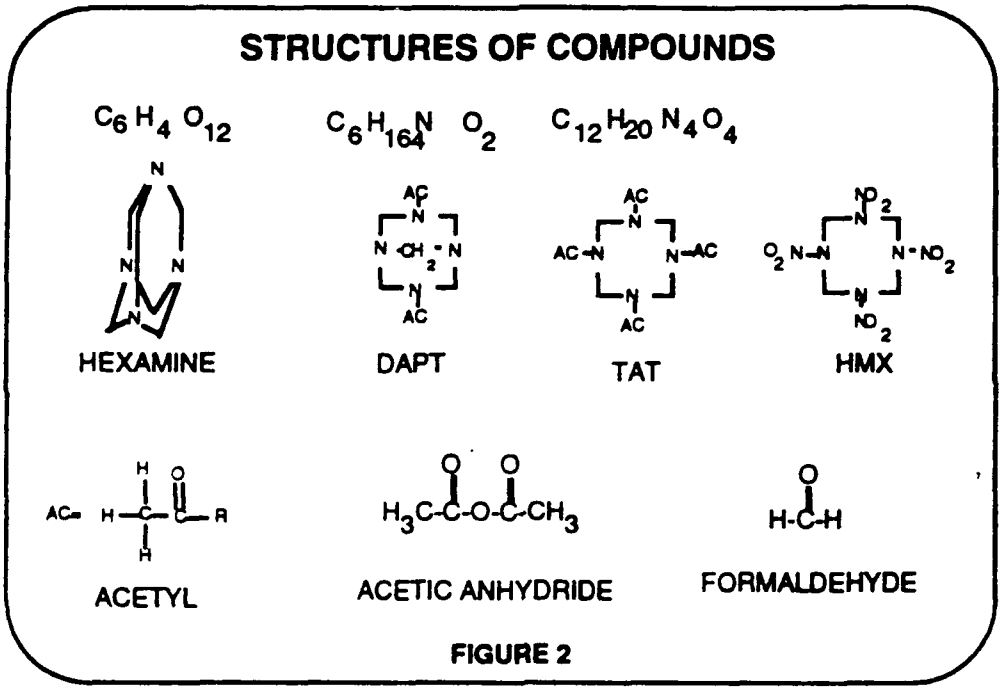


FIGURE 2

III. GARDEC DAPT Process

Prior state of the art:

Although DAPT has been successfully prepared in the past by many others, the GARDEC DAPT Process has several advantages over previous DAPT Processes.

The reaction of hexamine with acetic anhydride in the presence of water as a catalyst yields DAPT in a water/acetic acid medium (equation 1).

(1) $C_6H_{12}N_4 + 2\,(CH_3CO_2)O + H_2O \text{-----} > C_9H_{16}N_4O_2 + 2\,CH_3COOH + HCOH$

Actually, the theoretical yield of DAPT based on hexamine is 120%, since excess ammonia and acetic anhydride are added to the overall reaction as shown in equation 2.

(2) $C_6H_{12}N_4 + 2.4\,(CH_3CO_2)O + 0.8\,NH_3 \xrightarrow{H_2O} 1.2\,C_9H_{16}N_4O_2 + 2.4\,CH_3COOH$

In particular, the excess ammonia reacts with formaldehyde that is produced to form hexamine as shown in equation 3.

(3) $6\,HCOH + 4\,NH_3 \text{-----} > C_6H_{12}N_4 + 6\,H_2O$

Moreover, excess AC_2O is added to react with the freshly formed hexamine to form additional DAPT. Thus, the overall reaction gives a 120 % yield of DAPT based on hexamine. This in-situ reaction has been termed resynthesis.

Historically, hexamine reacts very rapidly and exothermically with acetic anhydride if a catalytic amount of water is present because water reacts more quickly as a catalyst to form DAPT than as a reagent with acetic anhydride to form acetic acid. Since the reaction is extremely exothermic, constant cooling of the reaction is necessary. Under prior synthetic methods the removal of heat has been the rate limiting step, thus necessitating several large recycle loops through heat exchangers. The need for rapid mixing and circulation, and several recycle loops thru the heat exchangers give rise to a large reactor residence time. This causes a degradation of the DAPT as it is not stable in the resulting acetic acid solution. Moreover, the side products, acetic acid and the water which had been added for catalysis, cannot be readily stripped from the DAPT. Therefore, in the past, DAPT could not be easily isolated in high yield or purity. Since it is important to minimize reactor residence time, it is necessary to increase the reaction rate. The GARDEC DAPT process provides a simple solution.

The new DAPT process:[1,2]

An extremely important energy barrier was breached when a "hot DAPT process" was developed. The exothermic reaction is controlled by injecting the

NASA HQS, Suite 4237, 400 Maryland Ave, SW, Wash, DC 20546
National Science Foundation, 1800 G Street, NW, Wash, DC 20550
United Nations Library, ATTN: Acquisition Section, Room L-138A, New York, NY 10017

US NAVY

Naval Air Systems Command, Code AIR-03-D, 1411 Jefferson Davis Hwy, Arlington, VA 22202-3000
Naval Rsch Lab, ATTN: Dir of Rsch, Washington, DC 20375
Office of Naval Rsch, Code 10, 800 North Quincy St, Arlington, VA 22217-5000
HQ US Marine Corps, Code RD-1, Washington, DC 29380-0001

US AIR FORCE

Air Force Systems Command, ATTN: Technical Director, Andrews AFB, Washington, DC 20334-5000
Air Force Office of Scientific Research/XOT, ATTN: Distribution, Bolling AFB, Washington, DC 20332-6448

PROCEEDINGS

OF THE

1992 ARMY SCIENCE CONFERENCE

ORLANDO, FLORIDA

22 - 25 JUNE 1992

VOLUME III

PRINCIPAL AUTHORS N THROUGH Z

TABLE OF CONTENTS

PROCEEDINGS OF THE 1992 ARMY SCIENCE CONFERENCE

Quantitative Synthesis of HMX via the GARDEC HMX Process

Mr. Steven Nicolich*
U.S. Army Armament Research, Development and Engineering Center
Picatinny Arsenal, NJ 07806-5000

Mr. William Lukasavage
Geo-Centers Inc.
Picatinny Arsenal, NJ 07806-5000

Dr. Jack Alster
U.S. Army Armament Research, Development and Engineering Center
Picatinny Arsenal, NJ 07806-5000

I. Introduction

HMX is currently the most powerful mass produced conventional explosive. The only large scale production process used in the U.S. to make HMX is the Improved Bachmann (Holston Batch) Process. As a variation of the original Bachmann Process, it is based on the nitrolysis of hexamine. It will be simply referred to as the Bachmann HMX Process. Although it has been in existence for many years, the current procedure for hexamine nitrolysis has problems with yield, purity, and reactant excesses.

A new process for synthesizing HMX, which has been recently developed by Geo-Centers Corp. in cooperation with the Armament, Research, Development, and Engineering Center (ARDEC) at Picatinny Arsenal, proceeds via an inert intermediate precursor known as TAT, 1,3,5,7-tetraacetyl-1,3,5,7-tetraazacyclooctane. For many years, researchers have been unable to synthesize TAT in good yield and had to use large excesses of costly reactants. Thus, large scale preparation of HMX from TAT was not feasible. Since then, several important fundamental discoveries have been made in the laboratory that resulted in essentially 100% yield of TAT via a very efficient reaction process.

Furthermore, the synthesis of DAPT, 3,7-diacetyl-1,5-endomethylene-1,3,5,7-tetraazacyclooctane, an important pre-TAT intermediate, has been simplified in several ways. Finally, the nitration of TAT to HMX has been dramatically improved. The work accomplished at Picatinny Arsenal encompasses the synthesis of HMX from basic starting materials. This newly developed HMX process is known as the GARDEC HMX Process.

In summary, the synthesis of DAPT, TAT, and HMX will be discussed highlighting important attributes that make the GARDEC HMX Process more appealing from the standpoints of low cost, efficiency, and friendliness toward the environment.

II. Overview of Bachmann and GARDEC HMX Methodologies

An overview of the two HMX processes, the Bachmann Process and the GARDEC Process can be seen in figure 1.

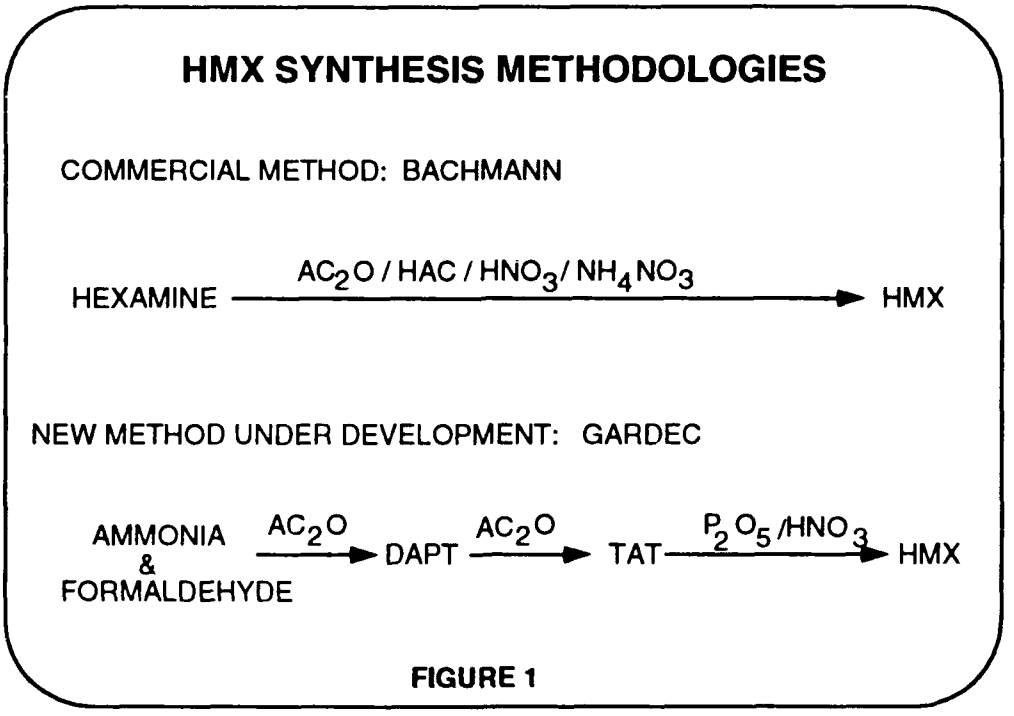


FIGURE 1

USA Communications-Electronics Cmd, ATTN: AMSEL-TDD, Ft Monmouth, NJ 07703-5001
Center for Command, Control & Communications System, ATTN: AMSEL-RD-C3, Ft Monmouth, NJ 07703-5202
Center for Night Vision & Electro-Optics, ATTN: AMSEL-RD-NV-D, Ft Belvoir, VA 22060-5677
Electronics Warfare/RSTA Center, ATTN: AMSEL-RD-EW-D, Ft Monmouth, NJ 07703-5303
Signals Warfare Lab, ATTN: AMSEL-RD-SW-D, Vint Hill Farms Station, Warrenton, VA 22186-5100
USA Laboratory Command, ATTN: AMSLC-TD, 2800 Powder Mill Road, Adelphi, MD 29783-1145
Atmospheric Sciences Lab, ATTN: SLCAS-DD, WSMR, NM 88002-5501
Ballistic Research Lab, ATTN: SLCBR-OD, APG, MD 21005
Electronics Technology & Devices Lab, ATTN: SLCET-D, Ft Monmouth, NJ 07703-5000
Electronic Warfare Vulnerability Assessment Lab, ATTN: SLCEW-M-D, WSMR, NM 8802-5513
Harry Diamond Labs, ATTN: SLCHD-D, Adelphi, MD 20783-1197
Human Engineering Lab, ATTN: SLCHE-H, APG, MD 21005-5001
Materials Technology Lab, ATTN: SLCMT-D, Watertown, MA 02172-0001
Army Research Office, ATTN: SLCRO-D, RTP, NC 27709-2211
Army Research, Development & Standardization Group (Europe) Box 65, FPO New York 09510
USA Missile Command, ATTN: AMSMI-R, Redstone Arsenal, AL 35898-5240
USA Tank-Automotive Cmd, ATTN: AMSTA-CG, Warren, MI 48397-5000
  Aberdeen Proving Ground, ATTN: STEAP-OC, APG, MD 21005-5001
  Dugway Proving Ground, ATTN: STEDP-CO, Dugway, UT 84022-5000
  Electronic Proving Ground, ATTN: STEEP-DO, Ft Huachuca, AZ 85613-7110
  Aviation Development Test Activity, ATTN: STEBG-CO, Ft Rucker, AL 36362-5276
  Combat Systems Test Act, ATTN: STECS-CO, APG, MD 21005-5059
  White Sands Missile Range, ATTN: STEWS-CG, WSMR, NM 88002-5000
USA Aviation Systems Cmd, ATTN: AMSAV-G, 4300 Goodfellow Blvd, St. Louis, MO 63120-1798
Belvoir RD&E Center, ATTN: STRBE-Z, Ft. Belvoir, VA 21005-5606
Natick RD&E Center, ATTN: STRNC-Z, Natick, MA 01760-5000
US Army Corps of Engineers
  Cold Regions Rsch & Eng Lab, ATTN: CRREL-TD, P. O. Box 282, Hanover, NH 03755-1290
  Construction Eng Rsch Lab, ATTN: CERL-Z, P. O. Box 4005, Champaign, IL 61820-135
  Engineer Topographic Labs, ATTN: ETL-TD, Ft Belvoir, VA 22060-5546
  Waterways Experiment Station, ATTN: CEWES-IM-MI, P. O. Box 631, Vicksburg, MS 39180-0631
USA Medical R&D Command, ATTN: SGRD-ZA, Ft Detrick, Frederick, MD 21701-5012
Aeromedical Rsch Lab, ATTN: SGRD-UAC, Ft Rucker, AL 36362-5292
Inst of Dental Rsch, WRAMC, ATTN:SGRD-UDZ, Wash, DC 20307-5300
Inst of Surgical Rsch, ATTN: SGRD-USZ, Ft Sam Houston, TX 78234-6200

Letterman Army Inst of Rsch, ATTN: SGRD-ULZ, Presidio of San Francisco, CA 94129-6800
Biomedical R&D Lab, ATTN: SGRD-UBZ, Frederick, MD 20701-5010
Medical Rsch Inst of Chemical Defense, ATTN: SGRD-UV-ZA, APG, MD 21010-5425
Medical Rsch Inst of Environmental Medicine, ATTN: SGRD-UEZ, Natick, MA 01760-5007
Medical Rsch Inst of Infectious Diseases, ATTN: SGRD-UIZ-A, Ft Detrick, Frederick, MD 21701-5011
Walter Reed Army Inst of Rsch, ATTN: SGRD-UWZ, Washington, DC 20307-5100
USA Health Services Command, Ft. Sam Houston, TX 78234-61000
USA Environmental Hygiene Agency, ATTN: HSHB-Z, APG, MD 21010-5422
USA Research Institute for the Behavioral and Social Sciences, ATTN: PERI-ZT, 5001 Eisenhower Ave, Alexandria, VA 22333-5600
ARI Field Unit, ATTN: PERI-IJ, POB 2086, Ft Benning, GA 31905
ARI Field Unit, ATTN: PERI-SB, POB 6057, Ft Bliss, TX 79906
ARI Field Unit, ATTN: PERI-SH, Ft Hood, TX 76544-5065
ARI Field Unit, ATTN: PERI-IK, Steele Hall, Ft Knox, KY 40121
ARI Field Unit, ATTN: PERI-SL, POB 290, Ft Leavenworth, KS 66027
ARI Field Unit, ATTN: PERI-IR, Ft Rucker, AL 36362-5000
ARI Field Unit, ATTN: PERI-IO, POB 5787, Presidio of Monterey, CA 93940
USA Training and Doctrine Command, ATTN: ATDO-ZT, Ft Monroe, VA 23651-5000
Combined Army Test Act, ATTN: ATCT-CG, Ft Hood, TX 76544-5065
COMMANDANTS:
  USA Air Defense Cen, ATTN: ATZC-CG, Ft Bliss, TX 79916-5000
  USA Armor Cen, ATTN: ATZK-CG, Ft Knox, KY 40121-5000
  USA Aviation Cen, ATTN: ATZQ-CG, Ft Rucker, AL 36362-5000
  USA Chemical Cen, ATTN: ATZN-CG, Ft McClellan, AL 36205-5000
  USA Combined Arms Cen, ATTN: ATZL-SC, Ft Leavenworth, KS 66027-5000
  USA Engineer Cen, ATTN: ATZA-CG, Ft Belvoir, VA 22060-5000
  USA Infantry Cen, ATTN: ATZB-CG, Ft Benning, GA 31905-5000
  USA Intelligence Cen, ATTN: ATSI-SA, Ft Hauchuca, AZ 85613-7000
  USA Ordnance Cen, ATTN: ATSL-CMT, APG, MD 21005-5201
  USA Signal Cen, ATTN: ATZH-CH, Ft Gordon, GA 30905-5000
  USA Field Artillery Center, ATTN: ATZR-C, Ft Sill, OK 73503
SUPERINTENDENT:
  US Military Academy, ATTN: Technical Library, West Point, NY 10996
COPIES FURNISHED:
  Defense Advanced Research Projects Agency, 1400 Wilson Blvd, Arlington, VA 22209-2308
  Defense Logistics Agency, Cameron Station, Alexandria, VA 22304-6183
  Defense Technical Information Center, Cameron Station, Alexandria, VA 22304-6145
  Lawrence Livermore National Lab, ATTN: L-191, POB 808, Livermore, CA 94550
  Los Alamos National Lab, ATTN: Dir for Energy, Rsch & Tech, Los Alamos, NM 87545

"DAPT soup" into a distillation still operated at reduced pressure. The large amount of heat is thus controlled and put to use by letting the reaction evaporate its own side products, acetic acid and water, at an acceptable temperature. The boiling point of the reaction mixture could be completely controlled by adjusting the partial pressure above it in the still. By reacting in this manner, no heat exchangers are required and no cooling is needed. The resulting DAPT is very concentrated, thick enough to turn solid if permitted to cool, and contains only a small residual quantity of acetic acid and water. Therefore, the bulk of the acetic acid, stripped from the DAPT, could be isolated and reclaimed as a pure substance for sale, or used as a feed stream for a ketene facility to produce more acetic anhydride, or for more convenient disposal. Concentrated DAPT can now be used or the DAPT could be easily isolated from the residual water and acetic acid as a white crystalline powder.

It was also found that the GARDEC DAPT process lends itself to making DAPT from more basic materials than hexamine, ammonia and formaldehyde, which at first form a solution of water and a pre-hexamine intermediate, pentamethylene-tetramine, PMT[3]. In particular, equations 5, 6 and 7 show how DAPT can be formed directly from ammonia and formaldehyde without specifically forming hexamine.

(5) $5\ HCH0 + 4\ NH_3 \text{ ------> } PMT + 5\ H_20$

(6) $PMT + 5\ H_2O + 2\ AC_20 \text{ ------> } 1\ DAPT + 5\ H_20 + 2\ HAC$

---

(7) $5\ HCH0 + 4\ NH_3 + 2\ AC_20 \text{ ------> } 1\ DAPT + 5\ H_20 + 2\ HAC$

Moreover, resynthesis does not occur since hexamine is not present. Quantitative yields of DAPT based on equations 5-7 have been experimentally verified, when the highly exothermic reaction is carried out instantaneously under vacuum at an elevated temperature limited only by the vaporization of acetic acid and water. In this scenario, the heat of reaction is sufficient to distill off the bulk of the acetic acid and water, yielding a concentrated solution of DAPT.

DAPT summary:

The new engineering approach used for the DAPT synthesis has led to substantially superior and cheaper process as compared with previous processes. This is due to the following.

• A substantial reduction in utility cost has been achieved by performing the reaction under vacuum. This has allowed for eliminating cooling and using the heat of reaction to drive off the reaction side products, acetic acid and water.

• Elimination of cooling eliminates the need for heat exchanger capitalization, and refrigeration capitalization.

• Choice of a given ratio of reactants and the high temperature reached by the reagents provide for instantaneous reaction time.

• Instantaneous reaction time results in 1. quantitative yield of DAPT, 2. high reactor throughput, and 3. small reactor volume.

• Using formaldehyde and ammonia in lieu of hexamine improves reaction efficiency by eliminating resynthesis and allows for reduced raw material cost.

IV. GARDEC TAT Process

Prior state of the art:

Known for over forty years, TAT until now has been a painstakingly prepared laboratory curiosity. Prior to work completed on the GARDEC Process, TAT had to be prepared a few grams at a time in procedures which required many steps and a large excess of reactants. Thus, TAT could not be made on a large scale. Efforts by other researchers have been as recent as 1985 and 1990[4,5].

The new TAT process:

The new GARDEC Process provides a fast and efficient method of preparing bulk quantities of pure TAT using only stoichiometric quantities of reactants. Currently, TAT is prepared in multi-pound quantities at Picatinny Arsenal, using no reactant excesses, with yields and purity approaching 100%. Multi-pound quantities of TAT which have been isolated were found to approach 100% purity. The process is carried out in few simple steps that require only common laboratory equipment.

The new, simplified pathway to TAT has been made possible by investigations leading to a more complete understanding of the underlying chemistry and physics. A rigorous examination of the acetylation of DAPT has revealed the overall reaction encounters chemical equilibria which seems to have been at the heart of the frustration experienced by earlier researchers. In

developing the GARDEC Process, it was found that the equilibria can be exploited by controlling the vapor phase above the reaction mass. Studies at Picatinny revealed that the reaction of DAPT to form TAT went through the following reaction sequence (figure 3).

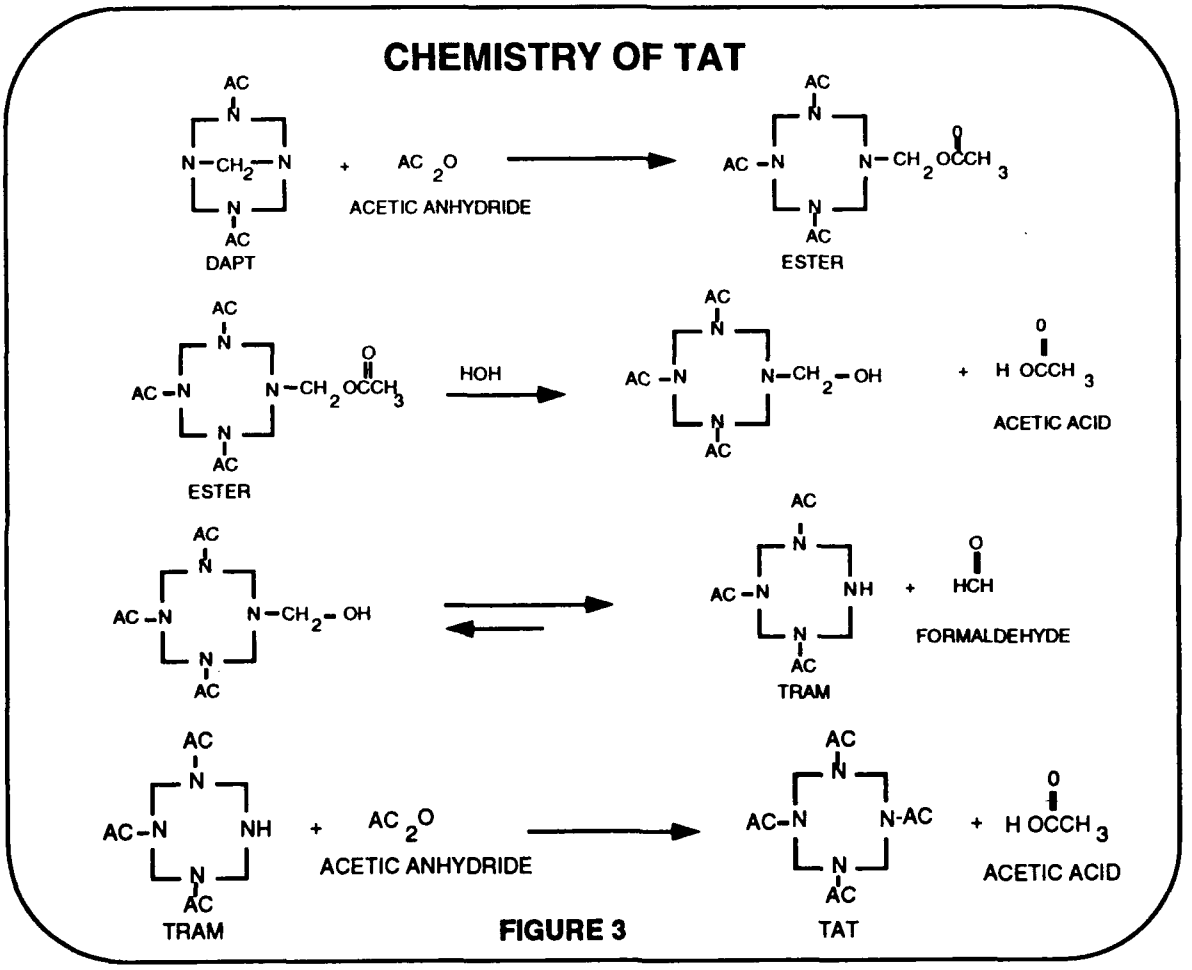


FIGURE 3

After proving that TAT could be made in a batch process in essentially 100% yield using no excess reactants, an understanding of the kinetics led to the achievement of a continuous process. Several interesting findings evolved which confirmed previous batch reaction studies and allowed for a complete demonstration of the engineering principles underlying continuous TAT synthesis. Two basic requirements must be met to continuously run the TAT

reaction. Acetic acid and formaldehyde, reaction byproducts, must be continuously removed. Acetic acid is formed from the reacted acetic anhydride, and the formaldehyde is formed during the hydrolysis of the ester. A build up of formaldehyde hinders the forward reaction completely, and a build up of acetic acid alters the course of the reaction towards TRAT as described by figure 4.

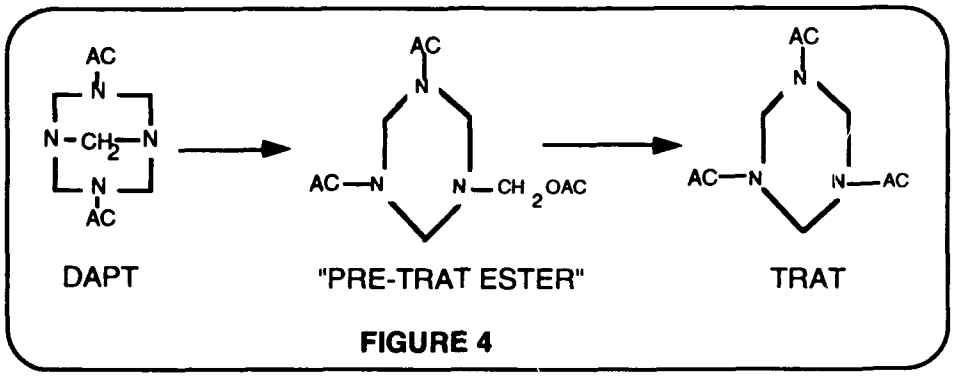


FIGURE 4

A report published by the Chinese in 1990 on tracing the mechanisms of acetolysis of hexamine and DAPT via NMR also supports these conclusions[6].

Furthermore, it was found that since the reaction of acetic anhydride with the triacetyl-amine (Tram) is extremely rapid as long as the equilibrium is pushed by moving the formaldehyde, the concentration of the acetic anhydride may be kept quite low and in this way be completely reacted before it can volatize from the reactor. Also the water present in the reaction for the needed hydrolysis does not appreciably attack the acetic anhydride. Furthermore, the inhibiting formaldehyde can easily escape from the surface of the stirred solution. The formaldehyde can then be recovered and reacted with ammonia which can then be subsequently reacted with acetic anhydride to form additional DAPT. DAPT is then reacted to form additional TAT. The TAT produced under this system was of high quality with a literature melting point of 161°C.

TAT summary:

The TAT synthesis, although it was very difficult to uncover, is very simple to operate. During the GARDEC Process reaction sequence it is important to point out that the TAT synthesis is nothing more than an acetylation of DAPT. To synthesize TAT requires no more introduction of new chemical

feedstock, no additional environmental concerns, and recovery of byproducts is similar to the DAPT step.

The overall GARDEC steady state reactions for TAT synthesis from basic starting materials can be written as shown in equation (8-11).

(8) $5\ HCOH + 4\ NH_3 \text{------>}\ PMT + 5\ H_2O$

(9) $PMT + 5\ H_2O + 2\ AC_2O \text{------>}\ 1\ DAPT + 5\ H_2O + 2\ HAC$

(10) $DAPT + H_2O + 2\ AC_2O \text{------>}\ TAT + 2\ HAC + HCOH$

---

(11) $4\ HCHO + 4\ NH_3 + 4\ AC_2O \text{------>}\ TAT + 4\ HAC + 4\ H_2O$

If starting with hexamine, then the equation can be written as follows.

(10) $C_6H_{12}N_4 + 6\ (CH_3CO_2)O + 2\ NH_3 \text{------>}\ 1.5\ C_{12}H_{20}N_4O_4 + 6\ CH_3COOH$

As can be seen, it is possible to utilize all methylene introduced into the reaction. It is theoretically possible to obtain a 150% yield of TAT based on hexamine as shown in equation 10.

V. Nitration of TAT to HMX

Prior state of the art:

Traditionally, TAT has been nitrated by using a large excess of dinitrogen pentoxide in nitric acid, as much as 12 moles per mole of TAT, or by the in-situ generation of dinitrogen pentoxide via phosphorus pentoxide in nitric acid. As much as forty moles of phosphorus pentoxide per mole of TAT have been used[7,8]. Although in some cases reasonable yield and purity are obtained, these methods entail considerable expense due to the large excess of nitrating agent.

Improved methods of TAT nitration:[9,10]

Studies were undertaken to examine this reaction more closely in an attempt to reduce the amount of dinitrogen pentoxide or phosphorus pentoxide to the lowest possible level. As a result of these studies an improved method was derived from a better understanding of the nitration mechanism and kinetics of each step. The process may be phenomenologically divided into two major steps,

the formation of SEX (1-acetyl-3,5,7-trinitro-1,3,5,7-tetraazacyclooctane) and the conversion of SEX to HMX. A procedure was developed which used only 2.25 moles dinitrogen pentoxide per mole of TAT in nitric acid and which gave quantitative yields of SEX. It had also been established that SEX could be nitrated to HMX with just concentrated nitric acid. By using the appropriate reaction conditions, it was possible to nitrate TAT in one reactor to HMX with only 2.25 moles of dinitrogen pentoxide or phosphorous pentoxide in nitric acid and obtain quantitative yields of HMX with a purity exceeding 98%.

VI. Summary and Conclusions

HMX is not only the most powerful non-nuclear explosive in general military use, but its value as a rocket propellant is unsurpassed. The preparation of HMX from elementary and inexpensive building blocks by use of a minimum of simple, safe, energy efficient, and environmentally sound steps has long been a highly desirable objective. With the GARDEC Process, we believe that this has finally been achieved. The GARDEC Process is a superior process as compared to other HMX processes. Scaled -up laboratory experiments strongly indicate that the process is less costly with respect to raw materials, utilities, labor, capitalization, maintenance, and product quality. Furthermore, environmental pollution is substantially reduced through the minimization of raw materials input, use of different process steps that allow for better recovery, and obtaining essentially 100% yield of HMX which circumvents the generation of hazardous waste streams. Specifically, when compared to the Bachmann Process, the total amount of raw material input due to stoichiometric use of reactants and a higher yield in the GARDEC Process account for a minimization of nitramine waste streams which has given GARDEC significant environmental advantages. Efficient use of acetic anhydride, unlike in the Bachmann Process, also minimizes reactant recovery.

Plans need to be developed to fully demonstrate on a pilot plant scale that the GARDEC HMX process can be incorporated into an existing production line such as at Holston AAP. Moreover, a tech data package needs to be developed so that the process can be available for off-the-shelf use.

References

1. Lukasavage, Nicolich, Alster, *Isothermal Method of Preparing DAPT via a Moving Bed Reactor*, US Patent Pending, File #07/687607 (1991)

2. Lukasavage, Nicolich, Alster, *An Improved Process for Preparing 3,7 Diacetyl - 1,3,5,7 - Tetraaza-[3.3.1]-Bicyclononane (DAPT) via Hexamine with no Cooling*, US Patent Pending, Docket #DAR-35-91 (1991)

3 Nielsen, et. al., *Structure and Chemistry of the Aldehyde Ammonias. Formaldehyde-Ammonia Reaction. 1,3,5- Hexahydrotriazine*, Org. Chem., Vol. **44**, No. 10, 1678-1684 (1979).

4. Gelber, et. al., *HMX Improvement Studies: The Preparation of TAT*, Technical Report , US Army Armament Research, and Development Center, ARAED-TR-85005, (1985).

5. Wamhoff, et. al., *New Synthetic Approaches to TAT, Technical Report*, Us Army European Research Office, R&D 4480A-CH-01, (1990).

6. Ju, et. al., *Study on the Mechanism of the Acetolysis of Hexamethylenetetramine*, Propellants, Explosives, Pyrotechnics **15**, 54-57 (1990).

7. Siele, et.,al., *Alternative Procedures for Preparing HMX*, Propellants and Explosives **6**, 67-63 (1981).

8. Siele and Gilbert, *Process for Preparing 1,3,5,7 - Tetranitro - 1,3,5,7-Tetraazacyclooctane*, United States Patent #3,939,148 (1976).

9 Lukasavage, Nicolich, Alster, *An Improved Process for Preparing 1-acetyl-3,5,7-Trinitro -1,3,5,7-Tetraazacyclooctane (SEX)*, US Patent Pending #07/775406 (1991).

10 Lukasavage, Nicolich, Alster, *An Improved Process for Preparing 1,3,5,7 - Tetranitro - 1,3,5,7 - Tetraazacyclooctane (HMX)*, US Patent Pending, File #07/775402 (1991).

Using Iterative Orthophoto Refinements to Correct Digital Elevation Models

F. Raye Norvelle
U.S. Army Topographic Engineering Center
Research Institute
Fort Belvoir, VA 22060

INTRODUCTION

Digital correlation methods provide a fast and effective means for automatically generating terrain elevation data from digital stereopairs of aerial photographs. Compilation rates of over two hundred points per second can be achieved on modern computers. The data produced have to be edited, however, and this can be a very tedious and time-consuming process. For example, the amount of correlation data produced in 15 minutes by the correlation method *match*, developed by the U.S. Army Topographic Engineering Center (TEC), can require up to 5 hours to inspect and edit using manual and computer-assisted techniques.

To minimize interactive editing requirements, TEC has developed a new DEM editing/generation technique named the "Iterative Orthophoto Refinements (IOR)" method. It is highly effective and is commensurate in speed with modern correlation methods. The IOR method, which is similar in concept to that reported by Schenk(1989), is based on the premise that, given a DEM and accurate exterior orientation parameters for a stereopair of digital images, it is possible to generate two, supposedly identical, orthophotos of the original stereoimages. If the orientation data are correct, any geometric mismatch between the two orthophotos will be caused by errors in the derived elevations. The mismatches can be measured automatically using a digital correlation method, converted to equivalent elevation errors and used to refine the original DEM. New orthophotos can then be generated from the updated DEM and the process repeated in an iterative manner until no measurable mismatches exist between the orthophotos.

A detailed description of the IOR method is given in the next section. This is followed by sections on test results, a discussion of the results and conclusions.

THE IOR METHOD

The IOR method mainly requires the use of a digital correlation routine, a digital orthophoto generation capability and a technique for converting geometric mismatches between orthophoto pairs to equivalent DEM error. Each of these processes will be discussed in the following sections.

THE DIGITAL CORRELATION METHOD

The correlation method used in the IOR process is named *match*. It is the offspring of a program developed by TEC in the late 1970's named the "Digital Interactive Mapping Program (DIMP)." In the late 1980's, DIMP, which was written in the FORTRAN language for a CDC Cyber computer, was modified to include various new capabilities, rewritten in the C-language for a Silicon Graphics Iris 4D/85 engineering workstation and renamed *match*. A full description of *match* is beyond the scope of this paper. The reader is referred instead to Norvelle (1981) which describes DIMP in detail and is sufficiently applicable to *match*.

Match is an area-based digital correlation method that operates in image space and is referenced to an evenly spaced grid of points on the left image of a stereopair. That is, for equally spaced image points (grid) on the leftmate, *match* attempts to determine, based on the maximization of the normalized cross correlation coefficient, the image coordinates of corresponding points on the rightmate. This routine is used in two ways in the IOR method. First, correlation is performed on a stereopair of images to determine a dense grid of conjugate points. These points are intersected to obtain a DEM. The DEM is then used in conjunction with the exterior orientation data for the stereoimages to generate an orthophoto of each image of the stereopair.

Match is used in a second mode to automatically determine the geometric mismatches between the two orthophotos. The procedure is the same as in the first mode but now the conjugate image points determined by *match* should have identical orthophoto image coordinates. The extent to which they are not identical (mismatches) can be transformed into equivalent elevation errors that are then used to correct the current DEM values.

ORTHOPHOTO TRANSFORMATION

Orthophotos are produced using a differential rectification technique. Projection equations are used to

rigorously compute corresponding image coordinates (xp,yp) for the ground coordinates (Xg,Yg,Zg) of a rectangular array of four DEM points. The Xg and Yg-ground coordinates of the DEM points are then mathematically related to their corresponding image coordinates according to Equations 1.

$$x_p = a X_g + b Y_g + c X_g Y_g + d \qquad \text{Eqs (1)}$$
$$y_p = e X_g + f Y_g + g X_g Y_g + h$$

Once the coefficients are determined, Equations 1 are used to compute image coordinates for ground points (orthophoto pixels) that fall within the boundaries of the four DEM values. For each computed image point (xp,yp), a gray level is extracted from the original image using bilinear interpolation and assigned to the orthophoto at coordinates Xg and Yg. The computations proceed from one rectangular array of DEM values to the next until the full orthophoto is generated.

The ground-sample value (meters) for each pixel of the orthophoto is selected by the user. Its value is normally chosen to have about the same ground resolution as a pixel on the input image. The orthophoto pixel size must also be divisible into the DEM spacing by some integer value. This is required so that the *match* routine will be capable of determining a mismatch between orthophotos at ground coordinates registered with the DEM values. For example, if an input image pixel is approximately equivalent to 2.12 meters on the ground, and the elevations in the DEM are spaced 25 meters apart, an orthophoto pixel will be chosen that has a 2.5 meter ground resolution. When *match* is used, conjugate points (mismatches) will be determined at a spacing of 25/2.5 = 10 pixels on the left orthophoto and will coincide exactly with a DEM value.

DEM ERROR COMPUTATION

Figure 1 shows the geometric relationship between mismatches in the orthophotos (dx) and the DEM error (dh) that caused the mismatch. P' is a measured DEM point (shown at the dot positions) and is in error by a value of dh. The orthophoto image of point P' falls at a ground position Xl on the left orthophoto and at Xr on the right. The mismatch between the two is dx. Since the coordinates of the camera stations and the ground positions of P'(Xl) and P" (Xr = Xl+dx) are known, it is possible to compute the position of P by intersection.

The error in elevation between P' and P can be computed

and, if the slope of the true ground surface is known, converted to the error dh directly beneath the DEM point P'. The true ground slope is not known, however, but can be approximated from the measured DEM surface. Alternately, the approximate error, dh', can be computed as dh' = dx(H/B), where H/B is the reciprocal of the base-height ratio between camera stations. The error dh (or dh') can then be used to update the DEM value of point P'.

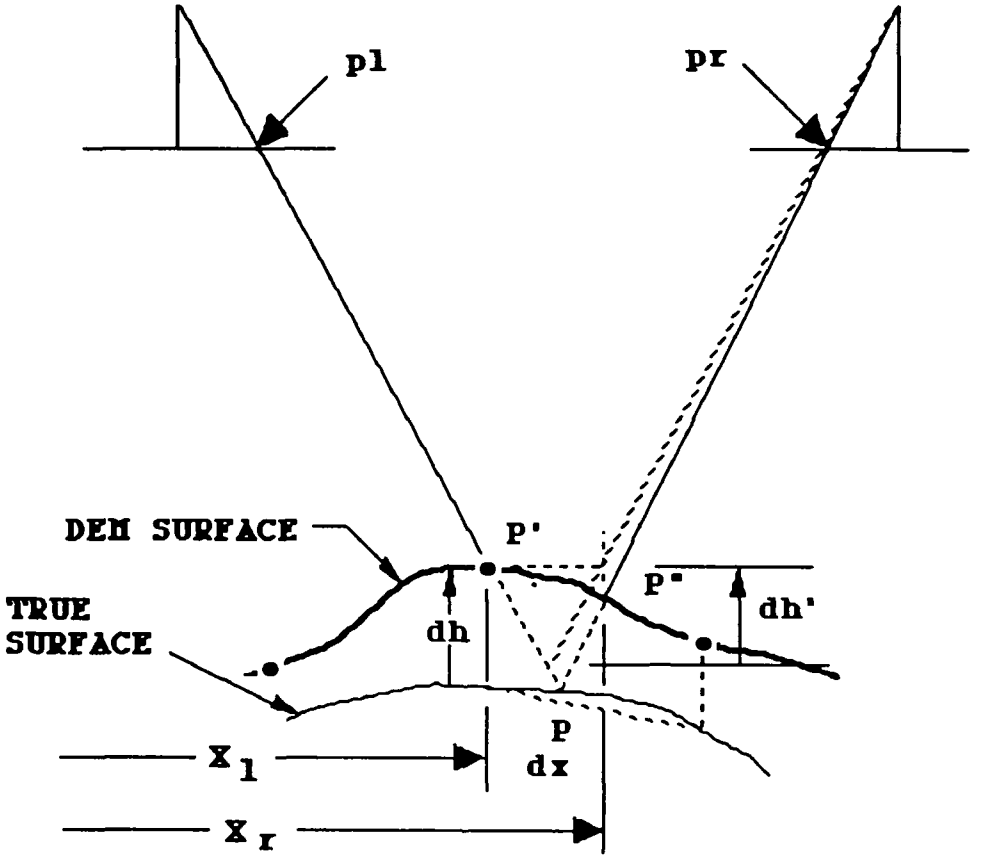


Figure 1. Relationship Between Orthophoto Mismatches and DEM Error.

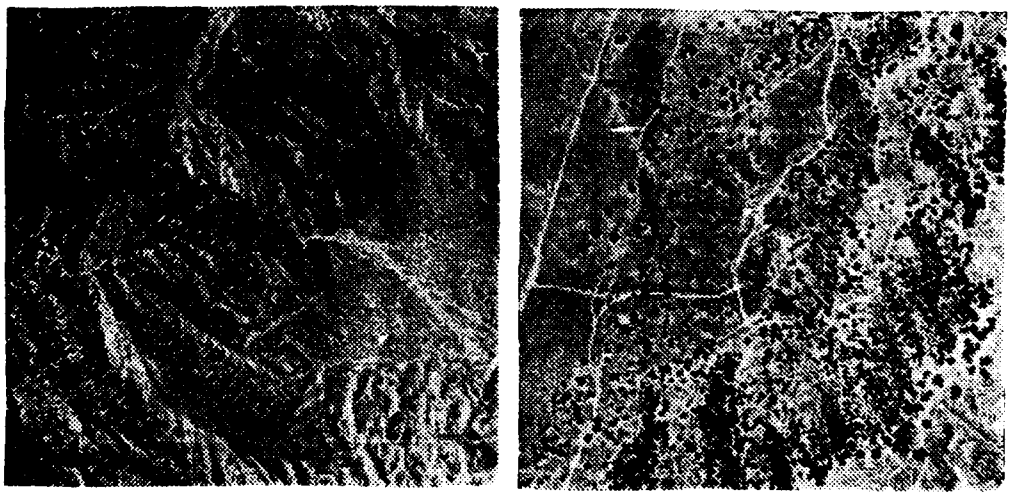
In cases where the DEM errors are erratic (individual "spikes" in the elevation data), the slope of the ground, as approximated by the slope of the DEM surface, may be in serious error. This will cause an exaggerated dh correction. In such cases, the approximate correction, dh', may be more dependable.

In either case, since the true ground slope is not known, the IOR process is necessarily an iterative one.

RESULTS

TEST IMAGES

Two examples of test results obtained with the IOR method are given in this section. The first test was made with 1280 by 1024-pixel stereoimages obtained by scanning 1:40000 scale photographs of the San Bernardino Mountains in California (Figure 2a) with a 50-micrometer spot size. The terrain is very rugged (elevations range from 1350 to 1950 meters above sea level) with slopes up to 45 degrees and slope changes (ridges little variance in gray shades. Ridge lines and slopes have a moderately-dense cover of small trees. The sharp slope changes and low radiometric variance in some image areas cause difficulties for a correlation method and large DEM errors will occur. These images, therefore, present a good test for the IOR method.



(a) (b)

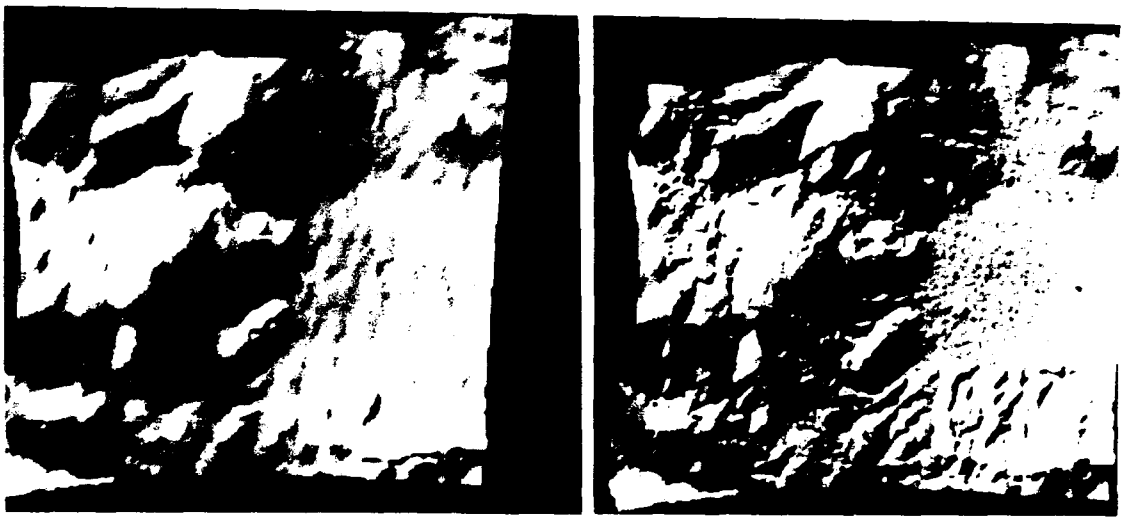
Figure 2. (a) San Bernardino Mountains and (b) Fort Hunter Liggett Test Images.

In the second test, 2048 by 2048-pixel stereoimages are used. They were obtained by scanning 1:20000 scale photographs of Fort Hunter Liggett, California, (Figure 2b) with a 25-

micrometer spot size. The terrain is dotted with tall, isolated trees whose shadows are prevalent on one image but occluded by the trees themselves on the other. The terrain, with elevations from 369 to 480 meters, is otherwise grass covered and provides very little radiometric variance in the images to support digital correlation and DEM generation. Large elevation errors can be expected, especially with regard to tree heights, making these images a difficult test of the IOR method.

SAN BERNARDINO MTS TEST RESULTS

The original stereoimages were initially reduced in scale by 4x to (1) minimize the large x-parallaxes caused by the rugged terrain and (2) reduce the time required to obtain an initial DEM. The reduced images were correlated to provide a DEM with a 40-meter spacing. The spacing was changed to 10 meters by bilinear interpolation. The resulting DEM is shown as a shaded relief image in Figure 3a. The "sun" is in the upper-left corner at an altitude of 50 degrees from the nadir.



(a) (b)

Figure 3. San Bernardino Mtns. Shaded Relief Images of (a) The Approximate DEM and (b) The DEM After 2 Iterations of the IOR Method.

The initial DEM, although approximate, was used to generate a pair of orthophotos from the original, full-resolution stereoimages. The geometric mismatches (x-parallax)

between the orthophotos were measured automatically with *match*, converted to elevations errors and used to refine the initial DEM. The mismatches equated to elevation errors ranging from -100 to +43 meters.

A second set of orthophotos were made and, when viewed in 3-D, were found to contain new, but smaller, mismatches. A second iteration of the IOR method was performed and detected elevation errors ranging in value from -11.0 to +13.6 meters. These errors were used to further refine the once-corrected DEM. The refined DEM is shown as a shaded relief image in Figure 3b. The total elapsed time for the above steps was 30 minutes.

For validation purposes, a third set of orthophotos were generated. No significant mismatches were found in the new orthophotos except in one area of dense tree growth. The radiometric variance in this image area was so low that correlation was impossible. Any further DEM corrections would have to be made using manual editing techniques.

FORT HUNTER LIGGETT TEST RESULTS

Previous correlation experience with these images showed that it is nearly impossible for *match* to determine the conjugate images of the tall, individual trees because of the large x-parallax (16 pixels) between tree images and the adverse affects of large tree shadows. Consequently, the original stereoimages were reduced in scale by 4x to (1) reduce the large x-parallaxes, (2) reduce the size of tree shadows and (3) to minimize the time required to obtain an initial DEM.

The reduced images were correlated at a 5-pixel spacing on the left image to create a DEM with a 10-meter spacing. This initial DEM is shown as a shaded relief image in Figure 4a. The "sun" is in the upper-left corner at an altitude of 65 degrees from the nadir.

Orhtophotos of the reduced images, with each pixel equal to a 2-meter ground-sample distance, were generated and viewed in 3-D to determine the efficacy of the initial DEM. Many significant mismatches were noted among tree images. Therefore, instead of immediately using full-scale orthophotos as in the test above, orthophotos of the reduced images were incorporated in the IOR process to further refine the DEM, including tree heights. Refinements were made for elevation errors that ranged in value from -11.6 to +9.0 meters.

Next, the original steroimages were reduced by 2x and used with the current DEM to create orthophoto pairs with each pixel equal to a 1-meter ground-sample distance. Four iterations of the IOR method were required to correct the current DEM and

produce acceptable orthophoto pairs. The ranges of detected elevation errors with each iteration of the IOR method were -6.9 to +11.1, -17.9 to +11.3 , -6.7 to +12.3 and -3.2 to +4.1 meters, respectively.

The previously corrected DEM was next used to create orthophotos of the original, full-resolution stereoimages. Each pixel of the orthophotos had a ground-sample distance of 0.5 meters. These orthophotos were correlated at every 20 pixels on the left orthophoto to register precisely with the 10-meter spacing of the DEM. After one iteration of the IOR method, elevation errors from -4.9 to +4.2 meters were detected and used to refine the current DEM. The IOR procedure was repeated a second time at full resolution and provided DEM corrections for elevation errors ranging from -2.0 to +2.0 meters. The resulting DEM is shown as a shaded relief image in Figure 4b.

(a) (b)

Figure 4. Shaded Relief Images of the Fort Hunter Liggett Test Area. (a) Initial DEM. (b) Final DEM.

Using the final version of the DEM, new orthophoto pairs were created and inspected in 3-D to validate that the DEM was correct. No mismatches were found in the terrain imagery but

in some of the grass areas it was not possible to observe mismatches even if they existed. This is because there was not enough variance in the gray shades to provide adequate stereoviewing. Although the orthophotos may appear correct, the DEM may actually be in error.

Mismatches were still evident in the orthophotos where tall trees existed. This indicates that the true height of the trees had not been established by the IOR process. These trees are typically 12 meters in height and produce 16 pixels of x-parallax in the original stereoimages. The mismatches on the orthophoto pairs were generally 4 pixels in size (2 meters on the ground) which equates to an elevation error of about 3 meters.

DISCUSSION OF RESULTS

Figures 3 and 4 clearly illustrate that the IOR method is successful in adding high-resolution corrections to a DEM. The Figures do not show the accuracy of the corrections, however, but this is judged by 3-D viewing of the final orthophotos. The 3-D error surface, represented by mismatches in the orthophotos, should be flat. This, with some notable exceptions, was found to be the case in both tests.

SAN BERNARDINO MTNS TEST

The DEM of the San Bernardino Mountains was fully corrected except in one small area of the stereoimages. The image area is a very sharp, tree-lined valley. When generating the initial DEM, the abrupt change in slope direction, coupled with the low radiometric variance of the trees, caused the correlation process to fail and produce significant DEM errors in this region. Further DEM improvements by the IOR method were limited because, in the presence of adverse radiometric characteristics, the *match* routine could not accurately detect mismatches in the orthophotos. In adverse areas such as these, manual editing is required.

The IOR method was completed on the San Bernardino images in 30 minutes. In previous tests with these images, 5 hours were required to manually edit the data obtained by conventional correlation procedures on the full-resolution stereoimages. This is a 10 to 1 improvement in speed using the IOR method. Furthermore, orthophoto pairs made from the manually-edited DEM were not flat when viewed in 3-D, indicating that not all error had been sufficiently corrected.

FORT HUNTER LIGGETT TEST

The results of this test are considered to be very good in spite of the fact that the obtained tree heights are not totally accurate (± 3 meters). In previous tests, conventional correlation techniques were used on the full-resolution images and were not successful in detecting tree heights. The IOR method was successful because, by initially reducing the images and x-parallaxes by 4x, it was possible for *match* to correlate on individual trees and, consequently, obtain approximate tree heights. When orthophotos were generated, the positions of the tree tops were sufficiently corrected to allow the IOR process to detect mismatches and further refine the tree heights.

By stepping from 4x to 2x and finally to 1x-reduced images, it was possible to keep the x-parallaxes (mismatches) in the orthophotos small enough that they could be measured by *match* and corrected by the IOR process. If, as in the test above, the intermediate steps performed on the 2x-reduced images had been omitted, the mismatches at 1x would have been too large and the IOR method would have failed on these images.

GENERAL

In both test cases, the IOR method is considered more accurate and faster than conventional correlation on original stereoimages followed by manual or interactive editing. It is faster because the orthophotos are more nearly identical and less searching ("pull-in" range) is required to find the conjugate points. It is more accurate because dissimilarities between the original stereoimages, a potential cause of error in a correlation process, are essentially removed from the orthophotos.

Usually, two or three iterations are necessary to correct a DEM. Special cases, such as the Fort Hunter Liggett images, may require more. Generally, if the mismatches are not removed from the orthophotos in 3 iterations, conditions (occlusions,shadows,etc) probably exist which cause DEM errors that can only be corrected by manual methods.

CONCLUSIONS

- The IOR Method is a fast and effective technique for generating and editing DEM's.
- The orthophoto transformation process removes image dissimilarities and allows correlation on the orthophotos to be performed more accurately and faster than on the original stereoimages.
- Two to three iterations of the IOR method are typical for accurate DEM correction.
- The IOR method can be 10 times faster than manual editing methods.

ACKNOWLEDGEMENTS

The stereoimages and associated orientation data of the San Bernardino Mountains, CA, were furnished by the U.S. Geological Survey, Western Mapping Center, Menlo Park, CA.

The stereoimages and associated orientation data of the Fort Hunter Liggett area were furnished by the Test and Experimentation Command, Fort Ord, CA.

REFERENCES

Norvelle, F. Raye, October 1981. *Interactive Digital Correlation Techniques for Automatic Compilation of Elevation Data*, ETL-0272, U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Va.

Schenk, A. F., 1989. *Determination of DEM Using Iteratively Rectified Images*, Photogrammetry Technical Report No. 3, Department of Geodetic Science and Surveying, The Ohio State University, Columbus, Ohio

New Human Exposure Limits for Freefield Artillery Blast Overpressure

Dr. James H. Patterson, Jr.,* Mr. Ben T. Mozo, and MAJ William R. Nelson
U.S. Army Aeromedical Research Laboratory, Fort Rucker, AL 36362-5292

Dr. Daniel L. Johnson and Dr. Donald R. Richmond
EG&G Special Projects, Albuquerque, NM 87119-9024

LTC Gary Ripple and LTC Yancy Y. Phillips
Walter Reed Army Medical Center, Washington, DC 20307-5100

INTRODUCTION

Blast overpressure or high intensity impulse noise can injure many organ systems in the human body. The inner ear appears to be the most sensitive, being injured at the lowest pressure levels. Air containing organs such as the middle ear, respiratory tract, and gastrointestinal tract are next. Impulse noise exposure limits are intended to protect the soldier from these injuries, especially those produced by his own weapon systems.

The current military standard which defines the exposure limits for impulse noise is MIL-STD-1474C.[1] Over the past 15 years, the Army has developed or adopted numerous heavy weapons which can produce impulse noise levels which exceed these limits, e.g., M198, 155 mm towed howitzer; the M119, 105 mm towed howitzer; the M109A3E3, 155 mm self-propelled howitzer; the VIPER, AT4, and Ranger Anti-Armor Anti-personnel Weapon System (RAAWS) antiarmor weapons; the 120 mm and the improved 81 mm mortars. Many of these weapons have been fielded with firing restrictions in training because they exceed the impulse noise limits.

System specific studies to determine whether there is a hazard to hearing have been completed for two of these weapons, the M198[2] and the VIPER.[3] These studies demonstrated that, when personnel operating these systems are protected with properly inserted foam earplugs, these systems can be fired without significant risk of injury from the impulse noise. These studies called into question the validity of the impulse noise limits in MIL-STD-1474C. However, they did little to establish new limits.

In order to empirically determine valid exposure limits for personnel exposed to freefield impulse noise like that produced by heavy weapons, the U.S. Army Medical Research and Development Command undertook a major study of the effects of impulse noise on soldiers. For purposes of this study, heavy weapon noises were divided into three classes. The first class, typical of towed artillery, consists of long-duration impulses containing primarily low frequency energy. The second class, typical of mortars and some antiarmor weapons, consists of short-duration impulses containing primarily high frequency energy. The final class is one of intermediate-duration impulses with an energy distribution between the first two. This report covers only the results of the first of these three classes. From these results, new limits for artillery weapons have been derived.

METHODS

The methodology used in these studies is to expose human volunteers, under controlled conditions, to a progression of impulse noises while monitoring them for signs of adverse effects after each exposure. The primary measure of effect on hearing is called temporary threshold shift (TTS), a temporary elevation of the threshold of hearing. TTS was determined by subtracting a baseline audiogram from each of a series of audiograms measured after a noise exposure. The baseline for each volunteer was the average of at least 8 audiograms taken over several days before their first exposure to impulses in the study. After each impulse noise exposure, a series of audiograms were taken starting at 2 minutes after the last impulse. Additional audiograms were started at 20 and 60 minutes postexposure. The TTS resulting from each noise exposure is the difference between the baseline and each postexposure audiogram. If TTS did occur, additional audiograms were utilized until the threshold had returned to normal. The TTS derived from the 2-minute postexposure audiogram was used in the results to be presented here. The other postexposure audiograms were only used to assure complete recovery before the next exposure.

The indicators of nonauditory effects used in this study were direct laryngoscopic examination of the upper airway, stool guaiac, pre- and postexposure spirometry, and a questionnaire concerning general symptoms. We believed that the laryngoscopic exam was probably the most sensitive indicator of nonauditory effects.

The impulses used in the first phase of this study were produced by the detonation of C4 explosive charges about 10 feet above a concrete pad. Figure 1 shows a typical pressure signature at the volunteer's location. Charge weights were determined that would produce impulse noise levels ranging from the 100-shot limit of MIL-STD-1474 to the nonauditory limits (Dodd et al.[4]) at the volunteers' positions.

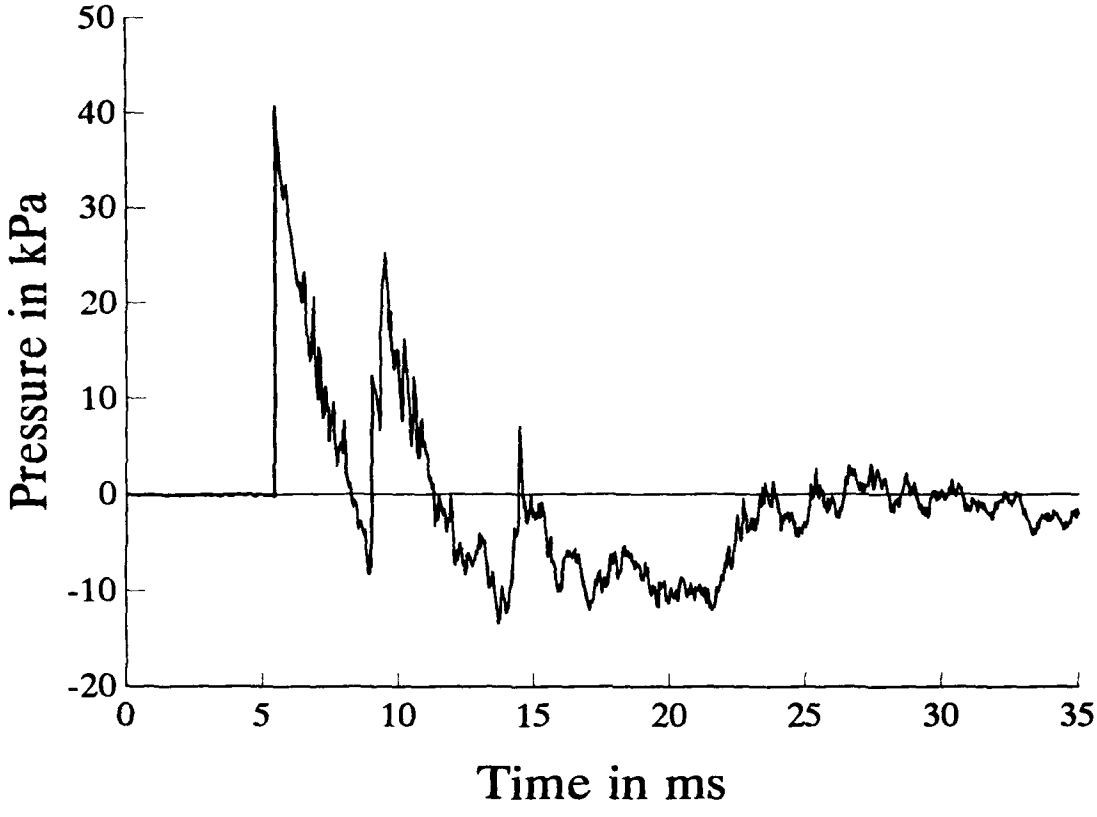


Figure 1. Pressure-time signature of a typical exposure impulse.

Table 1 summarizes the levels and other basic parameters of the impulses. An exposure condition consisted of 6, 12, 25, 50 or 100 impulses at one of the levels shown in Table 1. Volunteers were exposed to only one exposure condition per day. The first condition was always 6 impulses at level 1. On successive days, the level was increased for 6 impulses until either a significant effect was noted or until level 7 was reached. Then the number of impulses was increased at a lower level until 100 impulses at some level was reached. Any time a significant effect was noted, the general rule was to increase number and decrease intensity. This design allows for each volunteer to be exposed to progressively more energy by either increasing intensity or number, while eliminating exposure conditions which would produce no effect.

Table 1.

Average Blast Exposure Parameters

Intensity Code	Peak (kPa)	Peak (dB)	A-duration (ms)	B-duration (ms)	A-impulse (kPa*ms)
1	9.0	173.1	2.2	21.9	9.2
2	13.7	176.7	2.4	22.3	14.2
3	18.6	179.4	2.5	20.2	19.8
4	25.9	182.2	2.8	22.2	27.7
5	34.9	184.4	2.9	21.5	38.0
6	50.3	187.9	2.9	22.6	55.9
6B*	45.3	187.1	3.0	22.1	51.4
7	65.8	190.3	2.9	27.7	72.5

* Level 6B was used for 50 and 100 shots.

During an exposure condition, the volunteers were seated on chairs which positioned their test ear at 1.5 meters above the concrete pad, 5 meters from the center of detonation, and the right ear was oriented toward the detonation. They wore earmuffs, the PASGT helmet, and goggles to protect their eyes from flying debris. The right ear was protected by the muff only; the left ear was given the added protection of a foam earplug and the earmuff. Only data from the right ear were used in the analysis. The first group of 62 volunteers used the muffs as they came from the factory. The average real-ear attenuation[5] for these muffs is shown in Figure 2. The second group of 60 volunteers wore a modified muff. The modification consisted of 8 plastic tubes inserted through the cushion of the right earmuff to simulate a poorly fitted earmuff. The attenuation of the modified muff is also shown in Figure 2.

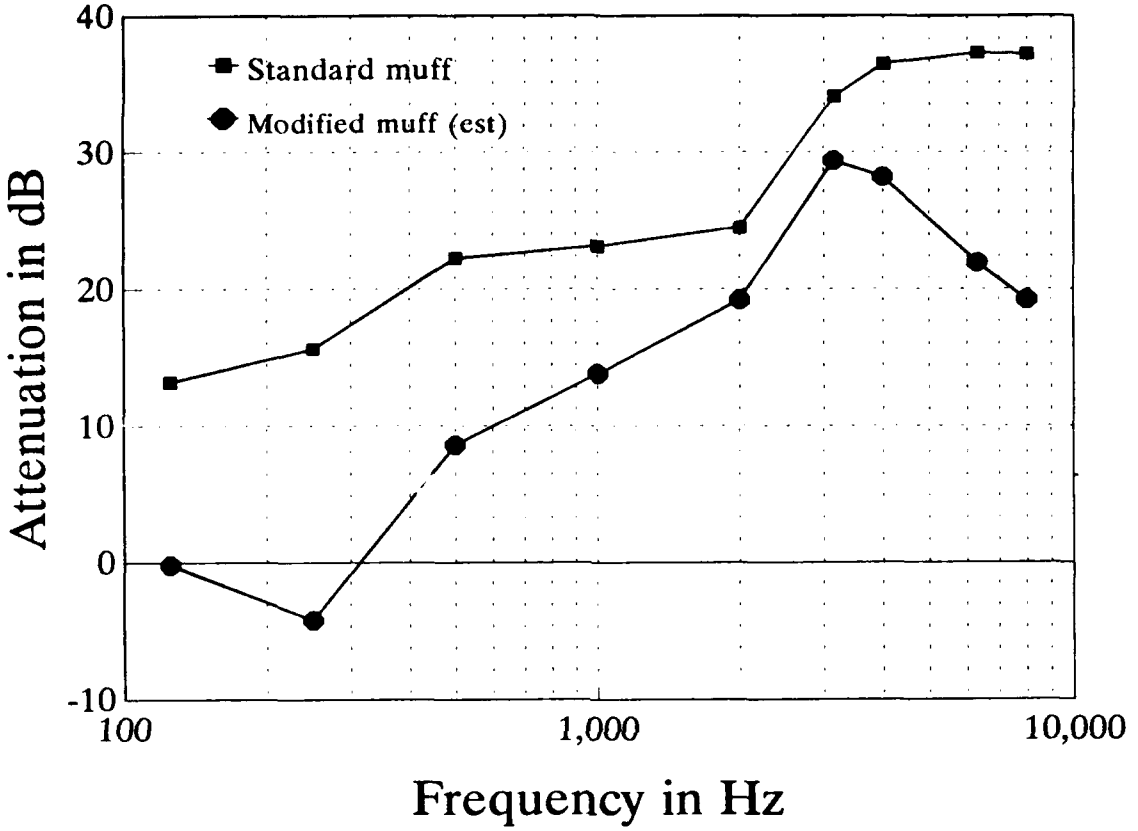


Figure 2. Average attenuation characteristics of the standard earmuffs and the modified earmuffs using real-ear attenuation test (REAT).

RESULTS AND DISCUSSION

The results of the study can be summarized by counting the number of individuals exposed to a condition and then noting the number who showed a significant TTS. A significant TTS was defined as one which exceeded 25 dB at any audiometric test frequency. Figure 3 shows the results for the 62 volunteers who used the standard muffs. As expected, all 62 showed no effect for the 6 impulses at level 1. As the level increased, the number of volunteers participating decreased. This is due to several factors. Some of the volunteers withdrew from the study in accordance with their volunteer agreement. Others asked not to be exposed at increased intensity, but agreed to continue in the study for exposures at increased numbers of impulses and lower levels. A few were

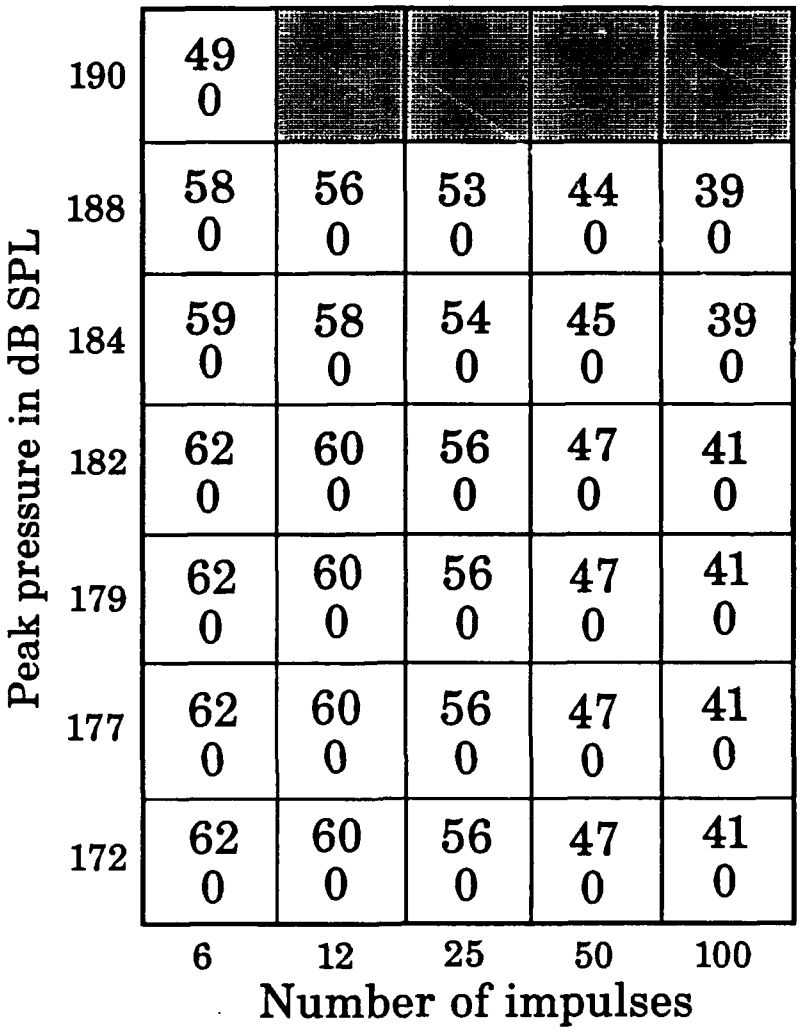


Figure 3. Number of individuals exposed (top number) and number of individuals showing an effect on hearing (bottom number) for each exposure condition when the standard earmuff was worn.

dismissed from the study for misbehavior. One of the first 62 subjects was discontinued in the study because of a positive laryngoscopic finding. This occurred at levels 5 and 6 for 6 impulses. While this finding may have been an artifact of initial procedural difficulties, this result could not be conclusively attributed to causes other than the exposure and was therefore counted as a nonauditory effect. It

should be noted that this was the only occurrence of a nonauditory effect in the first 62 volunteers. The next 60 volunteers had the same nonauditory exposure; only the earmuff changed. Only one positive laryngoscopic finding occurred in this group (100 impulses at level 6). Thus, we had an incidence rate of less than 2 in 100. No other nonauditory effects of medical significance occurred.

Among the group of volunteers wearing the standard muff, there were no TTSs exceeding 15 dB at any frequency for any exposure condition. The standard muffs provided enough protection so that even the most energetic exposure conditions produced no effect on hearing.

The results for the group of 60 volunteers using the modified muff are shown in Figure 4. The rate of voluntary withdrawals was much lower, less than 4 percent. Voluntary withdrawals dropped precipitously after some personnel changes at the test site.

With the reduced attenuation, we saw some significant TTSs begin to appear. At level 7, there was one significant TTS out of 56. At level 6 for 12 impulses, there was also 1 subject with significant TTS. As this subject moved through the matrix, he showed an effect at each step down in intensity and up in number. At level 6B, 100 impulses, there was one additional subject with significant TTS. Thus, this condition showed a total of 2. This is an incidence rate of less than 4 percent at 187 db peak pressure for 100 impulses using marginal hearing protection.

Exposure conditions are considered acceptable if they resulted in a TTS less than 25 dB in 95 percent of the population exposed. Statistically, we can use a binomial test[6] of the null hypothesis: 95 percent of the population show less than 25 dB TTS to test the acceptability of any exposure condition. The alternative hypothesis is that less than 95 percent of the population show an unacceptable TTS or equivalently that more than 5 percent of the population show an excessive TTS. Since very few exposure conditions resulted in TTSs exceeding 25 dB, we need to consider only those conditions which resulted in the largest percentage at TTSs exceeding this criterion. For the modified earmuff, 2 out of 58 individuals showed unacceptable TTS as a result of exposure to 100 impulses at level 6B. For a sample size of 58 individuals, we can use the normal approximation[6] to test the null hypothesis. This gives a $z=-0.24$ and a $p=0.595$ which is not significant at any reasonable significance level. Therefore, we cannot reject the hypothesis that 95 percent of the population would show a TTS less than 25 dB when only marginal protection was used.

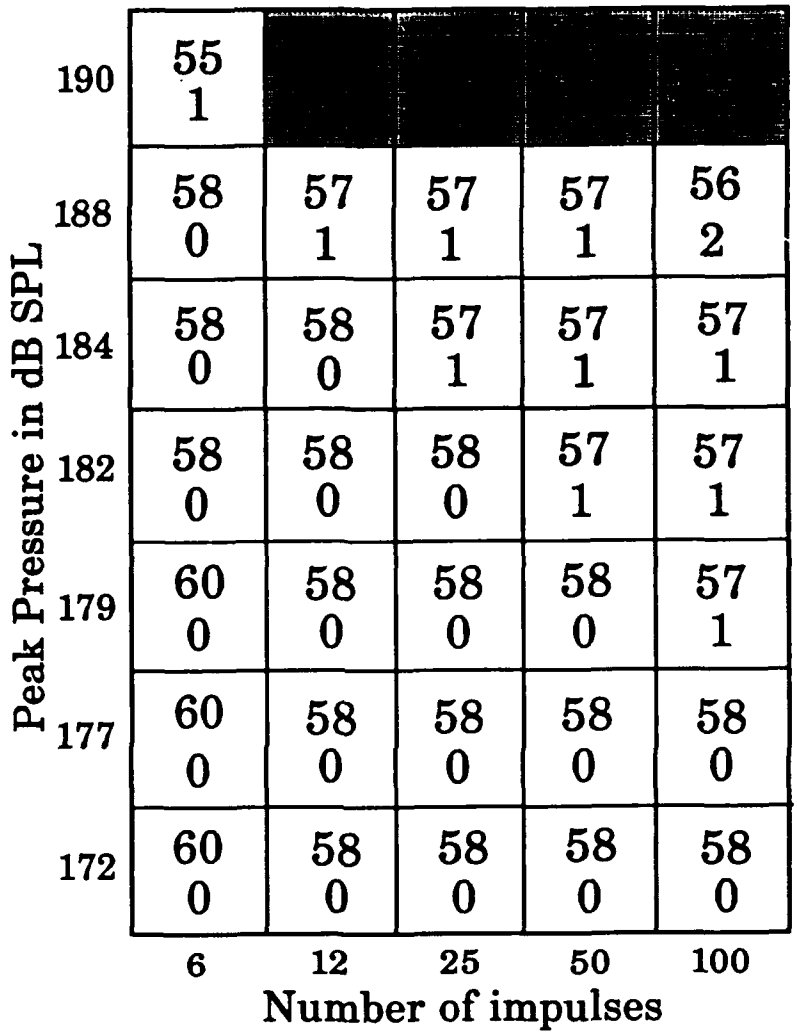


Figure 4. Number of individuals exposed (top number) and number of individuals showing an effect on hearing (bottom number) for each exposure condition when the modified muff was worn.

Using the results of this study, new exposure limits for freefield artillery impulse noise have been developed. Table 2 shows a comparison of these new limits with the limits under MIL-STD-1474C. The new limits increase the allowable peaks for a 20 ms B-duration by 10-14 dB. It should be noted that these limits are the maximum exposure conditions permitted by the nonauditory limits and may not represent the most severe exposure for which hearing can be protected.

Table 2.

Comparison of New Limits with MIL-STD-1474C

No. of rounds	Peak level limit New (dB)	1474C (dB)	Peak pressure limit Proposed (kPa)	1474C (kPa)	B-duration 1474C (ms)
6	190	179.5	65	19	20
100	187	173.5	45	10	20

CONCLUSIONS

Marginal hearing protection is adequate to protect 95 percent of the exposed personnel against the impulse noise used in this study up to 187 dB for 100 impulses. The protection need not be the best available; even an earmuff with a leaky seal will provide good protection. To the extent that heavy weapons produce pressure signatures which are similar to those used in this study, they should be safe to fire 100 rounds per day with levels up to 187 dB peak. It should be noted that the results reported here only apply to 155 mm and 105 mm towed howitzers and the area outside the turret around the 155 mm self-propelled howitzer. The pressure signatures characteristic of other heavy weapons will be addressed in later phases of this ongoing study.

References

1. Department of Defense. *Noise Limits for Military Ground Materiel.* Washington, D.C. MIL-STD-1474C. (September, 1990).

2. J.H. Patterson, B.T. Mozo, R.H. Marrow, R.W. McConnell, I.M. Lomba-Gautier, and D.L. Curd, *Direct Determination of the Adequacy of Hearing Protective Devices for Use with the M198, 155 mm Towed Howitzer*, USAARL Report 85-14 (1985).

3. J.H. Patterson and B.T. Mozo, *Direct Determination of the Adequacy of Hearing Protection for Use with the VIPER*, USAARL Report 87-9 (1987).

4. K.T. Dodd, J.T. Yelverton, D.R. Richmond, J.R. Morris, and G.R. Ripple, Nonauditory Injury Threshold for Repeated Intense Freefield Impulse Noise, *J. of Occup. Med.* **32(3)**, *261 (1990)*.

5. *American National Standard, Method for the Measurement of the Real-Ear Attenuation of Hearing Protectors*, Acoustical Society of America (1985).

6. S. Siegel, *Non-parametric Statistics* (McGraw-Hill, New York, Toronto, London, 1956), pp. 36-42.

Oral, Microencapsulated, CFA/II Vaccine Against E. coli Diarrheal Disease: Preclinical Evaluation

Robert H. Reid*, COL, MC
Edgar C. Boedeker, COL, MC
Charles E. McQueen, LTC, MC
Catherine L. Wilhelmsen, LTC, VC
Lee-Ying Tseng
David Davis, Ph.D.
James A. Kodak
Keya Sau
Walter Reed Army Institute of Research
Washington, DC 20307-5100

Hitesh R. Bhagat, Ph.D.
School of Pharmacy, University of Maryland,
Baltimore, MD 21202

During Operation Desert Shield 57 percent of U. S. Troops studied had at least one episode of diarrhea. In 50 percent of the cases studied a bacterial pathogen was found, with antibiotic resistant, enterotoxigenic E. coli (ETEC) and Shigella sonnei being the most common pathogenic species. The isolated ETEC were identified by their pilus Colony Factor Antigens (CFA's). 11% expressed CFA/I, 31% expressed CFA/II, 32% expressed CFA/IV, and 26% expressed none of these antigens.

The first step in pathogenesis is adherence to the small intestine epithelial cells by protein fimbrial (pilus) adhesins or fibrillar called Colonization Factor Antigens (CFA's). Three major CFAs have been recognized, CFA/I, CFA/II, and CFA/IV[1].

Ten human volunteers who were immunized orally twice weekly for 4 weeks with CFA/II developed a poor antibody response and did not show any significant protection when challenged with pathogenic ETEC[2]. This disappointing response

was attributed to adverse effects of gastric content, even at neutral pH, on fimbrial proteins[3]. When the vaccine was administered by inoculation directly into the duodenum, 4 of 5 immunized volunteers developed a significant rise in secretory IgA with CFA/II antibody[2].

D and L-lactic acid and glycolic acid, as homo- and copolymers, are biodegradable and permit slow and continued release of antigen with a resultant adjuvant activity. These polymers have been shown to be safe in a variety of applications in human beings and in animals[4-8]. Delivery of antigens via microspheres composed of biodegradable, biocompatible lactide/glycolide polymers[5-8] may enhance the mucosal response by protecting the antigen from digestion and targeting them to lymphoid cells in Peyer's patches[5-8]. McQueen et al[9] have shown that E. coli AF/R1 pili in PLGA microspheres, introduced intra-duodenally in rabbits, protected them against diarrhea and weight loss when challenged with the parent rabbit diarrheagenic strain of E. coli (RDEC-1). Only one vaccinated rabbit of six lost weight and only one had soft, pelleted stool. In contrast, all control unvaccinated animals became ill, lost weight, and shed soft pellets or unformed mucoid stool. Significant lymphocyte proliferation to AF/R1 from Peyer's patches and biliary IgA anti-AF/R1 antibody levels were seen.

In order to improve the CFA/II vaccine it was incorporated into PLGA microspheres under GMP in order to protect it from digestion and target it to the intestinal lymphoid system. The CFA/II BPM vaccine has undergone pre-clinical evaluation and has been found to be safe and immunogenic. Under Good Laboratory and Good Manufacturing Practices, E. coli. strain M424C1-06:H16 producing CFA/II were cultured in 75-80 CFA agar plates (24 x 24 cm) for 24 hr then harvested by scraping. The harvest was homogenized at slow speed for 30 min. in a Sorvall omni mixer with over head drive unit and cup immersed in an ice bath. The homogenate was centrifuged at 4° C at 16,500 x g for 30 min. The supernatant was saved the pellet rehomogenized and centrifuged with the supernatants pooled. The supernatant pool was centrifuged at 50,000 x g for 45 min. The resulting supernatant was treated with ammonium sulfate at 20% satuaration, stirred 30 min. at 4° C, than stored at 4° C for 16 hr then centrifuged at 19,700 x g for 30 min. The supernatant was saved and treated with ammonium sulfate at 45% saturation, stirred 30 min. at 4° C, stored at 4° C for 66-72 hr, then centrifuged at 19,700 x g for 45 min. The pellet was resuspended in about 100 ml of PBS containing 0.5% formalin, held at 22° for 18 hr, then dialyzed for 45-50 hr against PBS at 4° C using a total of 12 liters in 2 liter amounts. The dialysis was terminated when the PBS contained less then 0.03% formalin using Nessler's reagent and fuchsin sulfuric acid reagent. The final product, formalinized CFA/II, contained 1 mgm protein/ml PBS, was sterile

and passed the general safety test. It contained 90% CS3 and 10% CS1 by polyacrylamide SDS gel analysis[10]. The CFA/II was desalted and freeze dried.

Solvent extraction technique was used to encapsulate the freeze dried CFA/II into poly(lactide-co-glycolide) (Medisorb Techologies International, viscocity 0.73 dl/g) microspheres in the 1-10 um size range to achieve theoretical antigen loading of 1% by weight. The freeze dried antigen-sugar & matrix was dispersed in an acetonitrile solution of the polymer and then emulsified to achieve desired droplet size. Microspheres were solidified and recovered by using heptane as extracting solvent. Four microsphere batches were pooled and vacuum dried to remove traces of solvent.

Using a precalibrated eye-piece micrometer, the diameter of 150 randomly chosen microspheres was determined and the micropshere size distribution was calculated. The samples were examined with an Hitachi S-450 scanning electron microscope operated at 15-20 KV. Protein content was determined using the micro bicinchonic acid (BCA) method. Determinations were made of moisture, acetonitrile and heptane contents. Tests were conducted to determine release of CFA/II from microspheres and the general safety of the CFA/II microspheres.

Two naive rabbits were immunized with CFA/II microsphere vaccine at 25 ug protein in two different sites intra-muscularly on day 0. Sera were obtained from all animals before immunization on day 0 and days 7 and 14. The sera were tested by ELISA for IgG antibodies to CFA/II antigen and to individual coli surface (CS) proteins CS3 and CS1.

Five rabbits were vaccinated intra-duodenally with CFA/II microspheres containing either 25 or 50 ug of protein on days 0 and 7. The rabbits were sacrificed by chemical euthanasia at day 14 and subjected to complete necropsy. The Peyer's patches were removed and their cells suspended in a concentrated modified medium. The cells were placed in 2 ml of media at a concentration of 2.5 x 10^6 cells/ml for each well of a 24 well plate. These cells were challenged separately with BSA and the CFA/II antigen at doses of 500, 50 and 5 ng/ml in triplicate wells. The plates were incubated at 37° C with 5% CO_2. Spleen cells were obtained from immunized rabbits on day 14 following intra-duodenal immunization with CFA/II microsphere vaccine. The cells were placed in 96-well round bottom microculture plate at a final concentration of 6 x 10^5 cells/well and incubated for 0, 1, 2, 3, 4 and 5 days at 37° C with 5 CO_2. 96 well flat bottom microculture plates were coated with 3 ug/ml of CFA/II antigen overnight blocked with PBS with 0.05%

Polysorbate 60®.
On the harvest days, the cells were gently flushed out of the wells of the round bottom plates and transferred to the corresponding wells in the antigen coated, 96 well flat bottom microculture plates to be tested for the presence of antibody secreting cells using ELISPOT technique.

The results of size frequency analysis on CFA/II microsphere vaccine Lot L74F2 are shown in (Figure 1). The particle size distribution is plotted in % frequency against particle size in diameter (size) expressed in um. The average number fregency diameter is 4.5 um. The average volume frequency diameter is 9.4 um. The percent volume between diameters of 5-10 um is 63% and the percent volume less than 10 um diameter is 77%.

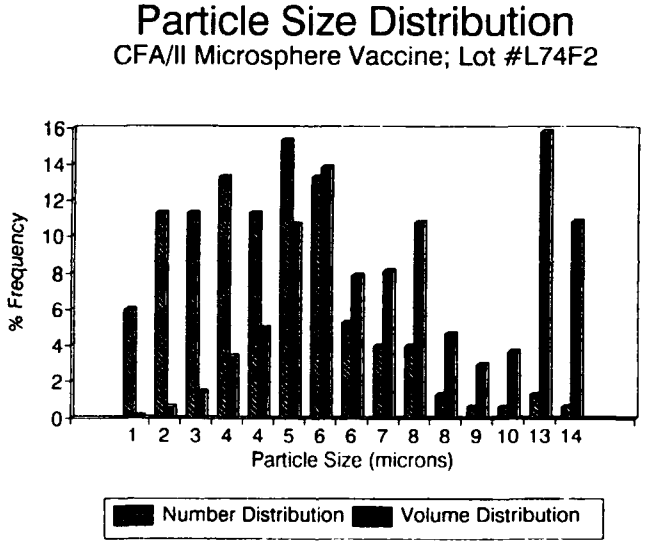


Figure 1

The microspheres of Lot L74F2 are seen in (Figure 2) which is a scanning electron photomicrograph. They appear as smooth, and non-porous, and nearly all the microspheres are less than 10 um as compared to the 5 um bar.

The protein loads of the four individual batches are the following: K62A8, 1.16% $\pm$ 0.10 SD; K63A8, 1.023% $\pm$ 0.17 SD; K64A8, 1.232% $\pm$ 0.13 SD; and K65A8, 0.966% $\pm$ 0.128 SD. The mean average protein load is 1.16% $\pm$ 0.15 SD. The protein load of the CFA/II microsphere vaccine in the final dose vial is the

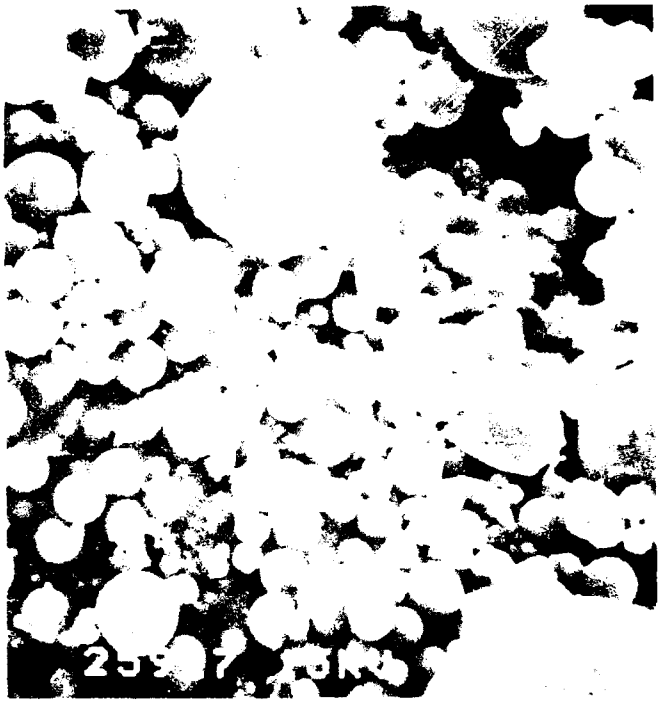


Figure 2

following: Lot L74F2, 1.175% ± 0.17 SD.

The CFA/II microsphere vaccine (Lot 74F2) percent water content was found using the Karl Fischer titrimeter method to be 2.154% using triplicate samples.

The acetonitrile residuals of the 4 individual CFA/II microsphere batches are the following: K62A8, <0.1%; K62A8, <0.1%, K64A8, <0.1%; and K65A8, <0.1%. The acetonitrile residual of the CFA/II microsphere vaccine in the final dose vial is the following: Lot L74F2, 0.07 ± 0.05%. The heptane residual of the 4 individual CFA/II microsphere batches are the following:K62A8, 1.9%; K63A8, 1.4%; K64A8, 1.6% and K65A8, 1.6%. Following pooling in heptane and subsequent drying, the heptane residual of the CFA/II microsphere vaccine in the final dose vial is the following: Lot L74F2, 1.6 ± 0.1%.

Twenty two colonies grew on the blood agar plates after 48 hr of culture. 21 were identified as coagulase negative staphlycocci and 1 as a micrococus species, all of which are considered to be nonpathogenic to humans. Three fungal colonies grew on the fungal culture plates and each was identified as A. glaucus.

The results of the release study on Lot L74F2 are shown in (Figure 3). The

results are plotted as percent release of CFA/II against time in hr for three samples: A, 33.12 mgm; B, 29.50 mgm; C, 24.20 mgm at 1, 3, 6, 8, 12, and 22 hr intervals. An average of 8% of CFA/II is released at 1 hr rising to 20% at 8 hr then a slower release to 25% at 22 hr. The average represents the mean ± 1 SD.

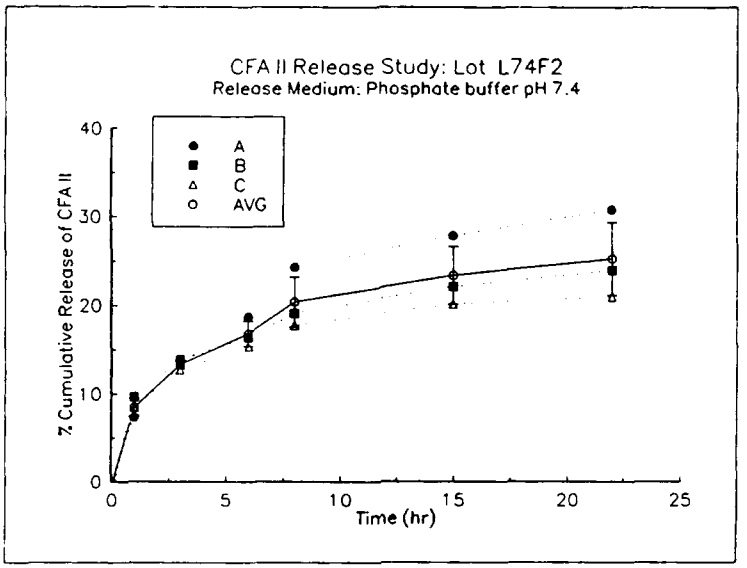


Figure 3

None of the animals displayed any signs of toxicity for 7 days. The mice gained an average of 2.3 g and the guinea pigs gained an average of 43 g.

The serum IgG antibody titers to CFA/II microsphere vaccine Lot L74F2, CS1, and CS3 proteins following (25 ug) protein IM immunization on day 0 in two rabbits were determined using ELISA and the results seen in (Figure 4). The results were determined on serial dilutions of sera and are expressed as mean antibody titers against the different antigens at 0, 7 and 14 days. Antibody titers greater than 1000 were seen at 7 days to both CS1 and CS3 protein which rose to greater than 10,000 by day 14. Rabbit 109 developed an antibody titer to CFA/II of 1,000 by day 7 rising to 3,000 by day 14. Rabbit 108 had a log higher rise at day 7 and 2 log higher rise at day 14 being 3×10^4 at day 7 going to 1×10^5 at day 14 (data not shown).

The lymphocyte proliferative responses for Peyer's patch cells of rabbits 65, 66, 83, 85, and 86 immunized intraduodenally with 50 ug protein of CFA/II

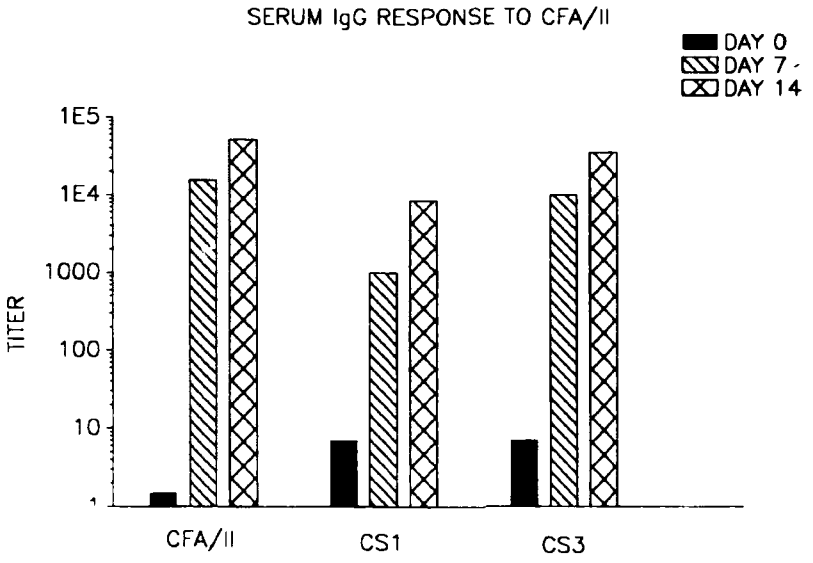


Figure 4

microsphere vaccine 14 and 7 days earlier are seen in (Figure 5, next page). The cells are challenged in vitro with CFA/II or BSA at 500, 50, or 5 ng/ml or media in triplicate. The lymphocyte transformation was determined by tritiated thymidine incorporation. The uptake of tritiated thymidine in kcpm is expressed as mean ± 1 SD. The results are expressed as kcpm against antigen dose. No response to BSA or media control is seen in any of the five rabbits. All rabbits responded by lymphocyte transformation in a dose dependent manner to the CFA/II. The highest dose responses were 3-10X's the media control and are highly significant with a p-value of <0.006. Using the paired student t-test, the p values of 500 ug/ml dose of CFA/II compared to media control are: 65, $p = 0.002$; 66, $p = 0.002$; 83, $p = 0.006$; 85, $p = 0.002$; and 86, $p = 0.0002$.

The lymphocyte proliferative responses from Peyer's patch cells of rabbits 77, 78, 80, 88, and 91 immunized intraduodenally with 25 ug protein of CFA/II microspheres vaccine 14 and 7 days earlier is seen in (Figure 6, next page). The cells are again challenged in vitro with CFA/II or BSA at 500, 50, or 5 ng/ml or media in tripiicate. The lymphocyte transformation was determined by tritiated thymidine incorporation. The uptake of tritiated thymidine in kcpm is expressed as mean ± 1 SD. Rabbit 80 gave no response probably due to a poor Peyer's patch cell population which did not respond well to Concanavalin A mitogenic stimulation either. The remaining 4 rabbits gave positive responses with the high CFA/II dose response being 2-8x media control and highly significant with p-values of <0.009.

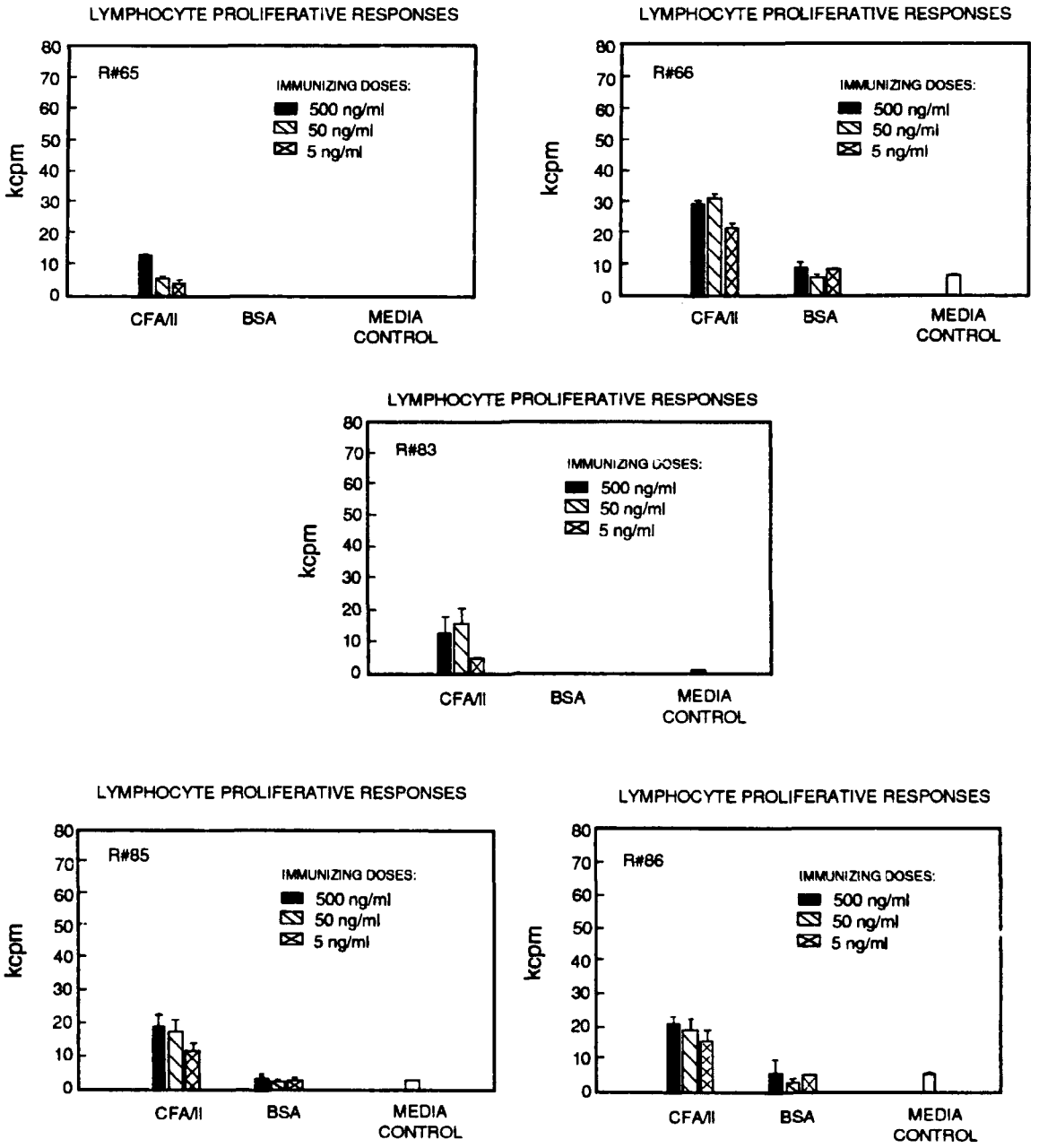

LYMPHOCYTE PROLIFERATIVE RESPONSES
R#65
IMMUNIZING DOSES:
500 ng/ml
50 ng/ml
5 ng/ml
kcpm
CFA/II
BSA
MEDIA CONTROL
LYMPHOCYTE PROLIFERATIVE RESPONSES
R#66
IMMUNIZING DOSES:
500 ng/ml
50 ng/ml
5 ng/ml
kcpm
CFA/II
BSA
MEDIA CONTROL
LYMPHOCYTE PROLIFERATIVE RESPONSES
R#83
IMMUNIZING DOSES:
500 ng/ml
50 ng/ml
5 ng/ml
kcpm
CFA/II
BSA
MEDIA CONTROL
LYMPHOCYTE PROLIFERATIVE RESPONSES
R#85
IMMUNIZING DOSES:
500 ng/ml
50 ng/ml
5 ng/ml
kcpm
CFA/II
BSA
MEDIA CONTROL
LYMPHOCYTE PROLIFERATIVE RESPONSES
R#86
IMMUNIZING DOSES:
500 ng/ml
50 ng/ml
5 ng/ml
kcpm
CFA/II
BSA
MEDIA CONTROL

Figure 5

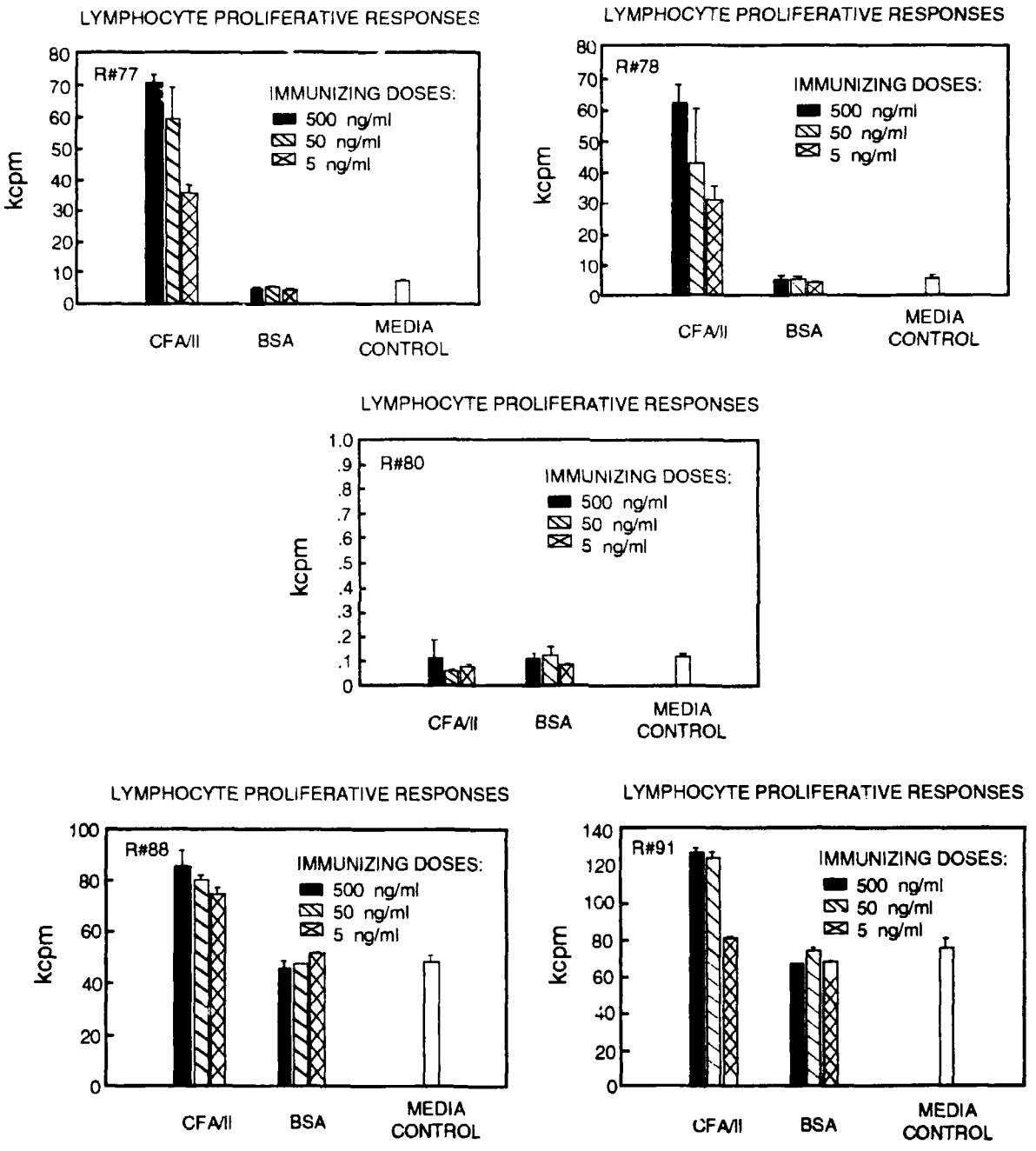

LYMPHOCYTE PROLIFERATIVE RESPONSES
R#77
IMMUNIZING DOSES:
500 ng/ml
50 ng/ml
5 ng/ml
kcpm
CFA/II
BSA
MEDIA CONTROL
LYMPHOCYTE PROLIFERATIVE RESPONSES
R#78
IMMUNIZING DOSES:
500 ng/ml
50 ng/ml
5 ng/ml
kcpm
CFA/II
BSA
MEDIA CONTROL
LYMPHOCYTE PROLIFERATIVE RESPONSES
R#80
IMMUNIZING DOSES:
500 ng/ml
50 ng/ml
5 ng/ml
kcpm
CFA/II
BSA
MEDIA CONTROL
LYMPHOCYTE PROLIFERATIVE RESPONSES
R#88
IMMUNIZING DOSES:
500 ng/ml
50 ng/ml
5 ng/ml
kcpm
CFA/II
BSA
MEDIA CONTROL
LYMPHOCYTE PROLIFERATIVE RESPONSES
R#91
IMMUNIZING DOSES:
500 ng/ml
50 ng/ml
5 ng/ml
kcpm
CFA/II
BSA
MEDIA CONTROL

Figure 6

Again no response were seen to BSA as compared to the media control. Using the paired student t-test, the p values of 500 ng/ml dose of CFA/II compared to media control are: 77, $p = 0.001$; 78, $p = 0.0015$; 80, p = insignificant; 88, $p = 0.0093$; and 91, $p = 0.001$.

The results of the ELISPOT assay of spleen cells from rabbits 65, 66, 83, 85, and 86 immunized intraduodenally with 50 ug protein of CFA/II microsphere vaccine 14 and 7 days earlier are seen in (Figure 7, next page). These cells were placed into microculture and tested on days 0, 1, 2, 3, 4, and 5 by ELISPOT for cells secreting antibodies specific for CFA/II antigen. The results are expressed as # of antibody secreting cells per 9×10^6 spleen cell against culture days. Positive responses were seen in all 5 rabbits on days 2-5. Days of maximum responses occurred on day 3 for rabbits 65 and 66; day 4 for rabbit 85; and day 5 for rabbits 83 and 86. The responses are highly significant being 7-115 times higher then the 1-2 cells seen on all days in 4 control rabbits (67, 69, 72, 89) (data not shown).

A consistent finding in the spleens of all rabbits both the 25 and 50 ug protein dose groups was minimal to mild diffuse lymphocytic hyperplasia the periarteriolar lymphatic sheaths (T cell dependent). 2/5 rabbits of the 50 ug dose group (83 and 86) also had mild lymphocytic hyperplasia of splenic follicular (B cell dependent) areas. The three rabbits in an untreated control group had histologically normal spleens.

Reactive hyperplasia of mesenteric lymph nodes was often seen in vaccinated rabbits. Two of five rabbits in the 25 ug human dose equivalent group (88 and 91) and 4/5 rabbits in the 50 ug dose group (65, 66, 83, 86) had diffuse, minimal to mild lymphocytic hyperplasia of paracortical (T cell dependent) areas of the mesenteric lymph nodes. Two rabbits in the 50 ug dose group (83 and 86) also had minimal to mild lymphocytic hyperplasia of cortical follicular (B cell-dependent) areas. The mesenteric lymph nodes of the other vaccinated rabbits and of the untreated control rabbits were within normal limits. Incidental or background lesions found in one or more rabbits of all three groups were acute minimal to mild pnuemonia and foreign body microgranulomas of the cecal gut-associated lymphoid tissue.

McQueen et al[9] have found that the AF/R1 adhesin of rabbit diarrheagenic Escherichia coli (RDEC-1) incorporated into biodegrable microspheres could function as a safe and effective oral intestinal vaccine in the rabbit diarrhea model. The AF/R1 was incorporated into poly (D,L-lactide-co-glycolide) microspheres and administered intraduodenally. Jarboe et al[11] reported that Peyer's patch cells obtained from rabbits immunized intra-duodenually with AF/R1 in microspheres

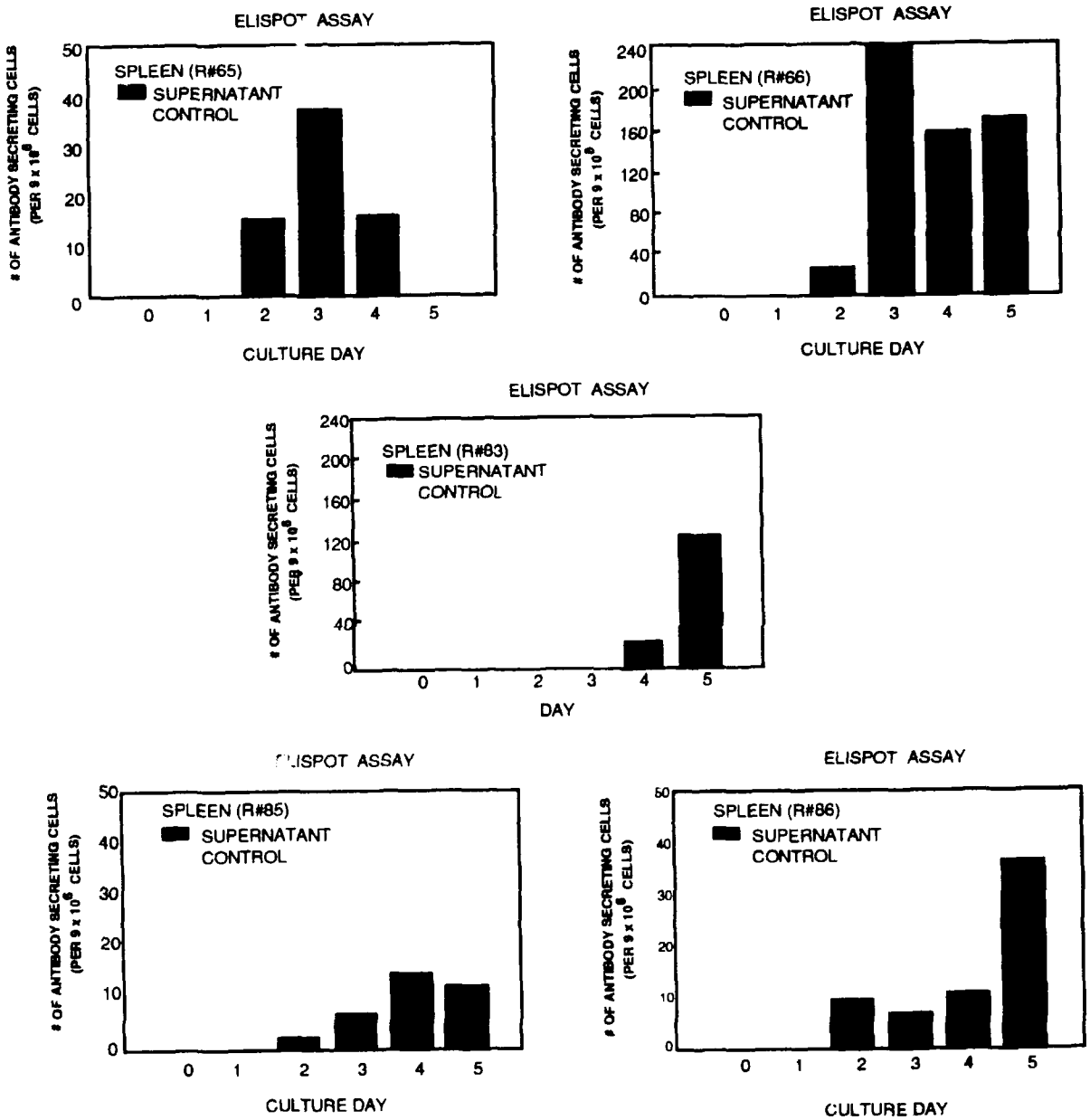

ELISPOT ASSAY
SPLEEN (R#65)
SUPERNATANT
CONTROL
OF ANTIBODY SECRETING CELLS (PER 9 x 10^6 CELLS)
CULTURE DAY
ELISPOT ASSAY
SPLEEN (R#66)
SUPERNATANT
CONTROL
OF ANTIBODY SECRETING CELLS (PER 9 x 10^6 CELLS)
CULTURE DAY
ELISPOT ASSAY
SPLEEN (R#83)
SUPERNATANT
CONTROL
OF ANTIBODY SECRETING CELLS (PER 9 x 10^6 CELLS)
DAY
ELISPOT ASSAY
SPLEEN (R#85)
SUPERNATANT
CONTROL
OF ANTIBODY SECRETING CELLS (PER 9 x 10^6 CELLS)
CULTURE DAY
ELISPOT ASSAY
SPLEEN (R#86)
SUPERNATANT
CONTROL
OF ANTIBODY SECRETING CELLS (PER 9 x 10^6 CELLS)
CULTURE DAY

Figure 7

responded with lymphocte proliferation upon in vitro challenge with AF/R1. This early response at 14 days gave a clear indication as to the immunogenicity of E. coli pili contained within the polymer microspheres.

CFA/II from enterotoxigenic E. coli, given as an oral vaccine was found to be ineffective, as the fimbrial and fibrillar proteins were found to be rapidly degraded when treated with conditions mimicking those contained in the stomach[3]. The CFA/II was found to be immunogenic when given in high doses intra-intestinally producing intestinal secretory IgA antibodies[2].

The CFA/II vaccine has been incorporated into poly(D,L lactide-co-glycolide) microspheres under Good Manufacturing Practices and tested under Good Laboratory Practices. The microspheres are spherical, smooth surfaced and without pores. The majority (63%) are between 5-10 um in diameter by volume, a size range which has been suggested to promote localization within the Peyer's patch in mice and perhaps enhance local immunization[5-8]. The protein content is 1.174%, and 0.62% was the core loading of the AF/R1 microspheres which were effective.

The organic residuals are at a safe level, with the heptane exposure being 1.7 mgm per vaccine dose and the acetonitrile exposure being 0.1 mgm per vaccine dose. This is compared to the occupational maximum allowable exposure of 1800 mgm/15 min for heptane and the human oral TDLO (any non-lethal toxicity) of 570 mgm\Kg for acetonitrile.

The CFA/II protein was produced under aseptic conditions, however the incorporation of protein into microspheres and microsphere pooling were done in a clean room, allowing entrance of microbes. The World Health Organization (WHO) Requirements of Typhoid Vaccine (Live Atttenuated, Ty2la oral) allow two hundred non-pathogenic bacteria as well as 20 fungi per dose. The CFA/II microsphere vaccine has only 22 non-pathogenic bacteria and 3 fungi per dose.

A general safety test (from the World Health Organization Requirements of Typhoid Vaccine; Live, attenuated, Ty21a oral) was performed in which the microsphere vaccine was given by gastric lavage to guinea pigs. Both mice and both guinea pigs demonstrated no toxicity & gained weight over the 7 day test clearly indiciating the innocuous nature of this vaccine by passing this safety test.

The CFA/II microsphere vaccine (Lot 74F2) is immunogenic giving high titer serum IgG antibody responses as early as 7 days following intra-muscular injection in rabbits. This test will be used as potency test for future lots of the CFA/II

microsphere vaccine.

The CFA/II microsphere vaccine was also immunogenic following intra-duodenal administration to rabbits. The highest lymphocyte proliferative responses from Peyer's patch cells were seen with the lower 25 ug dose. This is the human equivalent dose and suggests that higher doses of antigen in polymer microspheres may attenuate this immunological reponse.

The antibody secreting B-cells demonstrated in the rabbit spleen at 14 days is a clear indication that B-cells have been immunized. They may represent resident B-cells immunized in the spleen or B-cells immunized at the level of the Peyer's patches and are migrating through the spleen to return to the intestinal mucosal lamina propria[12]. The delay of several days before secreted antibody is detected suggests either maturation is required of the B-cells or that down regulation may be present initially and lost with time in culture.

Further evidence of immunization by the CFA/II microsphere vaccine given intra-duodenually is demonstrated by the lymphocytic hyperplasia in the spleen particularly in the T cell dependent areas in all vaccinated rabbits. On the other hand, greater T cell dependent area lymphocytic hyperplasia in the mesenteric lymph nodes were seen in rabbits receiving the higher 50 ug dose, 4/5 compared to 2/5 with the lower dose, indicating a dose response. These changes are most likely due to the vaccine since no similar changes were seen in three untreated control rabbits. Also no abnormal pathological changes attributable to the vaccine were seen.

The CFA/II BPM vaccine has undergone pre-clinical evaluation and has been found safe and immunogenic. This vaccine is ready for clinical Phase I safety testing following the Food and Drug Administration's approval of the Notice of Claimed Exemption for the IND.

REFERENCES

1. J.B. Kaper and M.M. Levine, *Vaccine* **6**, 197-199 (1988).
2. M. Levine et al. in *Molecular Biology of Microbial Pathogenicity, Protein-Carbohydrate Interactions in Biological System*, D. Lark et. al., Ed. (Academic Press, London, 1986), 143-145.
3. M. Schmidt, E.P. Kelly, L.-Y. Tseng, and E.C. Boedeker, *Gastroenterology* **88**, A1575 (1985).
4. D.L. Wise, T.D. Fellmann, J.E. Sanderson and R.R. Wentworth, in *Drug Carriers in Biology and Medicine*, G. Gregoriades, Ed. (Academic Press, London, 1979), 237-270.
5. J.H. Eldridge, R.M. Gilley, J.K. Staas, Z. Moldoveanu, J.A. Meulbroek and T.R. Tice, *Curr. Top. Microbiol. Immunol.* **146**, 59-66 (1989).
6. J.H. Eldridge, C.J. Hammond, J.A. Meubroek, J.K. Staas, R.M. Gilley and T.R. Tice, *J. Controlled release* **11**, 205 (1989).
7. J.H. Eldridge, J.K. Staas, J.A. Meubroek, J.R. McGhee, T.R. Tice and R.M. Gilley, *Mol. Immunol* **28**, 287-294 (1991).
8. Z. Moldoveanu, J.K. Staas, R.M. Gilley, R. Ray, R.W. Compans, J.H. Eldridge, T.R. Tice and J. Mestecky, *Curr. Top. Microbiol. Immunol.* **146**, 91-99 (1989).
9. C.E. McQueen, E.C. Boedeker, R.H. Reid, D. Jarboe, M. Wolf, M. Le and W.R. Brown, *Vaccine* (in press) (1992).
10. M.M. Levine, P. Ristaino, G. Morley, C. Smyth, S. Knutton, E. Boedeker, R. Black, C. Young, M.L. Clements, C. Cheney and R. Patnaik, *Infect. Immun.* **44**, 409-420 (1984).
11. D. Jarboe, R. Reid, C. McQueen and E. Boedeker, *Abstracts of the Annual Meeting of the American Society of Microbiology*, 121, (May 1990).
12. J. Tseng, *J. Immunol.* **132**, 2730-2735 (1984).

Kinetic Model for Predicting Bacterial Destruction from Intrinsic Chemical Marker Formation in Thermally Processed Foods

Ross, Edward, Dr., *Kim, Hie-Joon, Dr., Taub, Irwin, Dr.,
Kustin, Kenneth, Dr., and Shattuck, Edgar, Mr.
U.S. Army Natick Research, Development and Engineering Center
Natick, Massachusetts 01760-5018

INTRODUCTION

The safety of an operational ration component made shelf stable by thermoprocessing must be validated to the satisfaction of regulatory agencies, the Food and Drug Administration (FDA) and the United States Department of Agriculture (USDA). For conventional thermoprocessing involving batch loads of sealed containers in retorts, the validation is based on temperature measurements at the slowest heating point in the food, which can be converted into a factor corresponding to the number of log cycles of pathogenic Clostridium botulinum destroyed. The minimum required destruction is 12 log cycles. For the new thermoprocessing technologies, in which particulate foods flow continuously through the processing system until aseptically packaged in sterile containers, no temperature measurement can be made. Validation without adequate knowledge of thermal history would require the cumbersome use of a testor microorganism imbedded in the food and the subsequent determination of the level of destruction attained. Conceptually, if a chemical entity is produced in the food that correlates well with bacterial destruction, it should be possible to use the food itself as an indicator of the thermal history at the center of the fastest moving particulate, and thereby provide the validation required. As the following discussion will show, such intrinsic chemical markers have been found, the yield of which, when combined with a mathematical model correlating bacterial destruction to marker formation, can be used to validate the sterility of aseptically processed particulates.

MARKER DETECTION AND CHARACTERIZATION

One such marker detected in broccoli is shown in a three-dimensional representation in Fig. 1. It was obtained by extracting the heat-processed broccoli with an aqueous solution and then injecting the extract into an anion exclusion chromatographic column. As a consequence,

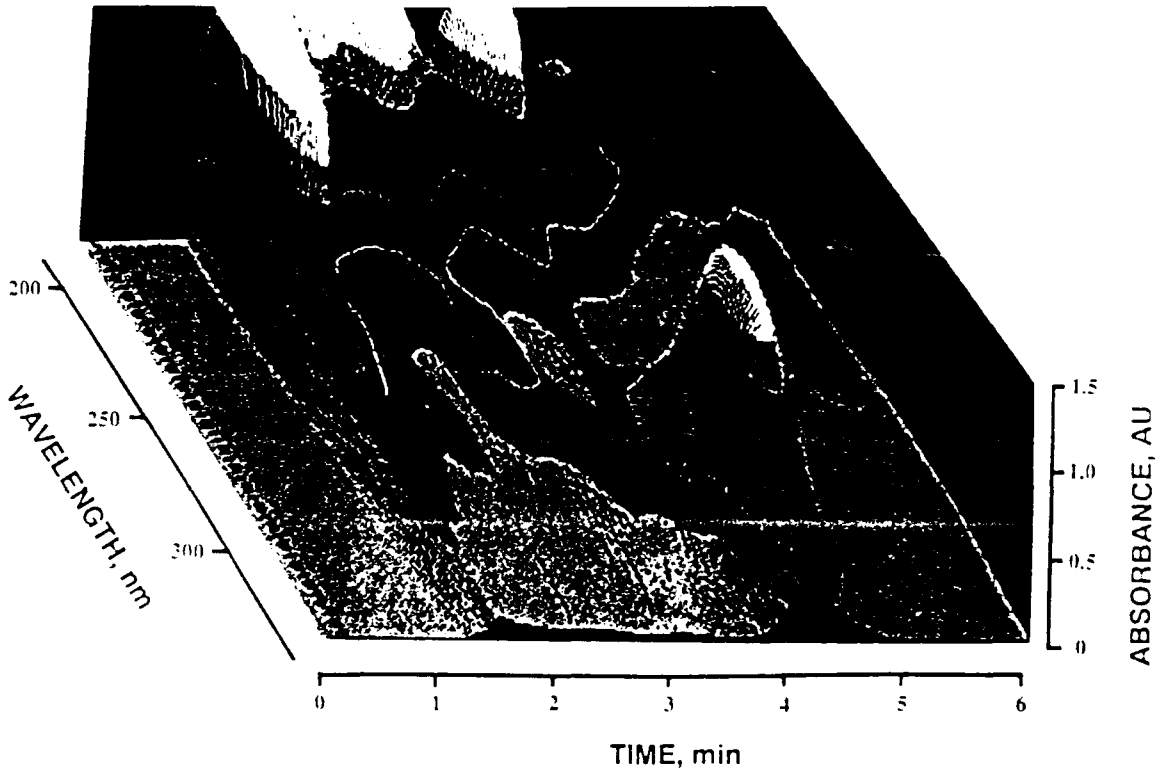


Fig. 1 Three-dimensional representation of a spectrochromatogram for partially purified marker, M-1, obtained from heated broccoli

the extracted components are separated into a distinctive order of elution and pass through a photodiode array detector. The detector takes an optical absorption spectrum of the eluted components every two seconds, the wavelength ranging in this case from 200 to 350 nm. This representation then defines a time-wavelength domain in the X-Y plane; the intensity of the absorption appears in the Z direction. The elution of the marker, M-1 in this case, occurs at 4 min, and the peak of its optical absorption spectrum is at 298 nm; both elution time and peak wavelength characterize the marker.

A two-dimensional representation, namely a contour map, is shown in Fig. 2 for beef samples. The upper portion is for the unheated control sample. In the time-wavelength domain out to 8 minutes, there are no peaks of interest. The lower portion is for the heat-processed beef sample. Marker M-1 appears at 4 min, with a 298 nm absorption maximum, and another marker, M-2, that is specifically associated with meats, appears at 5.5 min, with an absorption maximum of 285 nm.

Markers have also been found in fruits. If an extract of orange juice is heated, M-1 and another compound, M-3, eluting at 7.5 min, appear. Experiments have been done to show that the precursor of M-1 and M-3 is fructose. The identities of M-1 and M-3 have been established using combined gas chromatography and mass spectrometry (GC/MS). The mass spectrum of M-1 shows that it has a molecular weight of 144 and a cracking pattern that matches what exists in the literature for 2,3-dihydro-3,5-dihydroxy-6-methyl-4(H)-pyran-4-one.[1] M-3 is 5-hydroxymethylfurfural (molecular weight of 126). The identity of M-2 is not yet positively established.

To use these markers in demonstrating sterility, we must know their kinetic characteristics precisely. These were determined for M-1 by taking a broccoli extract, heating it at 121°C (250°F), and tracking the isothermal formation of M-1; it rises quite rapidly at the beginning, then begins to trail off, eventually reaching a plateau or limiting value (Fig. 3). Since there appears to be a limit to the amount of the precursor that can convert to the marker, that plateau is defined as the limiting marker yield at time infinity, or M_{∞}.

A plot of the logarithm of the marker yield as a function of time gives a straight line, indicating that this conversion is a first order process. If the formation is tracked at several different temperatures and if the logarithm of the observed rate constants is plotted against reciprocal temperature, a straight line is also obtained, indicating that the formation of this marker follows Arrhenius kinetics. The activation energy for the reaction can be obtained from the slope of this plot, which for M-1 is 23.7 kilocalories per mole.

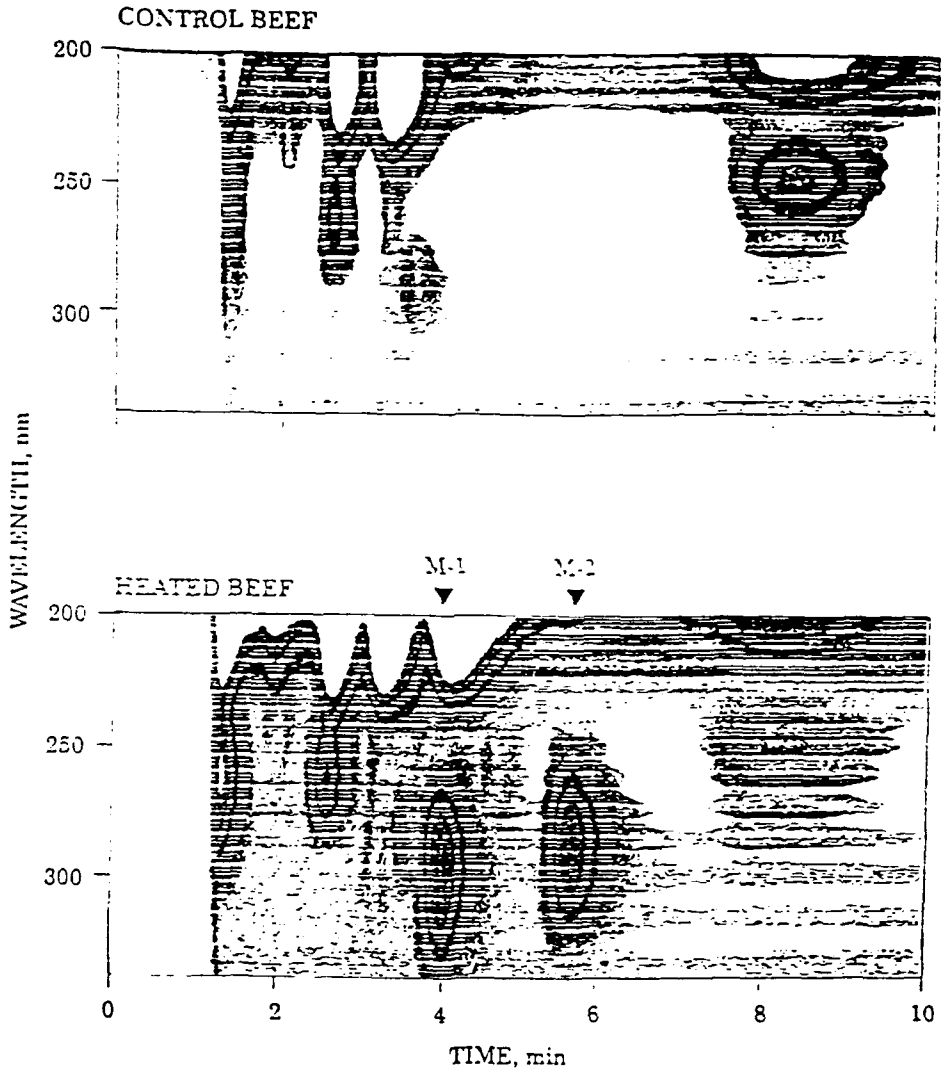


Fig. 2 Contour diagram for a control and a heated beef sample showing formation of two markers, M-1 and M-2, upon heating

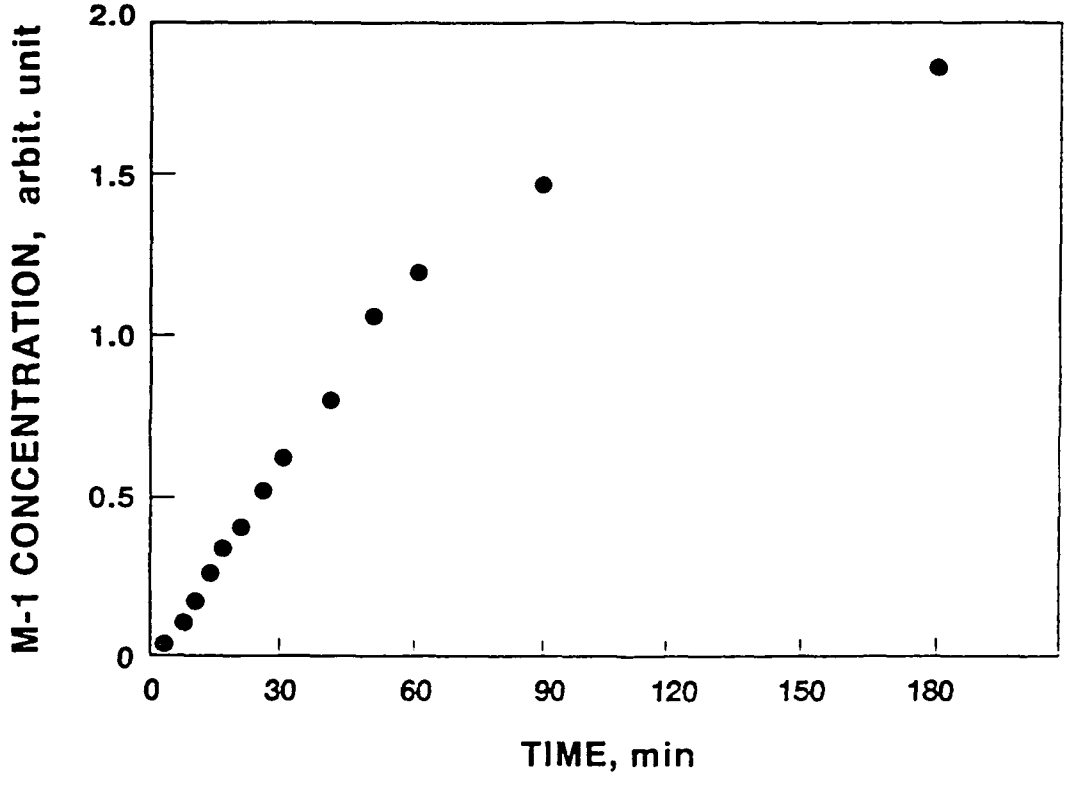


Fig. 3 Kinetic plot for isothermal formation of M-1 in broccoli at 121°C

MODEL RELATING BACTERIAL DESTRUCTION TO MARKER FORMATION

To use this information in a practical way requires that we mathematically relate the destruction of food-contaminating bacteria to the formation of intrinsic markers in the same system.[2] To understand the relationship, three points are emphasized. First, both the precursor of the marker and the bacterium of interest can be considered to be collocated and, therefore, experience the same thermal exposure. Secondly, since both kinetic processes are first order, we need only consider the fraction of convertible precursor converted to the marker, namely M/M_∞, and the fraction of the total bacterial population remaining at any time, namely N/N_o. And thirdly, since the rates of bacterial destruction and marker formation are temperature-dependent, the time-temperature profile influences these processes and the cumulative effects must be found by integrating over this profile. The final predictive equation relates log N_o/N to (M/M_∞) and involves several key kinetic parameters and the peak temperature in the profile. The derivation of the model is outlined briefly, but will be detailed elsewhere.

We know that N(t), the number of microorganisms remaining at time, t, behaves as a first-order process, so

$$N(t_e) = N_o\, e^{-\sigma(t_e)} \qquad (1)$$

$$\sigma(t_e) = \int_0^{t_e} k[T(t)]\, dt = \ln[N_o/N(t_e)]. \qquad (2)$$

k, the rate constant, depends on absolute temperature, T, and so on t. $N = N_o$ at $t = 0$. The dependence of the rate constant on temperature is assumed to be given by the Arrhenius relationship

$$k[T(t)] = k_o\, e^{[-E_a/\{R\, T(t)\}]},$$

where E_a is the activation energy, R the universal gas constant, and k_o the pre-exponential or frequency factor.

These relations are put in dimensionless form by introducing a reference temperature, T_r, and defining

$$E = E_a/(R\, T_r) \qquad (3)$$

$$x = T_r/T - 1 \quad \text{or} \quad T = T_r/(1 + x). \qquad (4)$$

E is a dimensionless activation energy, and x is a dimensionless inverse transformation of T, i.e., x decreases as T increases. Then

$$E_a/(RT) = [E_a/(RT_r)][T_r/T] = E(1+x),$$

and the Arrhenius relationship and Eqns (3) and (4) lead to

$$k = k_r\, e^{-Ex},\ k_r = k_o\, e^{-E}. \qquad (5)$$

Eqn (5) implies that k_r is the rate constant at the temperature, T_r. In dimensionless terms, Eqns (1) and (2) become

$$\sigma(t_e) = k_r\, F(t_e) \qquad (6)$$

$$F(t_e) = \int_0^{t_e} e^{-Ex(t)}\, dt\ . \qquad (7)$$

We assume that M(t), the marker concentration at any time, t, obeys first-order formation kinetics, so

$$M(t_e) = M_\infty\{1 - e^{-\alpha(t_e)}\},\ \text{where}$$

$$\alpha(t_e) = \int_0^{t_e} k_m[T(t)]\, dt = -\ln[1-M(t_e)/M_\infty]. \qquad (8)$$

M_∞ is the limiting marker concentration attained at maximum conversion of the precursor, and k_m is the rate constant for marker formation. The rate constant is also assumed to follow the Arrhenius relationship with activation energy E_{am} and pre-exponential constant k_{mo}. The dimensionless transformation of T for M is the same for N. k_{mr} is the rate constant when $T = T_r$, and

$$E_m = E_{am}/(R\, T_r).$$

By analogy with the formulas for N, we obtain from Eqn (7)

$$k_m = k_{mr}e^{-E_m x};\quad k_{mr} = k_{mo}e^{-E_m}$$

$$\alpha(t_e) = k_{mr}F_m(t_e) \qquad (9)$$

$$F_m(t_e) = \int_0^{t_e} e^{-E_m x(t)}\, dt. \qquad (10)$$

If the temperature pulse is quadratic in t, then we can calculate $F(t_e)$ exactly. The pulse is taken as symmetric with duration $t_e = 2\ t_p$, and the transformation of T into x gives

$$x(t) = x_p + (x_o - x_p)[(t-t_p)/t_p]^2.$$

x is parabolic, starting and ending with $x = x_o$; the peak temperature (minimum x) is attained for $x = x_p$ when $t = t_p$. Inserting this pulse-form in Eqn (7) gives

$$F(t_e) = t_e\ (C/g)\ \mathrm{erf}(g)\ e^{-Ex_p}, \text{ where}$$

$$C = \sqrt{\pi}/2,\ g = \sqrt{[E(x_o - x_p)]}.$$

It is usual that $g > 1.4$, in which case erf(g) is nearly 1,[3] so

$$F(t_e) = (C/g)\ t_e\ e^{-Ex_p}.$$

The estimates of $\sigma(t_e)$ and $\alpha(t_e)$ are found by combining Eqns (6) through (10) to obtain

$$\sigma(t_e) = k_r(C/g)\ t_e\ e^{-Ex_p} \qquad (11)$$

$$g = \sqrt{[E(x_o - x_p]}$$

$$\alpha(t_e) = k_{rm}(C/g_m)\ t_e\ e^{-E_m x_p} \qquad (12)$$

$$g_m = \sqrt{[E_m(x_o - x_p]}.$$

Dividing Eqns (11) and (12), we find

$$\sigma/\alpha = (k_r/k_{rm})\sqrt{(E_m/E)}\ e^{[(E_m - E)x_p]}. \qquad (13)$$

If x_p is known, Eqn (13) allows the calculation of σ from measurements of α. If x_p is not known, it can be found from measurements on two (or more) different markers, designated by subscripts m and q. Both markers receive the same temperature-pulse as the microorganism, and so obey the same formulas except for parameter-values. If Q is the concentration of the extra marker and Q_∞ its limiting concentration,

$$\alpha_q = -\ln[1 - (Q/Q_\infty)]$$
$$= k_{qr}(C/g)t_e\, e^{-E_q x_p} \qquad (14)$$
$$g_q = \sqrt{[E_q(x_o - x_p)]}$$

Eqns (12) and (14) can be solved for x_p,

$$x_p = \ln\{(\alpha_m/\alpha_q)(k_{qr}/k_{mr})\sqrt{(E_m/E_q)}\}/(E_q - E_m), \quad (15)$$

and σ can then be found from Eqn (13).

To illustrate these calculations, they are performed for two microorganisms, Clostridium botulinum (k_r=10.96 min^{-1}, E = 88.5) and B. stearothermophilus (k_r =1.1515 min^{-1}, E = 73.3), using T_r = 394.1° K and R = 1.987 cal/mol/(deg K). The M-1 marker has k_r = .0151 min^{-1}, E = 30.3. For Clostridium botulinum and the marker, and for B. stearothermophilus and the marker, Eqn (13) leads to

$$\log_{10}(N_o/N) = 184.4\, e^{-58.2x_p[-\ln\{1-(M/M_\infty)\}]}$$
$$\log_{10}(N_o/N) = 21.29\, e^{-43.0x_p[-\ln\{1-(M/M_\infty)\}]},$$

respectively. These relationships are graphed in Fig. 4 and 5 for various peak temperatures. Fig. 4 shows that at 250°F marker measurements in the range $0 < M/M_\infty < .15$ furnish estimates of N_o/N in the range $0 < \log_{10}(N_o/N) < 30$ for Clostridium botulinum. The marker is easily measurable in this range, and the log cycle reduction in the microorganism can be calculated to demonstrate that the food had been adequately processed to assure safety.

EXPERIMENTAL TEST OF THE MODEL

The validity of the model was tested using microbiological data from a simulated aseptic processing procedure. Spores of the thermophilic bacterium, B. stearothermophilus, are embedded in alginate beads,[4] placed at the center of a particulate food, and subjected to a thermal pulse. Surviving bacteria are counted and their decadic reduction was related to the marker concentration determined in the vicinity of the bacteria.

These microbiological experiments have several limitations. First, there is a limit to the number of spores (approximately 10 million) that can be loaded at the center of the food particulate. Second, the size of the alginate beads is difficult to reproduce, and these beads are cured before being exposed to high temperature; therefore, the initial bacterial

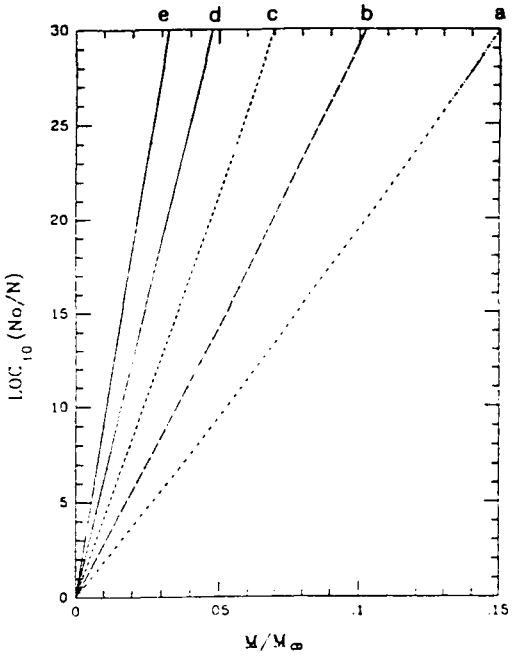


Fig. 4 Relation between $Log_{10}(N_o/N)$ for C. botulinum and M/M_∞ at several peak temperatures:
(a) -·-·-, 250°F;
(b) — —, 255°F;
(c) ——, 260°F;
(d) ·····, 265°F;
(e) ———, 270°F.

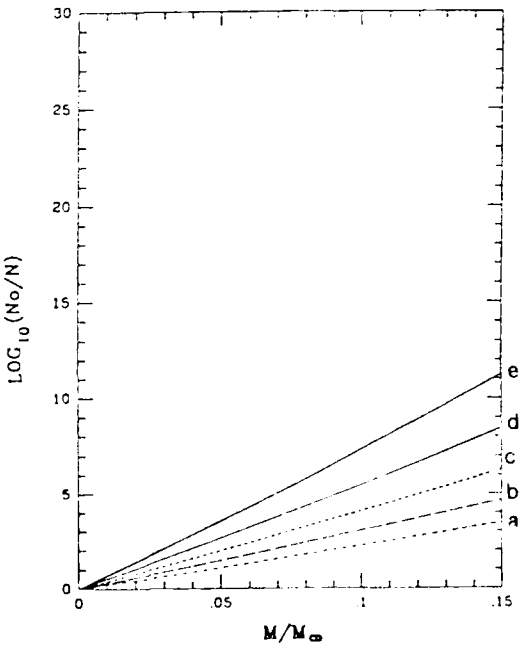


Fig. 5 Relation between $Log_{10}(N_o/N)$ for B. Stearothermophilus and M/M_∞ at several peak temperatures:
(a) -·-·-, 250°F;
(b) — —, 255°F;
(c) -----, 260°F;
(d) ·····, 265°F;
(e) ———, 270°F.

counts vary significantly, by as much as 100%. Third, since the spores are diluted while harvesting and culturing, it is difficult to count reliably less than 10 colony forming units related to the surviving spores. As a result, six decadic reduction in B. stearothermophilus is the maximum reduction one can hope to observe.

Despite these limitations, encouraging preliminary results were obtained. Alginate beads containing approximately $8x10^6$ spores were placed at the center of ham cubes of different sizes (1", 3/4", and 1/2"). Ten ham cubes of the same size were heated in a chamber with pressurized steam until 131°C was reached at the center of the cubes, as measured by a thermocouple in one of them. As expected, it took much longer for the larger cube to reach the same temperature at the center. For example, it took 1.5 min for the center temperature of a 1/2" cube to increase from 100°C to 131°C; the same temperature increase took 6 min for a 1" cube. The temperature profile was a good approximation of a quadratic function. When the final temperature was reached, the chamber was rapidly evacuated and the center temperature dropped to about 100°C instantly. The alginate bead was recovered and the surviving spore population was counted. The ham cubes were cut into three portions (surface, intermediate, and center). The marker, M-1, was extracted with 5-fold excess water and its concentration was determined by anion exclusion chromatography with photodiode array detection. The ham sample was also heated in an oven at 131°C until a plateau value was obtained.

In the 1/2" ham cube, M-1 was observed at a low concentration corresponding to a M/M_∞ value of approximately 0.05. The survivor population varied considerably; however, a decadic reduction of approximately one was observed on the average as shown in Fig. 6. A much higher value of M/M_∞, about 0.21, was observed from the center of the 3/4" cube. About 3.5 decadic reduction was observed from the 3/4" cube. The model described above predicts a linear relationship between log N_o/N and M/M_∞. The experimental results are consistent with the prediction of the model. A higher value of 0.43 was observed for M/M_∞ in 1" ham cubes. The linear relationship predicts about 7 decadic reduction and no bacterial survivor was observed, consistent with the marker yield.

In the 1" ham cube, a gradient in the marker concentration was observed decreasing from the surface to the center. The gradient is consistent with the direction of the heat transfer from the surface to the center.

This test shows that the model correctly describes the form of the relationship between bacterial destruction and marker formation. Additional experiments are underway using a more suitable temperature range and more precise microbial counting technique to further refine this test.

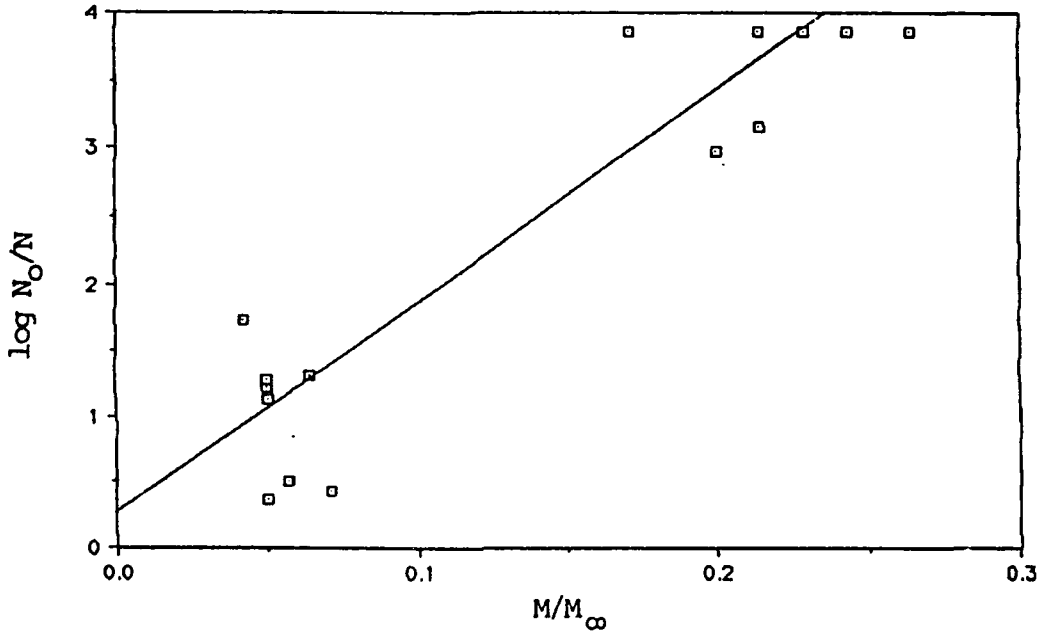


Fig. 6 A plot of decadic reduction of B. stearothermophilus vs. M/M_∞ at the center of 1/2" ham cubes (lower left) and 3/4" cubes (upper right) showing a linear relationship

CONCLUSION

The model developed and the test results obtained indicate that the markers could be used to validate the sterility of aseptically processed particulates. It should be possible to calculate the decadic reduction that would have been attained in microbial contaminants in the largest particulate moving at the fastest speed based on the marker yield. Some refinement in the technique and confirmation of key kinetic parameters are needed and are being made to achieve the precision desired. The approach is sensitive, being able to demonstrate a wide range of bacterial destruction with the accessible range of marker detection. It is also possible to determine the heat distribution within the particulate, by cutting up the particulate and analyzing for the marker in any section from the surface to the center. The distribution of residence times for particulates in aseptic processing systems can be determined as well. The marker approach can be used in general to demonstrate that the process is safe and, in particular, to gain approval for aseptic processing of particulates.

REFERENCES

1. F.D. Mills, D. Weisleder, and J.E. Hodge, Tetrahedron Letters. 15, 1243 (1970).
2. R.T. Toledo and S.-Y. Chang, Food Technol. 44(2), 72 (1990).
3. M. Abramowitz and I.A. Stegun, Handbook of Mathematical Functions, US National Bureau of Standards, US Printing Office, Washington, DC., 1964.
4. K.L. Brown, C.A. Ayres, J.E. Gaze, and M.E. Newman, Food Microbiol. 1, 187 (1984).

The Infiltration of Cotton Fibers with Sol-Gel Solutions, Forming Oxide and Non-Oxide Ceramics

Richard J Ruzga, Mr., and *Vincent F Hock, Mr.
U.S. Army Construction Engineering Research Laboratories
P.O. Box 9005
Champaign, Illinois 61826-9005

INTRODUCTION

Many investigators have used naturally occurring ceramic and organic materials to create various novel processing techniques. The objective of this research was to employ cotton to fabricate both oxide and non-oxide ceramics. Cotton fiber contains a high amount of carbon and carbon containing components. The capillarity and absorbency of the fibers make them ideal for infiltration with liquid ceramic materials. The method investigated dealt with producing silica sol-gel solutions and subsequently infiltrating the cotton fiber structures. The fibers themselves could be infiltrated in either single fiber form or fabricated into complex structures prior to the infiltration process. Depending on the atmospheric conditions during heat treatment, the cotton fibers can be oxidized or reduced, becoming a source of carbon.

Since the fibers were infiltrated with silica in the form of a sol-gel solution, oxidizing atmospheres yield SiO_2, while reducing atmospheres yield SiC. The production of SiC by this method is obviously significant due to the fact that most current procedures which produce SiC involves the pyrolysis of expensive polymeric precursors.

As a result of the unique processing, it is believed that the resulting array of carbon filaments is superior to the reinforcement obtained through synthetic fibers, by extending the ceramic matrix into the hollow structure of the cotton fiber. Conversely, due to a lack of porosity, a synthetic fiber bonds only with the ceramic on the surface.

Obviously, the potential for this material includes areas currently proposed for ceramic composite materials within the Department of Army (vehicle panels, machine parts, ballistic surfaces, radar dishes, personnel helmets, and high temperature woven and fibrous insulation). Given that the supply of advanced materials comes from either overseas sources or from petroleum products, this research is significant because the starting materials are domestic, militarily secure, readily available and beneficial to our country's agricultural base.

FIBER CHEMISTRY AND PROPERTIES

The cross section of a cotton fiber consists of a number of concentric layers, which can be broadly classified as belonging to either the outermost region (primary wall) or the inner region (secondary wall). It is the primary wall, in contact with the physical environment, that largely determines the processing behavior of the fiber. Nevertheless, the body of the fiber has a modifying effect on the periphery of the cotton hair and cannot be dismissed. The fiber morphology can be concisely summarized into six distinct regions, beginning with the outermost layer and continuing in toward the center of the fiber, as illustrated in Figure 1.

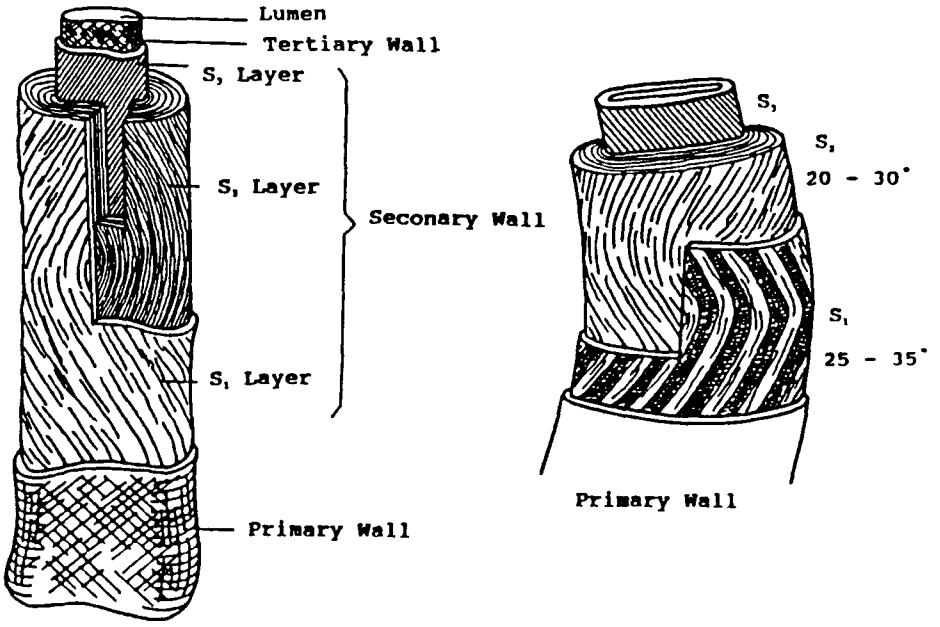


Figure 1: Cotton Fiber Morphology Showing Layers

Cellulose is the primary chemical component of the cuticle, although, in addition, other noncellulosics in the form of ash, pectic, and proteinaceous materials can be found distributed inhomogeneously throughout the fiber.[1] The interaction of liquids with the solid surfaces of cellulose fibers is facilitated by the substantial accessibilities of internal surfaces. Jeffries et al.[2] found the accessibility of liquids to the internal surfaces increases progressively from raw cotton to mercerized cotton. Mercerization is a technique used to increase luster, dye absorption, and strength in cotton materials. Molecules whose average dimensions are 30Å X 8Å X 3Å, have access through the network system of openings to the extremely large internal surfaces of the fiber.[3] The molecules cannot penetrate within the crystalline regions but are apparently well distributed through the more porous amorphous portion and over the large crystalline surface areas.

The mere presence of cotton in a sol-gel solution is not sufficient for infiltration. When infiltration does take place, it continues to an equilibrium between the cotton and the sol-gel solution if conditions are not disturbed. Ordinary infiltration depends upon a balance of effects but is controlled mainly by heat and the addition of sodium chloride or sulfate.

CONVERSION OF CELLULOSE TO CARBON

Carbon fibers were first produced by Edison[4] almost a century ago through the carbonization of bamboo and cotton cellulose materials which were formed into the required shape before being placed in the carbonizing chamber. Later, Edison made consistent filaments by dissolving cleaned commercial cotton in a solution of zinc chloride. Superior filaments were made by depositing pyrolytic carbon on to such a base filament by 'flashing' the filament in a hydrocarbon atmosphere.

The use of carbon fibers in the form of felt, wool, or cloth has found many applications in high-temperature furnaces, corrosion resistant materials, packings, electric heating elements, and refractory composite materials for use in the aerospace industry. Cellulose and rayon are the most widely used precursors for making carbon fibers. The chemical decomposition is very complex and not fully understood. Cellulose yields 15-30 wt% carbon fiber and does not melt during decomposition; hence the physical form of the precursor is maintained. However, much shrinkage accompanies the breakdown of the organic structure.

The breakdown of cellulose can be modified by varying the heating rate, the furnace atmosphere, and the moisture content. The type of cellulose used also has an influence on the end-product obtained. Additions of catalysts can alter the mode of breakdown; those which contain halides allow faster heating rates and increase the carbon fiber yield up to 40%. The temperature range 250-300°C is very critical in the process as the bulk of the weight loss takes place within this temperature range. A large shrinkage accompanies the release of volatiles, and the rate of release must be controlled for optimum results.

Ford and Mitchell[5] describe heating conditions suitable for making graphite fibers and textiles, the process being carried out in a protective atmosphere:

1. Temperature rise from 10°C/hr to 50°C/hr in the range 100°-400°C.
2. Temperature rise up to 100°C/h in the range 400°-900°C.
3. Heating to about 3000°C until graphitization occurs.

The heat treatment atmosphere can play an important part in determining the mode of breakdown of the cellulose and the final structural properties. Inert atmospheres such as nitrogen demand slow heating rates if good yields are to be obtained. Work carried out by Shindo, Nakanishi, and Soma[6] showed that non-oxidizing acid vapors such as Hcl allow faster heating rates and yields up to 40% to be obtained. These halides act as a catalyst and assist the conversion to carbon. Metal halides may also be used in the manufacture of active carbon fibers from cellulosic materials.

SOL-GEL PROCESSING TECHNIQUES

A sol-gel describes the dispersion of colloidal particles (10-1000 Å) in liquid solutions. When the viscosity of the sol increases sufficiently, usually through the partial loss of its liquid phase and polymerization (gelation), it becomes rigid. Advantages of sol-gel processing routes for the formation of ceramics include better homogeneity and purity, lower temperature of preparation, and novel procedures. Some disadvantages are large drying and firing shrinkages, lengthy processing times, and expensive starting materials.

Sol-gel solution processing techniques can lead to the formation of two distinct types of gels: polymeric and colloidal. Polymeric gels are formed from metal alkoxides which have at least three replaceable functional groups. These gels

polymerize, without precipitation, into three dimensional networks. Colloidal gels are formed from sols of stable colloidal particle suspensions. Gelation occurs as the agglomerates grow to consume the host liquor. Various starting compounds have been used to prepare silica gels. The most frequently used starting materials are silicic acid and alkoxysilanes. The gels prepared from different starting compounds have different volume shrinkages, oxide contents, and physical properties and microstructures after heat treatment.

The formation of a polymeric gel from metal alkoxides in the sol-gel process generally involves the following two steps:

(i) Hydrolysis (1)

$$M(OR)_n + xH_2O \longrightarrow M(OH)_x(OR)_{n-lx} + xROH$$

(ii) Polycondensation (2)

$$-M-OH + H-O-M \longrightarrow -M-O-M- + H_2O \quad \text{(dehydration)}$$

$$-M-OH + R-O-M \longrightarrow -M-O-M- + ROH \quad \text{(dealcoholation)}$$

The above reactions tend to occur simultaneously and their competitive/complimentary rates (depending on different factors) often determine the nature of the final product. Gelation can be brought about either by (1) bringing the polymer species into close proximity and increasing the probability of crosslinking between them by removing the solvent, or alternatively by (2) allowing the system to age, bringing about further dehydration and hydrolysis by the random collision of the polymeric species in solution. Continued crosslinking brings about further network formation, expulsion of organic groups, and finally, gelation. This investigation attempted to produce oxides and non-oxides combining processing routes shown in Figure 2.[7]

EXPERIMENTAL PROCEDURES

The experimental procedure used the following molar ratios:

2 mol H_2O to 1 mol alkoxide
5 mol ethanol to 1 mol alkoxide

These molar ratios were maintained throughout the processing, and the exact batch mixtures were experimentally determined to work best with the amount of catalyst used. The quantity of the acetic acid catalyst remained constant and was not changed with respect to the size of the sol-gel batch.

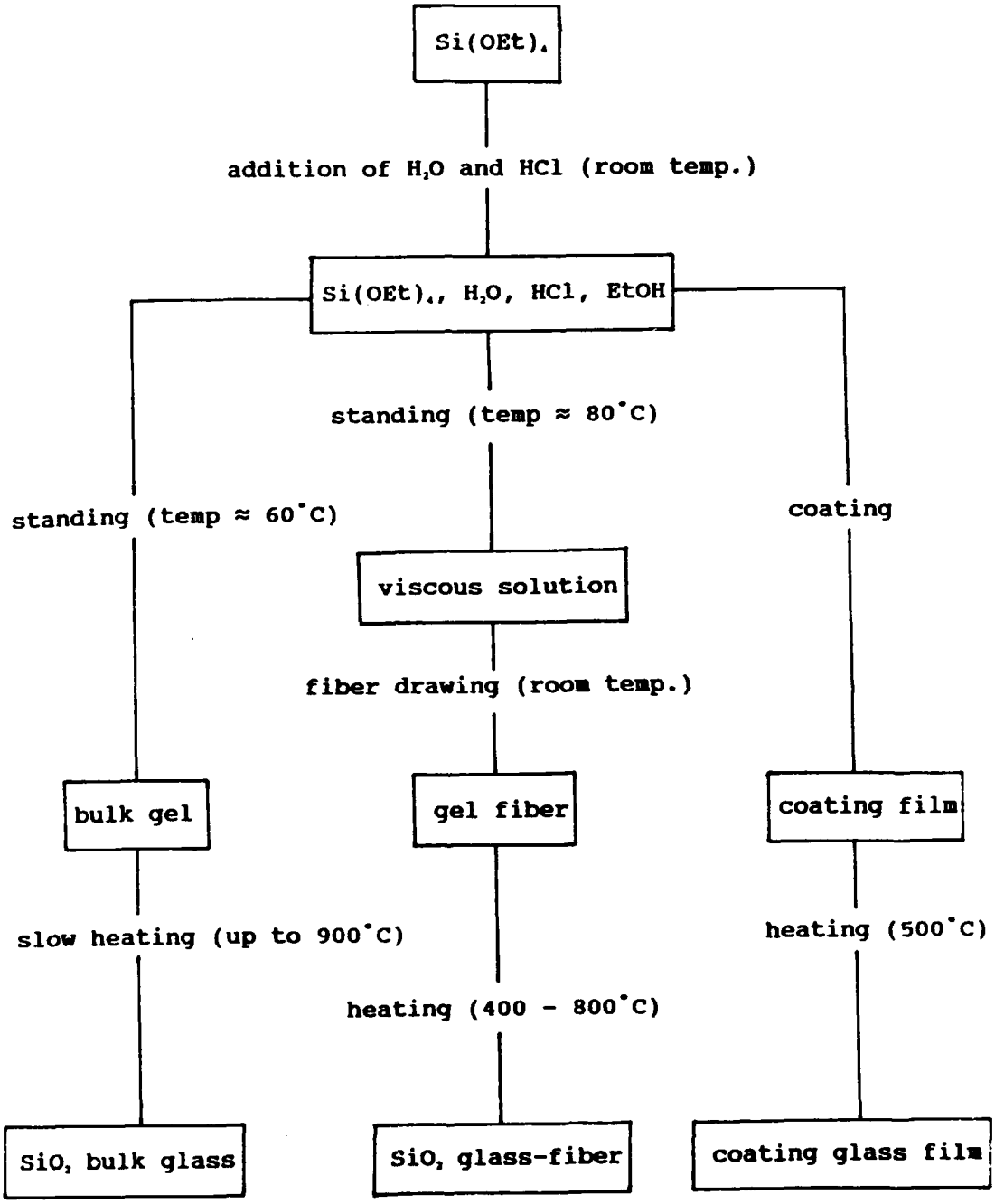


Figure 2: Commonly Used Sol-Gel Processing Routes

The approach involved mixing 730 ml of ethanol with 90 ml of deionized water in a reaction kettle for 10 minutes at room temperature on a magnetic stirring hot plate. This allowed the alcohol and water to become homogeneously mixed. Next, approximately 557 ml of tetraethylorthosilicate (TEOS) was added to the solution, and was heated to 50°-60°C. The solution was allowed to mix within this temperature range for 5.5 hr, and then 5 drops of concentrated acetic acid was added to solution. The solution was then stirred for 16 hr as the sol-gel solution was allowed to cool to room temperature.

A sample of the sol-gel solution was extracted, weighed, and then dried at approximately 100°C in a drying oven. This powder was calcined at 600°C for 24 hr to convert the mixture to pure silica powder. Again the powder was weighed, from which the SiO_2 yield of the sol-gel reaction could be calculated. The theoretical value for silica yields from sol-gel reactions is approximately 29%. The silica yields calculated for the sol-gel solutions prepared by the above method ranged from 8 - 12%, slightly less than half the theoretical value.

Two forms of cotton fiber mats were used: raw fibers and bleached, mercerized fibers. These mats were sectioned into samples which had the dimensions of 6" X 8". Care was taken to ensure that the edges remained clean cut and that there was no delamination between the fiber layers. Various sample thicknesses were created by stacking the individual sheets on top of one another. Some of the experiments performed used two sheet thicknesses while others used more. It was found that delamination occurs during heat treatments if the samples are stacked too thick due to insufficient infiltration. The sample was then sandwiched between two perforated sheets of stainless steel cut to the sample dimensions. This assembly was placed in a 6" X 8" X 3" high stainless steel 'cage' to support the sides of the sample during infiltration.

The sample, pressed between the two perforated stainless steel sheets and enclosed in the perforated stainless cage, was then placed in a 8.5" X 10.5" X 6.5" stainless steel tank. A stainless steel weight was placed on top of the assembly to compress the sample slightly. String was attached to the sample holder so that the fibrous sample could be lifted and removed easily from the tank after it had been infiltrated with the solution. The silica solution was then added to the tank in an amount which completely covered the sample. A rubber gasket was then placed between the tank and the top, and the top was secured to the tank by using C-clamps. The top of the tank was

machined with two standard 24/40 openings, into which a condensing tube and glass stopper were placed. A third opening was used for inserting a thermometer to monitor the infiltration temperature. A magnetic stirring bar was placed under the assembly to circulate the solution.

The samples were immersed in the solution for 24 hr, removed and pressed, then placed in a drying oven at 100°C for approximately 24 hr. The cotton samples generally showed a weight gain of approximately 120% due to absorption. The dried samples were then treated according to heat treatment schedules employed by Ford and Mitchell.[5]

Some cotton samples were oxidized in a furnace in ambient air. Other samples were converted to carbon by flowing argon gas into the furnace chamber as a protective atmosphere. The furnace chamber was purged for approximately 1.5 hr prior to heating with the argon, then the flow rate was controlled using a flowmeter. The argon gas channeled through the refractory base plates via an alumina tube, and the sample was placed between the inlet tube and outlet tubes. An alumina bell-shaped cover was placed over the gas tubes and the sample.

An alternate method investigated for the formation of silicon carbide involved the use of infiltrated cotton samples which were fired very quickly to temperatures less than 400°C. The cotton mat showed very little shrinkage when the conversion took place in argon atmosphere. This predominantly carbon structure was re-infiltrated with colloidal silica and dried in a drying oven. This composite sample was then heat treated under argon up to 1000°C.

An optimal atmosphere would be one created by a reducing gas such as 10% hydrogen in nitrogen. Another problem which possibly could be eliminated with a strongly reducing atmosphere is the action of the oxygen contained within the silica sol-gel solution, which tends to re-oxidize the carbon structure in the two-step process. The sample looses carbon in the form of carbon dioxide and becomes silica.

RESULTS AND DISCUSSION

The process of mercerization, as previously discussed, alters the physical and chemical make-up of the cotton fiber by increasing the absorbency of the cotton. It was believed that this form of cotton fiber would increase the uptake of the sol-gel solution. Figures 3 and 4 are scanning electron micrographs

of the two fiber forms, raw and mercerized, respectively. Figure 3 shows that the lumen is only slightly open due to the collapse of the fiber upon drying. It appears that the infiltration material must be on the order of a few hundred nanometers to impregnate this central core. Figure 4 reveals that the mercerized fibers are swelled and that the lumen is completely lost. This fiber form is less dense than the raw fiber form and contains more capillaries.

An attempt to examine the effects of mercerization included using a differential scanning calorimeter to determine the energy required to fully combust the fiber. Figure 5 is a differential scanning calorimetry (DSC) plot which compares the energy required to pyrolyze an equivalent amount of the two cotton fiber forms. Both forms of the cotton fiber degrade in the temperature range of 250°C-475°C. The curve corresponding to the raw fiber is more broad than that for the treated fiber because raw fiber contains waxes and other impurities which are removed during the process of mercerization. Since mercerized cotton cellulose has a more structured form than raw cellulose, the corresponding heat flow needed for degradation is greater, thus appearing graphically as a higher maximum point.

The cotton fiber samples which were infiltrated with the sol-gel solution and subsequently heat treated generally had shrinkages of 3 times their original dimensions and were very powdery. The samples had very little structural integrity due to incomplete infiltration. An investigation into whether the center of the fiber was actually infiltrated was difficult to conduct. Procedures were conducted with the aim to increase the strength of the resultant samples by adding up to 5 percent MgO to the sol-gel. This addition appeared to increase the strength when used in conjunction with alumina sol-gel mixtures, but proved to be of little value when employing silica sol-gels.

Two different approaches were investigated for the production of non-oxide fibers. The first consisted of using argon as a protective atmosphere. This atmosphere proved to be inadequate in reducing the oxygen in the fiber and sol-gel mixture thus leaving silicon carbide. A good candidate reducing gas would be 10% hydrogen in nitrogen. Care must be taken to insure proper ventilation occurs and that the possibility of combustion is minimized. The furnace used in this investigation had inlet and outlet gas ports on the set plate. Therefore the gas flowed most heavily near the bottom of the sample. Figure 6 is an scanning electron micrograph (SEM) of the composite material formed at the bottom region of the sample.

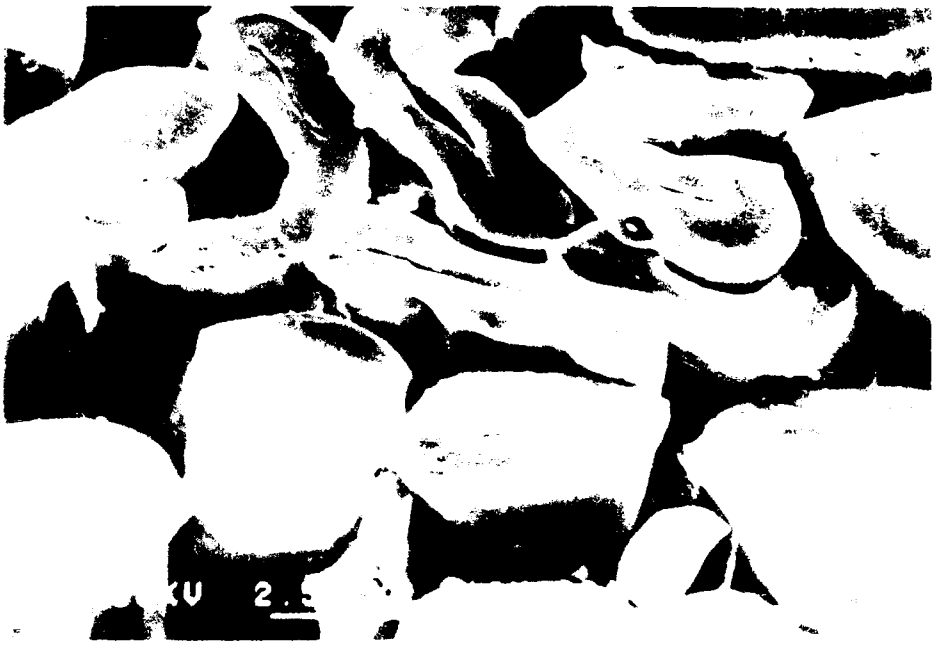


Figure 3: SEM Micrograph of Raw Cotton Fiber

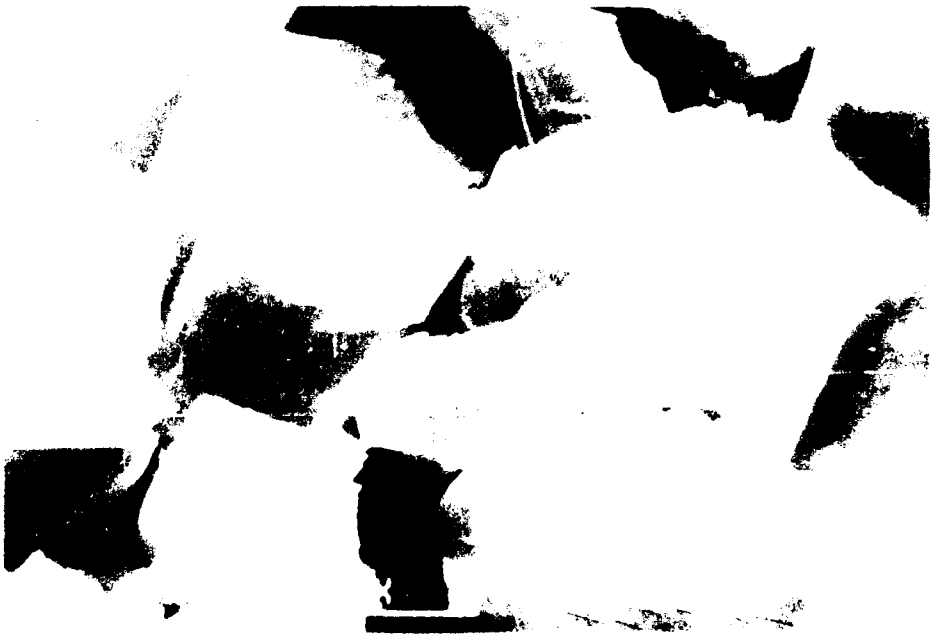


Figure 4: SEM Micrograph of Mercerized Cotton Fiber

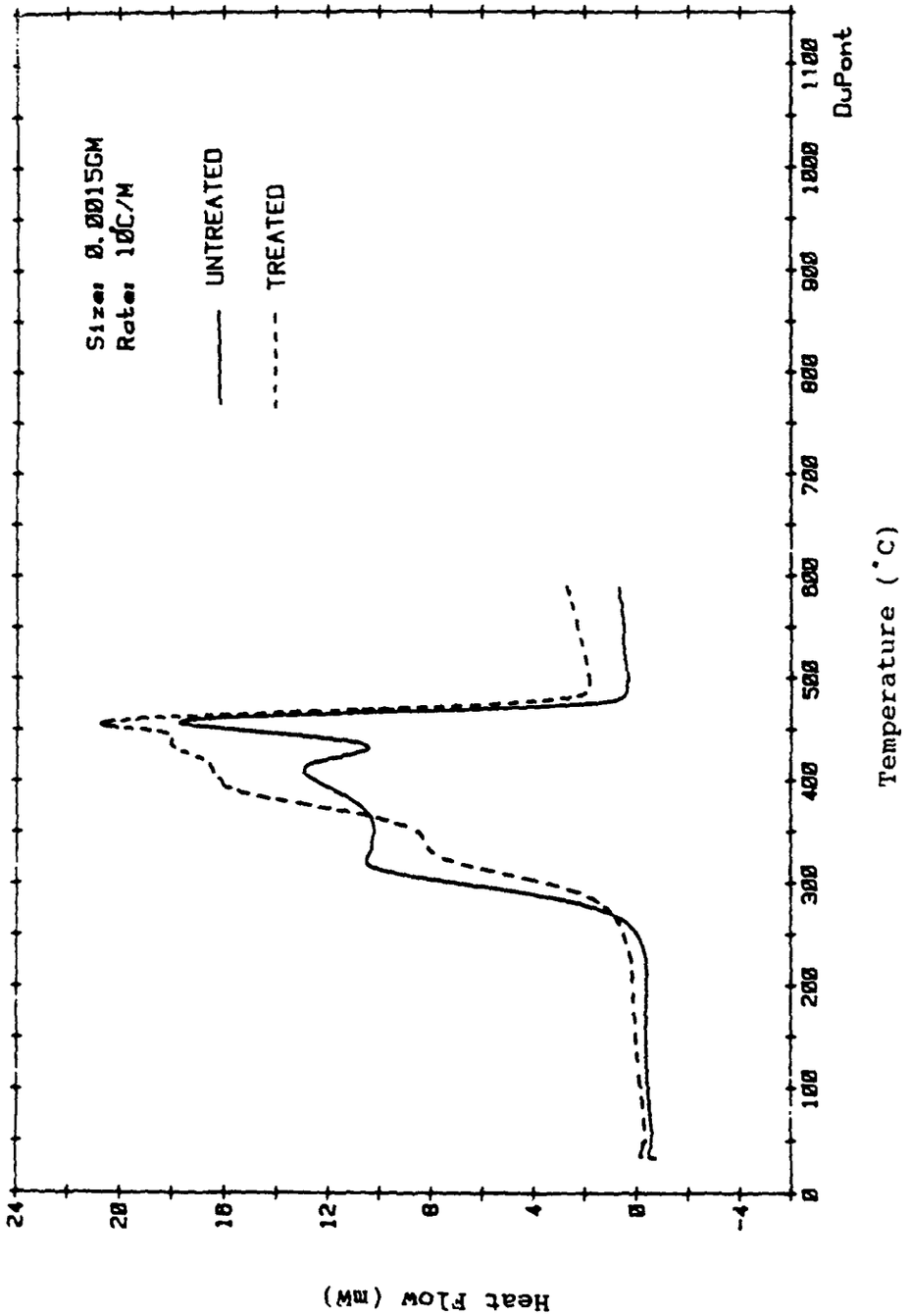


Figure 5: DSC Plot Comparing Raw and Mercerized Fibers

It is apparent that a fibrous structure remains and that the clumps of material which surrounds these fibers are silica. These two separate phases illustrate that silicon carbide did not form; rather, a composite of silica and carbon fibers did. This form may prove to be of use in some applications if either the silica existed in a continuous phase with carbon fiber running through it, or if the two phases were to react.

The second approach investigated for the formation of silicon carbide involved pyrolyzing the cotton fiber sample by running it up to 750°C at a rate of 2°C/min under argon. This sample was then reinfiltrated with the original sol-gel solution and then dried at 100°C for 24 hr. This method, as mentioned earlier, was also performed using colloidal silica for infiltration. The result, Figure 7, is similar to that which was formed from the one-step process. One difference is that the second infiltration appears to greatly increase the amount of the matrix silica phase. Figure 7 shows that many of the original cotton fiber structural characteristics are retained during the quick conversion of the cotton to carbon.

Figure 6: Bulk Ceramic with Fiber Composite

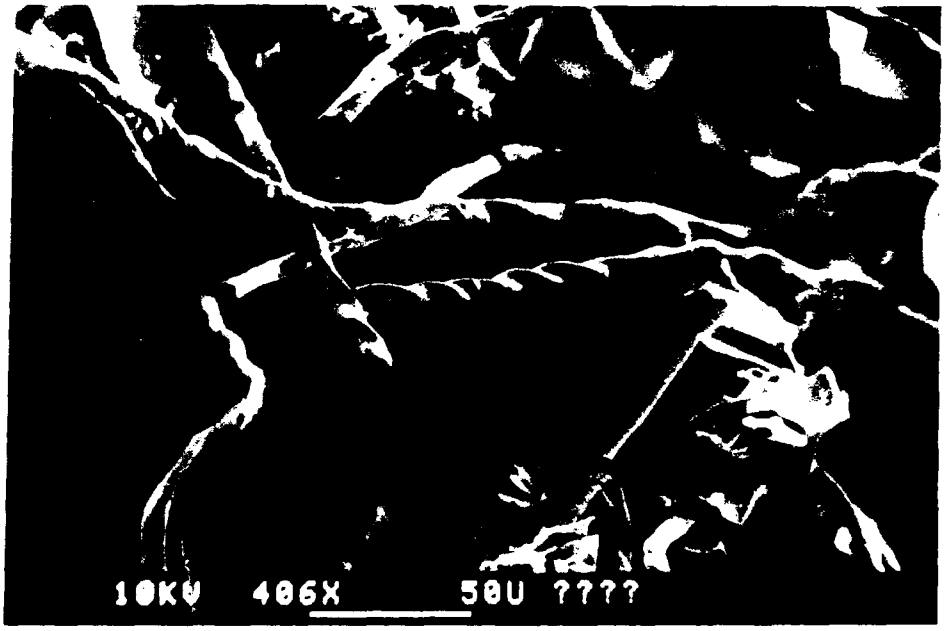


Figure 7: Quick Conversion of Cotton to Carbon

SUMMARY

The possibility of forming both oxide and non-oxide ceramic fibers by using cotton fibers and infiltrating them with sol-gel ceramic precursors was investigated. Silicon-carbide fibers were produced using the second approach in small quantities although the majority of the samples retained excess silica.

Heat treatments of the cotton fiber must emphasize critical time/temperature control through the temperature range of 250°C-450°C. Efforts to produce ceramics by this method should incorporate a heating rate similar to that of Ford and Mitchell.[5] The success of such investigations is strongly contingent upon the ability to maintain the necessary atmospheres and temperatures required for reaction throughout the entire sample.

Fibrous silica ceramics can be processed using this method but they are brittle and crack with the slightest force. It appears that complete infiltration does not occur; rather, the voids between fibers become filled with the sol-gel solution. The initial efforts of this investigation centered on using the mercerized cotton fiber samples, while the bulk of this work was conducted using raw fiber. This introduced the added problem of

fully infiltrating the fibers and not just coating them. The samples which were processed using the raw fiber exhibited less shrinkage when compared to those using the treated fiber.

It was found that in order to fully reduce the sample, the atmosphere must not only maintain an oxygen-free environment, but also reduce the additional oxygen which is present from the gel. The two-step process where the cotton is first converted to carbon and then re-infiltrated with either the initial sol-gel or with a colloidal silica offered possible solutions to this problem. The best situation would be to perform this two-step process under a highly reducing condition.

In order to fabricate near-net shapes, the processing conditions must be engineered to decrease the shrinkage which accompanies the heat treatments. From this investigation it was also concluded that other forms of silica should be used in the infiltration procedure, including colloidal suspensions.

Based upon these initial results, it appears feasible to fabricate an oxide or non-oxide fiber or woven mat via the infiltration of cotton fibers using sol-gel solutions. These fibers or mats can then be fabricated into usable shapes for Department of Army applications.

CONCLUSIONS

Fibrous oxide ceramics can be processed using this method but they are brittle and crack due to incomplete infiltration.

In the one-step process for forming non-oxide materials, it was apparent that a fibrous structure remained and that the composite consisted of a silica matrix with carbon fibers. The two-step process for forming non-oxides resulted in the formation of small amounts of silicon carbide with the remainder being silica. The major difference between the one-step and two-step processes was that the second infiltration results in an increase in strength due to a higher loading of the silica sol-gel solution.

REFERENCES

1. K.E. Duckett, "Surface Properties of Cotton Fibers", *Surface Characteristics of Fibers and Textiles*, **2**, (NY, 1975), 67-108.

2. R. Jeffries, D.M. Jones, J.G. Roberts, K. Selby, S.C. Simmens, and J.O. Warwicker, *Cellulose Chemical Technology*, **3**, 255-274, (1969).

3. Mauersberge, *Matthew's Textile Fibers*, 5th edition, (Wiley & Sons).

4. T.A. Edison, U.S. Pat. 470925 (15.3.1892)

5. C.E. Ford and C.V. Mitchell, U.S. Pat 3107152 (15.10.63).

6. A. Shindo, Y. Nakanishi, and I. Soma, "Highly Crystallite-Oriented Carbon Fiber from Polymeric Fibers", *High Temperature Resistant Fibers from Organic Polymers*, (Interscience, NY, 1969), 305.

7. S. Sakka, *American Ceramic Society Bulletin*, **64**, 1463-1466, (1985).

Cloning, Expression, Production and X-Ray Crystallographic Structure of Acetylcholinesterase

COL Jerald C. Sadoff*, Walter Reed Army Institute of Research, Washington, D.C. 20307-5100; Dr. Hermona Soreq, The Hebrew University, Jerusalem 91904, ISRAEL; Dr. Joel Sussman, Dr. Israel Silman, The Weizmann Institute of Science, Rehovot 76100, ISRAEL; Dr. Avigdor Shafferman, Israel Institute for Biological Research, Ness-Ziona 74050, ISRAEL

Chemical warfare nerve agents are one of the most serious threats to United States military operations worldwide. Their use, or even the potential for their use has enormous implications in terms of military operational capability, training and equipment requirements as well as civilian political and psychological factors. Current strategies to protect military and civilian personnel against these airborne agents, which can kill a human being in less than a minute, are based on physical barrier protection and pharmacologic prophylaxis and treatment. Physical barriers such as gas masks and protective suits obviously impair operations and are subject to failure during battle. The pharmacologic prophylaxis and treatment drugs we currently have are only partially effective against the more potent agents and may have undesirable side effects if used at the wrong time or in an inappropriate dose.

All chemical warfare organophosphate nerve agents work by destroying the ability of the human enzyme, acetylcholinesterase, (HuAChE) to inactivate the neurotransmitter acetylcholine (ACh).[1] During normal synaptic transmission, within about 1 msec, ACh is released from the presynaptic nerve cell, diffuses across the synaptic cleft, reversibly binds to the ACh receptor and is hydrolyzed by HuAChE.[2] When HuAChE is poisoned by organophosphate nerve agents ACh is not hydrolyzed and continues to stimulate nerves and muscles in an inappropriate manner leading to respiratory paralysis, seizures and rapid death.

Strategy. We believe that recombinant produced HuAChE can be given prophylactically as a scavenger to bind nerve agents prior to their contact with HuAChE in the nervous system. HuAChE is the target of all nerve agents and is also the molecule with the highest affinity for the nerve agents. HuAChE should not be immunogenic because it is a natural human protein and should not have physiologic effects because it will not cross the blood brain barrier. This concept has recently been validated in animal models where it has been shown that

fetal bovine serum derived acetylcholinesterase (AChE) can act as an effective prophylactic scavenger against the powerful nerve agent soman.[3] In order to use HuAChE as a prophylactic scavenger agent, it is necessary to clone, sequence and express the enzyme at a high level.

We also believe that HuAChE can be converted from an enzyme that is inactivated by nerve agents to an enzyme that can very rapidly hydrolyze them. Such an enzyme would be a non-pharmacologic prophylactic agent that could be given in much lower doses then HuAChE with the capability of preventing all nerve agent intoxication. In order to produce such a modified enzyme, it is necessary to understand the detailed structure of HuAChE and its interaction with AChE, nerve agents and reactivating agents at an atomic level. This can be accomplished by X-ray crystallography and computer modeling studies linked to functional studies of mutant proteins produced by site directed mutagenesis. These structure function relationships can then be used to guide site directed mutagenesis experiments in the construction of recombinant HuAChE capable of rapidly hydrolyzing organophosphate nerve agents. The information generated in these studies may also be useful in the rational design of new treatment drugs that can displace nerve agents from HuAChE by determining contact residues and structural requirements for drugs of this nature.

Cloning of HuAChE. At the time we began these studies HuAChE could not be obtained in large enough quantities to determine protein sequence or produce specific antibodies. Furthermore, most cells make extremely low levels of HuAChE specific messenger RNA. This made it difficult to construct cDNA libraries containing the full length sequence, hard to screen expression libraries with weakly cross reactive antibodies or to use the subtraction techniques that had been successfully used to clone other proteins under such circumstances.

Cloning of HuAChE was finally accomplished[4] by identifying tissues that expressed translatable HuAChE mRNA by microinjecting their mRNA into *Xenopus* oocytes and assaying for functional HuAChE. cDNA libraries from such tissues were then either constructed or purchased. A differential screening procedure using synthetic oligonucleotide probes was devised to find HuAChE specific clones in these libraries based on the assumption that there would be similarities between the known sequence of *Torpedo* AChE[6] and HuAChE. A 5'-end labeled 35-mer oligodeoxynucleotide GTAChE, which is complimentary to the sequence that encodes the peptide Tyr-Met-Met-His-Trp-Lys-Asn-Gln-Phe-Asp-His-Tyr found in the carboxyl-terminal region of *Torpedo* AChE was used for primary screening of the libraries. This probe was designed to not hybridize with cDNA for a similar human enzyme, butyrylcholinesterase (BuChE). False positives were eliminated by probing with OPSYNO which is a mixed oligonucleotide probe designed to hybridize with the active site region of *Torpedo* AChE and presumably all cholinesterases.

This differential screening procedure resulted in the isolation of two similar 1.5 kb clones designated NBG8A from a single library. Rescreening of several libraries with ^{32}P-labeled NBG8A resulted in the isolation of >25 positive clones, all encoding nucleotides downstream from the expected initiator AUG codon. One 500 base pair clone, FEMACHE, included a complete 3' nontranslated region ending with a polyadenylation site and a poly (A) tail. Another clone, designated ABGACHE, included most of the coding and 3' nontranslated sequence. The location of these sequences can be seen in figure 1.

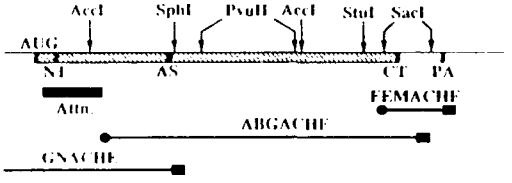


Figure 1

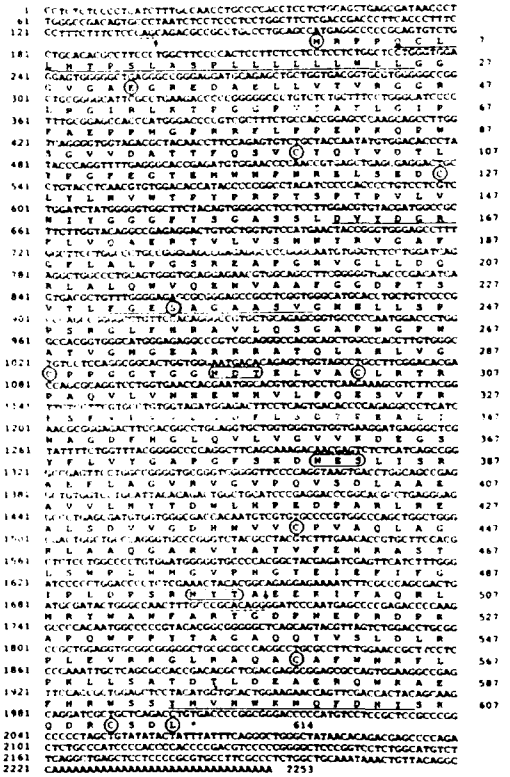
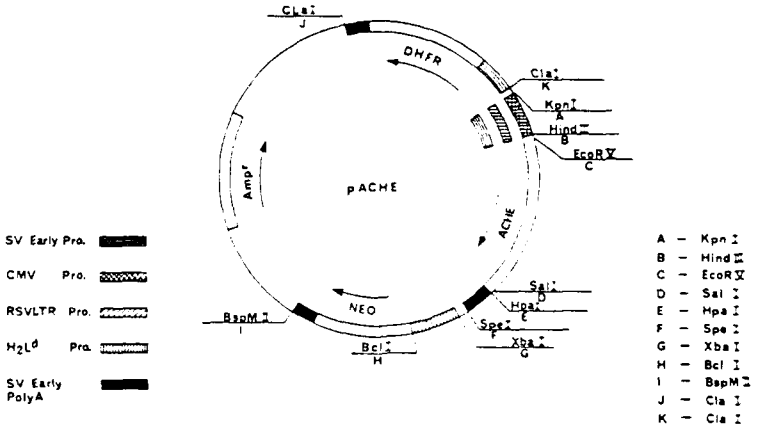


Figure 2

To derive a complete HuAChE coding sequence a 1/ mer probe, 5'-[d(CG-CCC-ATG-GTA-CAC-GTC)]-3', was designed according to the nucleotide sequence at the 5' end of the clone NBG8A and used to screen a human genomic DNA library. The resultant genomic DNA clone, GNACHE, (see figure 1) included part of the coding region for HuAChE, intronic sequences, the absent upstream sequence encoding the amino terminus of HuAChE, and a putative signal sequence preceded by the AUG codon which was imbedded in an appropriate consensus sequence for initiation of translation. The entire sequence derived from these overlapping clones is shown in figure 2. Nucleotides are numbered at left, predicted amino acids (one letter symbols) are shown below the corresponding codons and are numbered at right. The active site

sequence 231-240 is underlined. When a synthetic 2.2-kb mRNA transcript of the recombined sequence was microinjected into *Xenopus* oocytes catalytically active HuAChE was produced.

Expression of recombinant HuAChE. Once we had cloned and sequenced HuAChE, we began attempts to develop an efficient mammalian cell system for its expression. A multicassette mammalian expression vector for human recombinant acetylcholinesterase (rHuAChE) was designed and is shown in figure 3. A-K shown in figure 3 are unique restriction sites engineered into the vector to allow for maximum flexibility in construction.



...gn of multicassette mammalian expression vector of rHuAChE

The unique restriction sites (A-K) were engineered into a pUEM derivative to allow for maximal flexibility in construction/replacement of genetic element (e.g. neo, dhfr, ache enhancer-promoters and poly A signals). The actual sites were selected by a computer aided sequence elimination procedure (see text).

Figure 3

The main genetic elements in this vector were: a. a HuAChE cassette which includes the coding region of HuAChE, a SV_{40} poly A signal (from SV_{40}), and one of three efficient promoters of transcription - SV_{40} early, CMV early or RSV-LTR; b. a Neo^r cassette which includes the Neo^r gene, a poly A signal (from SV_{40}) and a weak transcription promoter - the mouse histocompatibility $H2L^d$ promoter. This cassette is used for selection of cells that have incorporated the vector thereby making themselves resistant to the antibiotic G418 through expression of the Neo^r gene which codes for C418 resistance; c. a DHFR cassette which includes the DHFR gene, an SV_{40} poly A signal and the SV_{40} early promoter. This cassette was inserted in order to select for cells that have undergone gene duplication as evidenced by increased expression of the DHFR gene which codes for resistance to the cytotoxic agent Methotrexate.

The main considerations in the construction of the HuAChE cassette were: 1. elimination of all sequences 5' to the coding

region except for the translation initiation signals which in rHuAChE appear to conform with known efficient signals; 2. elimination of superfluous sequences 3' to the coding region; 3. provision of sites upstream and downstream from the coding region which allow insertion and shuffling of promoters and poly A signals; 4. introduction of additional translation stop codons at the 3' end; 5. insertion of the minimal essential sequence of the polyadenylation signal of SV_{40} early poly A signal.

Over 40 plasmids were designed and constructed to produce expression plasmids with the elements shown in figure 3. A transient transfection method was used for screening plasmids and cell lines for production of rHuAChE. A method for producing a medium with no intrinsic acetylcholinesterase activity was developed for this purpose. It was found in these experiments that the human embryonic kidney cell line 293, when transiently transfected with a number of our expression vectors efficiently expressed catalytically active rHuAChE. Eighty percent of the enzymatic activity appeared in the cell growth medium as soluble rHuAChE in the monomeric, dimeric and tetrameric globular hydrophilic forms of this enzyme.

Experiments were then performed to make stable cell lines producing high levels of rHuAChE. Plasmids utilizing the CMV and RSVLTR promoters were transfected into 293 cells. Transfected cells which had integrated the plasmid into their chromosome were selected by exposure to the neomycin analog G418. Clones secreting rHuAChE at high levels were identified and expanded. The CMV promoter appeared superior in these experiments. From five independent transfection experiments, we obtained several recombinant 293 cloned cells that produced approximately 10 micrograms/10^6 cells/24 hours. Generally, synthesis of over 1 microgram of foreign protein per 10^6 cells/24 hours is considered to be a high level of expression[7] and very efficient systems display synthesis of 5-10 micrograms/10^6 cells/24 hours.[8] One of our cell lines, 293-R11, expressed 27 micrograms/10^6 cells/24 hours. Therefore, our level of production is within or exceeds the range of maximal levels reported in other mammalian expression systems. To the best of our knowledge these engineered cell lines have a production potential exceeding by at least a 1000 fold any other HuAChE producing cell line.

No increases in production by selection for gene duplication using MTXr-dhfr selection have been found. The copy number of chromosomally integrated cDNA, as revealed by Southern blotting of a number of stably transfected cell lines, was, however, related to rHuAChE production levels. We believe that in our stably transfected cell lines, 293 cells are expressing at the limit of their capability.

Using our expression system, we have been uniquely able to produce a large amount of highly purified rHuAChE. With this highly purified material we were able to accurately determine: Km (120uMK)

and kcat values (3.9×10^5) min^{-1}); substrate specificity (Acetylthiocholine : Butyrylthiocholine = 90:1, Acetylthiocholine : Propionylthiocholine = 2:1); sensitivity to various inhibitors (IC_{50} of BW264G51, which is a specific inhibitor of HuAChE = 0.008 uM and the IC_{50} of BW284C51, which is a specific inhibitor of butyrylcholinesterase = 200 uM; substrate inhibition (>1 mM Acetylthiocholine); and a specific activity of 6000 units/mg. Thus, our expressed rHuAChE has all the characteristics of a true AChE.

Amino acid sequencing of the NH_2 end of the purified protein revealed that the protein consisted of two molecular species differing at their amino end. The terminal 14 amino acid residues of one of these species corresponded to codons 32-44 of the HuAChE coding region, which suggests that the first 31 amino acid residues code for a signal peptide which is subsequently removed. The sequence of the second species originates at a distance of two amino acid residues. An arginine residue occupies the +1 site of this species while alanine and glyoine residues are located at -3 and -1 relative to the peptidase cleavage site, respectively. Thus, this minor cleavage site conforms with other known NH_2 processing sites of secreted proteins. To the best of our knowledge, this is the first demonstration of the existence of the previously postulated bidirected processing of a native protein in a eukaryotic system.

X-ray crystallography of *Torpedo* AChE. Because HuAChE could not be obtained, efforts to crystallize AChE from *Torpedo californica* were carried out at the same time as our cloning and expression efforts. In *Torpedo*, a major form of AChE is a homodimer bound to the plasma membrane through covalently attached phosphatidylinositol (PI). This dimer can be selectively solubilized by a bacterial PI-specific phospholipase G.[9] Combining this mild procedure with affinity chromatography we were able to obtain large amounts of pure, unnicked enzyme for crystallization. Extensive hanging drop experiments yielded conditions using ammonium sulfate in which single crystals grew within a few weeks to dimensions of 0.8 mm by 0.4 mm by 0.4 mm. They belong to space group $P3_121$, contain one monomer per asymmetric unit, and diffract to 2.6 Angstrom resolution.[10]

The structure was solved by standard methods (see ref 10). The sulfhydryl of Cys^{231} is the only free thiol in the enzyme and seemed to be a reasonable attachment site for a mercury derivative. We looked near the buried mercury atom for densities whose shape corresponded to the sequence CYS^{231}-Pro^{232}-Trp^{333}. When this sequence proved to fit the observed density satisfactorily, we built out from the tripeptide in both directions. The amino acid sequence could be fitted accurately to the electron density at all points. After crystallographic refinement, the model consisting of 526 residues and 71 water molecules had a R factor of 0.186.

The molecule has an ellipsoidal shape with dimensions approximately 45 Angstroms by 60 Angstroms by 65 Angstroms. It belongs to the class of alpha/beta proteins and consists of a 12 stranded central mixed beta sheet surrounded by 14 alpha helices. The overall structure is shown in the ribbon diagram in figure 4 and the secondary structure cartoon in figure 5 where the beta sheets are represented by gray arrows and the alpha helices by rods.

Figure 4

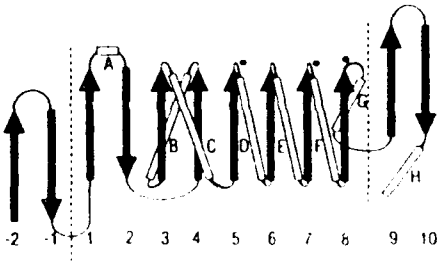
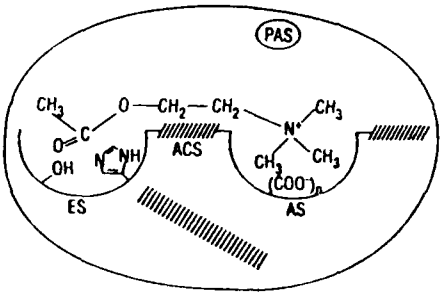


Figure 5

Early pharmacologic studies indicated that the active site of AChE contained two subsites, the "esteratic" and the "anionic" subsites,[11] which correspond to the catalytic machinery and the choline binding pocket respectively. The esteratic subsite was believed to resemble the catalytic subsites of other serine hydrolases.[12] The acetylcholine molecule is shown in figure 6 spanning these esteratic and anionic sites of this widely accepted model. We have made surprising discoveries concerning the nature of this model and each of these two subsites.



Schematic representation of the binding sites of AChE based upon previous kinetic, spectroscopic, and chemical modification studies: ES, esteratic site; AS, anionic substrate binding site; ACS, active site-selective aromatic cation binding site; and PAS, peripheral anionic binding site or sites. The hatched areas represent putative hydrophobic binding regions. The ACh molecule is shown spanning the esteratic and anionic sites of the catalytic center. Imidazole and hydroxyl side chains of His and Ser are shown within the esteratic site. Within the anionic site, $(COO^-)_n$ represents six to nine putative negative charges.

Figure 6

Organophosphate nerve agents poison AChE by forming a covalent bond to Ser^{200} in the esteratic active site of *Torpedo* AChE.[13] His^{440} has recently been implicated as part of the active site as well.[14] Our structure shows three residues His^{440}, Ser^{200}, and Glu^{337} that are all near one another and form a planar array that resembles the catalytic triad of chymotrypsin and other serine proteases.[15] The catalytic triad of *Torpedo* AChE, which we have discovered, differs from other serine proteases in that it contains Glu instead of the classical Asp in the triad and is of the opposite "handedness" i.e. is a mirror image of that seen in serine proteases. These new and unexpected findings predict that the 'oxyanion hole',[15] which is formed by the amide NH of the active site Ser in the serine proteases (N-terminal) would differ in AChE. Molecular modeling studies suggest that the 'oxyanion hole' of HuAChE would be formed by the main chain nitrogens of Ala^{204} and Gly^{122} interacting with the carbonyl oxygen and that of Gly^{121} interacting with the ester bond oxygen of ACh or its analogues.

The most remarkable feature of the three dimensional structure is a narrow, deep and convoluted gorge, about 20 Angstroms long, which penetrates halfway into the enzyme and widens out close to its base. We have named this cavity the "active site gorge" because it contains the AChE catalytic triad of amino acids centered around Ser^{200}. This gorge can be seen in figure 7A which is a stereo van der Waals surface dot drawing of a thin slice through the AChE monomer. The Ogamma atom of Ser^{200}, which is located approximately 4 Angstroms above the base of the gorge, can be seen from the surface of the enzyme. This is shown in figure 7B which is a space filling stereo view of the AChE molecule looking down into the active site gorge. Ser^{200} appears as a black space in the middle of this view.

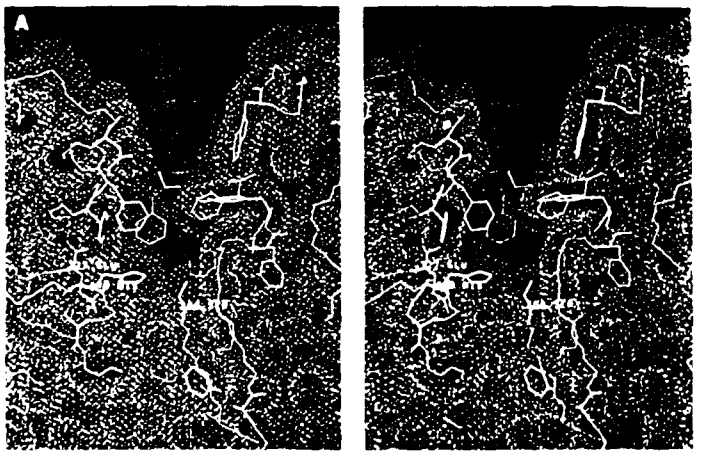


Figure 7A

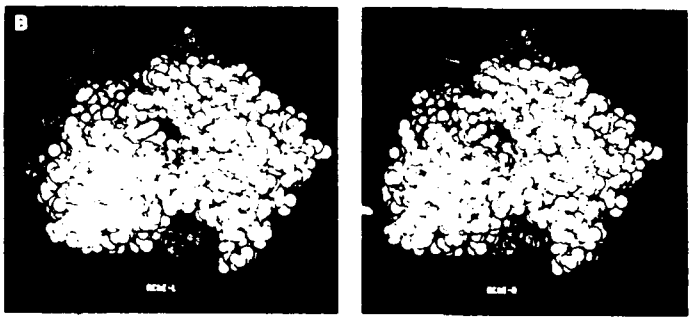


Figure 7B

Despite the structural complexity of the gorge and the flexibility of acetylcholine, manual docking experiments yielded a good fit of the extended all-trans conformation of acetylcholine into the gorge. As shown in figure 8, the acyl group of acetylcholine is positioned to make a tetrahedral bond with O_{gamma} of Ser^{200} while the quaternary group of the choline moiety of acetylcholine was within a van der Waals distance (approximately 3.5 Angstroms) of Trp^{84}. An affinity labeling study has also recently identified Trp^{84} as part of the putative anionic (choline) binding site.[16] The thin solid line in the figure represents the bond produced in the transition state between Ser^{200} O_{gamma} of AChE and the carbonyl carbon of acetylcholine, the dotted lines indicate the putative hydrogen bonds.

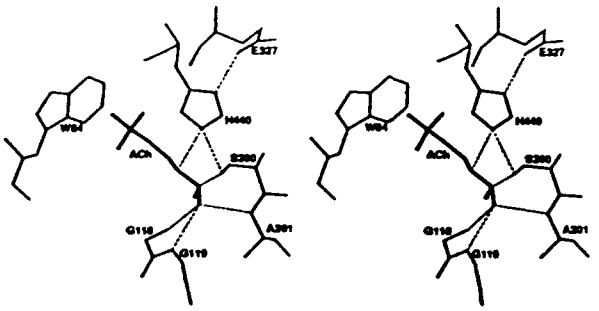


Figure 8

Fourteen aromatic residues line approximately 40% of the surface of the gorge. Only a few acidic residues are contained in the gorge, Asp^{285} and Glu^{273} at the very top, Asp^{72} which is hydrogen bonded to Tyr^{334} about half way down and Glu^{199} near the bottom. The high aromatic content of the walls and floor of the gorge, together with its dimensions help explain why previous biochemical studies have revealed a variety of hydrophobic and anionic binding sites distinct from, or overlapping, the active site. The gorge is so deep and its aromatic surface so extensive that there are many ways and places for

substrata, agonists, and inhibitors to bind. Certain ligands may be too bulky to penetrate the gorge but could partially block its entrance, while potent inhibitors such as the longer bisquaternary compounds might attach at one end to a peripheral site and at the other end to an aromatic residue in the wall of the gorge. Shorter bisquaternary inhibitors and oxime reactivators (which are the drugs we field for use in treatment of organophosphate nerve agent poisoning) might bind entirely within the gorge.

We believe the choline binding anionic subsite (figure 6) is clearly misnamed as it contains at most one formal negative charge. It was thought to be an anionic site because of the positive charge of acetylcholine and numerous other potent ligands and the six to nine putative negative charges in the binding site.[17] AChE is one of the most active enzymes known with an extremely fast on reaction. If "electrostatic guidance" is not responsible for this high catalytic activity, how can it be explained. We postulate the existence of an "aromatic guidance mechanism". Recent data and theoretical considerations support a preferential interaction of quaternary nitrogens with the pi electrons of aromatic groups.[18] After initial binding of AChE to low affinity sites at the top of the gorge, the aromatic lining could then function by providing an array of low affinity binding sites which would permit rapid diffusion to the active site.[19]

These X-ray studies have, therefore, demonstrated that the previous concepts of the esteratic site needed to be modified in terms of its physical orientation and the existence of Glu and not Asp in the catalytic triad, that the concepts of the anionic substrate binding site and peripheral anionic binding sites needed to be modified in favor of a deep aromatic lined gorge, and that "electrostatic guidance" should be replaced with the concept of "aromatic guidance".

Site directed mutagenesis of rHuAChE. A replacement method which is of higher fidelity than methods involving replication by polymerases[19] was utilized for site directed mutagenesis studies. HuAChE sequence variants which conserve the wild type coding specificity but carry new unique restriction sites were created so that a synthetic DNA duplex carrying the mutated codon could replace the appropriate HuAChE fragment. Following transient transfection, specific activity of secreted enzyme was determined by directly measuring enzymatic activity and utilizing specific antibodies to determine rHuAChE protein mass. Expression levels were normalized by comparison to a *cat* reporter gene.

Evidence for the involvement of Ser^{203}, His^{447} and Glu^{334} in the catalytic triad of rHuAChE was provided by substitution of these amino acids by alanine residues. Of 20 amino acid positions mutagenized thus far these three were unique in abolishing detectable enzyme

activity (<0.0003 of wild type levels) in the presence of proper production, folding and secretion. This is the first biochemical evidence for the involvement of a glutamate in a hydrolase triad and supports our X-ray crystal structure of *Torpedo* AChE. Attempts to convert rHuAChE to a Cys-His-Glu or Ser-His-Aps did not yield effective rHuAChE. Ser and Cys differ only by substitution of a Sulfur in Cys for an Oxygen in Ser but this one atom difference eliminates all activity. Substitution of Asp^{74} by Gly or Asn generated an active enzyme with increased resistance to succinylcholine and dibucaine; thus mimicking in an rHuAChE molecule the phenotype of an atypical BuChE natural variant (Asp^{70} to Gly mutation). This demonstrates our ability to change the natural properties of rHuAChE with site directed mutagenesis. Complete absence of secreted rHuAChE polypeptide was observed when Asp^{175} or Asp^{404} were substituted by Asn. These two aspartates play a role in the folded state of the polypeptide. This is supported by the X-ray structure that reveals the participation of these residues in salt bridges between neighboring secondary structure elements.

We have for the first time cloned, sequenced, expressed at high levels and performed site directed mutagenesis of HuAChE, which is the target of all organophosphate chemical nerve threat agents. We have also crystallized and solved the X-ray structure of *Torpedo* AChE and are currently working on crystallization of rHuAChE. These studies have already provided us a much deeper understanding of the specific structure function relationships of this enzyme at an atomic level, and demonstrated our ability to manipulate this molecule and produce it at high levels. We believe we have laid the foundation for use of rHuAChE in prophylaxis against nerve agent intoxication and for the development rHuAChE with hydrolytic activity for these organophosphate poisons.

REFERENCES

1. P. Taylor in *Pharmacologic Basis of Therapeutics*, A.G. Goodman et al., Ed. (MacMillan, New York, 1985), pp 110-129.

2. T.L. Rosenberry, *Adv. Enzymol. Relat. Areas Mol. Biol.* **43**, 103 (1975).

3. Y. Ashani et al., *Biochemical Pharmacology* **41**, 37 (1991).

4. H. Soreq et al., *Proc. Natl. Acad. Sci. USA* **87**, 9688 (1990).

5. Schumacher et al., *Nature (London)* **319**, 407 (1986).

6. C. Prody et al., *Proc. Natl. Acad. Sci. USA* **84**, 3555 (1987).

7. S.B. Yan, B.W. Grinell, F. Wold, *TIBS*. **14**, 264 (1989).

8. M.B. Hendricks, C.A. Luchette, M.J. Banker, *Biotechnology* **7**, 1271 (1989).

9. I. Silman and A.H. Futerman, *Eur. J. Biochem.* **179**, 11 (1987).

10. J.L. Sussman et al., *Science* **253**, 872 (1991).

11. D. Nachmansohn and I.B. Wilson, *Adv. Enzymol.* **12**, 259 (1951).

12. T.L. Rosenberry, *Adv. Enzymol.* **43**, 103 (1975).

13. K. MacPhee-Quigley, P. Taylor, S. Taylor, *J. Biol. Chem.* **260**, 12185 (1985).

14. G. Gibney et al., *Proc. Natl. Acad. Sci. USA* **87**, 7546 (1990).

15. T.A. Steitz and R.G. Shulman, *Annu. Rev. Biophys. Bioeng.* **11**, 419 (1982).

16. C. Weise et al., *EMBO. J.* **9**, 3885 (1990).

17. H.J. Nolte, T.L. Rosenberry, E. Neumann, *Biochemistry* **19**, 3705 (1980).

18. D.A. Dougherty and D.A. Stauffer, *Science* **250**, 1558 (1990).

19. L. Hedstrom et al., *Methods in Enzymol.* **202**, 671 (1991).

Laser-Based Sensitive Detection of Trace Atmospheric Vapors of Military Interest

Dr. Rosario C. Sausa*, Dr. George W. Lemire[a],
Dr. Josef B. Simeonsson[a], and Dr. Andrzej W. Miziolek
U.S. Army Ballistic Research Laboratory
Aberdeen Proving Ground, Maryland 21005-5066

INTRODUCTION

The detection and monitoring of trace atmospheric vapors of military interest is a major technological task of growing importance. Examples include propellant and explosives vapor detection for mine countermeasures and anti-terrorist applications, and the detection of chemical warfare (CW) agents. The increasing interest in the environmental aspects of military operations such as pollution prevention and compliance require the development of high-performance sensors for various toxic and non-toxic pollutant gases. Laser-based vapor sensors are attractive candidates for these applications since they are highly selective and specific and are capable of rapid real-time monitoring. The ability to manufacture them into rugged and power-efficient devices makes them attractive candidates as well.

Many molecules including CW agents and various energetic materials are difficult to detect spectroscopically by ultraviolet-visible (UV-VIS) absorption or emission due to the lack of distinguishing structure or absence of any features in the UV-VIS spectral region. As a result, an approach based on the use of UV laser radiation to photodissociate the parent molecules into characteristic fragments is being pursued (see references 1-8 for a partial list of representative studies). The photofragments typically include di- and triatomics, as well as atoms, which generally have structured, readily identifiable transitions in the UV-VIS spectral region. These

---

a - National Research Council Postdoctoral Research Associate

fragments may in turn be detected either by their prompt emission, if electronically excited, by laser-induced fluorescence (LIF), or by resonance-enhanced multiphoton ionization (REMPI). This approach has been demonstrated previously in our laboratory in a two-laser experiment using simple organophosphonates which are nerve agent simulants containing organic esters of substituted phosphoric acids.[1,8-9] Selected organophosphonates were dissociated in a simple flow cell using the focused output radiation of an excimer laser at 193 or 248 nm and the characteristic PO moiety, present in all nerve agents, was detected via its $A^2\Sigma^+ - X^2\Pi_r$ transition near 247 nm.

Propellants and explosives also contain molecules with characteristic functional groups, namely the NO_2 moiety bonded to either a carbon or nitrogen atom. This functional group is weakly bound to the main skeletal portion of the molecule by approximately 40-50 kcal/mole, depending on the molecule, and is responsible for the weak and structureless absorption feature observed in the UV near 230 nm.[10,11] A perusal of the literature reveals that the UV laser photodissociation of energetic molecules, such as RDX,[12] and model compounds,[3,12,13] under collision-free conditions results in the production of NO_2 in the initial step of the photolysis. NO_2 is usually detected by its prompt emission or by LIF.[3,12-13] However, both of these methods are very inefficient since the absorbed radiation is radiated over a large spectral region, visible to near infrared. The radiative lifetime for these transitions is also very large, typically ~ 50 -120 μsec,[14] and is indicative of a weak oscillator strength. Moreover, predissociation to NO + O predominates over fluorescence at wavelengths less than 400 nm . The NO is better suited for detection since it is readily formed from the predissociation[15,16] of NO_2 and its radiative lifetime[17] (~ 200 nsec) is significantly shorter than the NO_2 states excited in the visible region. The NO fragment can be detected with a high degree of sensitivity and selectivity by (1+1) REMPI and/or LIF via its $A^2\Sigma^+ - X^2\Pi$ transition near 226 nm. Thus, one laser tuned to 226 nm can be used both for parent photofragmentation and fragment detection.

The laser photodissociation/fragment detection technique is most effective when coupled with molecular beam (MB) sampling and time-of-flight (TOF) mass spectrometric detection. A TOF mass spectrometer can detect all ions simultaneously in less than 100 μsec with resolution ($m/\Delta m$) greater than 500, while molecular beam sampling greatly improves the sensitivity and selectivity over ambient samples. The molecules in the beam are colder than those at ambient conditions since they are subject to a supersonic expansion into vacuum through a pulsed, small-orifice, nozzle. They are thus produced in a few distinct ro-vibrational states, and as a result, their absorption spectrum is considerably less congested and simpler to analyze. Interference from other molecular rotational bands is also minimized, if not eliminated.

The pulsed nozzle also greatly reduces the total gas throughput entering the analysis chamber while still maintaining a high density of the sample. Thus, the pumping requirements are reduced allowing for a more compact design.

In this paper, we report on the sensitive detection of atmospheric vapor of NO_2 containing compounds based on the 226 nm photolysis of target molecule with subsequent detection of the NO fragment by REMPI and/or LIF via its $A^2\Sigma^+ - X^2\Pi$ transition. The analytical utility of this novel detector is demonstrated using trace energetic materials and model compounds in a MB/TOF apparatus. Limits of detection are reported for NO_2, dimethyl nitramine (DMNA), nitromethane, nitrobenzene, trinitrotoluene (TNT), and 1,3,5-trinitrohexahydro 1,3,5-triazine (RDX). The 193 nm and 247 nm excitation of diethyl methylphosphonothioate (DEMPS), a simulant for CW agents, was also studied and the results are presented.

EXPERIMENTAL

A schematic diagram of the experimental apparatus used for these experiments is depicted in Figure 1. The apparatus is a general purpose molecular beam apparatus that has been constructed for both laser-induced fluorescence and resonance-enhanced multiphoton ionization studies. There are two main chamber bodies to this apparatus. The first chamber consists of an eight inch Tee with smaller ports added for laser excitation, fluorescence detection, and pressure recording. This chamber is pumped out through the bottom by a 1000 liter per second turbo molecular pump (Leybold-Heraeus, TurboVac 1000).

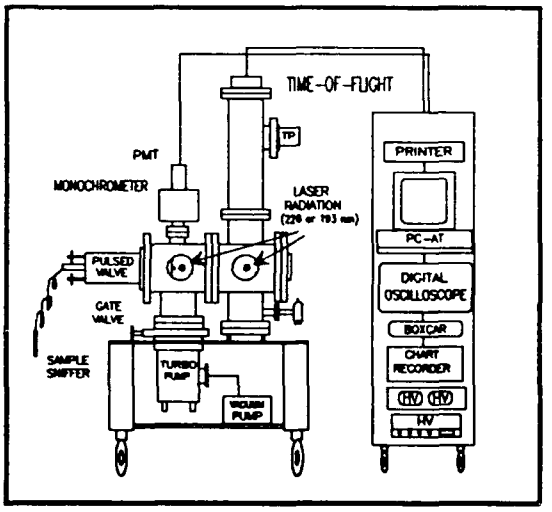


Figure 1. Schematic of the Experimental Apparatus.

At the rear of the Tee is a pulsed supersonic valve (R.M. Jordan Co., PSV) that can be translated the length of the chamber and can be positioned for either LIF or REMPI experiments. At the front of the Tee is mounted an eight inch 4-way cross serving as the second chamber. These two chambers are differentially pumped and are separated by a skimmer with a 3 mm orifice (Beam Dynamics, Inc., Model 2). Mounted on the top of the cross chamber is a commercial time-of-flight mass spectrometer (R.M. Jordan Co.) with reflectron.

The TOF mass spectrometer distinguishes masses by their different arrival times at the microchannel plate detector located at the end of the flight tube. The kinetic energy of the ions after acceleration is equal and is given by the equation,

$$1/2\ mv^2 = Ve \qquad (1)$$

where m is the mass of the ion, v is the velocity, V is the acceleration voltage, and e is the ion charge. Thus, every ion has its own unique velocity, inversely proportional to the square of the mass. The time, t, necessary for an ion to reach the end of a flight tube of length, L, is given by the equation,

$$t = L/v \qquad (2)$$

and the difference in time, Δt, which separates ion 1 from another ion 2 is,

$$\Delta t = L/(2Ve)^{1/2} (m_1^{1/2} - m_2^{1/2}) \quad . \qquad (3)$$

As shown in Eqn. 3, the difference in times of the ions depends on the difference in the square roots of the masses. For the apparatus employed in this study, a complete mass spectrum of a packet of ions of which the largest is 500 amu (1 amu = 1.67×10^{-24} g) can be recorded in approximately 72 μsec. This calculation is based on a flight tube of length 100 cm (L=200 cm since it is a reflectron) and an acceleration voltage of 2000 V (1 eV = 1.60×10^{-12} g.cm^2/sec^2).

Samples of nitromethane (Aldrich), nitrobenzene and trinitrotoluene (Eastman-Kodak), DEMPS (Alpha), and DMNA and RDX (ARDEC) were selected for this study. All were used without further purification. They were introduced into the analysis chamber either in an effusive flow, so that the gas in the sampled volume was replenished at a rate sufficient to avoid accumulation of photolysis products, or in a pulsed molecular beam. For the latter, vapors of the analytes were seeded in an atmospheric pressure of air (Potomac), nitrogen (Matheson), or argon (BRL) gas and the mixture expanded into the analysis chamber using a pulsed valve with a 0.5 mm diameter nozzle. The valve could be heated up to 150°C and was operated at 10 Hz. In the case of RDX and TNT, the sample was heated to approximately 100°C before analysis. Mixtures of argon doped with 0.1% NO (Matheson) and air with 6.2 ppm of NO_2 (Scott-Marrin) were used for calibration.

The pulse pressure in the analysis chamber was calculated employing the following equation,

$$P_{pulse} = P(r/R)^2(1-\cos^3\theta) \quad (4)$$

where P is the backing pressure, r is the radius of the nozzle, R is the radius of the skimmer orifice, and θ is the skimmer transmission angle. θ is equal to the arctangent of the ratio of the radius of the skimmer orifice (R) to the distance from the skimmer to the nozzle (D). A pulse pressure of 180 mTorr was obtained when using a backing pressure of 770 Torr, a nozzle diameter of 0.5 mm, a skimmer diameter of 3.0 mm, and a skimmer to nozzle distance of 2.0 cm.

The samples were photolyzed and/or ionized using unfocused or focused (f=250 mm or 500 mm lenses) UV radiation. A broad-band excimer laser (Lambda Physik, EMG 150MSC) was employed to generate 193 nm radiation, while an excimer pumped dye laser (Lumonics Hyper EX-400 and Hyper DYE-300) with second harmonic generator (Lumonics TRAK-1000) was used to generate the tunable UV radiation near 226 nm. Both lasers were operated at 10 Hz. Pulsed energies of the order of 5-10 mJ and 100-150 μJ for the 193 nm and 226 nm radiation, respectively, with duration of 15-20 nsec were employed for detection. The linewidth of the tunable UV radiation was approximately 0.16 cm^{-1} (fwhm).

The fluorescence signal whether prompt or induced was collected 90° to the photolysis beam and focused onto the entrance slit of a 0.25 m monochromator (McPherson, model 218) acting as a broad-band filter (≈ 8 nm fwhm). The monochromator was tuned to approximately 237 nm and 500 nm for the detection of NO and NO_2, respectively. Schott filters UG-11 and GG-435 were also used in place of the monochromator. The signals were detected with a photomultiplier tube (EMI 9789QA) and directed into a gated integrator (Stanford Research Systems). Both fluorescence and ion signals were displayed and monitored in real-time on a 125 MHz digital oscilloscope (LeCroy 9400). A PC-AT computer was employed for data acquisition and analysis.

RESULTS/DISCUSSION

A. Energetic Materials:

Presented in Table I are the names and structural formulas of molecules commonly used in explosives and propellants, as well as in CW agents. The molecules which are used in this study are also listed. A cursory inspection of the table reveals that the energetic materials and model compounds all contain one or more weakly bound -NO_2 functional group. The physical processes underlying our

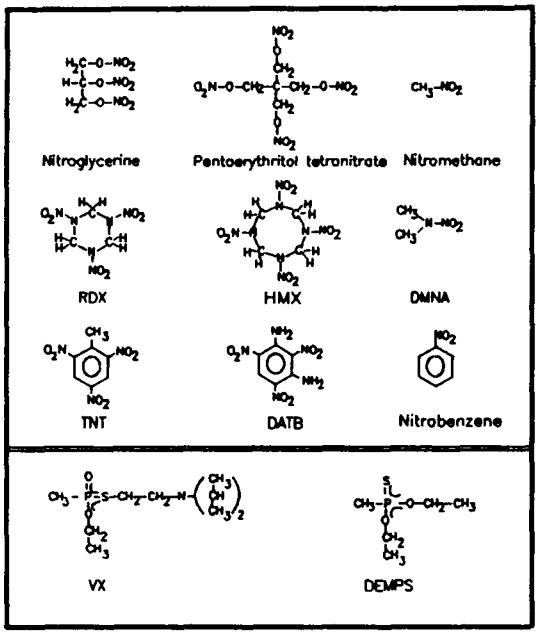


Table I. Structural Formulas and Names of Selected Compounds of Military Interest.

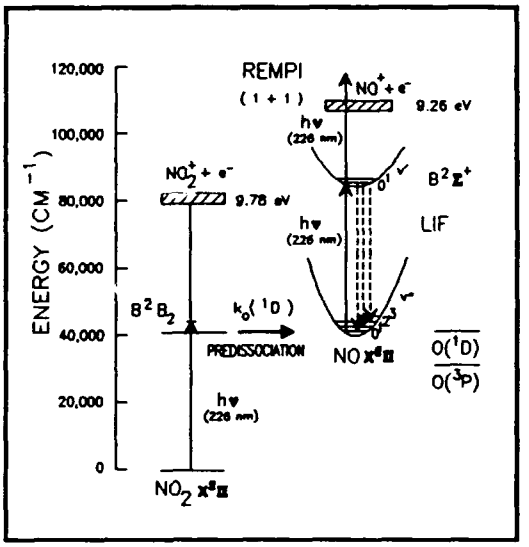


Figure 2. Potential Energy Diagram of NO_2 and NO Showing the Multiphoton Scheme Employed for the REMPI and LIF Detection of the NO Photofragment at 266 nm.

approach to the detection of these molecules may be understood by referring to Figure 2 which shows potential level diagrams of NO_2 and NO. In our approach, the target molecule (R-NO_2) is first photolyzed to NO_2, along with its corresponding companion radical (R), via the process

$$R\text{-}NO_2 \xrightarrow[226\ nm]{h\nu} R + NO_2 \quad . \tag{5}$$

The NO_2 fragment is then detected by monitoring its predissociative product NO using (1+1) REMPI or LIF employing its $A^2\Sigma^+ - X^2\Pi$ (0,0) band at 226 nm.

Presented in Figure 3 is a typical mass spectrum obtained when photolyzing target molecules in a molecular beam using 226 nm radiation. The molecular beam was obtained by flowing argon gas at approximately one atmosphere over the vapor

of the target compound and expanding the resulting gas mixture into the analysis chamber. The spectrum shown in Figure 3 reveals a prominent peak whose arrival time to the detector is measured as 23.2 μsec. This time corresponds to a mass to charge ratio (m/e) of 30 and is attributed to the NO^+ ion. The observed peak is clearly enhanced as a result of the REMPI process. Similar results were obtained when using air or nitrogen as the carrier gas. No such peak was observed, however, when the laser was tuned to an off-resonance transition.

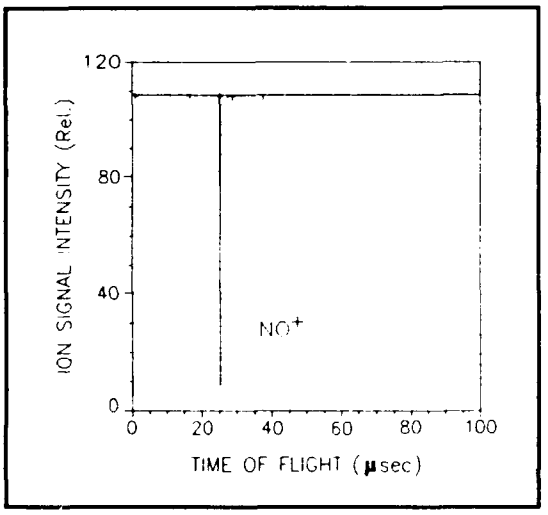


Figure 3. Time-of-Flight REMPI Spectrum of NO^+ Generated from DMNA. Signal generated from 226 nm excitation of DMNA (~100 μJ).

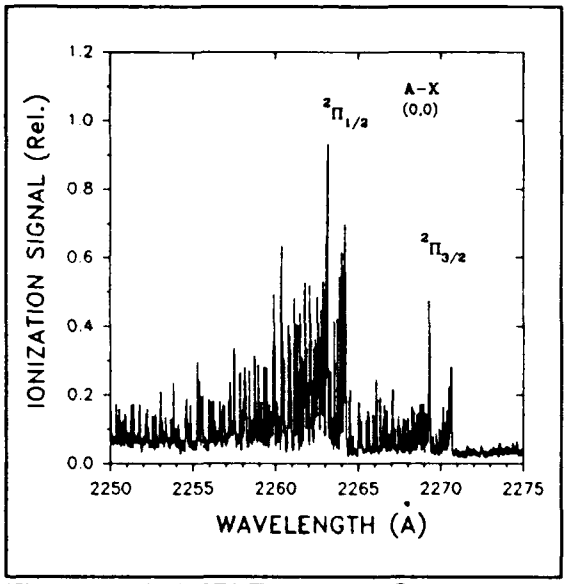


Figure 4. REMPI Excitation Spectrum of NO^+ in the Region of 226 nm. The NO^+ was generated from a molecular beam of DMNA.

An excitation spectrum obtained by scanning the tunable UV radiation while monitoring the m/e = 30 peak is shown in Figure 4. The spectrum which reveals numerous rotational lines attributed unequivocally to NO $A^2\Sigma^+$ - $X^2\Pi$ transitions corroborates our mass spectrum assignment.

The detection limits of the various compounds studied employing the fragmentation/REMPI technique using 226 nm radiation are presented in Table II. The REMPI signal production was measured using our present experimental setup, which is far from being optimized, under an identical set of conditions. The limits of detection are defined by the equation,

$$C_L = 3\,\sigma/S \qquad (6)$$

where C_L is the concentration (ppb) that produces a signal S (mV/ppb) equal to three times the standard deviation σ (mV) of the background when no analyte is introduced through the nozzle. The limits of detection calculated refer to the gas concentration of the gas prior to introduction into the analysis chamber. It should be noted that the estimates of the detection limit for the precursors depend on a number of factors including laser energy, sample introduction, voltage settings of the ion optics and detector, and photodissociation efficiency, all of which effect the rate of signal production. As seen from Table II, detection limits in the ppb can be obtained with a relatively small amount of laser energy, approximately 100 μJ. The difference in vapor pressure of the various precursors was taken into account when determining these limits. Thus, any difference shown in Table II results from the different absorption cross sections and oscillator strengths that the target molecules have at 226 nm, the number of NO_2 functional groups present, as well as the efficiency in the photodissociative pathways leading to the formation of NO in its $X^2\Pi$ ($v''=0$) vibronic state. It is clear that RDX, TNT, and NO_2 are particularly efficient in generating REMPI signals. In the case of NO_2, this was not too surprising since Morrison and Grant[15,16] observed that by far the most prominent ion fragment generated by the multiphoton excitation of NO_2 at wavelengths in the region from 425 to 455 nm was the NO^+. Their data suggest that the course of excitation in this system is dominated by predissociation into $NO(^2\Pi) + O(^1D)$ at the level of the second photon within the B ($^2B_2$) state. They contend that this high energy dissociation pathway becomes important soon after this channel becomes energetically accessible.

The selectivity of the laser photodissociation/REMPI technique for the NO^+ in the presence of other species that compete either by forming ions during the REMPI excitation process or from isobaric interferences is given by the following equation,

$$S_T = (R_{REMPI}/\Delta r_{REMPI})(R_{TOF}/\Delta r_{TOF}) \qquad (7)$$

where S_T is the total selectivity, R_{REMPI} is the range of UV-VIS spectrum, Δr_{REMPI} is the laser bandwidth used in the excitation process, R_{TOF} the range of all possible masses, and Δr_{TOF} is the resolution of the NO^+ ion signal at m/e = 30. A total selectivity value of 5×10^8 is obtained when using values of 500 nm and 0.001 nm for R_{REMPI} and Δr_{REMPI}, respectively, and 72 μsec (masses 1-500 amu) and 0.07 μsec for R_{TOF} and Δr_{TOF}, respectively.

Table II. Limits of Detection for Various Compounds Studied Using the Laser Photofragmentation/REMPI Technique with NO Detection at 226 nm

Compound	Limit of Detection (ppb)
NO	8
NO_2	240
CH_3NO_2	1000
DMNA	450 (2000)[a]
RDX	8[b]
TNT	24
Nitrobenzene	2400

a. Limit of detection obtained by laser-induced fluorescence of NO via its $A^2\Sigma^+ - X^2\Pi$ transition.

b. An uncertainty of 10°C in the temperature measurement would result in a factor of two in the limit of detection.

DMNA was chosen for further study, particularly to investigate the dependence of the concentration and laser energy on signal intensity. Figure 5 shows a plot of concentration vs. ion signal intensity of NO from DMNA. The plot reveals a linear relationship with the slope having a sensitivity value of 1.34 ± 0.04 (2SD). This linearity implies that the NO is formed in a collision-free environment and that this technique is suitable for quantification purposes. A detection limit of 450 ppb is obtained when using a background value, σ, of 200 μV.

The dependence of ion signal intensity on laser energy was also determined using the equation,

$$S_I = C(I_{226})^n \qquad (8)$$

where S_I is the ion signal intensity, I_{226} is the laser beam fluence, C is a system-dependent constant, and *n* the number of photons required to produce the NO^+ signal. The value of *n* was determined from the slope of a log-log plot. *A priori*, one might expect the value of *n* to be as high as four, two photons required to generate NO in its $X^2\Pi$ state and two photons to ionize it (see Figure 2). However, the slope of the plot yielded a value of $n = 1.62 \pm .01$ (2SD), indicating that the formation of the ion requires at least two photons. This near quadratic dependence suggests that one or more of the intermediate photochemical steps necessary to generate NO^+ from R-NO_2 is saturated. If this power dependence holds for higher laser energies, then an extrapolation of the signal intensity using a value of 20 mJ for the laser energy, not unreasonable with existing technology, yields a sensitivity value of 5000 greater than that measured for DMNA. This corresponds to a limit of detection of approximately 90 ppt.

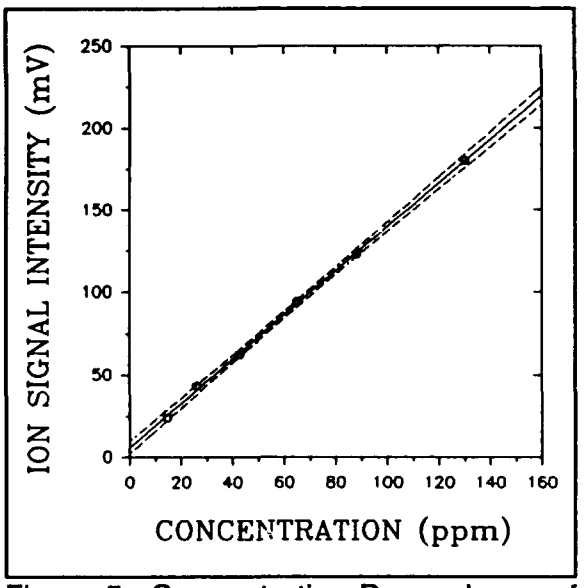


Figure 5. Concentration Dependence of the REMPI NO^+ Ion Signal. Solid line is a least squares fit of data (o) with 95% confidence (...).

For comparison, NO was also monitored by LIF via its $A^2\Sigma^+ - X^2\Pi$ (0,0) band following the photolysis of DMNA with 226 nm radiation. The limit of detection was estimated as 2 ppm, a factor of approximately 4 worse than that obtained with the REMPI technique. This difference results mainly from the higher background noise inherent in the LIF technique. The selectivity of the LIF technique was also found to be 10-100 times lower than the REMPI technique and is attributed to the relatively large bandwidth of the monochromator and interference filter used.

B. Chemical Agent Simulant:

The CW agent simulant DEMPS was also studied using the MB/MS apparatus. DEMPS was excited using UV radiation resonant with the PO $A^2\Sigma^+ - X^2\Pi$ transition

near 247 nm. Under our experimental conditions, the PO radical could not be detected. Focusing the laser beam was to no avail. This is not too surprising since the absorption cross-section of DEMPS is relatively small at this particular wavelength and the laser energy used, approximately 100 μJ, is a factor of 10-100 less than that typically required for the photofragmentation of other organophosphonates.[1, 8] The PO radical was detected, however, when using 193 nm radiation with a pulse energy of ~ 10 mJ. The spectra obtained under molecular beam and effusive flow conditions are presented in Figures 6a and 6b, respectively. At this wavelength the ionization process competes well with photodissociation and, as a result, fragment ions are produced. In both spectra the peak located at 29.0 μsec, corresponding to a mass-to-charge (m/e) of 47, is attributed to the PO^+. In Figure 6a, the prominent peaks at 23.6, 33.6, and 47.0 μsec correspond to m/e values of 31, 63, and 123, respectively. These are due to the fragmentation of the parent ion and are assigned to P^+ (m/e = 31), PS^+ and/or PO_2^+ (m/e = 63), and $[CH_3\text{-}(PS)\text{-}O\text{-}C_2H_5]^+$ (m/e = 123). The two peaks at 18.9 and 26.8 μsec correspond to m/e values of 40 and 20, respectively. They are attributed to Ar^+ and Ar^{++}, respectively, generated by the ionization of the carrier gas. The limit of detection for DEMPS employing the PO^+ was calculated as 6 ppm.

CONCLUSION

A novel ultra-sensitive detector for trace NO_2 containing compounds has been developed at the Ballistic Research Laboratory for the purpose of monitoring energetic materials such as propellants and explosives. It employs one laser operating at 226 nm and is based on the photofragmentation of the target molecule with subsequent detection of NO by REMPI and/or LIF via its $A^2\Sigma^+ - X^2\Pi$ (0,0) band origin. The analytical utility has been demonstrated on a number of compounds, including NO_2, DMNA, TNT, and RDX using a MB/TOF apparatus. Detection limits in the ppb were shown for RDX and TNT with extremely low laser energies, ~ 100 μJ/pulse. The ultimate sensitivity is no doubt much better than that shown here. Values in the low ppt are projected by just employing a laser

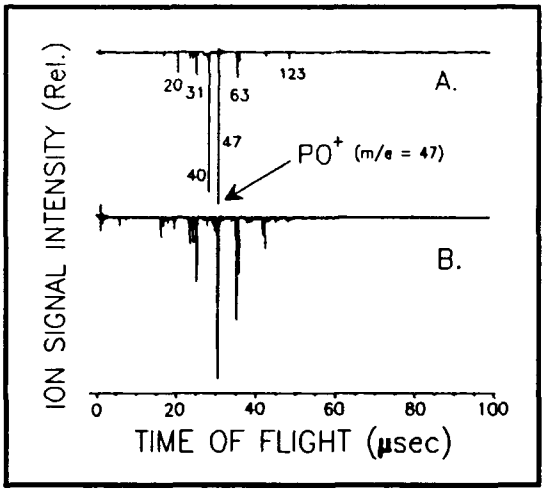


Figure 6. Time-of-Flight Spectra of DEMPS using 193 nm Radiation.
A. Molecular beam.
B. Effusive flow.

generating an output of only 20 mJ/pulse. Sample introduction and background noise could also be improved, thus yielding a lower limit of detection.

Ionization using 193 nm radiation coupled with MB/TOF spectrometric detection was employed in the detection of trace atmospheric vapors of DEMPS by monitoring the characteristic PO^+ fragment. A detection limit in the ppm was demonstrated. The ultimate sensitivity is projected in the sub-ppb with the application of higher laser energy. It is difficult, however, to predict the results for CW agents since their photoionization/fragmentation patterns may be quite different. Thus, direct examination will ultimately be required.

ACKNOWLEDGEMENTS

We thank Drs. S. Bulusu (ARDEC), R.S. Long (CRDEC), and R.A. Pesce-Rodriguez (BRL) for providing us with some of the sample compounds, and Dr. B.E. Forch (BRL) for helpful discussions. Support from the BRL/NRC Postdoctoral Research Program (GWL & JBS) and the PIF/OSD Capital Investment Program (RCS & AWM) is acknowledged.

REFERENCES

1. R.C. Sausa, A.W. Miziolek, and S.R. Long, *J. Phys. Chem.* **90**, 3994, (1986).
2. R.C. Sausa, A.J. Alfano, and A.W. Miziolek, *Applied Optics* **26(17)**, 3588, (1987).
3. M.J. McQuaid, A.W. Miziolek, R.C. Sausa, and C.N. Merrow, *J. Phys. Chem.* **95**, 2713, (1991).
4. P.J. Dagdigian, W.R. Anderson, R.C. Sausa, and A.W. Miziolek, *J. Phys. Chem.* **93**, 6059, (1989).
5. J. Schendel, R. Hohmann, and E.L. Wehry, *Applied Spectroscopy* **41(4)**, 640, (1987).
6. E. Wehry, R. Hohmann, J. Gates, L. Guilbault, P. Johnson, J. Schendel, and D. Radspinner, *Applied Optics* **26(17)**, 3559, (1987).
7. D.B. Moss, K.A. Trentelman, and P.L. Houston, *J. Chem. Phys.* **96(1)**, 237, (1992) and references therein.
8. S.R. Long, R.C. Sausa, and A.W. Miziolek, *Chem. Phys. Lett.* **117**, 505, (1985).
9. A.W. Miziolek, R.C. Sausa, B.F. Forch, and S.R. Long, US Army Ballistic Research Laboratory Technical Report, BRL-TR-2792, April 1987.
10. M.J. McQuaid and R.C. Sausa, *Applied Spectroscopy* **45(5)**, 916, (1991).
11. K.J. Smit, *Journal of Energetic Materials* **9**, 81, (1991).
12. C. Capellos, P. Papagiannakopoulos, and Y. Liang, *Chem. Phys. Lett.* **164**, 533, (1989).
13. J.-C. Mialocq and J.C. Stephenson, *Chem. Phys. Lett.* **123**, 390, (1986).
14. V.M. Donnelly and F. Kaufman, *J. Chem. Phys.* **69(4)**, 1456, (1978).
15. R.J.S. Morrison and E.R. Grant, *J. Chem. Phys.* **77(12)**, 5994, (1982).
16. R.J.S. Morrison, B.H. Rockney, and E.R. Grant, *J. Chem. Phys.* **75**, 2643, (1981).
17. G.B. McKendrick, C. Fotakis, and R.J. Donovan, *J. Photochemistry* **20**, 175, (1982).

TITLE: Accurate Measurement of Metal Diffusion in Electronic Materials using Secondary Ion Mass Spectrometry and Scanning Auger Microscopy

Dr. Stephen N. Schauer,
Geo-Centers Inc., Fort Monmouth, NJ 07703-5601.

Dr. Richard T. Lareau* and Dr. Howard S. Lee
U.S. Army Electronic Technology and Devices Laboratory
LABCOM, Fort Monmouth, NJ 07703-5601.

INTRODUCTION

Secondary ion mass spectrometry (SIMS) and scanning Auger microscopy (SAM) are two analytical techniques which permit measurements of three-dimensional elemental distributions [1,2,3]. SIMS is capable of measuring concentrations varying from values as high as a percent to as low as the part-per-billion (ppb) level. The spatial resolution can resolve sub-micron features, and in optimum circumstances depth resolution is better than 50 angstroms. SAM has the advantages of excellent spatial resolution and semi-quantitative information without standards at concentrations as low as 0.1 %, and excels in surface specificity.

Conventional ohmic contacts are typically alloyed in order to achieve low contact resistance between the contact metal and the semiconductor. These contact metals must not penetrate through the active device regions, which are typically less than 100 nm thick, hence, the necessity to include a thicker buffer or cap layer. The ability to accurately determine even trace amounts of diffusion is necessary for fundamental studies of contact formation, in order to develop devices for future applications, as well as failure analysis of devices presently in application.

We often encounter lateral inhomogeneities when analyzing device structures. Until recently, when depth profiling using SIMS or Auger electron spectroscopy (AES) averaged the signal over the entire analyzed area, so spatial information was very limited. We are now able to collect data in an image format, pixel by pixel, and deconvolute concentration gradients in a post analysis data processing system.

SIMS and SAM both rely on an ion beam to sputter the sample in order to measure elemental concentrations as a function of depth. This ion beam causes mixing of the sample at least to the depth of penetration, typically on the order of 50 angstroms. When going from a high concentration to a low concentration, this depth resolution is further limited by knock-on effects, memory effects from the sample chamber surfaces contaminated during sputtering in the high concentration region and crater edge effects, where the element of interest is being sputtered from the crater walls and re-deposited in the sample crater. Depth resolution is further limited by any surface topography or roughness, which cause different depths to be

analyzed simultaneously, resulting in a signal from an average depth, instead of a more finite depth. Metal interfaces are typically rough, and alloys inhomogeneous in the lateral and depth directions. Roughening can be enhanced by the ion beam, particularly in metals. An innovative technique was pioneered at ETDL to circumvent this problem [4,5] and since that time has become internationally recognized as an effective analytical method [6,7,8]. This technique is referred to as SIMS back-side depth profiling, indicating that the sample is profiled from the back surface instead of the front. Since semiconductor wafers are too thick to sputter through, the semiconductor substrate must be carefully thinned (wafers are typically ~600 μm thick; a realistic sputter depth is <5 μm).

Metal contacts on semiconductor substrates are an excellent example of the benefits of back-side depth profiling. Conventional, front-side depth profiling will show an apparent depth of penetration of the contact metal which is much greater than the actual diffusion length. Backside profiling can deconvolute true diffusion from ion beam induced artifacts.

Several studies are discussed which indicate the potential these analytical techniques have when analyzing semiconductor device structures, and illustrate the three-dimensional analyses which these techniques are capable of obtaining. The first two studies involved extensive materials investigations on ohmic contacts to both n^+ and p^+ III-V electronic device structures. The first study involves n^+ and p^+ contacts being studied for use in HBT device structures, where again the metal penetration depths into the active device areas are of importance. The second contact structure to be studied incorporated a quasi-epitaxial layer of Ge followed by a thicker Au outer contact. In this case, investigations center around understanding the role of Ge and Au in forming a non-alloyed ohmic contact, and the elemental interaction between the metal atoms and substrate Ga and As atoms, in addition to the metal penetration depths achievable. The third study involves an ohmic contact to a novel device, the dual-channel high electron mobility transistor (DCHEMT) fabricated at ETDL. The success of the final device design was achieved after changes were suggested as a result of back-side depth profiles from the initial device structures. The final example involves a failure analysis study at ETDL on a fully processed active device which typically is implemented in critical military systems, such as the Phoenix Missile System. The success of this study was only possible with application of the techniques and capabilities detailed here.

EXPERIMENTAL

A Cameca IMS-3f secondary ion mass spectrometer was used in these studies. This double-focussing mass spectrometer is capable of providing a chemical image with sub-micron spatial resolution when used in a micro-probe mode, or rapid analyses when used as an ion microscope. The primary ion beams used were O_2^+, to provide high sensitivity for electropositive elements, and Cs^+, to provide high sensitivity for electronegative elements and for sub-micron ion imaging.

Scanning Auger microscopy (SAM) and secondary electron microscopy (SEM) were performed with a Perkin-Elmer PHI 660 instrument. These techniques provided elemental and topographical maps, with approximate quantitation. In this case, the sputter ion beam used was Ar^+, which provided good surface removal rates and the possibilities of monitoring carbon and oxygen.

Samples which were analyzed using the back-side SIMS technique were prepared following the procedure developed at the US Army Electronic Technology and Devices Laboratory (ETDL) [4,5]. The metal contacts were grown on GaAs substrates which had 500

angstrom marker layers each alternately doped with silicon at concentrations of 5 X 10^{18} and 5 X $10^{16}/cm^3$ to provide an accurate depth calibration to correlate metal penetration depths. Beneath the marker layers was a buffer layer of GaAs followed by a chemical etch stop of $Al_{0.3}Ga_{0.7}As$. The samples were mounted to a glass slide with crystal bond, then mechanically lapped to a thickness of ~30 microns. The samples were chemically etched to the AlGaAs etch stop in a H_2O_2 solution buffered with ammonium hydroxide to a pH of 8.4, which selectively etches GaAs 20 times faster than AlGaAs [9]. The etching was done in a jet etching system to provide uniform etching; this system permitted both continuous rotation of the etched sample and a continuously pivoting and recycling jet spray system. The end point was monitored by the surface morphology; the polishing scratches disappeared when the $Al_{0.3}Ga_{0.7}As$ layer was reached. Past experience with thinner samples showed a change in color when the etch stop layer was reached, due to a change in the index of refraction. This process is shown schematically in Fig. 1. The etch stop was reached faster at the edges of the sample, and etching was stopped when enough area was exposed for an analysis. Since the samples were mounted on a glass slide, they were electrically isolated, which would cause a charge build-up during analysis. The samples were coated with ~300 angstroms of gold to provide a conductive path from the sample to the sample holder.

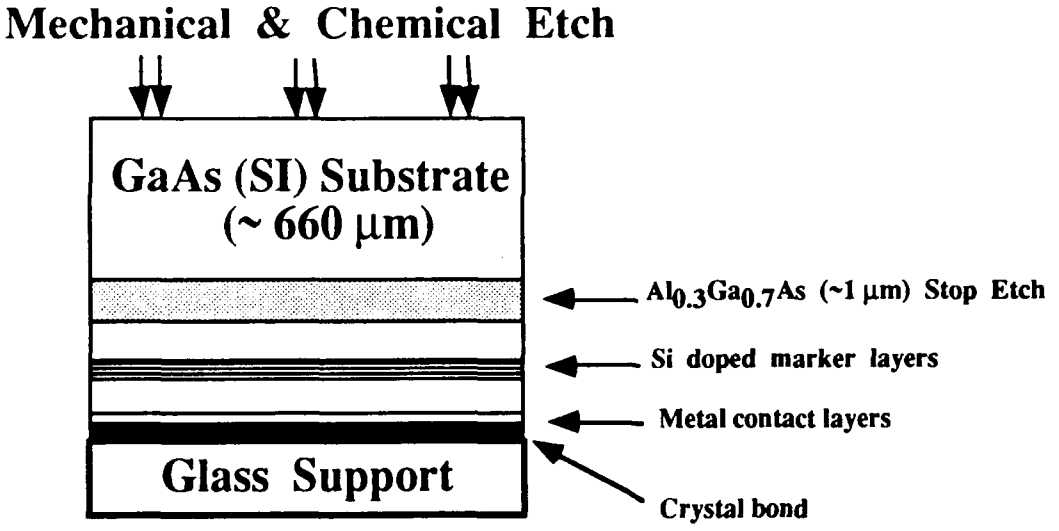


Figure 1. Schematic illustration of the preparation of a sample for the back-side depth profiling technique.

RESULTS

Excellent data can often be obtained on contacts with relatively routine analytical techniques. GeMoW contacts are being studied as ohmic contacts to p^+ and n^+ GaAs [10]. These contacts have been demonstrated to form ohmic contacts to both n and p type GaAs over a limited temperature range. The mechanism of this contact formation, and degradation at

higher annealing temperatures is currently being studied. Figure 2 shows the elemental distribution of W, Mo, Ge, Ga and As as a function of depth in an as-grown contact area on Be-doped GaAs. This same structure is shown in Figure 3 after annealing in an InAs overpressure at 780 C. It is readily apparent that considerable diffusion of Be, Ge, Mo, Ga, As and In occurs when these contacts are annealed.

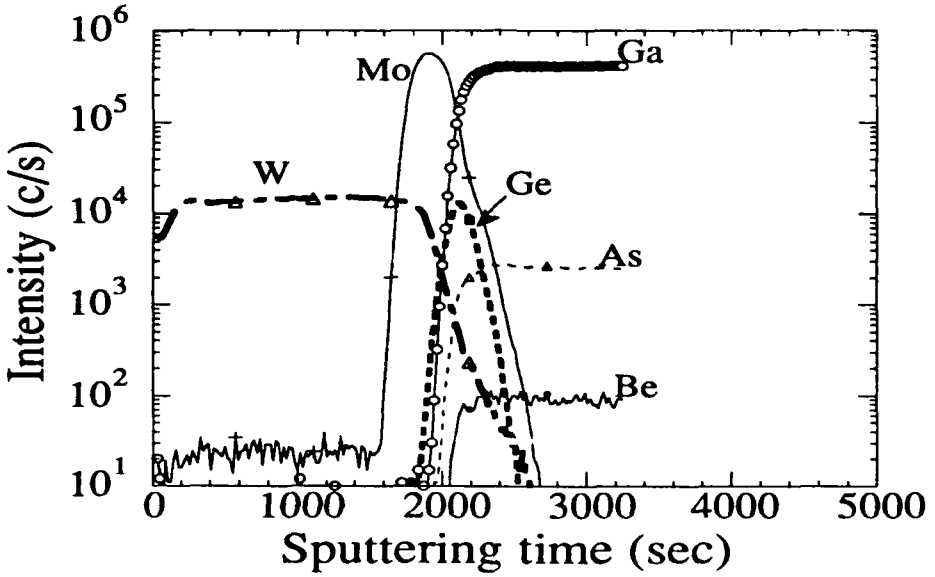


Figure 2. SIMS depth profile of an as deposited GeMoW ohmic contact.

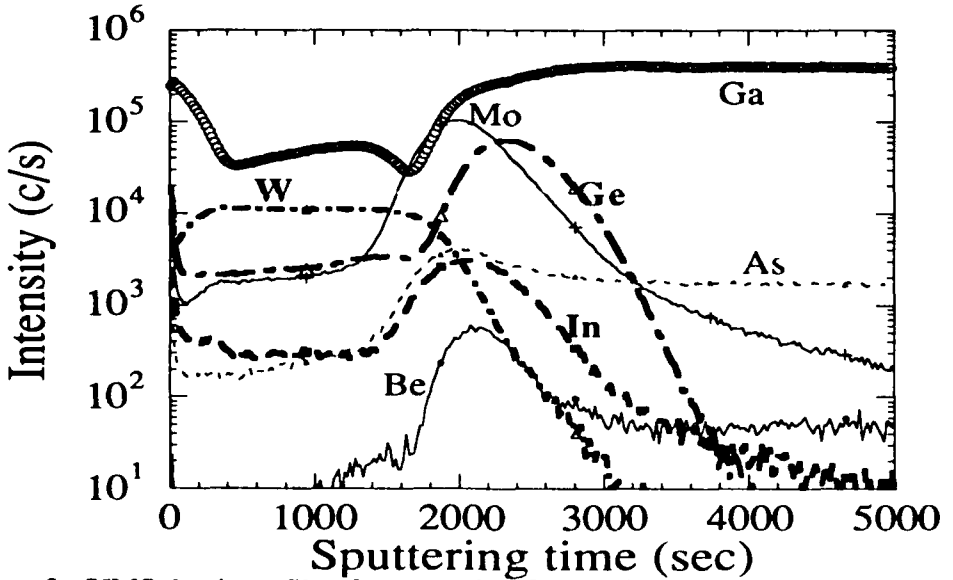


Figure 3. SIMS depth profile of an annealed GeMoW ohmic contact.

SIMS quantitation is notoriously difficult, since ion yields can vary by orders of magnitude, depending on the sample matrix and analyzed element. The yield of Auger electrons is less variable, so semi-quantitative data is possible without standards. The SIMS signal intensities of Ga, As, Be and In in the metal contact region were extremely high in the annealed sample, therefore, these same structures were analyzed using AES, in an attempt to measure the diffusion of Be, Ga, As, Ge, Mo, In and W. The As moved significantly into the Ge and Mo, but was below the detection limit for W, indicating the concentration was $\leq 0.1\%$. The concentration of Ga, Be and In were below the detection limits, putting an upper limit to the diffusion of these elements at ≤ 0.1 %.

Many structures contain important details, which can be very complicated to analyze. One example of this is a Au-Ge ohmic contact to n-GaAs grown in a UHV Molecular Beam Epitaxy (MBE) system [11]. The Ge was epitaxial with the GaAs and the Au was highly oriented. Ohmic behavior of the contacts was observed after annealing at 320 C for 3 hours. For these conditions specific contact resistivity and contact measurements on TLM patterns yielded average values as low as 7×10^{-6} Ω-cm^2 and 0.28 Ω-mm, respectively. Figure 4 is a SEM picture of the surface of an annealed Au/Ge contact on GaAs. Non-uniformities are evident across the surface. Figure 5 is an Auger depth profile of the as-grown sample. Figure 6 and 7 are depth profiles of the annealed sample measured in the dark and light regions, respectively. These two figures are from the same crater, and the data were extracted by image processing after the analysis. It is clear from the two distinctly different profiles that diffusion is occurring non-uniformly in a lateral direction, possibly forming nucleation sites.

Figure 8 compares SIMS front-side depth profiles of the as-grown and annealed samples. Considerable out-diffusion of both As and Ge is revealed in these profiles. The gold and germanium appear to have diffused considerably into the GaAs substrate in both samples. It is unlikely that diffusion will occur in the as-grown sample, so the apparent diffusion is due to ion beam mixing and interface roughness. Figure 9 presents ion images of the annealed sample. Similar distributions are observed here as in the previous SEM image (see Fig. 4). Again using image processing techniques, a cross-sectional image of the annealed sample is shown in figure 10. Note both lateral and in-depth elemental distributions are not uniform, possibly indicating metal 'spiking' or 'pin-hole' formation. This is essentially a pictorial format of figure 8. The ability to present the data in several different formats often gives a better qualitative understanding of the material interactions taking place.

To determine the actual amount of diffusion, the samples were analyzed using the back-side depth profiling technique. The depth profiles were collected using the imaging capability, in order to determine if any diffusion was laterally non-uniform. Figure 11 shows the back-side depth profiles of the as-grown and annealed samples. The metal penetration depths, calibrated to the marker layers, is approximately 4000Å. Lateral images of the annealed sample are shown in Fig. 12. It is apparent that the Au and Ge did not diffuse uniformly into the GaAs substrate. The select area depth profiles of the areas with high, low and total Ge and Au concentrations are compared in Fig. 13. It is important to realized that all three profiles, for both elements, were obtained simultaneously in the same crater, and deconvoluted during the post-analysis data reduction. The areas to compare were determined from the images in Fig. 12. Notice that almost all the diffusion occurs in the localized region. This would be impossible to determined without the imaging capability. This study is the first application of the back-side imaging depth profiling technique.

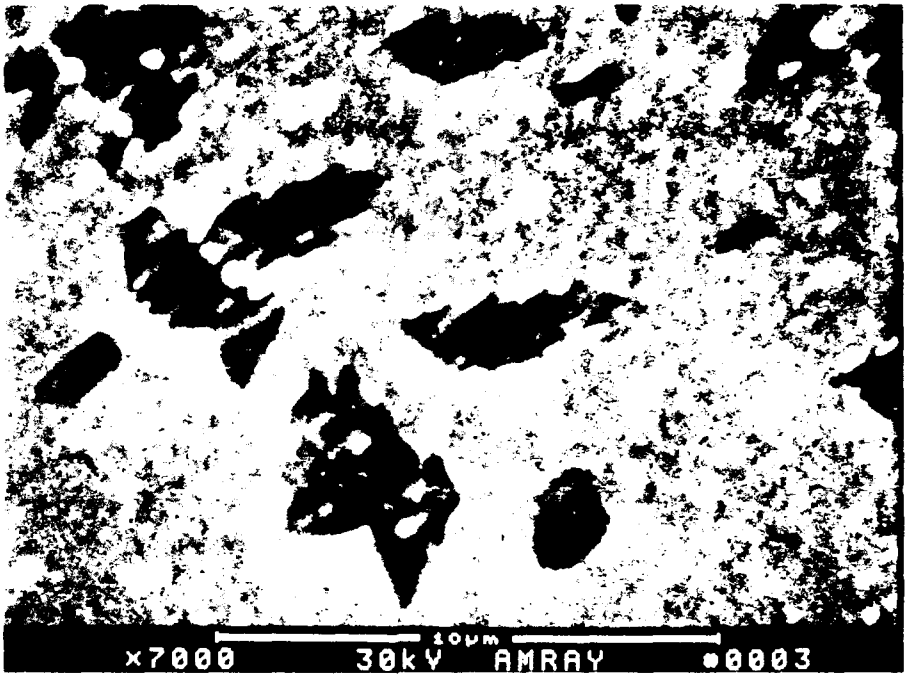


Figure 4. SEM picture of the surface of a AuGe contact on a GaAs substrate.

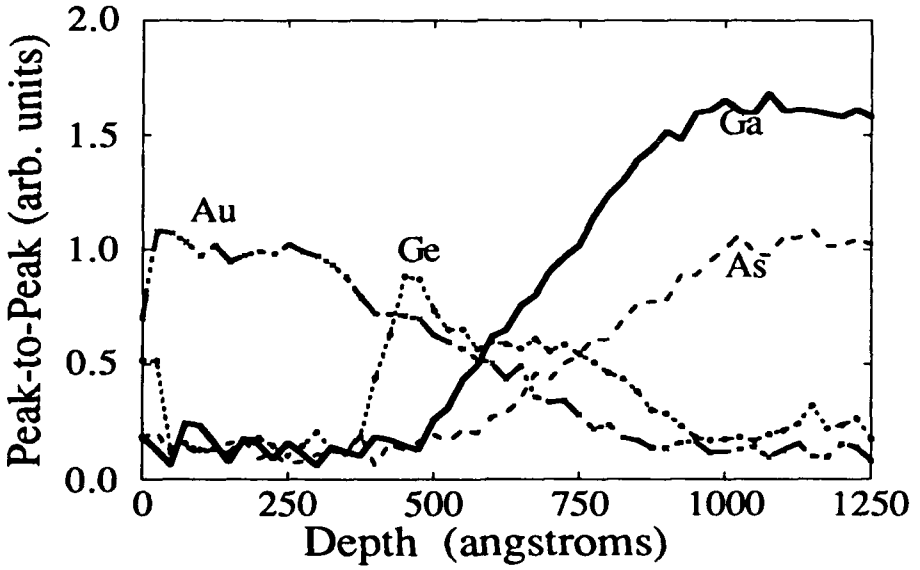


Figure 5. SAM depth profile of an as-grown AuGe contact.

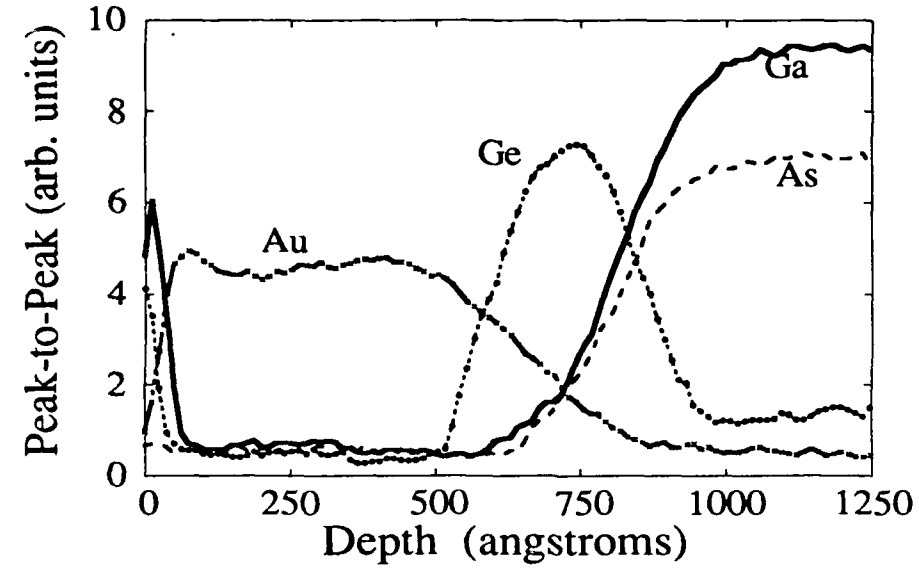


Figure 6. SAM depth profile of an annealed AuGe contact in light regions.

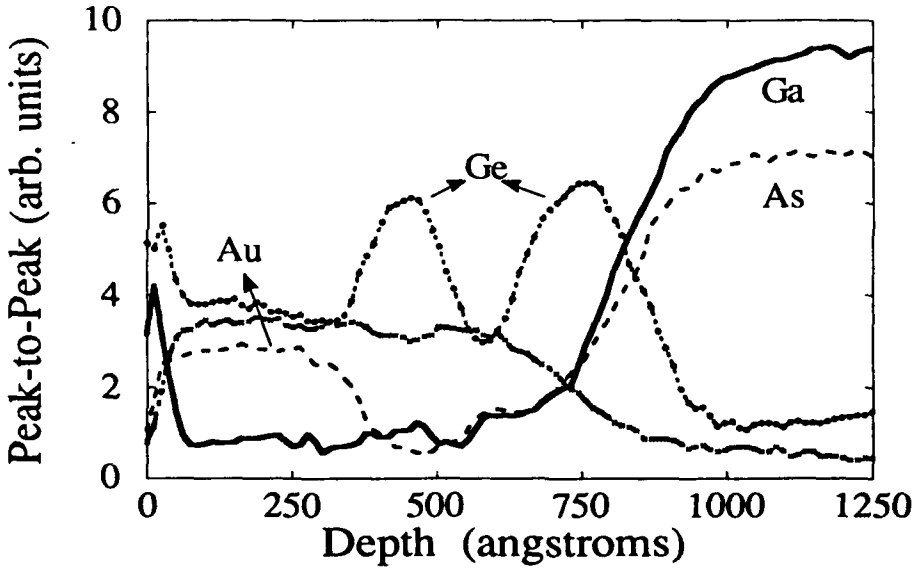


Figure 7. SAM depth profile of an annealed AuGe contact in dark regions.

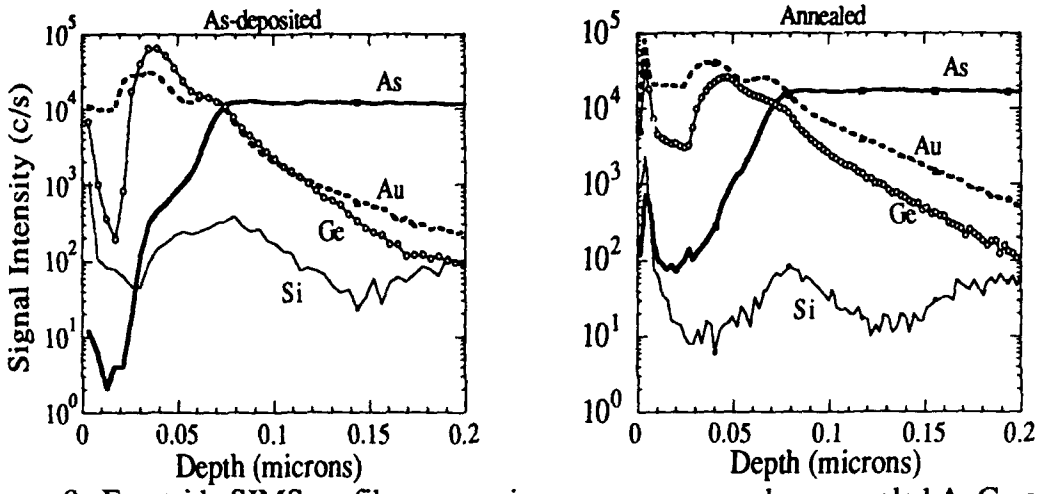


Figure 8. Frontside SIMS profiles comparing an as-grown and an annealed AuGe contact.

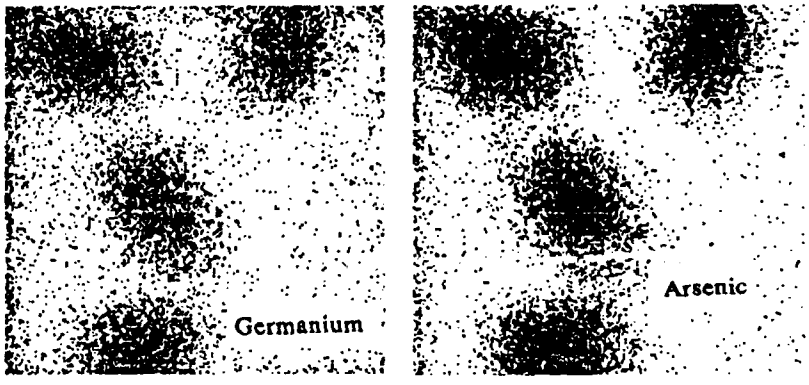


Figure 9. Ge and As ion images obtained during the depth profile in figure 8 (annealed).

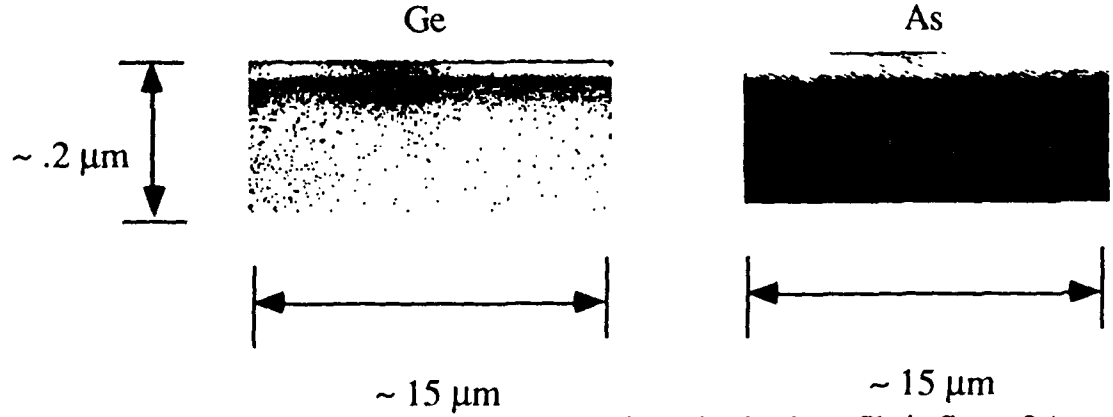


Figure 10. Cross sectional images of Ge and As from the depth profile in figure 8 (annealed).

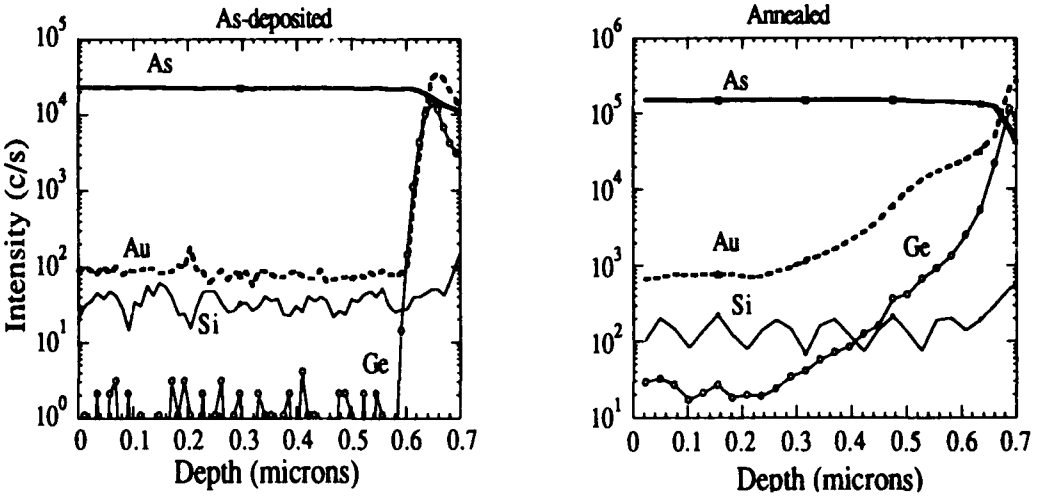


Figure 11. Backside SIMS depth profile comparing an as-grown and an annealed AuGe contact.

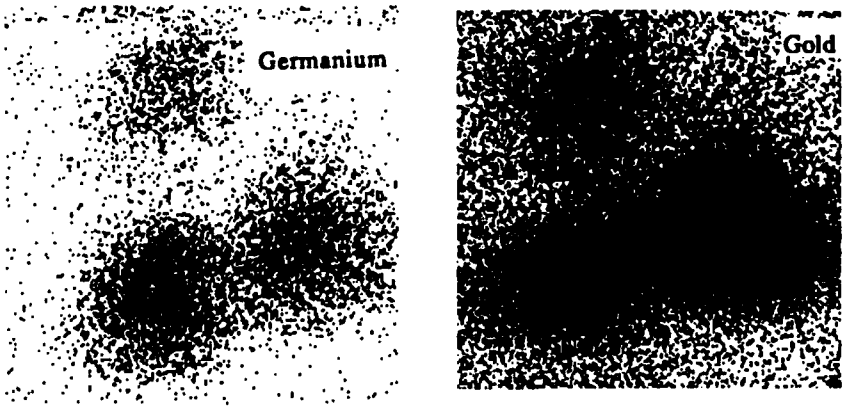


Figure 12. Ion images of Ge and Au from the back-side depth profile of the annealed AuGe contact in figure 11.

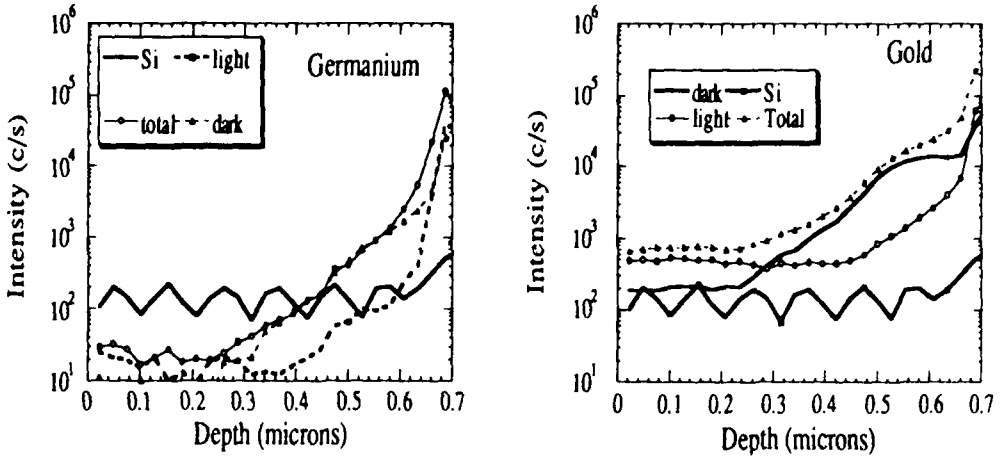


Figure 13. Selected area profiles of Ge and Au correlating the light, dark and total regions of figure 12, along with the Si marker layers used to calibrate the depth.

The back-side SIMS depth profiling technique has also been applied for trouble-shooting prototype electronic devices. Figure 14 schematically illustrates ETDL's novel design for a DCHEMT. This device is the result of an attempt to produce an ultrahigh-speed inverter, as conceived by Iafrate et al [12] and fabricated and tested at ETDL [13]. It employs a dual-channel structure in which a two-dimensional electron gas (2DEG) is formed on either side of a 50 nm GaAs quantum well that is sandwiched between two n-AlGaAs layers with appropriate spacers. Under zero bias, band bending forces electrons from the upper channel into the lower channel, allowing conduction only through the lower channel. Under forward bias, electrons are drawn into the upper channel from below, which reverses the effect. Biasing allows current-steering behavior to be obtained from a single device. This operation is possible due to the ability to contact each channel individually.

Figure 15 shows the SIMS front-side depth profile which confirms the initial electrical measurements and reveals the metal contact penetration into this 2DEG region and presence of severe Al and Si diffusion. Some of these results could be explained as SIMS artifacts, particularly when rough interfaces between each metal and contacting semiconductor surfaces are concerned. From previous work involving similar ohmic contact structures and back-side sputter depth profiling, it is clear that significant metal penetration from the contacts to the active device layers does occur, dictating the use of thicker n+ GaAs contacting layers to be included in the electronic device layer designs [4]. Figure 16 shows the back-side depth profile of the same device structure with a thicker (~3000Å) n+ GaAs contacting layer. The 2DEG region has little diffusion and is not effected by the contact metal penetration. The front-side depth profile of this sample, which has the thicker n+ GaAs contacting layer, still indicates metal penetration deep into this region. The back-side profile demonstrates that the contact metals do not actually diffuse into the device structure. The back-side sputtering approach avoids sputtering effects due to surface roughness between the MBE layers and contact

metallization layer(s), while at the same time providing the desired reverse concentration gradient of the penetrating metal layers.

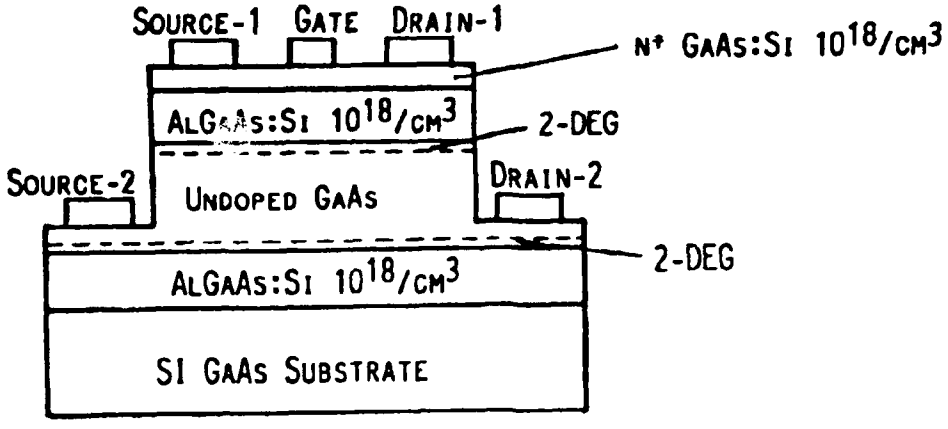


Figure 14. Schematic representation of a dual-channel HEMT device.

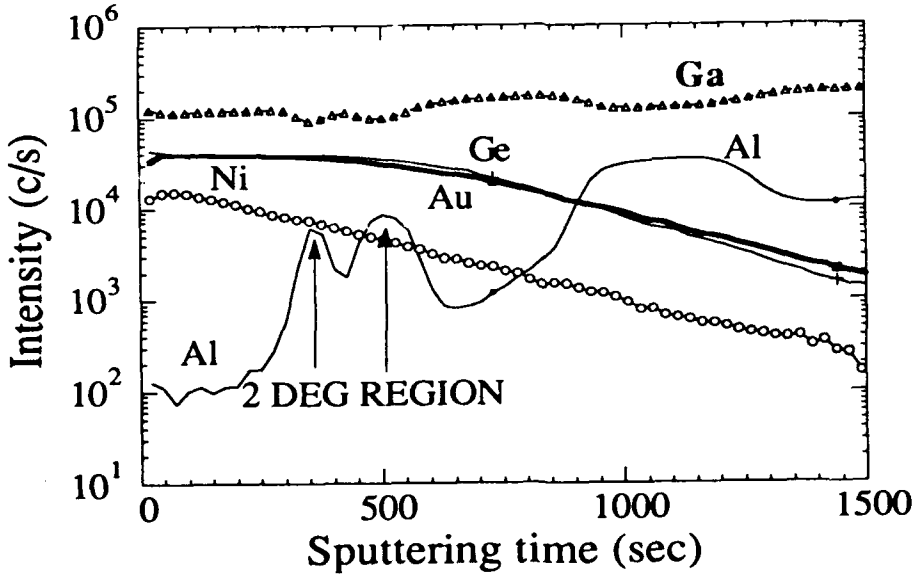


Figure 15. SIMS front-side depth profile of a dual-channel HEMT device.

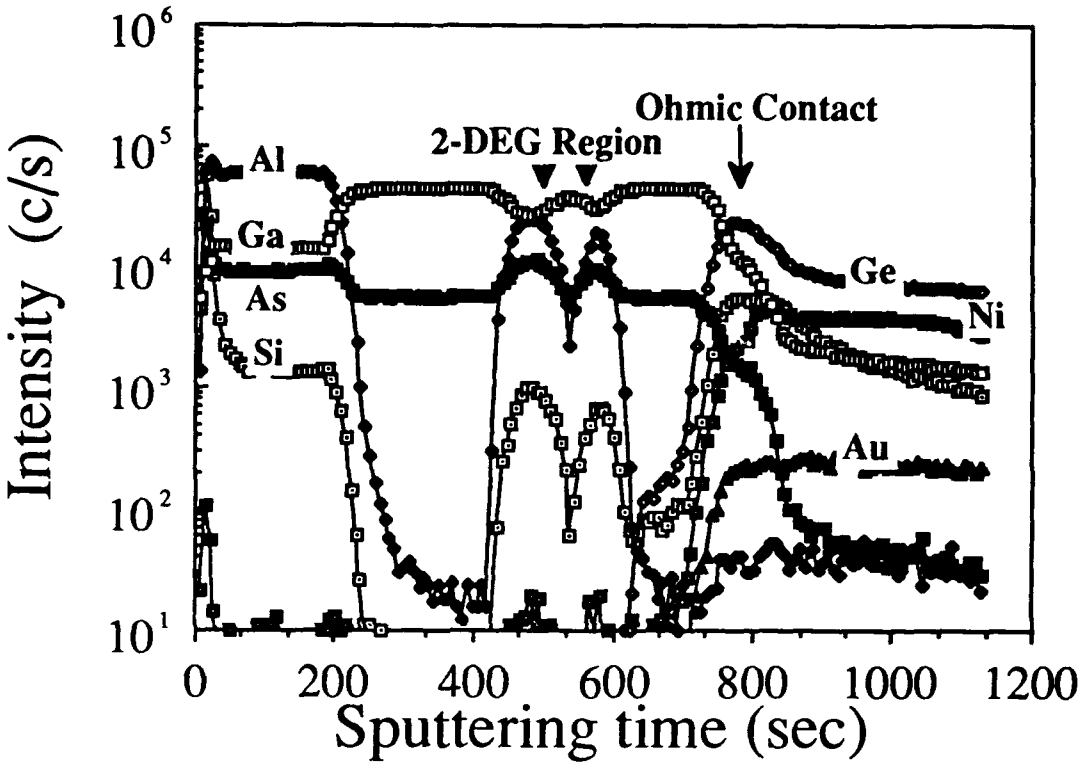


Figure 16. SIMS back-side depth profile of a dual-channel HEMT device.

We have also encountered samples that are of critical military importance which are well suited to the back-side sputter depth profiling method. IMPATT diodes, which were being used for the Phoenix Missile System, were believed to be failing due to a diode "burn-out" phenomena, whereby the sidewall of the mesa structure would blister and permit gold to flow from top to bottom contact, hence, shorting the electrical path. Although most of this work was performed by secondary electron microscopy at ETDL, the low level doping and metal penetration profiles provided additional and desirable information easily obtainable from SIMS. The device structure of the IMPATT diode (used in the Phoenix Missile System) required little sample preparation to measure a depth profile from the back of the contact structure since this diode is contacted from each side. Simple chemical etching of the top ohmic contact bonding pad allowed profiling through the entire III-V semiconductor structure as well as the metal contact structure from low to high concentration gradient. Presented in Figure 17 is a schematic representation of the intended IMPATT diode structure. The back-side depth profile is presented in Figure 18.

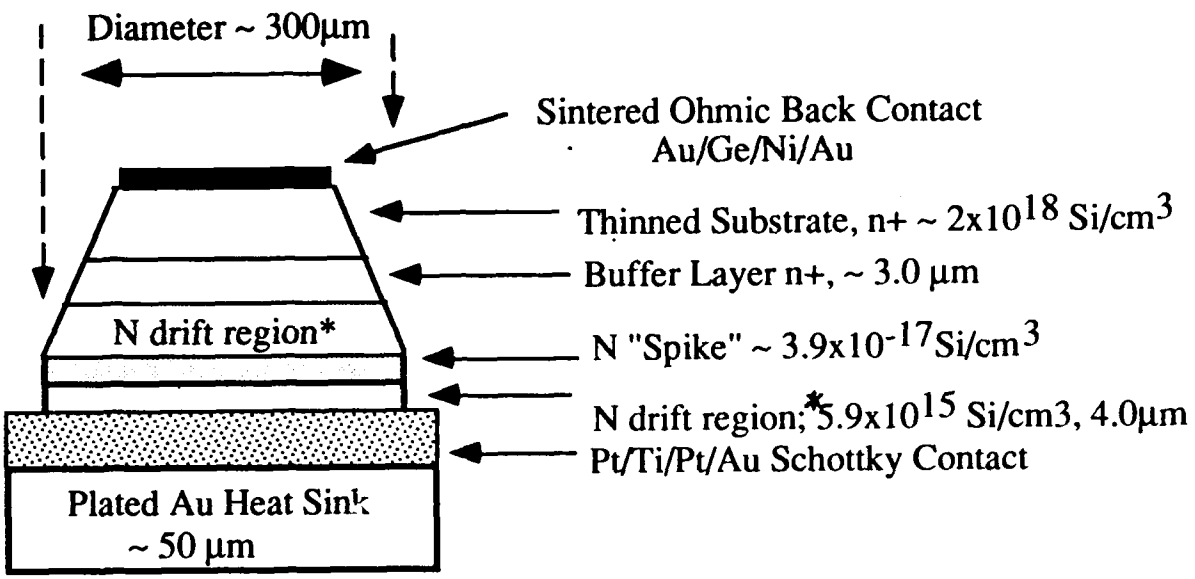


Figure 17. Design structure of the Phoenix Missile/IMPATT diode.

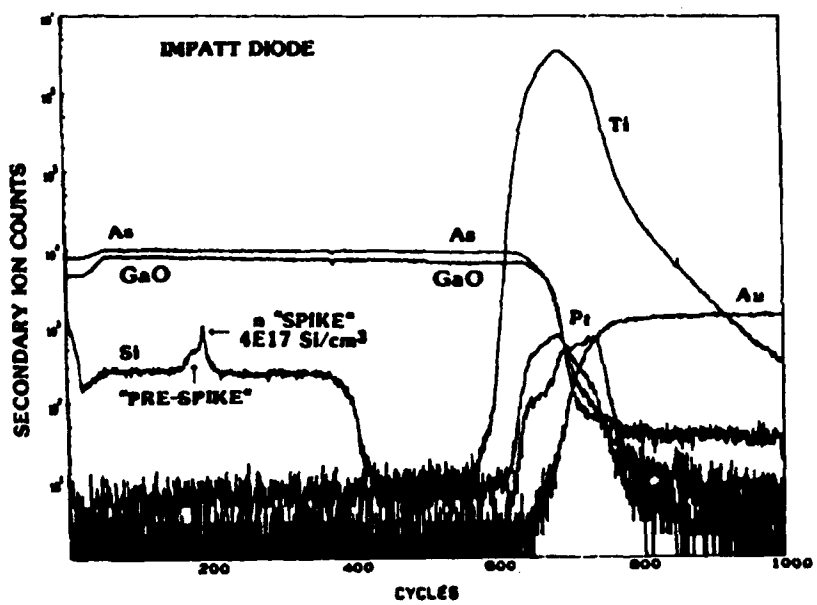


Figure 18. SIMS back-side depth profile of a failed IMPATT diode.

The back-side depth profile reveals two important features: (1) The Ti from the lower contact is penetrating very deep into the device layer and (2) The doping layer(s) consisting of Si at $5.9 \times 10^{15}/cm^3$ level was quite higher than the original design structure, and the "spike" area was proceeded by a Si diffusion front. As in the previous studies, the additional information gained from performing back-side depth profiling and SEM is invaluable in designing and fabricating electronic devices.

DISCUSSION

The accurate measurement of elemental distribution in semiconductor devices is essential in developing and trouble-shooting advanced designs. Without sensitive analytical techniques, scientists and engineers are often operating blindly when trying to explain and understand device properties without accurate correlation to material properties and behavior.

Future work in this area includes expanding the back-side sputter technique to Auger electron spectroscopy and Rutherford backscattering spectroscopy.

ACKNOWLEDGEMENTS

We would like to thank Capt. Ken Merkel of Wright Laboratory (WPAFB) for his ongoing collaboration on novel contact structures through the Tri-Service Reliance Program.

REFERENCES

[1] R.G. Wilson, F.A. Stevie and C.W. Magee, *Secondary Ion Mass Spectrometry* (John Wiley and Sons, New York, 1989).

[2] A. Benninghoven, F.G. Rudenauer and H.W. Werner, *Secondary Ion Mass Spectrometry* (John Wiley and Sons, New York, 1987).

[3] D. Briggs and M.P. Seah, Eds., *Practical Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, (John Wiley and Sons, New York, 1983).

[4] J.R. Shappirio, R.T. Lareau, R.A. Lux, J.J. Finnegan, D.D. Smith , L.S. Heath and M. Taysing-Lara, *J. Vac. Sci. Technol.* **A5**, 1503 (1987).

[5] R.T. Lareau, in *Secondary Ion Mass Spectrometry, SIMS VI*, A. Benninghoven, A.M. Huber and H.W. Werner, Eds. (John Wiley and Sons, New York, 1988) pp. 437-440.

[6] J.A. Jackman, A. Kular, L. Weaver, D. Mayer, T.E. Jackman and C. MacPherson, *Surf. Interface Analysis* **15**, 451 (1990).

[7] C.J. Palmstrom, S.A. Schwarz, E. Yablonovitch, J.P. Harbison, L.T. Florez, T.J. Gmitter , E.D. Marshall and S.S. Lau, *J.Appl. Phys.*, **67**, 334 (1990).

[8] J. Herniman, J.S. Yu and A.E. Staton-Bevan, *App. Surf. Sci.*, **52**, 289 (1991).

[9] K. Kenefick, *J. Electrochem. Soc* , **129**, 2380 (1982).

[10] K. Merkel, G.J. Trombley, R.W. Dettmer, C.I. Huang, G.D. Robinson, R.T. Lareau and S.N. Schauer, *IEEE proceedings*, 1991.

[11] H.S. Lee, R.T. Lareau, S.N. Schauer, R.P. Moerkirk, K.A. Jones, S.Elagoz, W.Vavra and R. Clarke, in *Advanced III-V Compound Semiconductor Growth, Processing and Devices*, S.J. Pearton, J.M. Zavada and D.K. Sadana, Eds. (Mater. Res. Soc. Proc. 240, Pittsburgh, PA, 1992) in press.

[12] G.J. Iafrete, L.C. Poli, T.R. AuCoin and L.F. Heath, "Dual Channel High Electron Mobility Field EffectTransistor", S.N. 228795, July 29, 1988.

[13] R. Khanna, M.B. Das, D.D. Smith, G.J. Iafrate and P.G. Newman, *IEEE Elect. Device Letters*, **10**, 531 (1989).

Basic Fibroblast Growth Factor Accelerates Repair and Prevents Degeneration in Retinal Laser Lesions

Mr. STEVEN T. SCHUSCHEREBA,* Dr. PHILLIP D. BOWMAN, Mr. DAVID J. LUND, Ms. JULIE A. QUONG, Mr. JOSEPH A. VARGAS, Mr. RONALD E. FERRANDO, AND Mrs. JANET M. YANG.
Division of Ocular Hazards, Letterman Army Institute of Research, Presidio of San Francisco, CA 94129-6800.

INTRODUCTION

In future military conflicts, temporary or permanent blindness caused accidentally or intentionally by laser exposure from laser range finders, designators, or anti-eye laser weapons[1,2,3] will produce medical emergencies, and long term care problems that have not been previously encountered. Experience with ocular injury from mustard gas exposure during WW-I [4,5] and the recent Iran-Iraq war[6] indicates that care of the blinded demands large numbers of personnel. Lasers, however, penetrate deep into the eye and may produce not only immediate blindness but also long term complications that can lead to further or total blindness.

The biomedical implications of the use of laser devices by the U.S. Army have been the subject of study for the last three decades[1,3]. For laser systems operating in the visible and near-infrared spectral region, the retina is the most vulnerable end organ because the cornea and lens focus light upon the absorbing layers (usually photopigments and melanin) of the retina. The focusing properties of the ocular optics cause a collimated laser beam to be focused to an extremely small retinal image (10s of microns), onto which the concentrated energy produces thermal, mechanical, and chemical processes that alter the retina. Injury produced by lasers to the retina of animals has been extensively characterized ophthalmoscopically, histologically, and functionally[7,8,9]. These changes vary from severe damage, accompanied by massive retinal and vitreo-retinal hemorrhaging[10] to small burns characterized by focal retinal opacities[11]. All ophthalmoscopically visible and some subvisible burns can permanently affect vision[9,12,13] because they destroy a certain number of photoreceptors which, as part of the central nervous system, cannot be replaced.

Numerous reports on accidental laser injury to human eyes have been documented since the advent of lasers[7,9,14,15,16,17,18]. Laser ocular injury to humans has shown that retinal scarring, retinal traction, holes, and large scale photoreceptor loss can occur, depending on the laser type and the amount of energy absorbed[3,9,18]. Currently, no treatment for laser eye injuries exists. Understanding the nature of the injury and devising a means to treat it has been a major focus of this research.

The existence in the eye of polypeptides with growth stimulatory properties has been recognized for some time[19,20,21,22]. They have been shown to stimulate mitogenic activity in a wide variety of cell types and are thus termed growth factors[23]. This growth-promoting activity is largely attributable to the family of so-called fibroblast growth factors (FGF) (first assayed for stimulation of proliferation of fibroblasts in cell culture, hence the name), which are concentrated in ocular and brain tissue. In addition to mitogenic activity, FGFs are now recognized as having multifunctional properties affecting mesoderm induction, angiogenesis, and tissue repair[21, 24,25,26]. Evidence also indicates that FGFs affect the maintenance of neural and glial cell functions[27]. Various studies have shown that basic fibroblast growth factor (bFGF) administration delays genetic retinal dystrophy[28] and promotes retinal regeneration[27,29], suggesting that it is also a trophic factor for neural tissue[19,20,21,22]. With recent advances in molecular biology it is possible now to produce relatively large amounts of polypeptide substances of sufficient purity for therapeutic use.

MATERIALS AND METHODS

Thirty-five New Zealand Red (NZR) rabbits were used and all procedures involving animals conformed to policies set forth in the NIH Guide[30]. NZR rabbits were anesthetized with a mixture of ketamine, xylazine, and acepromazine[31] and eyes were dilated with neo-synephrine and tropicamide.

The power at the cornea produced by the multi-line argon laser (454-514.5 nm) ranged from 30 mW/sec to 267 mW/4 msec and was calibrated so that delivered energy was at or slightly below levels which would cause subretinal hemorrhaging. The corneal irradiance diameter was 3 mm and the beam divergence was 1 mrad. Each rabbit received 8 lesions (2 rows of 4 each in the central-inferior fundus of each eye (Figure 1). These irradiances initially caused reproducible lesions that were dense white in appearance and about 500 μm in diameter.

Within 15 minutes after laser exposure, rabbits were injected with 1 ml

of 10% sodium fluorescein (Alcon Labs, Inc., Fort Worth, TX) through the ear vein, and fundus photographs and fluorescein angiograms were made 10 min later. Immediately after photography, eyes were injected with either vehicle control (10 µl; 10 rabbits) or bFGF (10 µg/10 µl; 15 rabbits). An additional ten rabbits received bFGF in the right eye and vehicle in the left eye. A single lot of recombinant human bFGF (Amgen Biologicals: Thousand Oaks, CA) was used in all experiments. After the intravitreal injection procedure, local anesthetic was again applied to the surgical site, antibiotic (gentamycin sulfate jelly, 3 mg/g) was packed under the eyelids, and the lids taped closed until the animal recovered from anesthesia. After recovery, animals were returned to their cages.

On days 1-4 after irradiation and treatment, fluorescein angiogram and fundus photography procedures were repeated. After treatment and photography procedures on day 4, animals were terminated with an overdose of Nembutal and the retinas were processed for routine transmission electron microscopy (TEM). Both eyes were enucleated and immersion-fixed within 5 min of death at 4 ^{0}C in a pH 7.4 cacodylate-buffered glutaraldehyde and paraformaldehyde solution[32]. Individual strips of retina (2 X 8 mm) containing one row of four lesions were dissected out and processed for TEM [33]. The effects of treatment were evaluated in fundus photographs, fluorescein angiograms, and by light and TEM. By light microscopy, the center of the lesion was determined by following serial sections until the maximum diameter of the lesion was found. At this point, the length of missing outer nuclear layer (ONL) gaps were measured and outer nuclear layer (nuclei) counts made. Normal dehydration and processing of tissues caused shrinkage; therefore, lesion diameters from histologic sections show smaller values than those observed by funduscopy.

For cell proliferation studies in the injury site, bromodeoxyuridine (BrDU), an analog of thymidine that is incorporated into DNA of dividing cells (Amersham, Arlington Hts, IL) was injected intraperitoneally in 2 rabbits at 2.5 days and in 8 rabbits at day 4 after bFGF treatment. Animals were terminated 1 hr after BrDU injection and the eyes fixed in buffered formalin. BrDU immunocytochemistry was performed to locate proliferating cells [34,35].

RESULTS

With the exception of an initial and occasional inflammatory reaction at the injection site in both control and bFGF treated eyes, no adverse effects of either the route or the effect of the injection were noted. By day 4, the injection site was relatively quiescent and the small entry hole had healed uneventfully.

Fundus Photographs:

Immediately after laser exposure and on day 1, almost all of the lesions in both control and bFGF-treated eyes appeared equal in size and opacity (Figure 1A and B). By day 2 after bFGF treatment, however, most lesions were reduced in size and opacity compared to controls. In some animals noticeable reduction in size and opacity occurred as early as day 2 (Figure 1C and D) after bFGF-treatment. Control lesions usually began to show reduction in size and opacity by day 3. By day 4 after bFGF treatment, most lesions were markedly smaller and less opaque than controls.

Fluorescein Angiograms:

Immediately after laser exposure and on day 1, fluorescence intensity indicated that all lesions were about the same size and leaked fluorescein to the same extent (Figure 2A and B). By day 2 after bFGF treatment, however, compared to controls most lesions showed a marked decrease in fluorescence intensity which continued through day 4 (Fig 2C and D). When maximum lesion diameters were measured on fluorescein angiograms at day 4, lesions treated with bFGF were significantly smaller than the corresponding control lesions (Figure 3).

Light Microscopy:

Sections taken through the center of control lesions at day 4 showed a large zone of missing outer nuclear layer (ONL -photoreceptor nuclei), large vacuoles in the lesion periphery, and an absence of proliferating cells at the lesion periphery (Figure 4A). In contrast, sections taken through the center of bFGF-treated lesions (Fig 4B) showed smaller gaps in the missing ONL zone, an absence of large vacuoles and the presence of large numbers of cells at the lesion periphery. In some animals, a large central fibroblastic (cellular) plug that extended into the subretinal space was observed. These plugs appeared to be covered anteriorly with at least a single layer and sometimes multiple layers of proliferating RPE-like cells (Figure 4B). The histologic changes in bFGF-treated lesions were correlated with improved wound healing that was observed funduscopically.

A comparison of the maximum diameters at the center of the lesions at day 4 indicated no statistically significant differences between control and bFGF-treated groups, confirming that the initial lesions were highly reproducible at the level of the RPE (Figure 5). When the number of nuclei in the ONL above the lesion center were counted, more nuclei were present above the bFGF-treated lesions (Figure 6A) than in controls. In addition, the width of the missing ONL gap above the center of the lesions was smaller in the

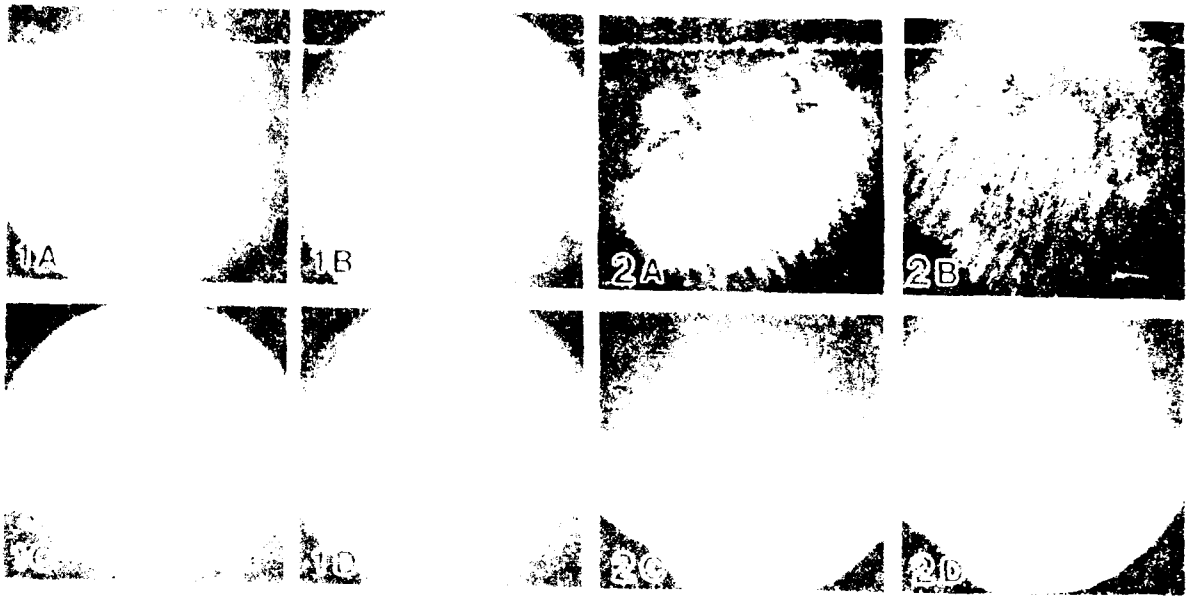


Figure 1. Fundus photographs of laser irradiated New Zealand Red rabbit eyes. A. Vehicle-treated eye immediately after laser exposure. B. bFGF-treated eye immediately after laser exposure. Both groups of lesions appear similar in size and severity. C. The same vehicle-treated eye represented in Figure 1A, but at day 2 after laser exposure. D. The same bFGF-treated eye represented in Figure 1B, but at day 2 after laser exposure. At day 2 the size and opacity of the lesions in the control are much larger and more visible than lesions in the bFGF-treated retina.

Figure 2. Fluorescein angiograms of the same eyes represented in Figure 1. A. Vehicle-treated eye immediately after laser exposure. B. bFGF-treated eye immediately after laser exposure. Both groups of lesions leak fluorescein to the same extent. C. The same vehicle-treated eye represented in Figure 2A, but at day 2 after laser exposure. D. The same bFGF-treated eye represented in Figure 1B, but at day 2 after laser exposure. At day 2, fluorescein leakage is readily apparent in the control but greatly reduced in the bFGF-treated lesions. Reduced fluorescein leakage in bFGF-treated eyes indicates accelerated repair in the blood-retinal-barrier.

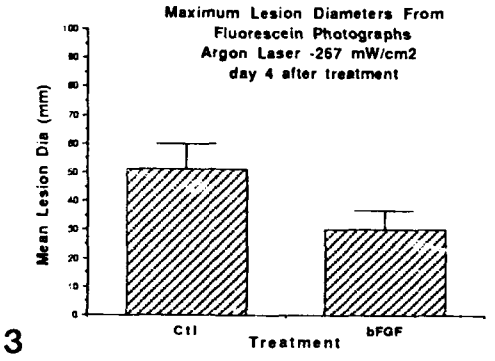


Figure 3. Histogram of maximum lesion diameters measured on representative fluorescein photographs day 4 after treatment. Graph indicates a significant decrease in maximum lesion diameter (i.e. a reduced area of leakage) after bFGF treatment.

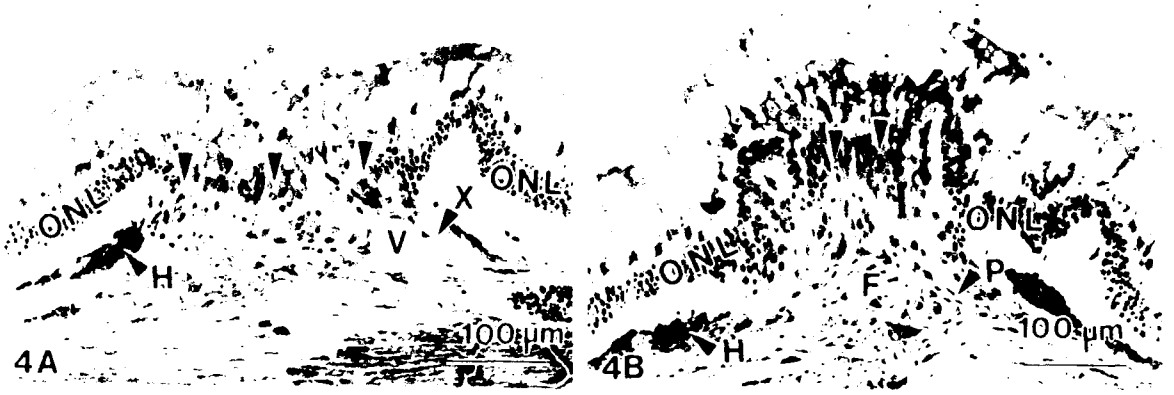


Figure 4 Light micrographs of control (A) and bFGF treated (B) lesions day 4 after laser exposure. In vehicle-treated lesions (A), a large portion of the outer nuclear layer (ONL) is missing (arrows) and a subretinal hemorrhage (H) is present. RPE cells are severely vacuolated (V) on either side of the lesion and little evidence exists for cell proliferation (X). B. In the bFGF-treated retina day 4 after exposure, a smaller portion of the outer nuclear layer (ONL) is missing in the center of the lesion (arrows). Although a subretinal hemorrhage (H) is present, there are no vacuolated RPE cells. A plug of fibroblasts (F) is present in the lesion center and extends into the subretinal space. Cell proliferation (P) of RPE-like cells is suggested by multiple layers. Enhanced cell proliferation of the retinal pigmented epithelial cells could explain the reduced fluorescein leakage because the blood-retinal-barrier in the rabbit is provided by these cells. Bars = 100 µm.

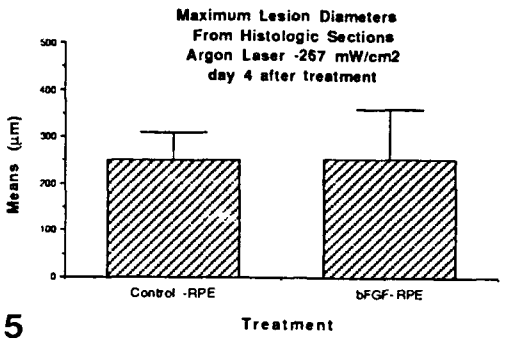


Figure 5. Maximum lesion diameters taken from semi-serial sections at day 4 after treatment at the level of the retinal pigmented epithelium indicate that lesion sizes were almost identical for both treatment groups. This evidence indicates that the initial lesions were uniformly produced.

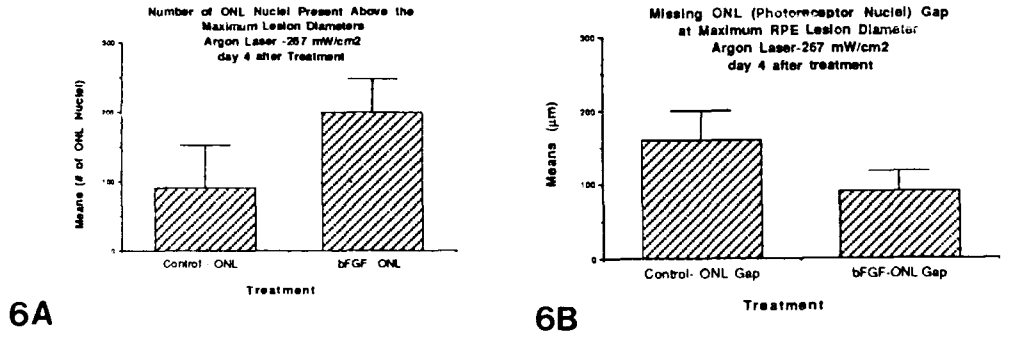


Figure 6. Graphs, at day 4 after exposure, indicating vehicle versus bFGF treatment effects on photoreceptor degeneration. A. Mean values for the number of photoreceptor nuclei (outer nuclear layer -ONL) above the maximum lesion diameters which were established in Figure 5. Evidence supports the concept that photoreceptor degeneration is prevented by bFGF treatment. B. The length (µm) of ONL gap created by the missing nuclei was measured on sections through the maximum diameters of the lesions. After bFGF treatment, more photoreceptor nuclei were present in the lesion centers than in controls. Conversely, the length of the missing photoreceptor nuclear zone was smaller after bFGF treatment than in controls.

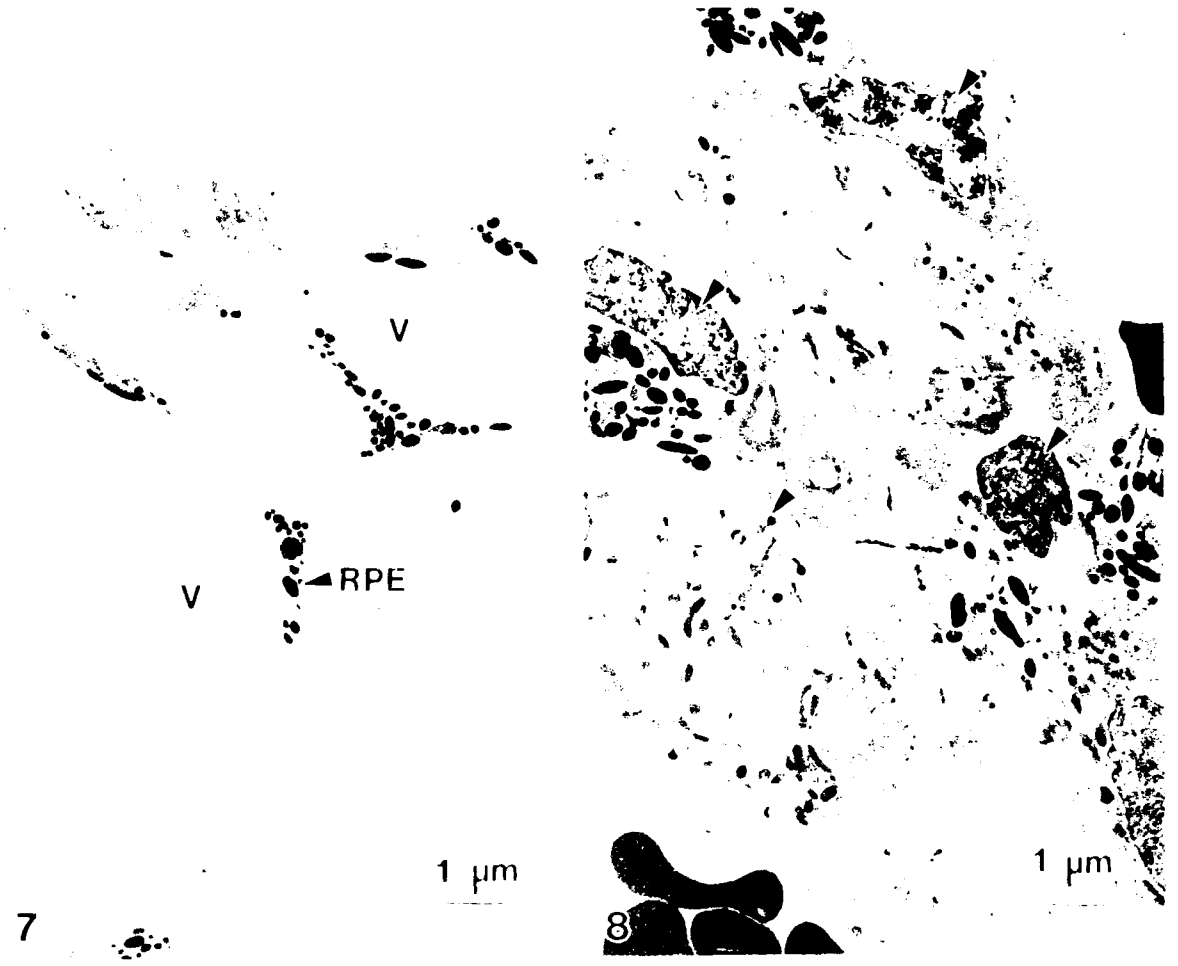


Figure 7. Transmission electron micrograph from the periphery of a lesion day 4 after laser exposure and vehicle treatment (see right side of Fig. 4A). Retina shows large vacuoles (V) and an absence of retinal pigmented epithelial (RPE) cell proliferation. Bar = 1.0 µm.

Figure 8. Transmission electron micrograph from the periphery of a lesion day 4 after laser exposure and bFGF treatment (see right side of Fig. 4B). Retina shows the absence of vacuoles and extensive multilayering of RPE-like cells (arrows). Bar = 1.0 µm.

Figure 9. Transmission electron micrograph from the periphery of a lesion day 4 after laser exposure and vehicle treatment. Retina shows short and stubby photoreceptor inner segments (IS) and an absence of outer segments. Retinal pigmented epithelium (RPE). Bar = 1.0 µm.

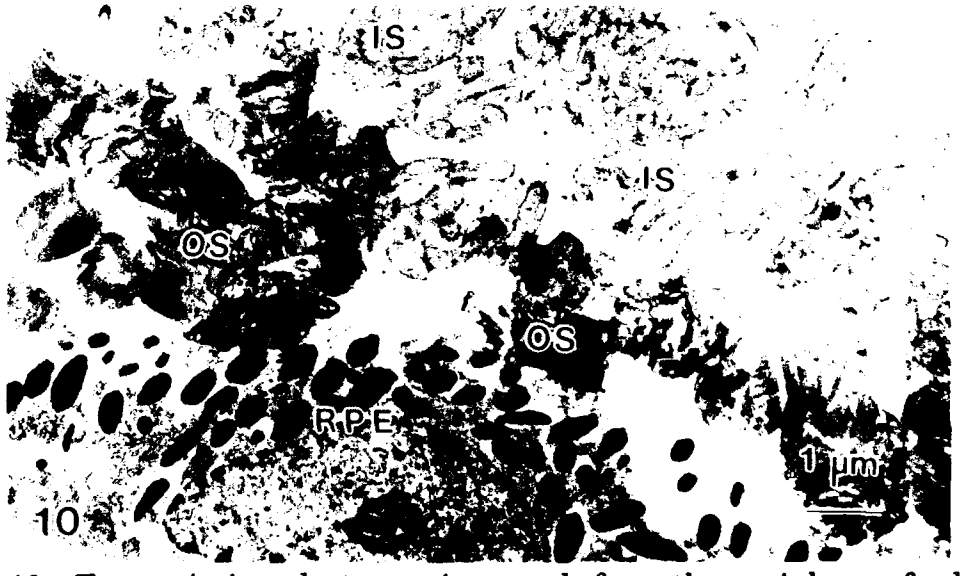


Figure 10. Transmission electron micrograph from the periphery of a lesion 4 days after laser exposure and bFGF treatment. Retina shows more normal inner segments (IS) and the presence of outer segments (OS) in the lesion periphery. The presence of OSs after bFGF treatment suggests that photoreceptor degeneration has been prevented. Retinal pigmented epithelium (RPE). Bars = 1.0 µm.

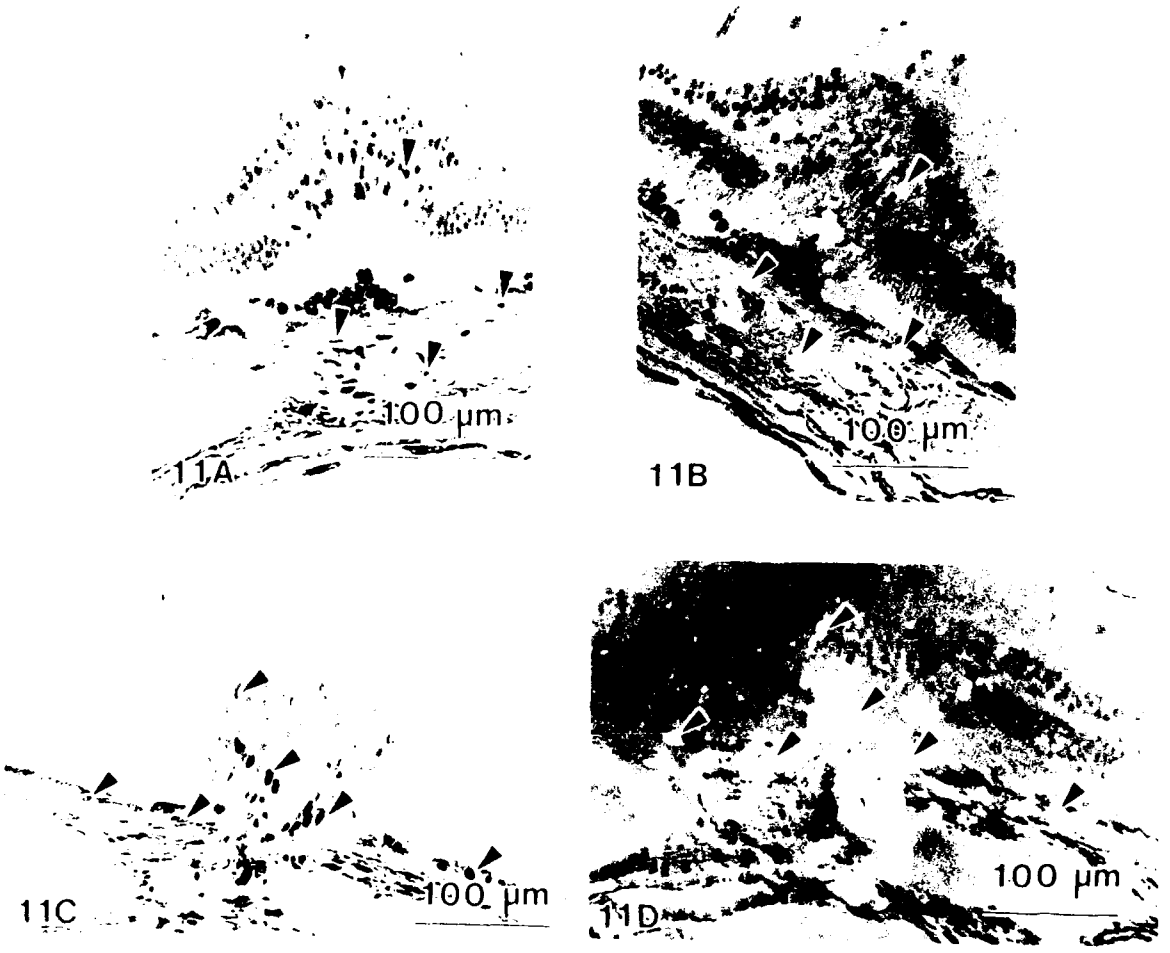


Figure 11 Retina day 4 after laser exposure and treatment. A. Phase-contrast photo of vehicle control retina after BrDU immunocytochemistry. BrDU immunolocalization is present as a dark precipitate (arrows). B. Epipolarization (enhances visualization of BrDU label) photograph of the same area in Fig. 11A showing BrDU labeling of cells (arrows). C. Phase-contrast photograph of bFGF-treated lesion. More BrDU immunolocalization is present (arrows) than in control. D. Epipolarization photo of the same area in Fig. 11C showing BrDU incorporation in cells (arrows). Note that more cells are shown proliferating in the choroid, RPE and outer retina (arrows) than in the control. Bar = 100 µm.

bFGF-treated group than in controls (Figure 6B). The data presented in Figure 6 indicate that photoreceptors were spared from degeneration in the periphery of the laser lesions.

Transmission Electron Microscopy:

The most striking finding by TEM was a consistent vacuolization of RPE cells at the periphery of control lesions (Figure 7). Sections taken from similar areas in bFGF-treated lesions showed an absence of large vacuoles and the presence of multiple layers of RPE-like cells at the lesion periphery (Figure 8). The layers of cells observed at the light microscopy level contained basal infoldings when observed by TEM, suggesting their derivation from RPE cells. In some of the control lesions, the photoreceptors in the immediate periphery appeared short and stubby with missing outer segments (Figure 9). In contrast, after bFGF treatment, photoreceptor inner and outer segments in the immediate periphery of the lesion appeared more normal (Figure 10). In the immediate periphery of bFGF-treated lesions, RPE cells were multilayered. In addition, some of the RPE cells in lesions showed elongated and thin processes that made contacts with normal and adjacent RPE cells. Normal desmosomal-like structures were present along the inner membranes of the thin cytoplasmic processes of the bFGF-treated RPE cells. Ultrastructurally, cell-cell junctional complexes were characteristic of those found in normal RPE cells.

BrDU Immunochemistry

BrDU-labeling of control lesions 2.5 days after laser radiation and treatment showed little label uptake in the retina, with the exception of an occasional nucleus taking up label in the INL (inner nuclear layer). Little labeling occurred in the choroidal layer. By 2.5 days after bFGF-treatment, however, several labeled nuclei were observed in the INL, most of them in the posterior portion of the INL and peripheral to the lesion centers. The location and the elongated shape of these nuclei suggests that they may be Müller or glial cells. Fibroblasts were also labeled in the choroid after bFGF treatment. Data indicate that 2.5 days after bFGF-treatment, slightly more RPE cells were seen labeled than in controls. By day 4 after treatment, controls showed BrDU labeling of only a few nuclei in the choroidal fibroblasts and vascular endothelial cells. An occasional nucleus was labeled in the ONL (Figure 11A and B). By day 4 after bFGF treatment, however, BrDU labeling was increased over controls in the INL glial cells, RPE cells, and choroidal fibroblast and vascular endothelial cells (Figure 11C and D). Glial cells that incorporated BrDU label at 2.5 days in the INL suggested that migration occurred into the ONL by day 4 after bFGF treatment.

DISCUSSION

New Zealand Red rabbits were used because the retina is homogeneously pigmented, similar to the human retina. This homogeneous pigmentation characteristic facilitated production of reproducible laser lesions as the absorbed energy was dissipated largely at the level of the pigmented layers. By funduscopy, lesions were readily observed against the dark background which provided information about the size and opacity of the lesions.

Fluoroscein angiography provided information about the functional integrity of the blood-retinal-barrier (BRB) which resides in the tight-junctions between retinal vascular endothelial cells and RPE cells. This structure acts as a permeability barrier and selectively transports nutrients into the retina while preventing the inward movement of many other blood-borne substances. The environment it maintains is vital to proper neural function. Sustained damage to the BRB leads to loss of function and ultimate degeneration of neural tissue, so restoration of the BRB after injury is critical for minimizing further damage. Because the rabbit retina is relatively avascular, the BRB is largely provided by the RPE and damage to it is the principal source of fluorescein leakage. Therefore, histologic changes to this cell layer would explain changes in leakage of fluorescein. Fluorescein angiography indicated that the amount of leakage through the BRB was clearly reduced after a single bFGF treatment.

TEM showed that after bFGF treatment, RPE-like cells in the lesion periphery were multilayered and exhibited more desmosomal-like junctional complexes, indicating cellular reassociation. After bFGF treatment, close apposition of RPE cell membranes containing dense zones suggested that junctional complexes were reformed more rapidly than those in the controls.

By TEM, photoreceptors on the periphery of bFGF-treated lesions retained more of their outer segments than control retinas evaluated from similar regions. These data indicate that photoreceptor degeneration was slowed or that normal function was maintained by bFGF treatment. The possibility that bFGF provides some important maintenance function to photoreceptor outer segments is suggested by studies that have demonstrated the presence of fibroblast growth factor receptors on outer segment membranes[36] and the selective binding of bFGF to the extracellular matrix surrounding photoreceptors[37].

Preliminary proliferation studies with BrDU incorporation indicate that choroidal fibroblast and vascular endothelial cells, RPE cells, and glial cells (including Müller cells) are stimulated to proliferate after a single bFGF treatment. In fact, Müller cells are known to provide nutritive support

functions for photoreceptors[38] and bFGF has previously been shown to be sequestered in close proximity to the apical surface of Müller cells[37]. This suggests that Müller cell activity may be regulated by bFGF and may contribute to accelerating repair and preventing photoreceptor degeneration.

In conclusion, treatment of argon laser lesions with a single intravitreal injection of recombinant human bFGF showed accelerated restoration of the BRB and healing. Stimulation of cellular proliferation of RPE cells, glial cells and choroidal cells and their enhanced function through either increased activity or numbers is a possible explanation for the beneficial effect. Prevention of photoreceptor degeneration is evident by the presence of more photoreceptor nuclei and more normal photoreceptors at the lesion periphery. For the first time, results suggest a potentially useful treatment regimen to improve the prognosis of retinal laser injuries that soldiers may encounter on the modern battlefield. Future work in our laboratory will focus on evaluating combinations of growth factors for additive or synergistic effects.

ACKNOWLEDGMENT: We gratefully acknowledge the direction and support of Bruce E. Stuck, Chief, Division of Ocular Hazards, the editorial comments of Susan E. Siefert, and the skillful technical assistance of Frank Lezotte and Andre A. Akers.

REFERENCES

1. B. E. Stuck, *Proceedings of Conference on Combat Ocular Problems*, (LAIR, Presidio of San Francisco, CA, 1980), pp. 57-73.
2. B. Tengroth and B. Anderberg, *Laser and Light in Ophthalmol.* **4**, 35-39 (1991).
3. J. Mellerio, et al., *Laser and Light in Ophthalmol.* **4**, 41-67 (1991).
4. M. W. Ireland, in *The Medical Department of the US Army in the World War.* (Government Printing Office, Washington, DC, 1926), vol. 14. pp 557-559.
5. W. F. Hughes, in *Ophthalmologic Reviews of the Wilmer Ophthalmological Institute of Johns Hopkins University and Hospital Archives of Opthalmology.* (Baltimore, MD, 1929).
6. M. Balali, *Proc. 1st World Congress on Biol. and Chem. Weapons,* (1984), pp. 254-259.
7. A. S. Rathkey, *Arch. Ophthalmol* **74**, 346-348 (1965).
8. J. Marshall, A. M. Hamilton, and A. C. Bird, *British Journal Opthalmology* **59**, 610-630 (1975).
9. H. Zwick, et al., *LAIR Inst. Report 463* , (LAIR, Presidio of San Francisco, CA, 1991) pp 1-22.

10. W. D. Gibbons and R. G. Allen, *Health Physics* **35**, 461-469 (1978).
11. D. J. Lund and E. S. Beatrice, *Health Physics* **56**, 631-636 (1989).
12. J. Marshall, *Health Physics* **56**, 617-624 (1989).
13. A. I. Goldman, W. T. Ham Jr., and H. A. Mueller, *Exp. Eye Res.* **24**, 45-56 (1977).
14. T. L. Curtin and D. G. Boyden, *Am. J. Ophthalmol.* **65**, 188-189 (1968).
15. G. I. Osipov and M. M. Pyatin, *Vestnik oftalmologii* **1**, 50-51 (1978).
16. B. J. Fowler, *Ann. Ophthalmol.* **15**, 481-483 (1983).
17. V.-P. Gabel, R. Birngruber, and B. Lorenz, *Health Physics* **56**, 705-710 (1989).
18. L. Haifeng, et al., *Health Physics* **56**, 711-716 (1989).
19. D. Barritault, C. Arruti, and Y. Courtois, *Differentiation* **18**, 29-42 (1981).
20. D. Caruelle, et al., *J. Cell Biochem.* **39**, 117-28 (1989).
21. D. Gospodarowicz, G. Neufeld, and L. Schweigerer, *Cell Different.* **19**, 1-17 (1986).
22. J.-C. Jeanny, et al., *Exp. Cell Res.* **171**, 63-75 (1987).
23. D. Barritault, et al., *Journal of Neuroscience Research* **8**, 477-490 (1982).
24. P. Buntrock, K. D. Jentzsch, and G. Heder, *Exp. Path.* **21**, 62-67 (1982).
25. J. M. Davidson, et al., *J. Cell Biol.* **100**, 1219-1227 (1985).
26. A. Y. Fourtanier, et al., *Invest. Dermatol.* **87**, 76-80 (1986).
27. C. Pittack, M. Jones, and T. A. Reh, *Development* **113**, 577-588 (1991).
28. E. G. Faktorovich, et al., *Nature* **347**, 83-6 (1990).
29. C. M. Park and M. J. Hollenberg, *Dev Biol* **134**, 201-5 (1989).
30. National Institutes of Health, Guide for the Care and Use of Laboratory Animals (1980).
31. J. G. Fox, B. J. Cohen, and F. M. Loew, in *Laboratory Animal Medicine* (Academic Press Inc., San Francisco, CA, 1984) pp. 545.
32. M. J. Karnovsky, *J Cell Biol* **27**, 137A-18A (1965).
33. S. T. Schuschereba, et al., *Toxicol. Path.* **18**, 103-123 (1990).
34. N. J. Gonchoroff, *J. Immunol Methods* **93**, 97-101 (1986).
35. H. G. Gratzner, *Science* **218**, 474-475 (1982).
36. F. Mascarelli, D. Raulais, and Y. Courtois, *The EMBO Journal* **8**, 2265-2273(1989).
37. G. S. Hageman, et al., *Proc. Natl. Acad. Sci. USA* **88**, 6706-6710 (1991).
38. T. Kuwabara and D. G. Cogan, *Arch. Ophthalmol.* **66**, 96-104 (1961).

Non Contact, Non Destructive Optical Method for Measuring Electric Fields in GaAs, AlGaAs and InGaAs heterostructures

Hongen Shen[1], Dr., Mitra Dutta*, Dr., and Wayne Chang, Mr.
U.S. Army Electronics Technology and Devices Laboratory, SLCET-ED
Fort Monmouth, N.J. 07703-5601

Photoreflectance (PR)[1-4] is used extensively in the study of semiconductor microstructures[4-8], the sharp derivative-like spectra used to determine various quantum features. There has been recent growing interest in the use of Franz-Keldysh oscillations (FKO)[1-3,9] to determine the surface and or interface electric field [9-12], crystal quality[12,13], doping concentration[9], surface Fermi level (E_{pin}), and/or its relative changes ($\Delta E_{pin}/E_{pin}$) under different surface treatments[14,15]. While $\Delta E_{pin}/E_{pin}$ can be accurately determined [15], the determination of the absolute value of the E_{pin} is difficult as it depends on doping concentration and its distribution. We have done measurements on GaAs, AlGaAs and InGaAs samples using a special structure first proposed by Van Hoof et al.[16] with a surface electric field almost independent of doping concentration. However in the Van Hoof experiments the electric field deduced from FKO was only half the value of the "real field" derived from an electrostatic calculation using E_{pin} = 0.7-0.8 eV. Other studies[15,17] measuring E_{pin} using PR also reported values less than 0.7 eV. We have thus reexamined the FKO technique for measuring built-in field, and our measurements reported below reaffirm the usefulness of FKO. Our experiments demonstrate that illumination in a normal PR experiment can significantly affect the measurement, leading to an erroneously reduced value of electric field. Extreme caution must be exercised in the determination of the low field condition and previous anomalous results reassessed. In this paper we shall also discuss several applications for this technique : 1) determination of Fermi level (E_{pin}) of bare GaAlAs surface as a function of Al mole concentration; 2) measurement of the strain induced piezoelectric field in (111)B InGaAs-GaAs, a potential material for electro-optic devices; 3) the Fermi level pinning mechanism in low temperature MBE grown GaAs, a important material for insulating buffers in GaAs devices and ICs.

An undoped top layer of GaAs (or $Ga_{1-x}Al_xAs$) of thickness d was grown by

[1]GEO-Centers, Lake Hopatcong, N.J. 07849. Work performed at US army ETDL.

molecular beam epitaxy (MBE) on n^+ (100) GaAs substrates with a n^+ doped 1.0 μm thick GaAs (or $Ga_{1-x}Al_xAs$) buffer layer. Samples with different d and x as well as different growth conditions such as various growth rates, different arsenic to gallium fluxes, As_2 and As_4 sources, a range of substrate temperatures, indium and indium-free mounting were studied. The conduction band configuration is schematically shown in Fig. 1. The electric field in the top undoped layer given by the slope of the band edge is created by positive charges in the depleted region of n^+ layer and negative charges trapped in the surface states (s-i-n^+ structure). Neglecting the deviation from constant slope due to small background doping in the undoped layer, the field is E_{pin}/qd, where q is the electron charge. The small band bending near the undoped/n^+ interface was found to be small. The major advantages of this s-i-n^+ structure are (a) a combination of high electric field and small broadening parameter, leading to a slow damping[2] of FKO and allowing more accurate determination of the electric field; (b) the surface Fermi level derived from the field is insensitive to the n^+ concentration; (c) the field in the top layer, the major cause of error in E_{pin} determination, is controllable by the thickness d which is known to better than 3% from RHEED.

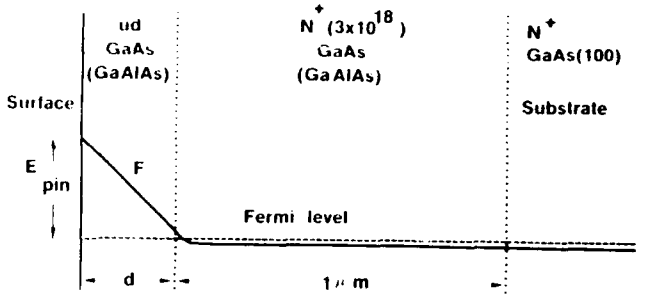


Figure 1 Conduction band diagram of a GaAs s-i-n^+ structure.

Photoreflectance spectra[19] are taken at room temperature. The probe beam is from a tungsten lamp and quarter-meter monochromator combination attenuated by neutral density filters. This beam is defocused on the sample to further reduce the power density. The pump beam is from either the 5145 Å line of an air cooled Ar^+ laser or from the 6328Å line of a He-ne laser with attenuation and defocusing. The power density for the probe and pump beam is less than $0.1\mu W/cm^2$. Use of such low power density reduces photon induced flattening of band bending and allows determination of E_{pin} under the unperturbed condition. Shown in Fig. 2 (a) is the PR spectra from a s-i-n^+ GaAs structure with d=2000 Å. FKO from E_0 and $E_0+\Delta_0$ gap are observed up to the 20^{th} extrema. The large number of extrema enable accurate determination of the electric field. The reduction of the amplitude at the 6^{th} extrema is due to the interference between the oscillations arising from heavy- and light-holes subbands. We do

not fit the experimental spectra with theoretical lineshapes, unlike Refs. 16 and 28, since FKO lineshape derived by Aspnes[1-3] is from flat band condition and that derived by Bhattacharya et al[9] is from depletion layer under a non-uniform electric field; neither lineshape being suitable in our case[20]. That the period of FKO under a small modulation is a function of the built-in electric field can be proved rigorously, and the light hole contribution to it can be neglected.[21] The extrema in FKO are given by[9]

$$m\pi = \phi + \frac{4}{3\pi}\left(\frac{E_m - E_g}{\hbar\theta}\right)^{\frac{3}{2}} \qquad (1)$$

where m is the index of the m^{th} extrema, E_g is the energy gap, E_m is the photon energy of the m^{th} extrema, ϕ is an arbitrary phase factor and $\hbar\theta$ is the electro-optical energy:

$$(\hbar\theta)^3 = \frac{e^2\hbar^2F^2}{2\mu} \qquad (2)$$

with e the electron charge and μ the reduced interband effective mass for electron and heavy hole pair in the direction of electric field F. The electron and hole effective mass along (100) is $M_{el}(100) = 0.067m_0$, and $M_{hh}(100) = 0.34m_0$, where m_0 is the free electron mass Plotted in Fig 2 (b) by squares is the quantity $(4/3\pi)(E_m-E_g)^{3/2}$ as a function of m for the spectrum Fig. 2 (a). The solid line is a linear fit to Eq (1), which yields $h\theta$ from which the electric field can be deduced. The surface Fermi level for all GaAs samples were 0.72±0.05 eV, in good agreement with other determinations[22]. Therefore, the electric field measured from FKO in our samples, under our experimental conditions, is in agreement with the value derived from electrostatic calculations.

Photoreflectance has also been performed on s-i-n^+ GaAlAs structures with different Al mole concentrations. Similar spectra with FKO from E_0 and $E_0+\Delta_0$ are observed.

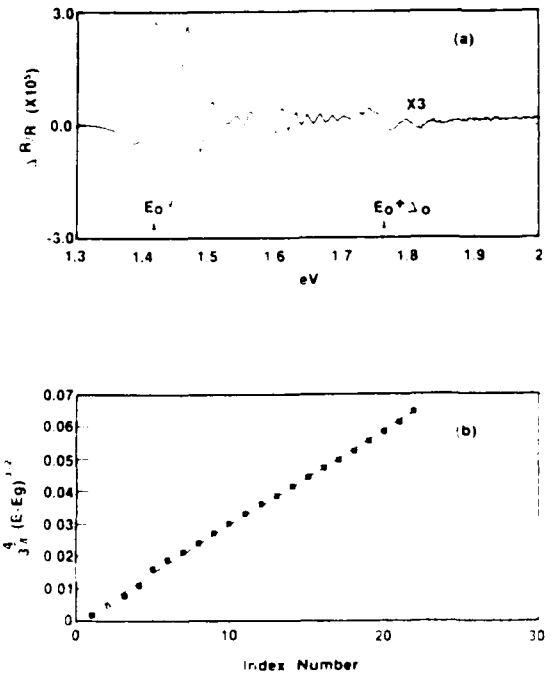


Figure 2 a) PR from a 2000 Å GaAs s-i-n^+ structure, b) Quantity $(4/3\pi)(E_m-E_g)^{3/2}$ as a function of FKO index m.

The built-in electric field is deduced from Eq (1) using[23] $m^*_{el}=0.067+0.083x$, and $m^*_{hh}=0.45+0.0.14x$, where m^*_{el}, m^*_{hh} are the effective masses for electron and heavy-hole. The surface Fermi level $E_{pin}(x)$ and the energy gap $E_g(x)$ as a function of Al mole concentration x is plotted in Fig. 3. We find $E_{pin}(x)=E_g(x)/2$, i.e. the surface Fermi level on bare AlGaAs surface is at mid-gap.

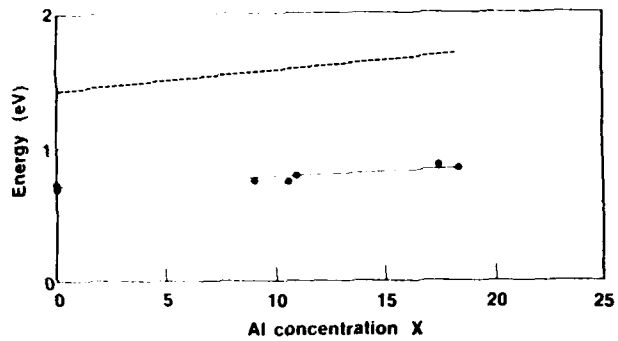


Figure 3 Energy gap E_0 (dotted line) and the surface Fermi level (dashed line) from bare $Ga_{1-x}Al_xAs$ surface as a function of Al mole fraction x.

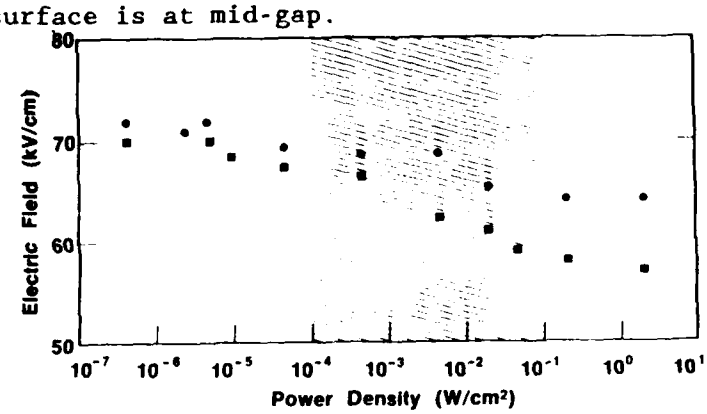


Figure 4 Electric field measured from FKO as a function of illumination power density from two nominally identical s-i-n^+ structures with d = 1000Å.

The effect of the illumination density on the electric field has been studied using three-beam photoreflectance. Both the probe and the pump beams are kept at low intensity. The third beam from a 20 mW He-Ne laser, provides the major illumination on the sample. The intensity of this illumination beam is adjusted by neutral density filters over seven orders of magnitude. The electric field derived from FKO, as a function of illumination power density for two nominally identical s-i-n^+ GaAs structures with d = 1000 Å is shown in Fig. 4. Circles represent data from a sample grown on n^+ GaAs (100) substrate using As_4 and indium mounting, while squares are those from one grown on n^+ GaAs 2° off (100) towards (110) using As_2 and indium-free mounting. The electric field is reduced as the intensity of the illumination beam increases. This is due to light generated photovoltage similar to that reported recently[24] in low temperature photoemission experiments. The difference in the photovoltage in the two nominally identical samples may be related to the MBE growth condition and will be reported elsewhere[25].

Increasing the intensity of either the pump or the probe beam has a similar

effect. In majority of the photoreflectance studies, the total power density (both the probe and the pump beam) is about 10^{-4} to 10^{-1} W/cm2, which is indicated at Fig. 4 by the hatched area. The generally accepted criterion for low field modulation is that one obtains identical lineshapes by varying the pump beam intensity by one or two orders of magnitude. Our results indicate that a slight change in FKO over one or two orders of magnitude does not always mean that the experiment is in the low field region. Extreme caution must be used when optical techniques are used to study electric field in semiconductor structures.

We next discuss the results from the InGaAs samples. In this study four similar structures were used, two of them (samples #1 and #3) were grown on (100) substrates and the other two (samples #2 and #4) were grown on (111)B substrates. These structures had an undoped top GaAs/InGaAs layer of thickness d grown by molecular beam epitaxy (MBE) on an n^+ GaAs substrate with an n^+ buffer of 3000 Å, doped at 2 x 10^{18}. To obtain a strain generated electric field (SGEF), a 100 Å undoped $In_{0.15}Ga_{0.85}As$ quantum well was inserted in the middle of the top undoped GaAs layer in samples #3 and #4. The top layer of the other two samples was an undoped GaAs layer only. Sample structure, growth orientation, and top undoped layer are summarized in Table I. Shown in Fig. 5 are the conduction band diagrams for the four different samples. By symmetry considerations, only sample #4 exhibits a piezoelectric field. Samples #1-3 were references.

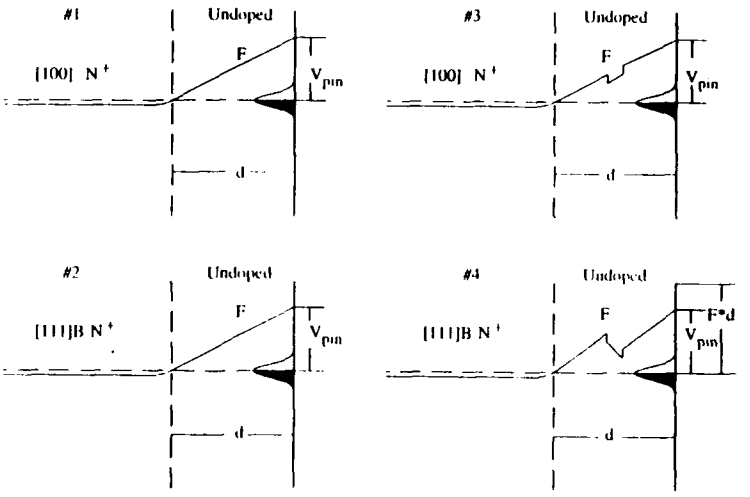


Figure 5 Band diagram of GaAs s-i-n^+ structures grown along [100] (samples 1,3) and along [111]B (sample 2,4). Samples 3 and 4 have an inserted 100Å InGaAs quantum well.

Sample	Growth Orientation	Inserted QW	Thickness d (Å)	Field (10^4/cm)	F*d (mV)	τ (ms)	V_{pin} (mV)
#1	[100]	No	750±40	9.3±0.5	700±50	2.0±0.2	690±20
#2	[111]B	No	850±40	8.1±0.5	690±50	1.5±0.2	690±20
#3	[100]	Yes	850±40	8.3±0.5	710±50	1.5±0.2	690±20
#4	[111]B	Yes	700±40	13.0±0.5	910±50	2.0±0.2	690±20

Table I: Built-in electric field and surface potential barrier from GaAs s-i-n^+ structure grown on [100] and [111]B substrate with and without an inserted InGaAs quantum well.

FKOs were observed in all four samples. We, however, did not observe any spectral features from the InGaAs quantum well. This is the result of the localization of electron and hole wave function on opposite sides of the quantum well leading to a loss of oscillator strength in the exciton feature. Therefore the FKOs are from the top undoped GaAs layer and not the InGaAs layer. These extrema in FKO are again given by Eqn. 1. Plotted in Fig. 6 is the quantity $(4/3\pi)(E_m-E_g)^{3/2}$ as a function of m for all four samples. The lines are linear fits to Eqn. 1 which yield $h\theta$. Using Eqn. 2, we deduced the electric fields in the top GaAs layer. The electron and hole effective mass[26] along (100) and (111) were $M_{el}(100)=0.067m_0$, $M_{hh}(100)=0.34-0.45m_0$, and $M_{hh}(111)=0.57-0.91m_0$, where m_0 is the free electron mass. The measured electric fields F, are listed in Table I. The uncertainty in the hole mass introduces an error to the electric field of approximately 2%. The electric fields in samples 1, 2, and 3 were essentially the same. However the undoped GaAs region of sample 4 had a much higher electric field than the other three samples. This can be understood as follows. In the top undoped layer, the field is induced by the negative charges trapped in the surface states (the

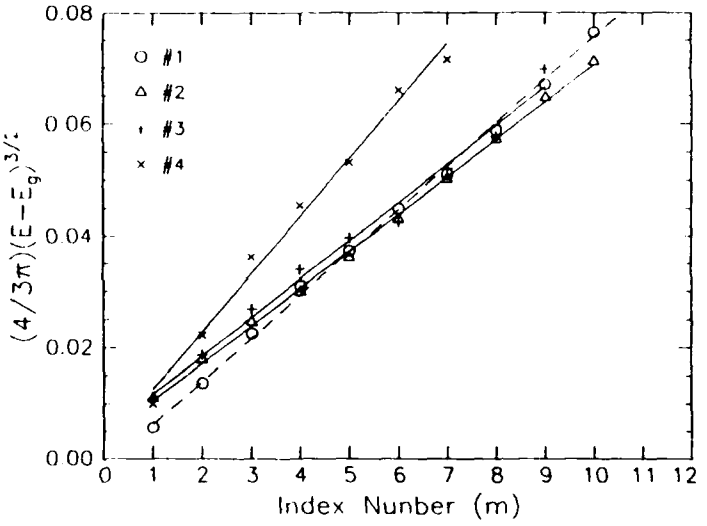


Figure 6 Quantity $(4/3\pi)(E_m-E_g)^{3/2}$ as a function of FKO index m for GaAs/InGaAs samples. Circles, sample 1; triangles, sample 2; crosses, sample 3; and x, sample 4.

surface Fermi level pinning) and by the polarization field in the strained layer,

$$F=\frac{V_{pin}+F_{pz}\ d'}{d} \tag{3}$$

where V_{pin} is the surface Fermi level, d' is the thickness of the InGaAs quantum well, d is the total thickness of the undoped layer, and F_{pz} is the SGEF. This equation is only valid when no interface charge is involved. Since sample 4 is the only sample which has the piezoelectric field in the InGaAs quantum well, the electric field in the undoped GaAs region of sample 4 is stronger than in samples 1, 2, and 3. In Eqn. 3 we have neglected the small band bending in the n^+ GaAs buffer and assumed that there are no free carriers in the undoped layer. A computer simulation has confirmed that this is accurate in our case. Surface Fermi level V_{pin} for a sample without SGEF can be determined by V_{pin}=F*d. However an independent measurement is needed for samples with SGEFs. In Ref. 27 we have demonstrated that the characteristic time constant for the photoreflectance signal can be used to determine the potential barrier height, and hence the surface Fermi level. In this work we used this method again, employing a He-Ne pump beam as in Ref. 27. The time response of PR signal was monitored by a data averaging digital oscilloscope. The probe beam was set at the first extremum of the FKO. The result is shown in Fig 7. The time constant was obtained by a least square fit to an exponential decay. The resulting time constants τ are listed in Table I. Also listed are the barrier heights V_{pin} calculated from[27]

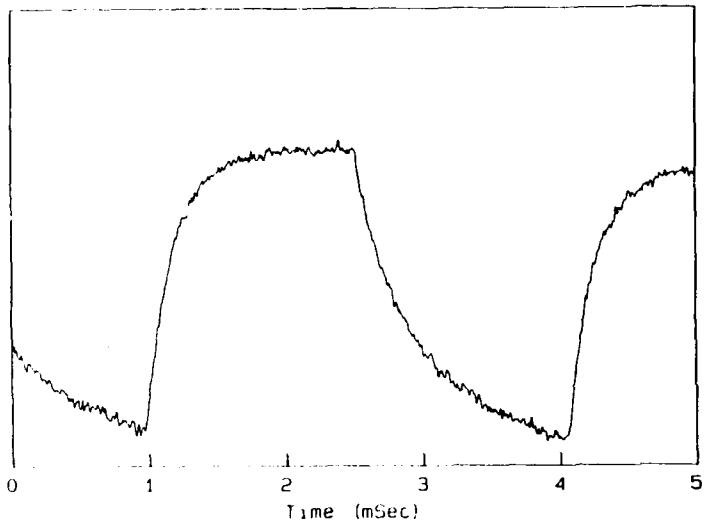


Figure 7 PR signal rise and decay curve from sample 4.

$$V_{pin}=\frac{kT}{q}\ln\left(\frac{\tau}{\tau_0}\right) \tag{4}$$

where k is the Boltzmann's constant, τ is temperature, q is electron charge and τ_0 is given by[27]

$$\tau_0=\left(\frac{\epsilon\epsilon_0}{d}\right)\left(\frac{kT}{q}\right)\left(\frac{1}{A^{**}T^2}\right) \tag{5}$$

with ϵ, ϵ_0 the dielectric constant of GaAs and the permittivity of free space respectively, and A^{**}=8.16A/cm^2-K^2, the effective Richardson constant. The surface Fermi levels (V_{pin}) calculated from Eqn. 4 are essentially constant at 0.69 V for all four samples. This result is consistent with Ref 28, in which the stress coefficient of V_{pin} is determined to be 0.67 mV/kbar. A net (111) uniaxial stress component of 10 kbar in the $In_{0.15}Ga_{0.85}As$ quantum well induces a change of Fermi level pinning of less than 0.01 eV. Also V_{pin} from time constant measurements is in good agreement with F*d measured by FKOs for sample 1, 2, and 3. However in sample 4 the piezoelectric field altered the electric field as measured by FKOs, but the Fermi level pinning measured by the characteristic time constant remained essentially the same. From Eqn. 3 and the difference between V_{pin} and F*d, the intrinsic SGEF is determined to be 2.2 x 10^5 V/cm.

In a single strained quantum well with thickness less than the critical thickness, the lattice mismatch is accommodated by biaxial strain in the InGaAs quantum well, therefore the piezoelectric field is given by[29-31]

$$F_{pz}=2\sqrt{3}\left(\frac{e_{123}}{\epsilon\epsilon_0}\right)\left(\frac{\Delta a}{a}\right)\left(\frac{C_{11}+2C_{12}}{4C_{44}+2C_{12}+C_{11}}\right) \tag{6}$$

where $\Delta a/a$ is the lattice mismatch and C_{ij} is the elastic stiffness. The parameters for $In_{0.15}Ga_{0.85}As$ are obtained from a linear interpolation of GaAs and InAs parameters. Using the numbers given by Ref. 21 we obtained F_{pz}=2.1 x 10^5 V/cm. The agreement between our experimental result and theoretical calculation is excellent. This agreement indicates that the GaAs/InGaAs interface is relatively free of interface trap states.

We now discuss the result from LT-GaAs samples. In this study the surface Fermi level pinning of a s-i-n$^+$ structure is replaced by a LT-GaAs Fermi level pinning layer (LT-i-n$^+$). The samples, grown by MBE on an n$^+$ GaAs substrate with a 1 μm thick n$^+$ buffer doped to 3X10^{18} cm^{-3}, have a 1000 Å thick undoped GaAs layer grown at 600°C and a 1000 Å thick LT-GaAs cap grown at 200°C. A reference 3 μm thick LT-GaAs sample, grown at 200°C directly on a SI GaAs substrate, is used for x-ray diffraction and EPR measurements. Neither sample is annealed in the MBE machine. Each sample is cleaved into several pieces for an annealing study. A rapid thermal anneal is performed at 600°C for 30,

60, 120 or 300 seconds.

PR spectra from as-grown and annealed samples show different FKOs. Different annealing times (30-300 s), however, result in only a small difference of the PR spectra. From both the as-grown and annealed samples, we have observed 16 to 17 extrema in the FKO indicating a uniform electric field over a large region. Plotted in Fig. 8 by open circles (squares) are the quantity $(4/3\pi)(E_m-E_g)^{3/2}$ as a function of m for the as-grown (annealed) sample. The line in Fig. 8 is a linear fit to Eq. 1 which yields $\hbar\theta$. Using Eq. 2, the electric fields of F = 47 ± 1 kV/cm and F = 65 ± 1 kV/cm are deduced for as-grown and annealed samples respectively.

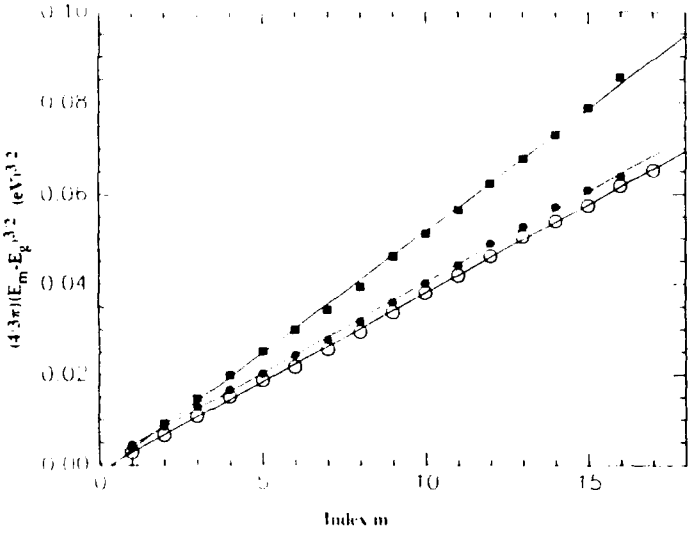


Figure 8 The quantity $(4/3\pi)(E_m - E_g)^{1/3}$ as a function of index m for LT-GaAs samples: open circles, as grown; closed circles, as-grown-etched; squares, annealed.

To prove the pinning of Fermi level in both the as-grown and annealed samples, we have etched the as-grown (annealed) cap LT-GaAs layer of about 700 Å (600 Å) using $1H_2SO_4:8H_2O_2:100H_2O$ solution at room temperature. The thickness of the remaining LT-GaAs layer is measured to be 300 Å (400 Å) for the as-grown (annealed) cap by a Dektak II profilometer. This also reveals a rougher etched surface for the annealed LT-GaAs sample than for the as-grown sample. We believe the roughness is related to the metallic arsenic precipitates. The measured electric field from the annealed sample is again 65 ± 1 kV/cm. The field in the as-grown sample is 48 ± 1 kV/cm. The results from the as-grown-etched sample are plotted in Fig. 8, where the closed circles represent the experimental data and the line represents the fit. Within the experimental error, there is essentially no change of field due to the etch, therefore, the Fermi levels in both as-grown and annealed samples are firmly pinned.

The Fermi level at the LT-GaAs can be calculated from solving Poisson's equation with degenerate Fermi-Dirac statistics in the n^+ and undoped layers by matching the electric field to the measured value. The calculated Fermi level is 0.47 eV (0.65 eV) below the conduction band at the LT-GaAs/undoped interface for the as-grown (annealed) sample. This calculation does not need

any prior knowledge of the LT-GaAs parameters.

The pinning of the as-grown LT-GaAs can not be explained by single or even multiple donor levels, however, the fact that the as-grown LT-GaAs is pinned is consistent with the infrared (IR) absorption[32,33] and EPR measurements[34]. IR absorption measures the concentration of neutral As_{Ga}^{0} defects, which is about 10^{20} cm^{-3}, while EPR measures the singly ionized As_{Ga}^{+} concentration, which is 5 X 10^{18} cm^{-3}. Since As_{Ga}^{+} is the major deep donor in the LT-GaAs, N_d^+ is not sensitive to temperature up to 300 K. The magnetic circular dichroism in absorption EPR (MCDA-EPR)[34] measurement also shows a gradual reduction of the As_{Ga}^{+} signal when the annealing time increases from 30 to 120 sec. All the EPR measurements are performed on the 3 μm reference sample grown under the same conditions. These results are plotted in Fig. 9 by a solid line. To maintain charge neutrality, the same number of negative charges must exist. At the present time the detailed nature of the negative charge is not fully understood, however, they must be intrinsic since extrinsic unintentional doping such as carbon can not be that high. We can only list several possibilities: V_{Ga} or Ga_{As} may act as an acceptor. Since the sample is grown at a low temperature with 10^{20} cm^{-3} As_{Ga} defects, statistically 5 X 10^{18} cm^{-3} of Ga_{As} is possible, although the experimental data for Ga_{As} is not available. Another possible candidate is As clusters or precipitates as often seen in annealed LT-GaAs. Regardless of the nature of the negative charges, we can deduce the As_{Ga} (0/+) level from:

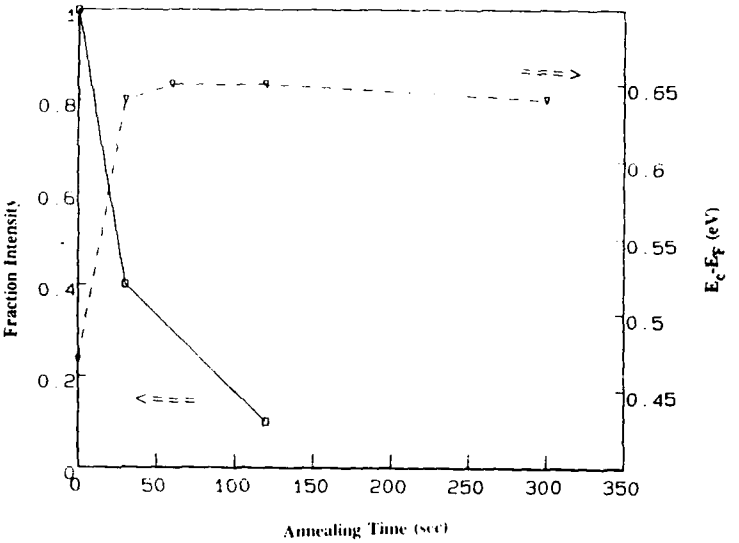


Figure 9 Fractional intensity (solid line) of As_{Ga}^{+} in LT-GaAs annealed at 600°C and Fermi level relative to E_c (dashed line) as functions of annealing time.

$$N_d^+ = N_d \frac{1}{1 + g\exp[(E_f - E_d)/kT]} \tag{7}$$

where g=1/2 for the singly ionized level of a double donor. Using N_d^+ = 5 X 10^{18} cm^{-3}, N_d=10^{20} cm^{-3} and E_c-E_f=0.47 V we obtained E_c-E_d=0.57 ± 0.03 eV, which is in good agreement with recent DLTS[35] and thermal spectroscopy[36] measurements

on LT-GaAs samples.

To explain this pinning, a computer model using Fermi-Dirac statistics to solve Poisson's equation was used. Shown in Figs.10 (a) and (b) by solid lines are the conduction band diagrams for the as-grown samples calculated using this model. In Fig. 10 (a), we use $N_d = 10^{20}$ cm^{-3} and at $E_c - E_d = 0.57$ eV below the conduction band for a single deep donor level of As_{Ga} as determined by IR absorption. In this case the Fermi level in the LT-GaAs is not pinned when an n^+ buffer is used. The electric field in this model depends on the thickness of the LT-GaAs cap layer. Shown in Fig. 10 (a) by squares is the band diagram when the LT-GaAs layer is only 300 Å. In Fig 10 (b), we assume $N_d = 10^{20}$ cm^{-3} and $N_a^- = 5 \times 10^{18}$ cm^{-3}. Due to the high degree of compensation of deep levels, the Fermi level in the as-grown LT-GaAs layer is firmly pinned. The electric field in the undoped layer does not depend on the thickness of the LT-GaAs cap, as shown by the triangles in Fig. 10 (b) for a 300 Å thick LT-GaAs cap.

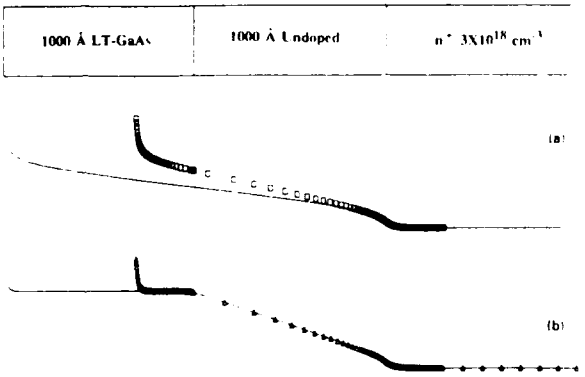


Figure 10 Conduction band diagram for as-grown sample: (a) Single deep donor level, (b) donors with compensation of acceptors.

The pinning of the Fermi level in the annealed sample can not be explained as compensation of deep donor levels with acceptors. The measured Fermi level, 0.65 eV below the conduction band, is several kT below the (0/+) energy level (0.57 eV) of As_{Ga} in the LT-GaAs. Therefore all the As_{Ga} are ionized. The MCDA-EDR measurement[10] shows an As_{Ga} concentration of about 10^{17} cm^{-3}. Then the total $N_d = N_d^+ \approx 10^{17}$ cm^{-3}. Such a low deep-donor level concentration is not enough to pin the Fermi level firmly. Recall the MCDA-EDR measurement shows a gradual reduction of the As_{Ga}^+ signal when the annealing time increases from 30 to 120 sec. However, the Fermi level pinning position drops to 0.65 eV immediately after the 30 sec anneal, and remains the same during further annealing. This result is even more difficult to explain by the compensation model. To maintain the Fermi level pinning position after the 30 sec anneal while allowing for the change of the concentration of N_d^+, an almost impossible balance of N_d/N_a would be needed.

On the other hand the metallic As precipitates Schottky pinning theory can explain our result. According to ref 37, a 3 nm radius As cluster at a density of 10^{16} cm^{-3} can compensate a doping concentration of 10^{17} cm^{-3} and result in a Fermi level pinning in the LT-GaAs layer. As long as enough As precipitates form during the first 30 sec annealing, the Fermi level is pinned at the same place. Further annealing results in more As_{Ga} being accommodated, in the precipitates, and less As_{Ga}^{+} being detected by EPR. We also notice the measured Fermi level of 0.65 eV is in good agreement with the barrier height of metallic As to GaAs conduction band. Therefore our results strongly support the theory of As precipitates.

In conclusion, we have studied the Franz-Keldysh oscillations from s-i-n^{+} GaAs and GaAlAs structures. We have demonstrated that the electric field obtained from FKO is in agreement with that derived from the electrostatic calculation. Our results indicate that illumination in a normal PR experiment can significantly affect the measurement and thus erroneously leads to a reduced value of electric field. Only measurements at extremely low illumination power are reliable. Different photoresponse from two identical samples were observed. We have demonstrated a number of utilizations of this non contact, non destructive optical method in the study of 1) Fermi level on the bare surface of GaAlAs, 2) piezoelectric field in a (111) B InGaAs/GaAs structure, and 3) the pinning mechanism in the low temperature grown GaAs.

References:

1. See, for example, M. Cardona, in Modulation Spectroscopy (Academic, New York, 1969).

2. See, for example, D. E. Aspnes, in Handbook on Semiconductors, Vol 2, Ed. T. S. Mos- (North-Holland, New York, 1980), p. 109 and references cited therein.

3. See, for example, F. H. Pollak, in the Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE, Bellingham, 1981), vol 276, p. 142 and reference cited therein.

4. O. J. Glembocki, B. V. Shanabrook, N. Bottka, W. T. Beard, and J. Comas, Appl. Phys. Lett., 46, 970, (1985).

5. H. Shen, P. Parayanthal, F. H. Pollak, M. Tomkiewicz, T. J. Drummond, and J. N. Schulman, Appl. Phys. Lett, 48, 653, (1986).

6. Y. R. Lee, A. K. Ramdas, F. A. Chambers, and J. M. Meese, in the Proceedings of the Society of Photo-Optical Instrumentation Engineers, (SPIE Bellingham), 794, 105, (1987).

7. O. J. Glembocki and B. V. Shanabrook, in the Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE Bellingham), **794**, 74, (1987).

8. X. L. Zhang, D. Heiman, B. Lax and F. A. Chambers, Appl. Phys. Lett., **52**, 287, (1988).

9. R. N. Bhattacharya, H. Shen, P. Parayanthal, F. H. Pollak, T. Coutts and H. Aharoni, Phys. Rev. **B37**, 4044, (1988).

10. R. Glosser and N. Bottka in the Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE Bellingham, 1987) Vol.**794**, 88, (1987).

11. N. Bottka, D. K. Gaskill, R. S. Sillmon, R. Henry, and R. Glosser, J. Electron Mat., **17**, 161, (1988).

12. X. Yin, F. Pollak, L. Pawlowicz, T. O'Neill and M. Hafazi, Appl. Phys. Lett., **56**,1278, (1990).

13. M. H. Herman, I. D. Ward, R. F. Kopf, S. J. Pearton and E. D. Jones, MRS 1989.

14. D. K. Gaskill, N. Bottka and R. S. Sillmon, J. Vac Sci. Technol., **B6**, 1497, (1988).

15. H. Shen, F. Pollak and J. M. Woodall, to be published in J. Vac. Sci. Technol.

16. C. Van Hoof, K. Deneffe, J. De Boeck, D. J. Arent and G. Borghs, Appl. Phys. Letts., **54**, 608, (1989).

17. T. Kanata, to be published in the Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE, Bellingham, 1990), Vol 1289.

18. J. S. Best, Appl. Phys. Letts., **34**, 522, (1979).

19. H. Shen, P. Parayanthal, Y. F. Liu and F. Pollak, Rev. Sci. Instrum., **58**, 1429, (1987).

20. The lineshape of FKO from uniform build-in field under small modulation has been derived. H. Shen, Ph. D Thesis, (1986), unpublished.

21. The contribution of the light hole band results in the interference and the reduction of the amplitude of the 6^{th}. It does not change the period of the FKO.

22. H. Hasegawa, H. Ishii, T. Sawada, T. Saitoh, S. Konishi, Y. Liu and H. Ohno, J. Vac. Sci. Technol., **B6**, 1184, (1988).

23. S. Adachi, J. Appl. Phys, **58**, R1, (1985).

24. M. H. Hecht, Phys. Rev. **B41**, 7918, (1990).

25. L. Fotiadis, H. Shen, M. Dutta, P. G. Newman, R. P. Moerkirk, W. H. Chang, R. N. Sacks, unpublished.

26. Semiconductors, Physics of Group IV Elements and III-V Compounds, edited by Landolt and Bornstein (Springer, New York 1982)

27.H. Shen, M. Dutta, R. Lux, W. Buchwald, L. Fotiadis, and R. N. Sacks, Appl. Phys. Lett., 59, 321 (1991).

28. Ki-Woong Chung, Z. Wang, J. C. Costa, F. Williamson, P. P. Ruden, and M. I. Nathan, Appl. Phys. Lett. 59, 1191, (1991)

29. D. L. Smith and C. Mailhiot, Rev. of Mod. Phys., 62, 173 (1990).

30. E. A. Caridi, T. Y. Chang, K. W. Goossen, and L. F. Eastman, Appl. Phys. Lett., 56, 659 (1990).

31. R. W. Martin, M. Lakrimi, C. Lopez, R. J. Nicholas, E. T. R. Chidley, N. J. Mason and P. J. Walker, Appl. Phys. Lett., 59, 659 (1991).

32. M. Kaminska, E.R. Weber, Kin Man Yu, R. Leon, T. George, F.W. Smith, A.R. Calawa, in Semi-insulating III-V Materials, Toronto 1990, ed. A.G. Milnes and C.J. Miner (Adam Hilger, Bristol, 1990) p.111.

33. M. Kaminska, E.R. Weber in Proc. 20th Int. Conf. Phys. Semicon., Thessaloniki, Greece, 1990, ed. E.M. Anastassakis and J.D. Joannopoulos (World Scientific, Singapore, 1991) p. 473.Figure Caption

34 H.J. Sun, G.D. Watkins, F.C. Rong, L. Fotiadis, E.H. Poindexter, Appl. Phys. Lett. **60**, 719 (1992).

35. D.K. Johnston, C.E. Stutz, K.R. Evans, unpublished.

36. K. Xie et. al., J. Electron Material, 20, 553 (1991).

37. A.C. Warren, J.M. Woodall, J.L. Freeouf, D. Grischkowsky, D.T. McInturff, M.R. Melloch, and N. Otsuka, Appl. Phys. Lett. 57, 1331 (1990).

High-Ratio Bandwidth Reduction of Video Imagery for Teleoperation

Mr. Dale R. Shires *
Major Frank F. Holly
Mr. Phillip G. Harnden

U. S. Army Human Engineering Laboratory, Behavioral Research Division
Aberdeen Proving Ground, Maryland 21005-5001

1. INTRODUCTION

To lessen the risks to the soldier and as a force multiplier, some of the Army's future combat vehicles will be remotely driven. Video or FLIR sensors will be mounted on the vehicle with visual information about the surrounding scene transmitted back to a remote control station. Since visual imagery is a large user of bandwidth, and bandwidth is very limited on the battlefield, there is a need to find a means of reducing this data transmission rate. This paper describes two techniques used to produced low-data-rate transmissions. In the first, image compression is performed by an algorithm designed to implement the new Joint Photographic Experts Group (JPEG) compression standard. This produces a bandwidth reduction ratio of 25-to-1. In the second step, a 90-to-1 bandwidth reduction is achieved by transmitting only one frame every 3 seconds rather than the usual 30 frames per second. By itself, this is unacceptable from a human factors viewpoint because of the large voids in the visual information presented to the remote operator. Therefore, we have developed an image processing algorithm that creates simulated dynamic imagery (synthetic optic flow) during the 3-second intervals. That is, each period consists of one "real" frame and 89 synthetic ones. The overall bandwidth reduction ratio is therefore 25 x 90 = 2250 to 1.

In computational vision, optic flow is often used to obtain information about the environment by analyzing the relationships among movement vectors. For example, motion parallax can be used to obtain information about

relative distances, and this information, in turn, can be used to segment a three-dimensional scene into its object components. In this instance, however, optic flow is used in the synthetic rather than analytic sense; this same relationship between scene information and movement vectors can be inverted to produce optic flow given certain information about the environment. The two pieces of information needed to create this optic flow are (a) the x, y, and z coordinates of every point in the environment corresponding to a pixel on the display screen, and (b) the instantaneous velocity and direction of the vehicle.

2. APPROACH

A schematic description of the approach used is depicted in Figure 1.

2.1 JPEG image compression

In this system, a 25-to-1 compression ratio is achieved by applying our implementation of the JPEG compression algorithm to the image that will be transmitted. The simulated optic flow algorithm described in this paper uses only gray scale images. Therefore, only those aspects of the JPEG standard that apply to luminance or brightness data compression are implemented. Parts of the JPEG procedure that apply to color or chrominance data, such as the Red-Green-Blue (RGB) to Luminance-Chrominance (YUV) conversions and subsampling of chrominance data, are not performed. JPEG compression proceeds in three steps: the Discrete Cosine Transform (DCT), quantization, and entropy encoding using Huffman encoding (Wallace, 1990).

2.1.1. Forward Discrete Cosine Transform

To begin, a digital image is formed from the camera system attached to the vehicle. This image is then divided into 8x8 pixel blocks where each pixel is some eight-bit gray scale intensity ranging from black (0) to white (255). Each of these pixel blocks is then transformed by the application of the DCT. The DCT, a relative of the Discrete Fourier Transform, transforms each of these blocks into 64 coefficients representing two-dimensional spatial frequencies. The formula to compute the DCT (Hamilton, 1991) is

$$F[u,v] = 0.25\ C[u]\ C[v] \sum_{m=0}^{7} \sum_{n=0}^{7} f[m,n] \cos\left[\frac{(2m+1)\,u\,\pi}{16}\right] \cos\left[\frac{(2n+1)\,v\,\pi}{16}\right]$$

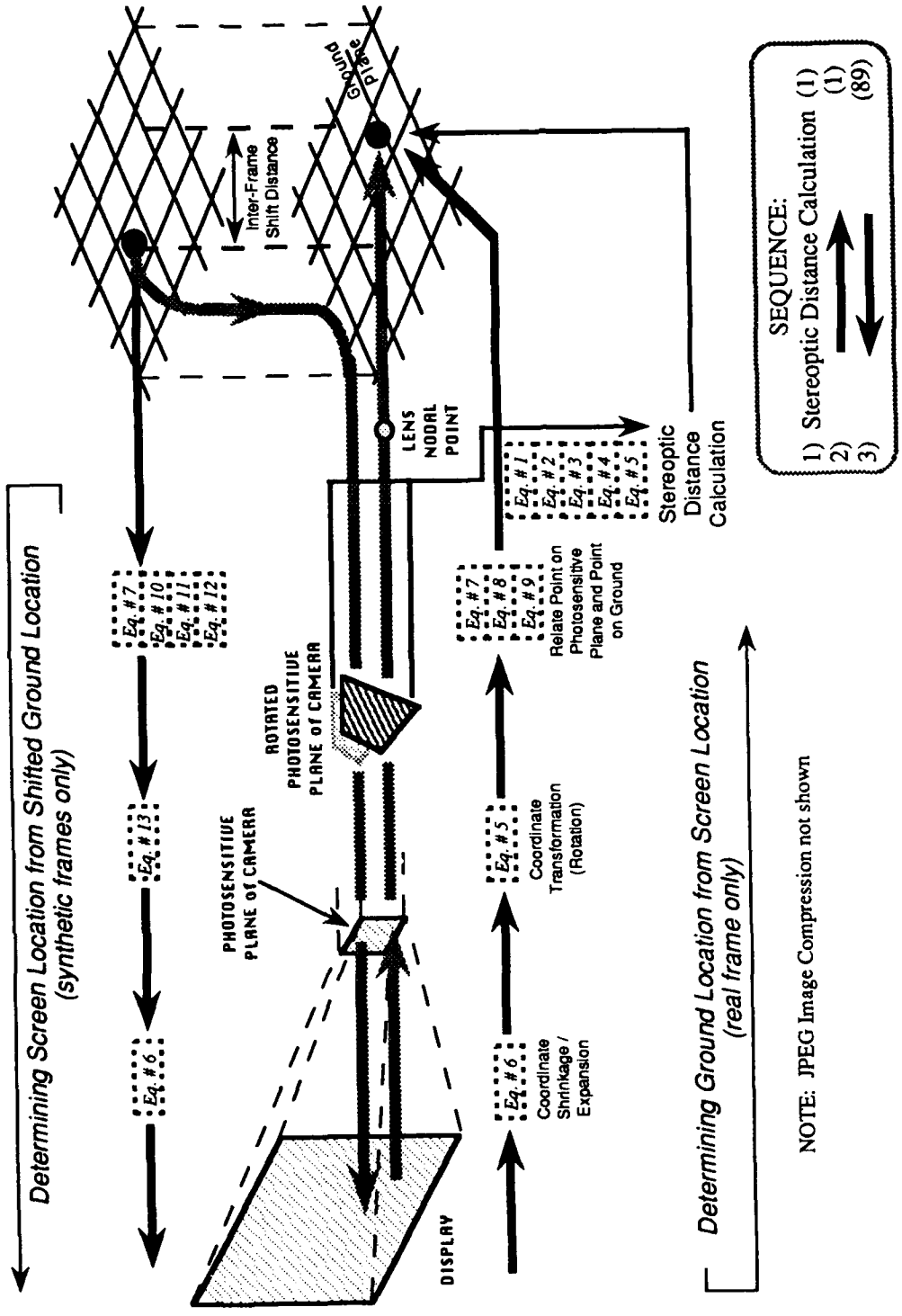


Figure 1. Steps for creating synthetic optic flow.

in which

u, v = discrete frequencies (0...7),
m, n = pixel horizontal and vertical indices in the 8x8 pixel block,

$$C[u], C[v] = \begin{cases} \frac{1}{\sqrt{2}}, \text{ for } u, v = 0 \\ 1, \text{ otherwise,} \end{cases}$$

and the formula for the inverse DCT is

$$f[m, n] = 0.25 \sum_{u=0}^{7} \sum_{v=0}^{7} C[u]\ C[v]\ F[u, v] \cos\left[\frac{(2m + 1)\ u\ \pi}{16}\right] \cos\left[\frac{(2n + 1)\ v\ \pi}{16}\right]$$

Since most pixel intensities change slowly across the image, the total energy of the 8x8 block concentrates mainly at the lower frequencies. In the 8x8 transform, this means that only a few coefficients are necessary to represent the frequency space. These coefficients are concentrated mainly in the top left corner of the DCT result matrix. Table 1 shows an 8x8 grid of gray scale values from an image and the DCT transformation result matrix (after JPEG mandated level shifting for a signed representation).

Table 1

Gray Scale to DCT Coefficient Transformation

Image intensity values								DCT result matrix							
152	149	142	135	129	123	120	119	2	84	18	10	-2	2	0	1
152	149	145	136	127	123	120	116	37	11	-6	-8	0	0	0	0
149	145	139	132	127	124	123	116	-3	0	0	2	1	0	0	0
147	145	140	132	127	123	121	118	-5	2	7	0	2	0	0	0
148	142	136	127	123	119	117	113	7	1	-5	-1	1	-1	1	0
146	136	124	118	115	115	116	113	-1	-4	3	1	0	0	0	0
142	132	121	115	115	117	116	112	-1	-1	1	2	1	0	0	0
140	133	126	122	121	120	115	109	0	0	0	2	0	0	0	1

2.1.2. Quantization

In this step, each coefficient resulting from the DCT transformation is quantized, that is, divided by its corresponding entry in a specified 8x8 quantization table (e.g., the number in row 4, column 3 of the DCT result matrix is divided by the entry in row 4, column 3 of the quantization table). The main reason for quantization is to reduce the magnitudes of the DCT coefficients. Many of the erroneous high frequencies are eliminated during this process, and a high number of zero coefficients are produced. These zero coefficients will be compressed in the final stage of JPEG.

JPEG is a "lossy" compression scheme. That is, it permanently discards information during compression so that the original image can never be exactly recovered. This is the step where the eradication of image data occurs. Careful selection of the quantization table is necessary since this step will ultimately determine image degradation and compressibility. Table 2 shows the quantization table used by this compression algorithm and the result of quantizing the DCT result matrix in Table 1.

Table 2

Results of Quantizing the DCT Result Matrix in Table 1

Quantization table								Quantization results							
24	17	15	24	36	60	77	92	0	4	1	0	0	0	0	0
18	18	21	29	39	87	90	83	2	0	0	0	0	0	0	0
21	20	24	36	60	86	104	84	0	0	0	0	0	0	0	0
21	26	33	44	77	131	120	93	0	0	0	0	0	0	0	0
27	33	56	84	102	164	155	116	0	0	0	0	0	0	0	0
36	53	83	96	122	156	170	138	0	0	0	0	0	0	0	0
74	96	117	131	155	182	180	152	0	0	0	0	0	0	0	0
108	138	143	147	168	150	155	149	0	0	0	0	0	0	0	0

2.1.3. Entropy encoding

The final step in JPEG compression is entropy encoding using Huffman encoding. Unlike the quantization stage which discarded image information, no image data is lost in this step. Huffman encoding identifies frequently

occurring data strings and replaces them with a single code. For example, the string "5555555555555" with an occurrence probability of .8 could be replaced by a single code such as "01." Strings with lower probabilities of occurrence will have longer codes. The Huffman code generation algorithm is based on the greedy method design principle which guarantees the creation of an optimal set of compression codes (Horowitz, 1978).

Huffman codes are generated after the DCT and quantization steps are performed on the 8x8 pixel blocks of the image. The coefficients in the 8x8 matrices left after quantization are ordered to form strings of as many zero coefficients as possible. The codes are then generated based upon the number of consecutive zeros in the quantization result matrices and their frequencies of occurrence. The Huffman code generation algorithm is performed only once using a typical image from the remote driving scenario. After that, all images to be compressed and decompressed will use the same Huffman codes.

Decompression of JPEG compressed images flows in the reverse direction. The steps are Huffman decoding, dequantization, and application of the inverse DCT.

2.2. Stereoptic range finding

Two video cameras are mounted on the vehicle. The first step in finding the x, y, and z coordinates of every point in the environment corresponding to a pixel on the display screen is to passively determine the range of each of these points using a stereopsis algorithm. (In this model, the x, y, and z axes refer to the horizontal, vertical, and distance axes, respectively.) The purpose of this range information is to assist in differentiating the y and z axes. There is an ambiguity between these two axes in a two-dimensional image such that a movement in the upward direction could signify a movement parallel to either.

The most difficult aspect of range finding using stereopsis is performing the stereo-correspondence match, that is, finding to which pixel in the left image a given pixel in the right image corresponds. For this, we have developed an algorithm to implement an approach described by Barnard (1987) and based upon simulated annealing (Kirkpatrick, 1983). Simulated annealing is a concept based upon principles of statistical physics and is well suited to problems which, like stereopsis, have many degrees of freedom and do not

have convex solution spaces. It simulates a physical system composed of a large number of coupled elements in which each element is considered to be a point on a lattice. These lattice points are the points on the camera photosensitive plate corresponding to pixels on the display. There are two lattices, one for the left image and one for the right image. A penalty function is written which expresses the appropriate (and often partially conflicting) constraints whose violation corresponds to energy. For this stereopsis algorithm, the two constraints are (a) a pair of left and right elements representing the same point in the scene should have the same brightness, and (b) the disparity of an element should be the same as that of its neighbors. Disparity is the difference in relative position between an element and its corresponding element in the other image caused by the slightly different position of the two cameras. This smoothness heuristic reflects the fact that surfaces are spatially coherent and is required because the first constraint is strictly local, and stereo-correspondences are locally ambiguous. In processing, the system is reduced to lower and lower energy configurations by reducing the "temperature" while keeping the system in thermal equilibrium. When a system is in thermal equilibrium, its states have a Boltzman distribution:

$$P(E) = \exp(\frac{-E}{T})$$

in which

E = energy,
$P(E)$ = probability of a state having energy E,
T = temperature of the system.

The system is taken to equilibrium by considering random, local state transitions on the basis of the change in energy that they imply. Since the system is stochastic, these local state changes can take the system away from convergence as well as toward it. This helps to prevent the system from sinking into local minima. The processing is complete when the system is in equilibrium at the lowest energy state achievable. This represents a maximization of the constraints to the extent possible for the system as a whole. In our implementation, the formula used is the one suggested by Barnard

$$E = \sum_k \| I_L(k) - I_R(i, j + D(k)) \| + \lambda \| \nabla D(k) \| \qquad (1)$$

in which the first term inside the sum represents the equal brightness constraint and the second term the smoothness constraint. The constant λ determines their relative importance. The $\|\nabla D(k)\|$ operator is the sum of the absolute differences between disparity $D(k)$ and the disparities of the kth lattice point's eight neighbors.

Once the stereo-correspondence matches are made using simulated annealing, a pair of equations of the form

$$A_L x + B_L y + C_L z = D_L \tag{2}$$
$$A_R x + B_R y + C_R z = D_R \tag{3}$$

are created for each pair of corresponding image points defining the two planes within whose intersection the corresponding point in the world must lie. The three points necessary for creating the plane for the point on the right camera photosensitive plate are obtained by using the coordinate values of the nodal point P_1 (0, 0, 0) of the right camera lens, the x, y, and z coordinates of the point P_2 on the photosensitive plane of the right camera, and a point having these same x and z values but a different y value, P_3. Similarly, the three points necessary for creating the plane for the corresponding point on the left camera are obtained by using the nodal point of the left camera lens, the corresponding point, and a point having the same x and z coordinates but a different y coordinate. The two equations defining the two planes are solved simultaneously to yield the equation defining the line of intersection of the two planes. The point in the world corresponding to the two points on the two photosensitive planes is now known to lie somewhere on this line. Thus, at this point, the x and z coordinates of this point in the world are known, but the y coordinate has yet to be determined. This y value is determined by first finding the angle between the plane $A_R x + B_R y + C_R z = D_R$ and the plane $x = 0$ by using the formula

$$\theta = \cos^{-1} \frac{\mathbf{N}_1 \bullet \mathbf{N}_2}{\|\mathbf{N}_1\| \, \|\mathbf{N}_2\|} \tag{4}$$

in which $\mathbf{N}_1$ is the normal vector to the first plane and $\mathbf{N}_2$ is the normal vector to the second plane. A coordinate transformation (rotation) is then performed by this amount using the formula

$$\begin{aligned} x' &= l_1x + m_1y + n_1z \\ y' &= l_2x + m_2y + n_2z \\ z' &= l_3x + m_3y + n_3z \end{aligned} \tag{5}$$

in which l, m, and n are the direction cosines of the x', y', and z' axes (new axes) relative to the x, y, and z axes (old axes), respectively. In this case, the new axes are those associated with the plane x=0. The line defined by the points P_1 and P_2 and the line created by the intersection of the planes defined by equations 2 and 3 now lie in the plane $x = 0$ and their point of intersection can be found using the standard method for lines in a plane. This provides the y coordinate of the world point whose x and z coordinates were previously known.

2.3. Synthetic optic flow

The first step in creating synthetic optic flow on the display is the creation of "brightness objects," which in the initial (i.e., "real") frame, are co-extensive with the image pixels. In the synthetic frames, each brightness object retains its initial brightness from the real frame throughout all 89 synthetic frames but is expanded and moved downward and outward in each frame. The amount of movement and expansion is determined by the location of the object point's corresponding world position and the speed and direction of the vehicle.

To accomplish this movement and expansion, the coordinates of the four corners of each pixel (the first steps about to be described correspond to the rightward flow in Fig. 1) are found. These coordinates are given by x+.5, y+.5; x+.5, y-.5; x-.5, y+.5; and x-.5, y-.5 in which x and y are the coordinates of the center of the pixel. The first step in finding the points in the world corresponding to these four corners is to map the origin of the display coordinate system into the center of the photosensitive plane and, since the origin of the world coordinate system is the nodal point of the right camera lens (only the right camera is used in creating the synthetic optic flow itself), assign it the values $x = 0$, $y = 0$, $z = -g$ in which g is the distance from the nodal point to the photosensitive plate. Next, the display coordinates (both the pixel centers and corners) are "shrunk" down to their corresponding world coordinates on the photosensitive plate of the right camera. This is accomplished by finding

$$x_c = \frac{H_c}{H_s} \text{ and } y_c = \frac{W_c}{W_s} \tag{6}$$

in which

H_c = height of the photosensitive plate of the camera,
H_s = height of the display screen,
W_c = width of the photosensitive plate of the camera,
W_s = width of the display screen.

Since the cameras on the vehicle will usually not be parallel to the ground but instead will have some downward tilt, the next step is to perform a coordinate transformation using Eq. 5. The world coordinates of the points on the camera photosensitive plate are determined by rotating the y and z axes in the direction opposite to the tilt using the nodal point of the lens as the pivot. A line is then defined using two points--the nodal point of the lens and the point on the camera photosensitive plate corresponding to the corner of the pixel. The equations defining this line in parametric form are

$$\begin{aligned} x &= t(x_1-x_2) + x_1 \\ y &= t(y_1-y_2) + y_1 \\ z &= t(z_1-z_2) + z_1 \end{aligned}$$

Since x_1, y_1, and z_1 are 0,

$$\begin{aligned} x &= t(-x_2) \\ y &= t(-y_2) \\ z &= t(-z_2) \end{aligned} \tag{7}$$

The value of t is found by substituting for y in the equation for the plane

$$y = S \tag{8}$$

in which S equals the y value found earlier for the world point corresponding to this pixel corner. This yields

$$t(-y_2) = S \text{ and } t = \frac{S}{-y_2} \tag{9}$$

This value for t is then substituted into equation 7 which yields the coordinates of the point at which this line intersects the plane $y = S$.

After this, we begin to work back the other way (these steps correspond to the leftward flow in Fig. 1), that is, finding the display screen points corresponding to ground points. First, the coordinate axes are rotated by the amount which the vehicle turned/inclined during the 1/30th second presentation of the real frame and translated by the amount which it moved during this frame using the formula

$$\begin{aligned} x' &= l_1(x - x_0) + m_1(y - y_0) + n_1(z - z_0) \\ y' &= l_2(x - x_0) + (a + m_2)(y - y_0) + (a + n_2)(z - z_0) \\ z' &= l_3(x - x_0) + (a + m_3)(y - y_0) + (a + n_3)(z - z_0) \end{aligned} \tag{10}$$

in which the coordinates x', y', z' are the new coordinates of the world point just found, a is the angle of downward tilt of the camera and the origin O' (the nodal point of the right camera lens) of the x', y', z' system has coordinates (x_0, y_0, z_0) relative to the x, y, z system. Next, the equation defining the plane of the right photosensitive plate is found, This is accomplished by taking two more points (x_3, y_3, z_3) and (x_4, y_4, z_4) on this photosensitive plate and writing the equation for this plane as

$$\begin{vmatrix} y_3 - y_2 & z_3 - z_2 \\ y_4 - y_2 & z_4 - z_2 \end{vmatrix} x + \begin{vmatrix} z_3 - z_2 & x_3 - x_2 \\ z_4 - z_2 & x_4 - x_2 \end{vmatrix} y + \begin{vmatrix} x_3 - x_2 & y_3 - y_2 \\ x_4 - x_2 & y_4 - y_2 \end{vmatrix} z = D \tag{11}$$

The goal now is to find where the photosensitive plate is intersected by a line beginning at a point in the world and passing through the nodal point. This is accomplished by first creating an equation for this line in the form of Eq. 7 and then substituting the parametric equivalents of x, y, and z into Eq. 11

$$\begin{vmatrix} y_3 - y_2 & z_3 - z_2 \\ y_4 - y_2 & z_4 - z_2 \end{vmatrix} t(-x_2) + \begin{vmatrix} z_3 - z_2 & x_3 - x_2 \\ z_4 - z_2 & x_4 - x_2 \end{vmatrix} t(-y_2) + \begin{vmatrix} x_3 - x_2 & y_3 - y_2 \\ x_4 - x_2 & y_4 - y_2 \end{vmatrix} t(-z_2) = D \tag{12}$$

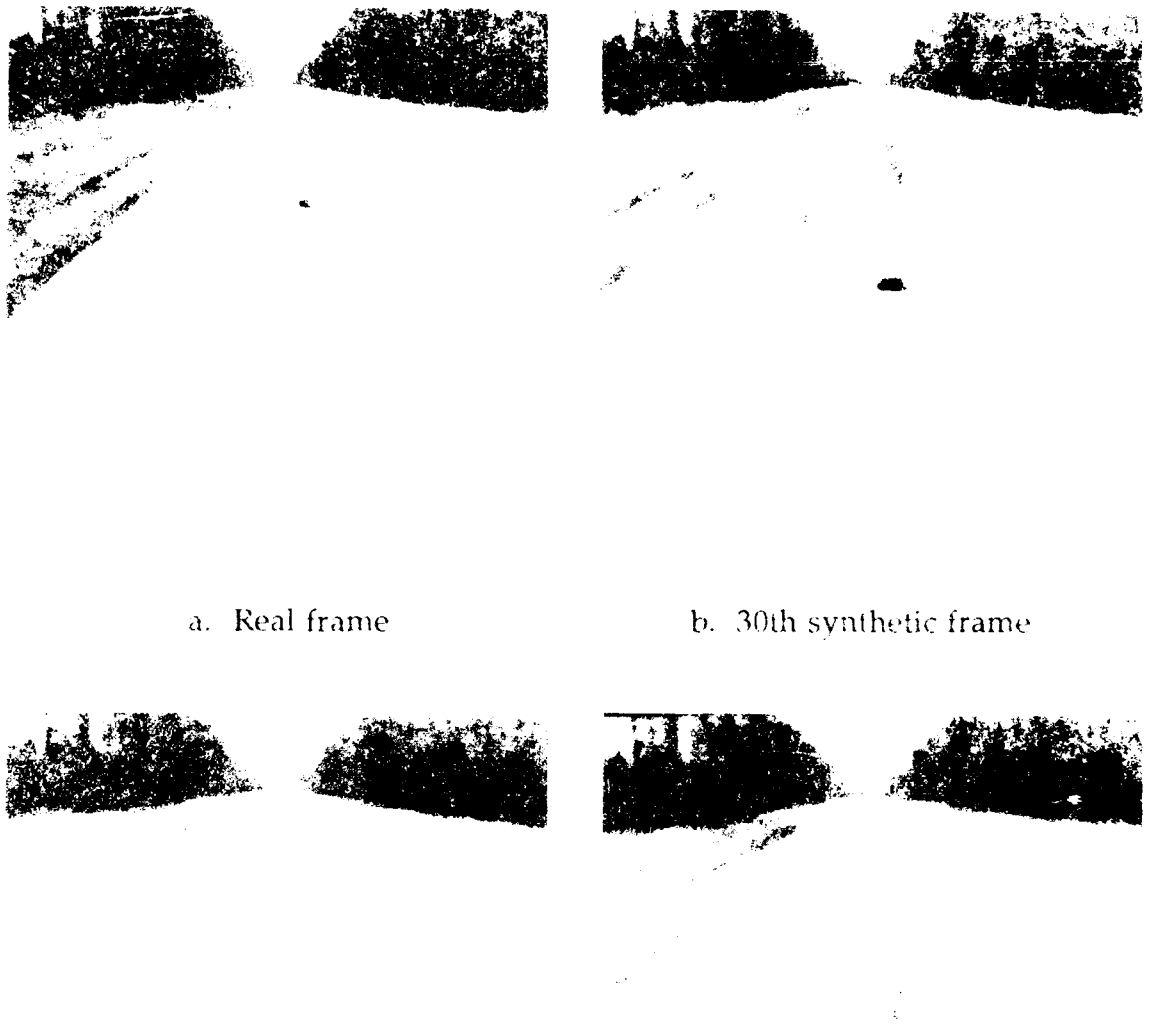
This equation is then solved for t as before and this t value is substituted into the equations for the line to find the coordinates at which the line intersects the photosensitive plane. The coordinates of the photosensitive plane are then rotated in the direction opposite to their previous rotation using Eq. 13 followed by the expansion of the coordinates using the reciprocal of Eq. 6.

$$\begin{aligned} x &= l_1x' + l_2y' + l_3z' \\ y &= m_1x' + m_2y' + m_3z' \\ z &= n_1x' + n_2y' + n_3z' \end{aligned} \tag{13}$$

When these steps are repeated for all corners of all brightness objects and the interiors of all brightness objects are assigned the brightness of their corresponding pixels in the real frame, the creation of the first synthetic frame is complete. Next, the coordinate axes are rotated by the amount which the vehicle turned/inclined during the 1/30th second presentation of this synthetic frame and translated by the amount which it moved during this frame. Then, the creation of the next synthetic frame is begun.

3. RESULTS

The result is smoothly flowing imagery which appears very close to that produced by full bandwidth transmission. However, the bandwidth reduction ratio produced by this system is 25 x 90 = 2250 to 1. Figure 2 depicts a sample of a real frame, the 30th synthetic frame after that, the 60th synthetic frame, and the 89th synthetic frame. It can be seen that, because of the faster movement and expansion in the foreground, there is a small resolution loss in the foreground in the later synthetic frames relative to the background, but the major features are still clear. Furthermore, this small loss of resolution does not appear to be troubling to the operator in viewing the actual dynamic imagery.



a. Real frame

b. 30th synthetic frame

c. 60th synthetic frame

d. 89th synthetic frame

Figure 2. Images generated by the synthetic optic flow algorithm. Only a small amount of resolution loss is noticeable in the synthetic frames.

4. REFERENCES

Barnard, S. T., "A Stochastic Approach to Stereovision" in *Readings in Computer Vision*, Addison-Wesley, New York, 1987.

Hamilton, E. R., "The Standardization of Still Picture Coding: The JPEG Image Compression Algorithm," C-Cube Microsystems, San Jose, California, October 15, 1991.

Horowitz, E. and Sahni, S., *Fundamentals of Computer Algorithms*, Computer Science Press, Rockville, Maryland, pp. 152-174, 1978.

Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P., "Optimization by Simulated Annealing," *Science*, vol. 220, no. 4598, pp. 671-680, May 13, 1983.

Wallace, G. K., "Overview of the JPEG Still Picture Compression Algorithm," *Proceedings of Electronic Imaging East*, pp. 66-71, 1990.

Safety and Kinetics of a Despeciated Equine $F(ab)_2$ Heptavalent Botulinum Antitoxin in Volunteers

*Maria H. Sjogren, LTC; Kelly T. McKee, Jr., LTC; Michael A. Clayton, Mr.; David R. Franz, COL; Richard M. Condie, Dr.; Clarence J. Peters, COL, U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, MD 21702-5011 and University of Minnesota, St.Paul, MN 55108

INTRODUCTION

Botulism is an intoxication caused by ingesting or inhaling any of seven types of neurotoxins produced by *Clostridium botulinum*. Humans are most often affected by strains expressing toxin types A, B, and E. Recently, there have been some reports of toxins type F and G causing botulism in humans (1,2). The toxins are large molecular weight proteins of approximately 150,000, which bind to the presynaptic membrane of neurons at peripheral cholinergic synapses to prevent release of acetylcholine and block neurotransmission. The blockade is most evident clinically in the cholinergic autonomic nervous system and at the neuromuscular junction. Human signs and symptoms of intoxication appear 12 to 36 hours after exposure (range 6 hours to 8 days). The clinical spectrum of disease includes food-borne, wound, and infant botulism (3). If used against our forces as warfare agents, botulinum toxins would most likely be delivered by aerosol to the respiratory tract, with the clinical presentation being very similar to that seen with food-borne botulism. General symptoms of weakness, lassitude, ptosis, and dizziness appear early. Symmetrical weakness of the extremities occurs, usually in descending fashion and with variable speed, and can progress to respiratory arrest with unexpected suddenness. Death is normally by paralysis of the respiratory muscles. Recognized cases of botulism before 1950 had a mortality rate of 60%. When current therapy, consisting of equine trivalent (A,B,E) botulinum antitoxins, tracheostomy, ventilatory assistance and intensive supportive care is initiated early, fatalities are expected to be less than 5% (4-6).

MATERIALS AND METHODS

Study Group:

The study was approved by institutional review committees. Before enrollment, informed consent was obtained from each volunteer.

In this clinical trial, each volunteer received a skin test for sensitivity to the product by inoculating 0.1 ml of a 1:10 dilution of antitoxin intradermally in the patient's forearm. The subject was observed for allergic reactions for 20 minutes. If no allergic response occurred, the volunteer received a full dose of antitoxin. The full dose of antitoxin consisted of 10 ml of the product, administered intravenously. The volunteer remained hospitalized for 48 hours after infusion for evaluation of immediate side effects. Sera were collected from the volunteers before the infusion, at 2 hours, 8 hours, and 12 hours during the first day and on a daily basis for the first week. Subsequent samples were collected on days 10, 14, and 21. Serum samples were tested for antitoxin levels. Serum chemistries consisting of liver tests (alanine aminotransferase, alkaline phosphate and bilirubin) and blood urea nitrogen (BUN) were run. Complete blood counts (CBC) were also conducted at the same time. Finally, complement cascade assays including C1q, C3, C4, and CH_{50} were tested in serum or plasma samples collected at the same time intervals.

Antitoxin Preparation:

USAMRIID investigators started the immunization of a thoroughbred horse with botulinum toxoid followed by booster doses of toxins (serotypes A-G) to induce the production of high-titer antitoxin serum (unpublished data, G. Lewis). Under contract with the U.S. Army, the horse was shipped to the University of Minnesota, ALG Program, for the purpose of completing the immunization program and performing large scale plamapheresis with subsequent fractionation and purification of despeciated IgG capable of neutralizing types A, B, C, D, E, F, G botulinum toxins (unpublished data, R. Condie).

Plasma collected from this horse was stabilized with SiO_2 in order to eliminate fibrinogen, plasminogen, and other proteins. The plasma was filtered twice during fractionation--once after stabilization and once after despeciation. Despeciation is designed to reduce the immunogenicity of horse plasma by eliminating extraneous proteins and the Fc portion of the immunoglobulin, leaving only the antibody- binding components. Despeciation was accomplished by digestion with pepsin under controlled conditions of temperature and pH. The inactivated pepsin was

removed from the plasma by batch filtration procedures. Protein solutions were concentrated and filtered at several points during fractionation with ultracentrifugation systems. After these procedures, the $F(ab')_2$ fragments were isolated from the plasma solution by column chromatography. The antitoxin was tested for sterility, pyrogenicity, and general safety with satisfactory results. In addition, final products were tested for purity and electrophoretic mobility by immune electrophoresis. The final product then contained despeciated equine IgG, from which the Fc portion had been removed, leaving only the $F(ab')_2$ portion. Immediately before human use, the antitoxin neutralization activity of the product was determined to be (for 10 ml): For type A:10,210 IU/ml, type B: 4,800 IU/ml, type C:1,480 IU/ml, type D:72,320 IU/ml, type E: 8,000 IU/ml, type F:880 IU/ml, and type G:3,200 IU/ml. Prior to initiating this clinical study, this product was tested for efficacy in a mouse model for botulinum poisoning. There was a 100% survival in mice challenged on days 1 through 7 after receiving the antitoxin intraperitoneally. The animals experienced decreased survival on each succeeding day to day 14, on which only 20% of the animals survived the toxin challenge. These results indicated that the equine heptavalent antitoxin was effective in neutralizing toxin *in vivo*. Clinical testing was the logical next step.

RESULTS

There were no side effects immediately after administration of the antitoxin. All laboratory tests, including CBC, liver tests, BUN, and the complement cascade assays, remained normal. The five volunteers remained healthy for the 4 months of follow-up. The antitoxin levels for type A are shown on Table 1. As observed, the antitoxin was detected on the first sample of infusion, which was taken at 2 hours, and the serum levels were still detectable at day 21, although at a much lower range.

TABLE 1

Neutralization Assay - Toxin Type A (IU/ml)

TIME	Vol 248	Vol 526	Vol 177	Vol 527	Vol 590
Day -1	0.0	0.0	0.0	0.0	0.0
Day 1					
0 hr	0.0	0.0	0.0	0.0	0.0
2 hr	2.56	2.5	3.23	6.45	4.06
8 hr	1.02	1.0	0.91	1.28	1.28
12 hr	0.64	0.64	0.81	1.28	1.28
Day 2	0.16	1.02	0.76	1.02	0.40
Day 3	0.57	0.23	0.32	0.29	0.32
Day 4	0.10	0.18	0.32	0.16	NA
Day 5	0.20	NA	0.11	0.16	0.23
Day 6	0.20	0.13	0.13	NA	NA
Day 7	0.11	0.11	NA	0.13	0.13
Day 8	0.06	0.11	0.11	0.13	0.02
Day 10	0.05	NA	NA	NA	0.11
Day 11	NA	0.11	0.17	0.10	0.10
Day 12	NA	0.17	0.11	0.10	0.07
Day 13	0.11	0.11	0.05	0.05	0.05
Day 14	0.04	0.17	0.04	0.04	0.05
Day 21	0.04	0.03	0.03	0.04	0.03

DISCUSSION

There is currently only one IND (Investigational New Drug) vaccine available in the United States for human use--a pentavalent, formalin-inactivated alum-adjuvanted toxoid for types A, B, C, D, and E (manufactured by the Michigan Department of Public Health), this product needs to be given before exposure to be effective. In addition, it needs to be administered multiple times and require annual boosters. It is not realistic to expect that civilian populations or all military populations would receive large-scale immunization against botulinum toxins.

The present medical armamentarium is limited to supportive therapy. For this reason, passive antibody as botulinum antitoxin is necessary to treat subjects exposed to botulinum toxins. The only antitoxin currently approved for use in the United States is an equine trivalent (A,B,E) antitoxin, manufactured by Connaught Laboratories (4,6). The recommended dose is two vials, and the hypersensitivity reaction rate is approximately 9% (7) with this product. Human pentavalent botulinum immunoglobulin (A,B,C,D,E) prepared from human botulinum immune plasma is available in very limited quantities as an IND drug in the United States (8,9).

In these circumstances, the heptavalent despeciated $F(ab')_2$ equine antitoxin offers two major advantages. One of them is that it is directed against the seven toxins and, therefore, in theory, is expected to protect against every known type of botulinum toxin. The other advantage is that the Fc portion of the immunoglobulin was removed and the $F(ab')_2$ fragments were isolated. This process minimizes the risk of serum sickness and other complications associated with the currently licensed trivalent equine product. Moreover, this product has been found to be efficacious in animal models.

In recent studies conducted at USAMRIID, six rhesus monkeys were given this product 48 hours before challenge with 4 to 7 inhalation LD_{50} of *C. botulinum* type A toxin and survived without signs of illness, while controls died in approximately 48 hours (unpublished data, D. Franz). At the time of challenge, the serum-neutralizing antibody titers in protected animals were from 0.6 to < 0.02 IU/ml. Therefore, these animals were protected with very low antitoxin titers. The neutralizing antibody assay used for these animals was the same method that we used to detect antibody in our volunteers. These data suggest that much less antibody is needed to protect against inhalation botulism.

Although the main disadvantage of using antitoxins as therapy is their rapid clearance by the immune system and the theoretical risk of serum sickness, we did not observe any evidence of serum sickness in these volunteers either clinically or when we assayed for complement cascade tests. The fact that the antitoxin was still detectable at 21 days after infusion is encouraging. The present work supports the concept that the administration of these antitoxins could lower the case fatality rate in this disease, if given early after exposure. Further evaluation of this antitoxin is warranted in clinical cases. It is likely that this preparation could replace the current trivalent antitoxin.

Bibliography

1. W.F. Sonnabend, O.A. Sonnabend, P. Grundler, E. Ketz, Intestinal toxicoinfection by *Clostridium botulinum* type F in an adult. Lancet 1987;357-360.

2. D.F. Gimenez, A.S. Ceccarelli, Another type of *Clostridium botulinum*. Zentralb. Bakteriol 1970;215:221-224.

3. W. Schaffner, *Clostridium Botulinum* (Botulism). In *Principles and Practice of Infectious Diseases* (Mandell, Douglas, Bennett, eds) John Wiley & Sons, Inc., New York, 1985 pp. 1359-62.

4. Centers for Disease Control. Botulism in the United States 1899-1977. Handbook for Epidemiologists, clinicians and laboratory workers. Atlanta, Centers for Disease Control 1979.

5. J.G. Morris, Current trends in therapy of botulism in the United States. In Biomedical Aspects of Botulism (Lewis, ed) Academic Press, New York, 1980 pp. 317-326.

6. C.H. Hatheway, J.D. Snyder, J.E. Seals at al. Antitoxin Levels in Botulism Patients Treated with Trivalent Equine Botulism Antitoxin to Toxin, Types A, B and E. J. Infect. Dis. 1984;150:407-412

7. R.E. Black, R.A. Gunn, Hypersensitivity reactions associated with botulinal antitoxin. *Am J Med* 1980;69:567-570.

8. G.E. Lewis, J.F. Metzger, Botulism Immune Plasma (Human). Lancet 1978;8090:634-635.

9. R.M. Condie, C.P. Drayton, M. Setterholm et al. Preparation of an Intravenous, Human, Pentavalent, Botulinum Immunoglobulin (manuscript in preparation).

ACKNOWLEDGMENTS

The authors wish to acknowledge the secretarial support of Ms. Judy Reichard.

USAMRDC contract No. DAMD17-82-C-2119 supported in part the research and preparation of the equine heptavalent botulinal antitoxin at the University of Minnesota.

Scaling Properties of the Magnetoexciton Problem

Doran D. Smith*, Dr.
U. S. Army Electronics Technology and Devices Laboratory
SLCET-ED, Fort Monmouth, New Jersey 07703-5601

I. BACKGROUND

An exciton is a quasi-particle that occurs in many solid state systems, especially optical ones. Excitons are the basis for several important optical devices and have been studied extensively by the scientific and engineering communities. An exciton is the bound state of an electron and a hole. The electrons negative charge and the holes positive charge experience a mutually attractive coulomb interaction, from which they form a bound state much like a hydrogen atom. The excitons binding energy is determined by the properties of the material in which it is contained. For many semiconductors of technological interest the exciton binding energy is dominated by the material's dielectric constant, ϵ, and the reduced mass of the exciton, μ, where

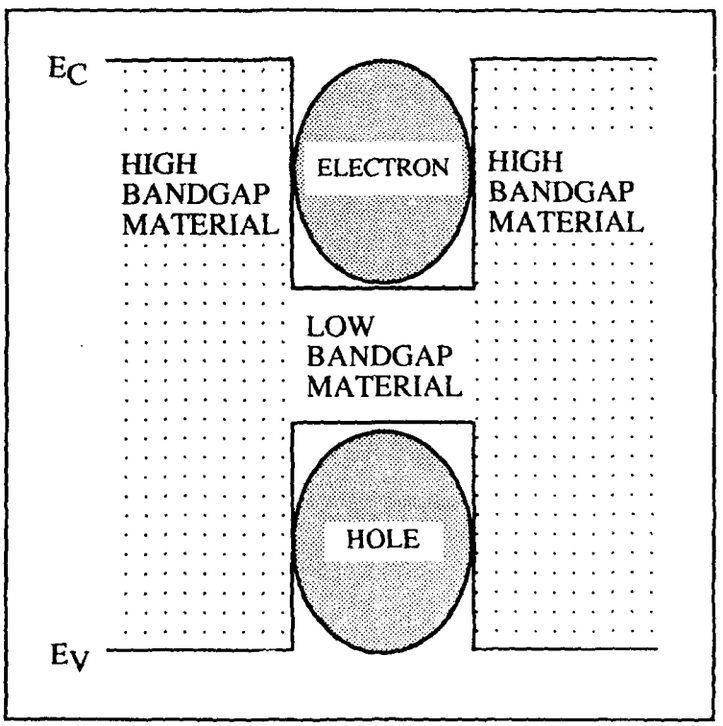


Figure 1. Squeezed electron and hole in a quantum well.

$1/\mu = 1/m_c + 1/m_h$. The greater the reduced mass the greater the binding energy.

As the exciton becomes more massive it will become smaller (more localized) and hence its binding energy will increase. The more localized the electron and hole are, the greater their charge overlap. (Charge density is proportional to the square of a particles' wavefunction.) The greater the overlap of the electron and hole, the greater their binding energy due to increased coulomb interaction. Another way to increase the electron and hole overlap is to physically squeeze them into a smaller volume. This can be done with the use of artificial materials, see Figure 1. By constructing a material composed of a thin slab of low band gap material surrounded on both sides by high band gap material (a quantum well), the electron and hole will be squeezed. When this electron and hole form an exciton, it will also be squeezed, see Figure 2. Modifying the optical properties of excitons by the use of quantum wells is a common practice in optical devices.

An exciton confined to a quantum well behaves much like a hydrogen atom with one major difference, it is confined to move in a plane instead of being able

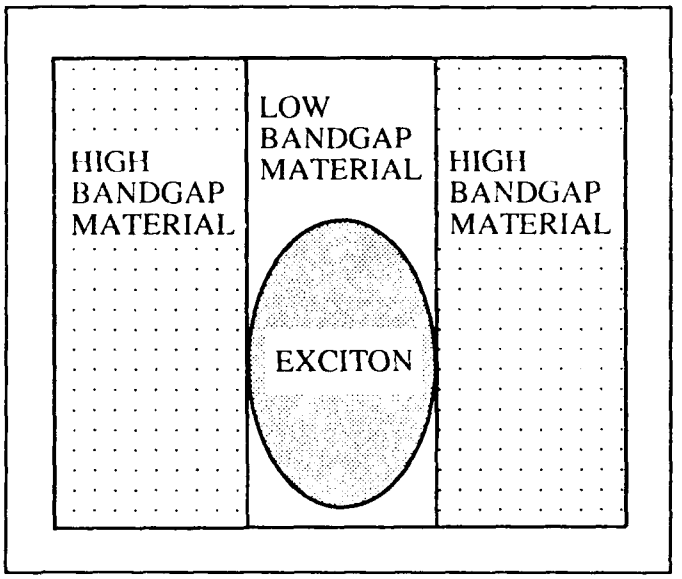


Figure 2. Exciton squeezed in a quantum well.

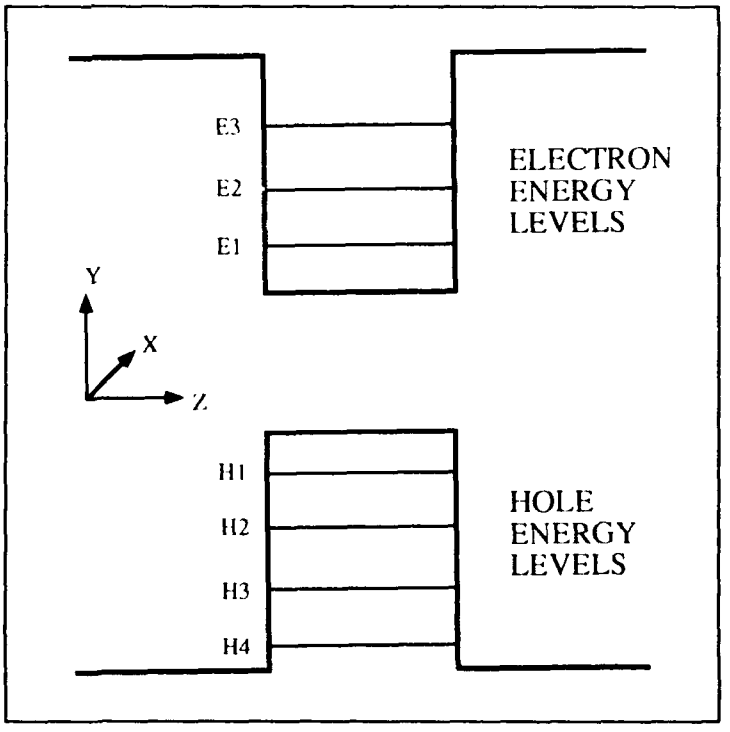


Figure 3: Allowed energy levels for an electron and hole in a quantum well.

to move freely in all three dimensions. The exciton in a quantum well is dynamically two dimensional. It is dynamically two dimensional because its motion is confined to a plane, but it is still electrostatically three dimensional. (The changing dielectric constant complicates the problem, but does not modify its dimensionality.)

A consequence of this confinement in the third dimension (assume z) is the allowed energy levels (subbands) in the z direction are quantized, see Figure 3. The electron and hole energy levels are each quantized separately, e_1, e_2, ... and h_1, h_2, ... These energy levels are functions of well width and particle masses. Excitons are formed from an electron in a given electron energy level, say e_1, and a given hole, say h_1.

In analogy with the hydrogen atom, the exciton has a ground state, 1S, and an infinite number of excited states, 2S, 2P, 3S, 3P, 3D, ... The energy of each of these levels is a function of the exciton's reduced mass and quantum well width. The reason for this dependance is the overlap of the electron and hole are effected by both the exciton reduced mass and the quantum well width. (Hole masses are also strongly effected by the quantum well width because of subband mixing.) The Bohr radius of the exciton is given by $a_o = \epsilon h^2/\mu e^2$, where ϵ is the material's dielectric constant, h is the Plank constant, and e is the electron charge. The Bohr radius is approximately the ground state (1S) radius of an exciton in bulk material. Therefore, this system can be thought of as having two length scales, the Bohr radius and the quantum well width. It is the existence of previously undiscovered scaling laws between these two length scales which is the subject of the rest of this paper.

II. SCALING LAWS

Although there have been numerous papers published on excitons in quantum wells[1] there is no simple way to determine the binding energy of the ground and excited states of an exciton in a finite width quantum well. In the two dimensional case (zero width quantum well with infinitely high barriers) and in the three dimensional case (bulk material) there are simple analytical solutions. It is the original contribution of this paper to show that due to scaling laws the quantum well width and Bohr radius (reduced mass) of the exciton are not independent. Once the binding energy is known for all exciton reduced masses and **one** quantum well width, then the binding energy is known for all well widths. It is expected that these scaling laws will useful to individuals seeking a simple way to determine binding energies in finite width quantum wells, instead of the less accurate method of scaling the closed form two or three dimensional results.

Duggan[2] presented a form of the Hamiltonian describing an exciton which accounts for the coulomb interaction of the hole and electron, the

finite width of the quantum well, and the effect of a magnetic field normal (z direction) to the quantum well walls, as shown in Eq. (1)

$$\left\{-\frac{h^2}{2\mu}\left[\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial}{\partial}\rho\right)+\frac{1}{\rho^2}\frac{\partial^2}{\partial}\phi^2\right]+\frac{e^2}{8\mu}B^2\rho^2+\frac{e}{2}\left[\frac{1}{m_h}-\frac{1}{m_e}\right]Bmh+V(\rho)\right\}\Phi(\rho,\phi) = \tag{1}$$

$$E\Phi(\rho,\phi)$$

where $V(\rho)$ is given by

$$V(\rho)=-\frac{e^2}{4\pi\varepsilon}\iint\frac{dz_e dz_h|Z_e(z_e)|^2|Z_h(z_h)|^2}{\{\rho^2+(z_e-z_h)^2\}^{1/2}}. \tag{2}$$

and m_e and m_h are the effective masses of the electron and hole respectively, B is the value of the magnetic field, and m is the azimuthal quantum number. Eqs. (1) and (2) assume that the exciton wavefunction can be written in the form,

$$\Psi(z,\rho,\phi)=Z(z)\Phi(\rho,\phi)=Z_e(z_e)Z_h(z_h)\Phi(\rho,\phi) \tag{3}$$

The Z's are the electron and hole wavefunctions in the z direction. This assumption is called the separable approximation. It has been shown to be valid for quantum wells no wider than 1.4 times the Bohr radius[3]. Simplifying the above Hamiltonian to the optically active S states (m=0) and converting it to a dimensionless form results in

$$\left\{\frac{\partial^2}{\partial x^2}+\frac{1}{x}\frac{\partial}{\partial x}-\frac{\gamma^2x^2}{4}+2V'(\rho)\right\}\psi(x)=E\psi(x) \tag{4}$$

where $V'(\rho)$ is given by

$$V'(\rho) = a_o N^4\int_{-\frac{L}{2}}^{\frac{L}{2}}\int_{-\frac{L}{2}}^{\frac{L}{2}}\frac{\cos^2(az_e)\cos^2(az_h)\,dz_e dz_h}{\{\rho^2+(z_e-z_h)^2\}^{\frac{1}{2}}} \tag{5}$$

and $a_0 = \epsilon h^2/\mu e^2$, $x = \rho/a_0$, $\gamma = h\epsilon B/\mu ce$, $a = \pi/L$, $N^2 = 2/L$ and infinite barriers have been assumed. By defining $x' = az_e$ and $y' = az_h$ Eq. (5)

becomes,

$$V'(x) = \frac{a_o}{L}\frac{4}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\cos^2(x')\cos^2(y')\,dx'dy'}{\left\{\frac{a_o^2}{L^2}\pi^2x^2+(x'-y')^2\right\}^{\frac{1}{2}}} \tag{6}$$

Inspection of Eqs. (4) and (6) show that when B=0, there are only two length scales in the Hamiltonian, a_0 and L, and the Hamiltonian depends only upon their ratio. Therefore, parameterizing the solution in terms of their ratio, the binding energy of any state can be found as a function of a_0/L.

There is a scaling law for the $B \neq 0$ case also, however, its use is more involved. It is convenient to rewrite Eq. (4) as,

$$\left\{\frac{\partial^2}{\partial x^2}+\frac{1}{x}\frac{\partial}{\partial x}-\left(\frac{he}{ce}\right)^2(\mu a_o)^2\left(\frac{B}{\mu^2}\right)^2\frac{x^2}{4}+2V'(x)\right\}\psi(x)=E\psi(x) \tag{7}$$

where the term μa_0 contains no μ dependence. Let a_0/L remain constant, but let each term change by a factor δ. $V'(x)$ is unchanged, but a_0 in the coefficient of the x^2 term is multiplied by δ. Two possible sources of change for a_0 are μ and ϵ. If the change occurred in μ then μ becomes μ/δ as shown in Eq. (8)

$$\left(\frac{he}{ce}\right)^2(\mu a_o)^2\frac{B^2}{(\frac{\mu}{\delta})^4} = \left(\frac{he}{ce}\right)^2(\mu a_o)^2\frac{(B\delta^2)^2}{\mu^4} \tag{8}$$

Hence, if B is scaled so that $B_{new} = B_{old}/\delta^2$, the solutions will remain unchanged. In other words, once a solution as a function of B field is obtained for some value of μ and L, the solution to a different L, δL is obtained by changing μ to μ/δ and scaling the B field axis by $1/\delta^2$.

Changes in ϵ are handled similarly. The ratio a_0/L is maintained constant, L becomes δL, and B is scaled so that $B_{new} = B_{old}/\delta^2$.

Although the $B \neq 0$ case is more inconvenient than the B = 0 case since a two dimensional table must be used, it is useful for doing calculations. Only a two dimensional study has to be done instead of a three dimensional one. Also, knowledge of how the Hamiltonian scales permits internal consistency checks on numerical solutions for both the B = 0 and $B \neq 0$ cases.

Table I lists the binding energy of the five lowest bound states for the system described by Eq. (4) in terms of the ratio a_0/L. The values chosen cover the technologically relevant range for both GaAs and InGaAs quantum wells. Also listed for ϵ = 12.5 (GaAs) and for a 10 nm quantum well, are the corresponding values of μ. The binding energy for any system

is obtained by using Eq. (9) and Table I with L in nm, μ as a fraction of the free electron rest mass (m_0), and the binding energy in Rydbergs.

$$\frac{a_o}{L} = \frac{\frac{\epsilon h^2}{\mu e^2}}{L} = \frac{0.66147}{\mu L} \qquad (9)$$

a_o/L	$\mu(m_o)$ L=10nm	Binding energy (Rydbergs) n = 1	2	3	4	5
66.147	0.001	-3.9059	-0.44092	-0.15924	-0.081354	-0.049251
33.073	0.002	-3.8247	-0.43782	-0.15856	-0.081108	-0.049136
22.049	0.003	-3.7501	-0.43493	-0.15793	-0.080877	-0.049027
16.537	0.004	-3.6812	-0.43221	-0.15734	-0.080659	-0.048924
13.229	0.005	-3.6171	-0.42964	-0.15677	-0.080452	-0.048826
11.025	0.006	-3.5570	-0.42720	-0.15623	-0.080254	-0.048732
9.4496	0.007	-3.5006	-0.42487	-0.15572	-0.080065	-0.048643
8.2684	0.008	-3.4473	-0.42265	-0.15523	-0.079883	-0.048556
7.3497	0.009	-3.3967	-0.42051	-0.15475	-0.079707	-0.048473
6.6147	0.010	-3.3487	-0.41845	-0.15429	-0.079538	-0.048393
3.3074	0.020	-2.9668	-0.40116	-0.15039	-0.078088	-0.047704
2.2049	0.030	-2.6956	-0.38770	-0.14730	-0.076929	-0.047149
1.6536	0.040	-2.4719	-0.37656	-0.14469	-0.075945	-0.046677
1.3229	0.050	-2.3195	-0.36700	-0.14249	-0.075087	-0.046263
1.1025	0.060	-2.1803	-0.35860	-0.14042	-0.074316	-0.045891
0.94495	0.070	-2.0620	-0.35109	-0.13860	-0.073621	-0.045554
0.82684	0.080	-1.9598	-0.34428	-0.13693	-0.072977	-0.045241
0.73497	0.090	-1.8702	-0.33805	-0.13539	-0.072387	-0.044954
0.66147	0.100	-1.7907	-0.33230	-0.13396	-0.071833	-0.044683

Table I Scale invariant exciton binding energies. All numbers were accurate to at least 100ppm before being rounded to the given values.

III Discussion

The theory described in the previous section needs no experimental verification in the sense that the scale invariance of the Hamiltonian is simply a property of the differential equation. However, whether the Hamiltonian is complete enough to accurately describe the physical system is a relevant question. For as simple as the theory is, it does an excellent job of predicting the properties of magnetoexcitons in finite width quantum wells, see Figures 4 (from ref. 2), 5 and 6 for examples[4]. These figures compare the experimentally measured energy levels (+) of excitons confined to quantum wells of widely varying widths to the theoretical predictions(•). All wells are GaAs. The barriers in figures 5 and 6 are AlAs, and the barrier in figure 4 is $Al_{0.35}Ga_{0.65}As$. For all of the data shown here the exciton is formed from an electron and hole deep in the quantum well $(e_1\text{-}hh_1)$[5], effectively the quantum well is infinitely deep.

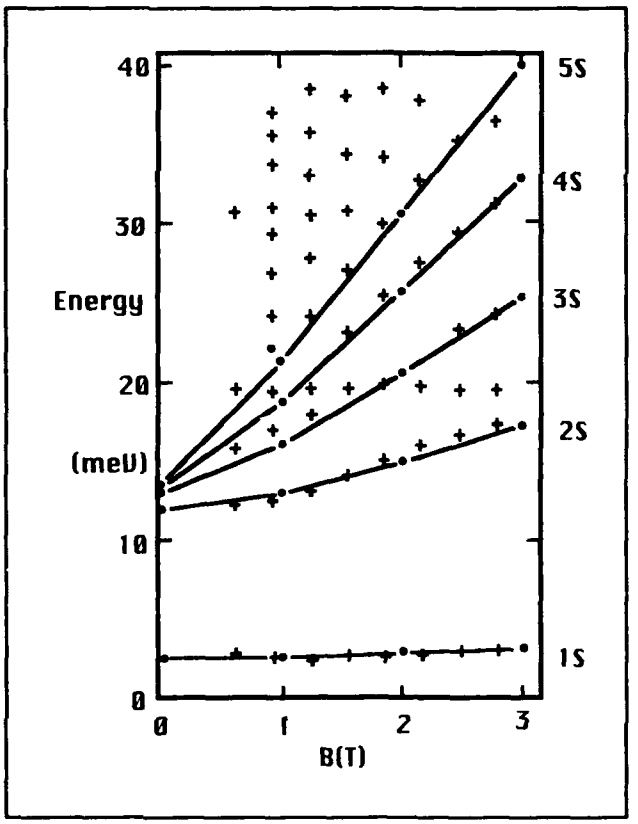


Figure 4 7.5nm wide quantum well, μ=0.0486, a_0/L=1.816.

D.D.S gratefully thanks Mitra Dutta for initially introducing me to the magnetoexciton problem, and gratefully acknowledges several useful conversations with Sergey Rudin.

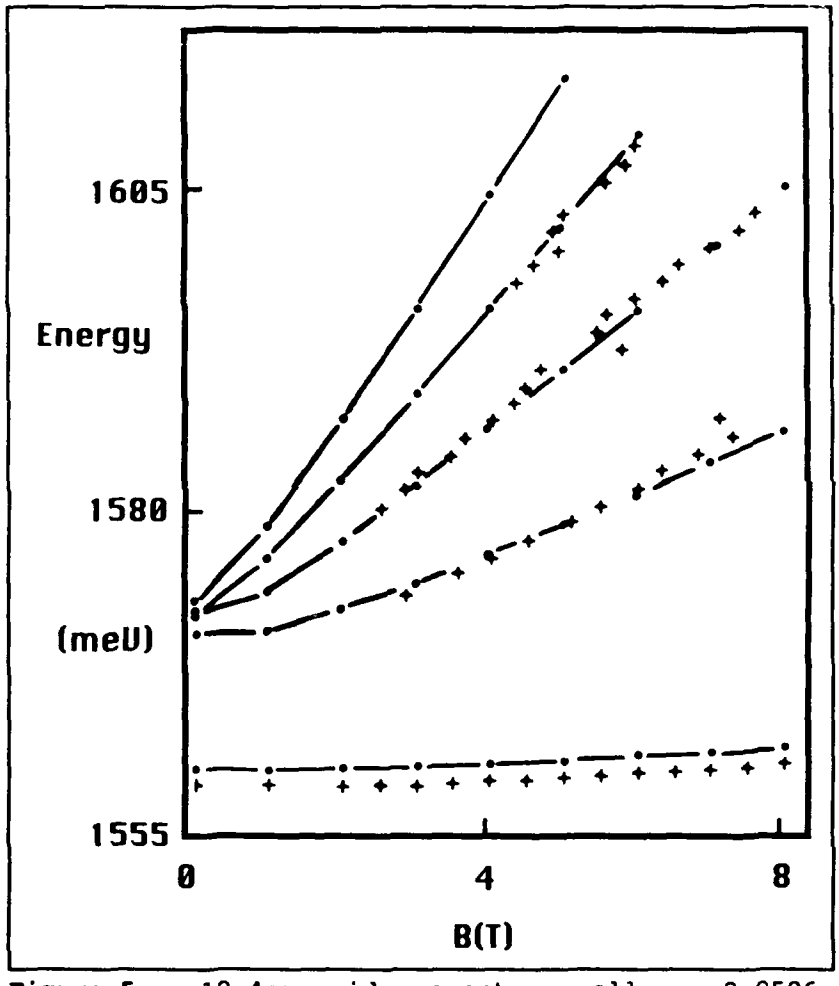


Figure 5 10.4nm wide quantum well, μ=0.0586, a_0/L=1.096.

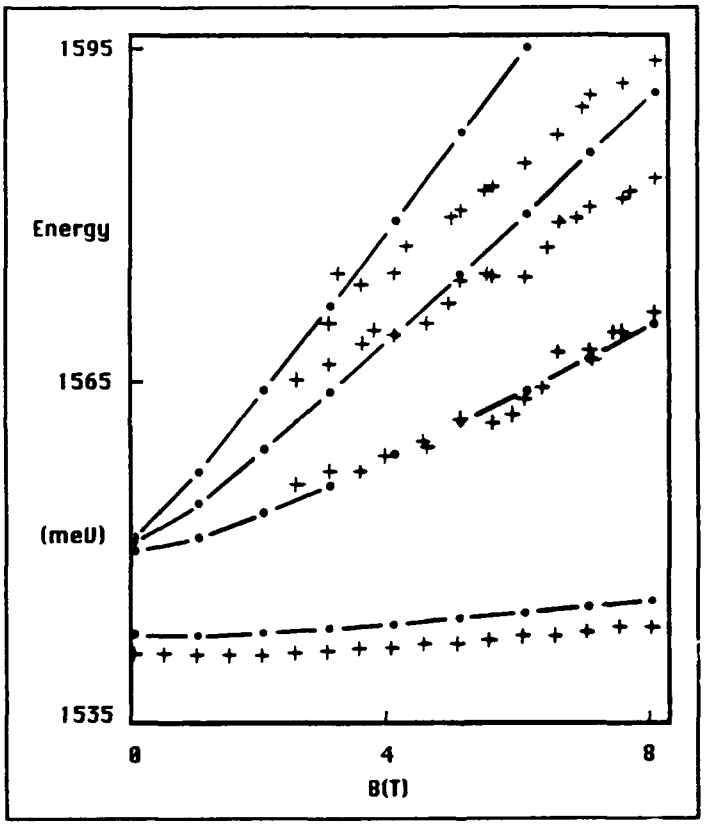


Figure 6 14.1nm wide quantum well, μ=0.043, a_o/L=1.091.

REFERENCES

1. Doran D. Smith, M. Dutta, X. C. Liu, A. F. Terzis, A. Petrou, M. W. Cole, and P. G. Newman, Phys. Rev. B, **40**, 1407, (1989) and contained references for a list of papers dealing with the magnetoexciton problem. For a special issue on semiconductor quantum wells and superlattices: physics and applications see, IEEE J. Quantum Electron., **QE-22**, Sept. 1986.

2. G. Duggan, Phys. Rev B **37**, 2759 (1988).

3. G. Bastard, E.E. Mendez, L. L. Chang and L. Esaki, Phys. Rev. **B** 26, 1974 (1982).

4. M. Dutta, Doran D. Smith, P. G. Newman, X. C. Liu, and A. Petrou, Phys. Rev. B **42**, 1474 (1990).

5. hh_1 means heavy hole one and corresponds to the h_1 introduced in section I of this paper. The reason for the designation hh is that there are two types of holes in GaAs, heavy holes and light holes (lh).

SMITH, L.

Cloning, Characterization, and Expression of Animal Toxin Genes for Vaccine Development

*Leonard A. Smith, Dr.
Department of Immunology and Molecular Biology
Toxinology Division
U.S. Army Medical Institute of Infectious Diseases
Fort Detrick, Frederick, MD 21702-5011, U.S.A.

INTRODUCTION

The advent of molecular immunology and recombinant DNA technology has engendered research in every area of the biological sciences. The role played by these technologies in vaccine development is becoming increasingly pivotal with the emergence of synthetic peptides; anti-idiotypes; and genetically engineered, live, attenuated recombinant viruses as potential vaccines. Basic research studies and vaccine development against many animal toxins has been slow for economical reasons. To date, investigative efforts in these areas have focused on the elucidation of structure-activity relationships of several animal toxins and the production of non-toxic antigens that could be used as vaccine candidates against these toxins.

Recombinant DNA technology has been used to clone and characterize animal toxin genes encoding proteins with variable pharmacological activities. Current endeavors have concentrated initially on the expression of cloned genes in host systems that can process the recombinant products into biologically active forms. After suitable hosts are obtained for the expression of toxin genes, site-directed mutagenesis may be employed as a means of altering the cloned genes in specific regions so the resulting recombinant products will have strategically replaced amino acid residues relative to the native toxin. Future research goals will involve the generation of cross-reacting materials (CRMs), which are mutant proteins that have lost their toxic or functional activities while still retaining immunogenicity. This research will yield insights into the structure-activity relationships of animal toxins and generate potential vaccine candidates against these same toxins.

SMITH, L.

The purpose of this paper is to present a brief background on animal protein toxins; to survey recombinant DNA techniques directly applicable to the cloning, characterization, and expression of genes encoding animal toxins; and to review published and ongoing work on the molecular biology of animal toxin genes.

CLASSES OF ANIMAL TOXINS

Many different kinds of animal protein toxins exist in nature. The sources from which they originate are equally diverse. They range from certain types of marine snails; stinging coelenterates and fish; to spiders, beetles, scorpions, snakes, insects, and even frogs. The pharmacological effects of purified venom components include neurotoxic, membrane-damaging, and blood coagulation/anti-coagulation-inducing effects. For example, various components of animal venoms exert their coagulant activity at various levels of the blood-clotting cascade, as in the case of the thrombin-like enzymes found in snake venoms. Membrane damaging toxins can be cytolysins that produce hemorrhage in tissues and are destructive to endothelial cells and blood vessels (hemorrhagins), or they can be hemolysins assaulting erythrocytes and interrupting their structural integrity resulting in the liberation of hemoglobin from red blood cells. Neurotoxins are substances that affect the normal functioning of excitable tissues attributable to a specific recognition and binding affinity to distinct sites in these tissues. Neurotoxins can be categorized according to effects and sites of action in the nervous system. For example, ion channel toxins modify ion conductance, presynaptic toxins affect neurotransmitter release, and postsynaptic toxins interfere with the binding and resulting action of neurotransmitters. Most of the molecular biology research that has been performed with animal toxins has pursued the study of snake and scorpion neurotoxins. The focus of this review will be on the genes encoding these toxins.

PROTEIN STRUCTURE OF ANIMAL TOXINS

Herpetologists present detailed and convincing explanations why some snakes evolved venom glands for survival competition (1). If the enigmatic evolutionary nature of snake venoms is captivating to naturalists, then the structural and

functional diversity of purified toxins should be as interesting to the protein chemist. Many of the presynaptically-acting neurotoxins isolated from snake venoms thus far have protein structures that vary greatly. Regardless of the differences in protein structure among presynaptically-acting snake neurotoxins, all possess phospholipase A2 (PLA2) activity, function at the prejunctional neuromuscular level, and ultimately cause the inhibition of acetylcholine release from presynaptic nerve terminals. In addition to structural diversity, many of the snake venom neurotoxins exhibiting PLA2 activity possess varying degrees of lethal potency in mice, as well as differences in enzymatic activity. Extensive research on the relationship between PLA2 activity and toxicity has not fully clarified the confusion among toxinologists. The structural heterogeneity among neurotoxins with similar or identical functions is not restricted to snake venom presynaptic neurotoxins. These differences are also observed with ion-channel and postsynaptic neurotoxins as well.

Although animal toxins are low molecular weight proteins, their protein structures are, nevertheless, complex. Most have multiple disulfide bonds and many have up to 14 disulfide bridges. Some have modifications at their carboxyl-terminus as in the case of many snail conotoxins, scorpion toxins and bee venom toxins. This modification consists of a C-terminal a-amide group (-X-NH2). About half of the bioactive peptides found in nervous and endocrine systems possess a C-terminal a-amide moiety, and for most of these, that modification is important to bioactivity. Animal toxins are secreted from cells as exocrine products, and as such, are generally synthesized with an N-terminal signal sequence necessary for initiating the exportation process from the cell. Usually, the signal peptide must be removed if the protein is to achieve its biologically active conformation. As with other protein modifications, removal of the signal sequence occurs as a cellular post-translation processing event. In order to express these toxins, one must be capable of manipulating the cloned gene and its host or its expressed recombinant product in such a fashion as to reproduce the processing functions as they occur in the animal. The next section will be devoted to the techniques that can be used in performing this formidable task.

RECOMBINANT DNA TECHNOLOGY

There are many publications describing in detail various techniques to be used in creating gene libraries; for screening gene banks in search of recombinant isolates of interest; and

characterizing cloned inserts by sizing, mapping, and nucleotide sequence analysis (2-5). Simply explained, genes can be cloned either from the genome, as reverse transcripts from messenger ribonucleic acid (mRNA), or as synthetic constructs from organic chemical reactions. Since most of the work in our laboratory has involved the construction and characterization of cDNAs, only that technology is described here.

All procedures used to construct cDNA libraries require the extraction and purification of mRNA from cells in which the most abundant poly(A)+ RNA is expected. In the case of mRNA encoding exocrine-secreted toxins, the glands themselves are the source of the mRNA. Since geographical variation of venom proteins among snakes of the same species has been documented(6), venom samples from individual animals to be used for molecular biology studies should be analyzed for the presence and quantity of individual constituents of interest prior to the removal of tissues from the animal. In addition, it is desirable to evacuate the venom gland(s) 1 day prior to their removal in order to increase the potential yield of poly(A)+ RNA (7). Glands removed are frozen in liquid nitrogen and stored at -700C until used. One of the most efficient methods for extracting total RNA from cells or tissues is the guanidinium isothiocyanate-hot phenol extraction method (8). Total RNA from this extraction is then subjected to oligo-dT column chromatography to enrich the mRNA pool (9). The integrity of the poly(A)+ RNA and its ability to serve as a template for full-length cDNA transcription can be evaluated on a small scale prior to library-sized reactions. Many commercial cDNA kits are available which contain buffers, dNTP solutions, oligo-d(T) primer, purified enzymes necessary for cDNA synthesis, and detailed manuals explaining each step.

Various methods exist for designing cDNA libraries. Essentially all the methods for constructing cDNAs rely on a primer-initiated reverse transcriptase to create the complement of the mRNA sequences. The newly formed cDNAs are then used as templates to synthesize double-stranded DNAs corresponding to sequences of the mRNAs. For molecular cloning, the duplex cDNAs are covalently linked to a plasmid or bacteriophage vector via complementary homopolymeric tailing or cohesive ends created with linker or adaptor segments containing appropriate restriction sites. Recombinant plasmids are used directly to infect *E. coli* cells, while bacteriophage constructs must be packaged into viable phage particles prior to infecting *E. coli* cells. Both methods result in the generation of primary cDNA libraries.

Primary or amplified libraries are then screened. Libraries or individual clones can be examined for their insert content by using appropriately radiolabeled nucleotide probes with sequence homology to a portion of the desired gene. There are other techniques available for screening gene libraries. Immunological methods rely on the assumption that the cloned gene is expressed and the product solubilized and in a conformational state that will be recognized by existing antibody reagents. Post-translational modifications necessary to promote the formation of biologically active animal toxins are complex and *E. coli* may not have the capability to perform those tasks. As the amino acid sequences for many animal toxins are available, the most efficient screening procedure for libraries containing animal toxin genes is hybridization with isotopically labeled oligonucleotide probes generated from known protein sequences or cDNA probes, if they are available.

Aside from expressing the gene and characterizing the recombinant product, DNA sequence analysis is the only method for unequivocally determining whether or not a positive clone from a library screen actually contains the desired gene. There are two different approaches to sequencing DNA: one is the Maxam and Gilbert method which is chemistry based (10); another Sanger or dideoxy chain termination sequencing, is enzymatic(11). Deoxynucleotide sequencing kits containing reagents and instructions for both types of sequencing are commercially available. Dideoxy chain termination has become extremely popular with the construction of cloning vectors, such as M13, pGEM, and Bluescript, that contain unique priming sequences for the polymerizing enzymes (e.g., Klenow, Taq DNA polymerase) used in this method. Inserts from positive library clones can be subcloned into these vectors by restriction enzyme excision of the insert from the DNA clone and reinserted into the specialized vector (2) or, in some cases, by using an in-vivo homologous recombination procedure (12).

After the nucleotide sequence of the cloned gene has been determined, ensuing experiments usually include expression of the cloned gene. There are a number of different expression systems available for gene expression. The choice of system to be used depends somewhat on the genes to be expressed. Prokaryotic hosts (e.g., *E. coli*) have been used to express many different kinds of heterologous genes as fused or unfused products (14). In addition, expression of cloned genes in *E. coli* can be regulated by thermal or chemical induction depending on the vector used. Sometimes, however, prokaryotic hosts express foreign genes poorly or their expressed products are biologically inactive. In addition, bacterial cells do not carry out post-translational modifications such as

glycosylation, amidation, phosphorylation, and cleavage of protein precursors. For these reasons, progress in molecular genetics has relied partly upon the availability of a broad and increasingly sophisticated array of cloning vectors and host systems.

The yeast strain, *Saccharomyces cerevisiae*, is a convenient host for the expression of heterologous genes. Benefits of using yeasts include high levels of secretion, the ability to express more protein per liter than mammalian systems although less than *E. coli*, and no additional protein refolding steps required which may be necessary in bacterial expression systems. Another yeast system, *Pichia pastoris*, has recently been shown to expressed very high yields of recombinant proteins by inserting multiply copies of genes into the chromosome of *Pichia pastoris* and using the strong inducible alcohol oxidase promoter to drive transcription (15). We are presently using this system to develop a safer and more efficacious vaccine against botulinum toxins (16).

Viral (17, 18) and mammalian (19) expression systems have also supplemented the genetic engineer's "biological tool chest." Because of their potential use in vaccine development, certain eukaryotic viruses, such as vaccinia virus (17) and baculovirus (18), have recently become popular as foreign gene expression systems and expression vectors. Vaccinia virus has been developed as a live, infectious expression vector for foreign genes inserted into the viral thymidine kinase gene(20). Insertion of a foreign gene into vaccinia virus can create a eukaryotic recombinant vector capable of expression in an *in vivo* animal model. In contrast to *in vivo* expression, the helper-independent baculovirus expression vector has been used to express a wide variety of heterologous genes in cell culture. A baculovirus vector was used to express the first recombinant HIV envelope proteins to receive Food and Drug Administration approval for clinical evaluation as a vaccine candidate for the acquired immunodeficiency syndrome (21). *Autographa californica* nuclear polyhedrosis virus is an insect virus (baculovirus) which can express foreign genes to remarkably high levels when a foreign gene is inserted into the modified polyhedrin gene promoter of the virus (18). One of the major advantages of this invertebrate viral expression vector over bacterial, yeast, and mammalian expression systems is the abundant expression of recombinant proteins, which are, in many cases, structurally and functionally similar to the authentic gene products.

A wide assortment of established cell lines and DNA vectors have been used to express cloned genes in mammalian cells. Some vector-cell systems, such as the Simian virus 40 (SV40) vectors in monkey kidney (COS) cells (22) have been developed for high levels of transient expression. Transient expression systems are valuable for determining rapidly if cloned genes or cDNAs can be expressed and whether or not the recombinant product has been processed properly. Production of larger quantities of protein from cloned genes or cDNAs, however, requires prolonged expression of the genes in stable cell lines. Thus, vector-cell systems have been developed for stable, long-term expression in mammalian cells. High- level expression in stable cell lines is based on the concept, that if foreign-gene copy number is elevated in each cell, this will lead to higher levels of protein production in those cells. Two methods of achieving high-gene copy number, and, therefore, high production levels, have been developed. In one method, the copy number of the recombinant gene is elevated by co-amplification of the number of desired protein-encoding genes with that of a selectable, amplifiable gene. Such a system based on co-amplification uses vectors containing the gene encoding dihydrofolate reductase as the selectable marker and Chinese hamster ovary (CHO) cells as the host cell line. The second method relies on viral DNA replication to achieve high-copy number of the cloned gene. Both methods serve to increase the copy number of the gene to be expressed. By coupling the protein-encoding sequences to strong promoters and enhancer regions, a high level of transcription can subsequently be achieved.

MOLECULAR BIOLOGY OF ANIMAL TOXINS

The first cloned and sequenced cDNA encoding a snake venom protein with sequence homology to known PLA2 enzymes was from the black-banded sea snake, *Laticauda laticaudata*, captured in New Caledonia (23). This full-copy gene specified a pre-protein with a 27-amino acid signal peptide and an 118-amino acid structural protein. Since that work, other cDNAs encoding snake venom PLA2 homologs have been cloned from the olive sea snake, *Aipysurus laevis*, collected at the Great Barrier Reef in Australia (24); from the tiger snake, *Notechis scutatus scutatus*, captured near Sidney, Australia (25); from the South American rattlesnake, *Crotalus durissus terrificus* (26); from the European long-nosed viper, *Vipera ammodytes*

ammodytes, collected around Slunj, Croatia (27); and from the Mojave rattlesnake, *Crotalus scutulatus scutulatus*, collected near Portal, Arizona (28).

Many of the phospholipase A2 enzymes isolated from snake venoms produce varying degrees of neurotoxicity in animals, while some are non-toxic. cDNAs cloned from the Mojave rattlesnake encode a snake venom phospholipase A2 enzyme (Mojave toxin) known to be a relatively potent presynaptic neurotoxin (29). Mojave toxin is a heterodimer containing two non-covalently complexed subunits. One subunit is a basic, single-chained protein with PLA2 activity and low toxicity, while the other is acidic, non-toxic, lacks PLA2 activity, and is actually composed of three peptide fragments held together by seven disulfide bonds. The amino acid sequence of the basic and acidic subunits of Mojave toxin have been determined (30, 31).

A Mojave rattlesnake cDNA library was constructed (28) and cDNAs encoding the basic and acidic subunits isolated and sequenced. The cDNA for the basic subunit encodes a precursor protein with a 16-amino acid signal peptide and a 122-amino acid structural protein identical to the amino acid sequence determined for the Mojave toxin basic subunit (31). The inferred amino acid sequence determined from the nucleotide sequence of clones harboring cDNAs encoding the acidic subunit of Mojave toxin was identical to that obtained from direct amino acid sequencing (30). Based on the sequence for the cDNAs encoding the basic and acidic subunits of Mojave toxin precursors (28), it is presumed that the signal peptide is 16 amino acids long, as this sequence is identical for both subunits and divergence begins subsequent to this sequence. Within the sequence following the putative signal peptide of the acidic subunit, three additional regions exist where post-translational processing occurs based on the known amino acid sequence for the acidic subunit of Mojave toxin (30). The acidic subunit of Mojave toxin enhances the pharmacological efficacy and, in particular, the lethal potency of the PLA2-containing basic subunit presumably by acting as the receptor-binding domain of the toxin.

The first snake venom postsynaptic neurotoxin cloned and sequenced was erabutoxin a, from the erabu sea snake *Laticauda semifasciata* (32). Since that original work, many cDNAs encoding postsynaptically-acting snake neurotoxins have been characterized from the Chinese cobra, the Thai cobra, the banded krait, and the black mamba (33). Four species of African mamba snakes have venom neurotoxins that differ from other snake neurotoxins (34). Postjunctional alpha-neurotoxins, which bind to nicotinic cholinergic receptors, appear to be the only

class of toxins mambas share with other snakes that have neurotoxic venoms. Typical mamba neurotoxins are the presynaptically-acting facilatatory toxins such as dendrotoxin, and anticholinesterase toxins such as fasciculin and toxin C. A black mamba cDNA library was constructed by using mRNA isolated from the glands of Dendroaspis polylepis polylepis (35). Fourteen clones harboring cDNAs encoding different snake toxin gene products were isolated and characterized. Two of the cDNAs were expressed in *E. coli* (39).

Other snake toxin genes that have been cloned and characterized include crotamine isoforms from the South American rattlesnake, *Crotalus durissus terrificus* (36). Crotamine appears to affect the functioning of voltage-sensitive, sodium channels of skeletal muscle sarcolemma, inducing a sodium influx resulting in a depolarization and contraction of skeletal muscle. In skeletal muscle, lesions from the effects of crotamine consist of necrosis of the muscle fibers characterized by extensive vacuolization of the sarcoplasmic reticulum and disruption of actin and myosin filaments. A bacteriophage cDNA library was constructed from mRNA isolated from the glands of the South American rattlesnake(36) in an attempt to clone and characterize these toxin genes. Four cDNAs isolated and characterized encoded precursor proteins with a 22-amino acid signal peptide, a 42-amino acid structural protein, and a terminal lysine which was proteolytically removed. Nucleotide sequence analysis of the cloned cDNAs predicted the existence of multiple variants of the crotamine toxin. The different forms, identified from the DNA sequences, displayed discrepancies in amino acid sequence for crotamine when compared with previously published reports. Direct amino acid sequencing of commercially purified crotamine and CNBr fragments thereof confirmed the structures predicted by the nucleic acid sequences.

Toxin genes from animal sources other than snakes have also been cloned and their genes sequenced. A gene library has been prepared from mRNA isolated from the North African fat-tailed scorpion (*Androctonus australis Hector*) (37). With mRNA isolated from the telsons of *Androctonus australis Hector*, a cDNA library was constructed and full-length cDNAs of about 370 nucleotides encoding precursors of toxins active on mammals or on insects were isolated and sequenced. Sequence analysis of the cDNAs reveal the precursors contained signal peptides of 19 amino acid residues for the mammal toxins and 18 residues for the insect toxins. In addition, precursors of toxins active on mammals have extensions on their C-terminal ends, Arg or Gly-Arg (37), while those active in insects did not possess these

extensions. The extensions are removed during post-translational processing and the removal of the dipeptide, Gly-Arg, results in the terminal amino acid becoming a-amidated. Clones isolated and characterized from this library included those encoding for mammal toxins AaH I, AaH I', AaH II, and AaH III, and insect toxins AaH IT1, and AaH IT2. The insect toxin cDNAs encoded mature toxins of 70 amino acid residues while those for the mammal toxins encoded 63 amino acid residues for AaH I and AaH I' and 64 amino acid residues for AaH II and AaH III. Only the mammal toxin AaH II was a-amidated on the C-terminus.

The final section will review work on the expression of animal toxin cDNAs in heterologous host systems. A cDNA encoding the postsynaptic neurotoxin, erabutoxin a, was fused to commercially-available expression vector, pRIT5 (38). This vector was designed to permit high-level expression of fusion proteins in both *E. coli* and *Staphylococcus aureus* cells. Foreign genes, inserted into a multiple cloning site, were fused to a vector sequence encoding a 31-Kd protein A moiety. The genes were then be expressed from the protein A promoter and translocated to the periplasmic space by *E. coli*, or secreted into the growth media by *S. aureus*. The fusion product itself was produced in a correctly folded conformation and was directly secreted as such into the periplasmic space of *E. coli*. Competitive binding experiments showed the hybrid protein blocked erabutoxin a binding to the acetylcholine receptor and to a monoclonal antibody that recognized short-chain neurotoxins. In addition, the fusion protein was not only more immunogenic and less toxic than the native erabutoxin a, it was capable also of inducing neutralizing antibodies as potent as those raised against the native toxin (38). This approach may be extremely useful for future development of serotherapy against envenomation by poisonous animals.

Structural genes encoding toxins K and E from the black mamba were amplified by polymerase chain reaction (PCR), inserted into the commercially available pMALp vector, and also expressed as a fusion proteins in *E. coli* (39). Toxin K binds to and alters potassium channels while toxin E is a strong trypsin inhibitor. Recombinant products were affinity purified, released from their maltose (MAL) binding protein escort by protease cleavage with Factor Xa, and the toxins purified by FPLC chromatography. The recombinant toxins were characterized by end-group amino acid sequence analysis and in biological activity assays. Toxin K was effective in binding to rat brain potassium channels while toxin E did show the ability to inhibit trypsin. Yields of expression were about 1 mg purified product per liter of induced *E. coli* cells (39).

Expression of scorpion toxin cDNAs was attempted in monkey kidney COS cells (37) and baculovirus (39). One of the cDNAs, designated pcD403, encoding the mammal toxin II from *Androctonus australis Hector* was used to transfect COS-7 (SV40-transformed African green kidney monkey) cells. The recombinant AaH II, expressed and secreted by the COS-7 cells, was characterized by an immunoassay, a receptor-binding assay, and a bioassay. In all three methods, the recombinant product behaved identically to native AaH II. The results obtained support the conclusion that the recombinant toxin monkey kidney cells transiently expressed and secreted upon their transfection with the recombinant plasmid, pcD403, was the mature form of AaH II. The yield of secreted product from expression in this system was 0.2 µg of affinity-purified recombinant per 10^6 cells (37).

CONCLUSION

The application of molecular biology to the study of animal toxins and their genes is entering a phase of rapid development. This technology will undoubtedly play an important role in many areas of animal toxin research in the future. Some of these areas will include studies on toxin structure and the relationship to activity and function; generation of recombinant toxins and their genetically-mutated derivatives for use in antibody production and vaccine development; and in basic research studies elucidating the organization of genes and the regulation of their expression in venomous animals.

The ability to express toxin genes in heterologous host systems holds the key to realizing the full potential of recombinant DNA technology as applied to animal toxin research and vaccine development. The structural complexity of these proteins and the fact that these are toxic to certain cells, can make their expression difficult. As we have seen, fusion proteins can be effective in reducing the lethal potency of some toxins, while still retaining or even increasing their immunogenic ability for producing neutralizing antibodies against the active toxin. In these cases, fused toxins will be extremely valuable for future development of serotherapy against envenomation. However, there may be many situations where fused toxins will not elicit neutralizing antibodies. In those cases, further work must be performed, either at the level of the expression system being used, or in the manipulation of the fused toxin in the removal of its escort or fused carrier.

One of the celebrated features of recombinant DNA technology is the wide array of techniques, expression vectors and hosts to choose from. Recently, it was said, "Expression systems are protein specific. You must be able to play around with each one, insert your gene of choice, tweak it, and then see what you've got" (40). This area of research is exciting, challenging, and holds much promise for future progress towards developing protective modalities against toxins.

REFERENCES

1. C. Gans, *Biology of the Reptilia*, Vol. **8** (Academic Press, London, 1978).

2. J. Sambrook, E.F. Fritsch, and T. Maniatis, Eds., in *Molecular Cloning: A Laboratory Manual*, Vols. **1**, **2**, and **3** (Cold Spring Harbor Laboratories, Cold Spring Harbor, New York, 1989).

3. F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, Eds., in *Current Protocols in Molecular Biology*, Vols. **1** and **2** (John Wiley and Sons, New York, 1989).

4. B. Perbal, in *A Practical Guide to Molecular Cloning* (John Wiley and Sons, New York, 1988).

5. D.M. Glover, Ed., in *DNA Cloning*, Vol.**1** (IRL Press, Oxford, 1986).

6. J.L. Glenn, and R.C. Straight, *Toxicon*, **16**: 81-84, (1978).

7. M.L. Vandenplas, S. Vandenplas, K. Brebner, A.J. Bester, and D. Boyd, *Toxicon*, **23**: 289-305, (1985).

8. J.R. Feramisco, D.M. Helfman, J.E. Smart, Burridge, and G.P. Thomas, *J. Biol. Chem.*, **257**: 11024-11031, (1982).

9. H. Aviv, and P. Leder, *Proc. Natl. Acad. Sci. U.S.A.*, **69**: 1408-1412, (1972).

10. A.M. Maxam, and W. Gilbert, *Methods in Enzymology*, Vol. **65**: 499-560, L. Grossman and K. Moldave, Eds. (Academic Press, New York, 1980).

11. F. Sanger, S. Nicklen, and A.R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.*, **74**: 5463-5467, (1977).

12. J.M. Short, J.M. Fernandez, J.A. Sorge, and W.D. Huse, *Nucleic Acids Res.*, **16**: 7583-7600, (1988).

13. S. Henikoff, *Gene*, **28**: 351-359, (1984).

14. T.J.R. Harris, in *Genetic Engineering*, **4**: 127-186, R. Williamson, Ed. (Academic Press, New York, 1983).

15. J.J. Clare, F.B. Rayment, S.P. Ballantine, K. Sreekrishna, and M.A. Romanos, *Bio/Technology*, **9**: 455-460, (1991).

16. L.A. Smith, unpublished results.

17. M. Mackett, G.L. Smith, and B. Moss, in *DNA Cloning: A Practical Approach*, D.M. Glover, Ed., **II**: 191-211 (IRL Press, Oxford, 1985).

18. D.W. Miller, P. Safer, and L.K. Miller, in *Genetic Engineering: Principles and Methods*, **8**: 277-298, J.K. Setlow, J.K. and A. Hollaender, Eds. (Plenum Press, New York, 1986).

19. C.R. Bebbington, and C.C. Hentschel, in *DNA Cloning: A Practical Approach*, D.M. Glover, Ed., **III**: 163-188 (IRL Press, Oxford, 1985).

20. B. Moss, and C. Flexner, *Annu. Rev. Immunol.*, **5**: 305-324, (1987).

21. M.A. Cochran, B.L. Ericson, J.D. Knell, and G.E. Smith, in *Vaccines*, **87**: 384-388 (Cold Spring Harbor Laboratory Laboratories, Cold Spring Harbor, New York, 1987).

22. H. Okayama, and P. Berg, *Mol. Cell. Biol.*, **3**: 280-289, (1983).

23. G. Guignery-Frelat, F. Duncancel, A. Menez, and J-C. Boulain, *Nucleic Acids Res.*, **15**: 5892, (1987).

24. F. Duncancel, G. Guignery-Frelat, C. Bouchier, A. Menez, and J-C. Boulain, *Nucleic Acids Res.*, **16**: 9048, (1988).

25. F. Duncancel, G. Guignery-Frelat, C. Bouchier, A. Menez, and J-C. Boulain, *Nucleic Acids Res.*, **16**: 9049, (1988).

26. C. Bouchier, F. Ducancel, G. Guignery-Frelat, C. Bon, J-C. Boulain, and A. Menez, *Nucleic Acids Res.*, **16**: 9050, (1988).

27. J. Pungercar, D. Kordis, R. Jerala, M. Trstenjak-Prebanda, M. Dolinar, V. Curin-Serbec, R. Komel, and F. Gubensek, *Nucleic Acids Res.*, **17**: 4367, (1989).

28. L.A. Smith, unpublished results.

29. R.L. Cate, and A.L. Bieber, *Arch. Biochem. Biophys.*, **189**: 397-408, (1978).

30. A.L. Bieber, R.R. Becker, R. McParland, D.F. Hunt, D.F., J. Shabanowitz, J.R. Yates III, P.A. Martino, and G.R. Johnson, *Biochim. Biophys. Acta*, **1037**: 413-421, (1990).

31. S. Aird, W. Kruggel, and I.I. Kaiser, *Crotalus scutulatus scutulatus*, (in press).

32. T. Tamiya, A. Lamouroux, J-F. Julien, B. Grima, B., J. Mallet, P. Fromageot, A. and Menez, *Biochimie*, **67**: 185-189, (1985).

33. L.A. Smith, unpublished results.

34. A.L. Harvey, A.J. Anderson, P.M. Mbugua, and E. Karlsson, *J. Toxicol.-Toxin Rev.*, **3**: 91-137, (1984).

35. L.A. Smith, *Nucleotide Sequence of a cDNA Encoding Toxin C from Dendroaspis polylepis* (Presented at the 21st Annual Meeting of the American Society for Neurochemistry, Phoenix, Arizona, 1990).

36. L.A. Smith, and J.J. Schmidt, *Toxicon*, **28**: 575-585, (1990).

37. P.E. Bougis, H. Rochat, and L.A. Smith, *J. Biol. Chem.*, **264**: 19259-19265, (1989).

38. F. Ducancel, J-C. Boulain, O. Tremeau, and A. Menez, in *Protein Engineering*, **3**: 139-143, (1989).

39. L.A. Smith, unpublished results.

40. M. Ratner, *Biotechnology*, **7**: 1129-1133, (1989).

Regression/Correlation Analysis of Ionospheric Parameters at Midlatitudes

* Haim Soicher, Dr., U.S. Army Communications-Electronics Command
Command, Control and Communications Systems Directorate
ATTN: AMSEL-RD-C3-TR-H
Fort Monmouth, NJ 07703-5203

Francis J. Gorman, U.S. Army Communications-Electronics Command
Command, Control and Communications Systems Directorate
ATTN: AMSEL-RD-C3-TR-H
Fort Monmouth, NJ 07703-5203

INTRODUCTION: HF radio communication depends on the ability of the ionosphere to return the radio signal incident on it back to earth. Prediction of ionization levels in the various ionospheric regions are derived from models and are used as a basis for planning and frequency management of HF radio systems worldwide. The models permit the calculation of system parameters such as operating frequencies, signal strengths, signal-to-noise ratios and multipath probability that can be used to describe the performance of HF radio systems. Uncertainties or inaccuracies in the models of the ionosphere have long been known to be one of the major, if not the major, cause for inaccuracies in the calculated propagation characteristics. This is particularly true for those applications of ionospheric predictions involving time scales that are less than the monthly average or monthly median.

To reduce average monthly RMS errors in predictions, adaptive techniques that use realtime observations to correct model biases have been devised. For mobile as well as fixed point communication modes applications, the determination of the correlation of propagation characteristics in the short, medium and long range oblique modes with those of a local vertical mode, is of particular interest. Thus determined correlation distances will permit the assessment of the reliability of HF tactical communication circuits within an extended area, based on monitoring of the vertical ionosphere at a central location.

To this end a study aimed at the determination of correlation distances of HF skywave propagation characteristics by recording both verticaland oblique ionograms was undertaken. A regionalized database is required to develop an accurate HF prediction algorithm, and a cooperative network of ionospheric sounder facilities was established in order to accumulate the propagation data. Preliminary results of the program based on vertical soundings taken by two stations in the network, the Royal Institute of Meteorology, Dourbes, Belgium, (50.1°N, 4.6°E) and the Ebro Observatory, Roquetas, Tarragona, Spain, (40.8°N, 0.3°E), separated by 1087 km of ground distance are reported here. A third station in Rome, Italy, (41.9°N, 12.5°E), is also currently on board.

THE DATA: Vertical ionograms using the Digisonde 256[1] were taken at hourly intervals from February 29 to July 3, 1988, at the Spanish location and from February 4 to June 17, 1988, at the Belgium location for a total of ~6200 ionograms. Digital ionograms are scaled automatically using the Automatic Real Time Ionogram Scaler with True Height (ARTIST)[2,3]. Amplitude, phase, incidence angle, polarization and Doppler shift measured by the Digisonde are analyzed to extract the overhead ordinary and extraordinary traces even during disturbed ionospheric conditions. The electron density profile may be calculated from the ordinary trace using profile-fitting methods. Whereas the ARTIST scales 18 parameters: f_oF_2, f_oF_1, h'F, h'F_2, MUF(3000)F_2, f_{min}, $f_{min}F$, f_xI, f_oE, h'E, f_oE, h'E, $f_{min}E$, range spread F, range spread E, frequency spread E; only a selected sample of these parameters will be reported here. Examples of data will be given for March 1988 and May 1988 when the International Relative Sunspot Number varied from 20 to 120, and 20 to 103 respectively, on a daily basis.

The diurnal variation of f_oF_2 values (ordinary wave critical frequency) at the Belgium and Spain sites were measured at hourly intervals for the month of March 1988 is shown in Figures 1a and 1b, respectively; and for May 1988 in Figures 2a and 2b respectively. The trend of the diurnal variation is different for the two months. For March, the ionospheric buildup after dawn is fairly sharp until f_oF_2 reached a maximum at 1200 UT for both sites. For May, the buildup is more flat and f_oF_2 reached a maximum in the afternoon. The values in Spain are larger during the day at both sites, reflecting solar zenith angle considerations. The standard deviation from monthly means varied at both sited from .4 to 1.0 MHz for March, and were slightly higher for May. Similar conclusions may be reached for the values of MUF (3000) F_2 (maximum usable frequencies) which is shown at Figures 3a and 3b for March 1988, and in Figures 4a and 4b for May 1988. The standard deviations from monthly means ranged from about 1 MHz to 4 MHz. Thus, in the mean the range of frequencies available for communications to the more Southern station (Spain) is larger than that of the Northern station for both months.

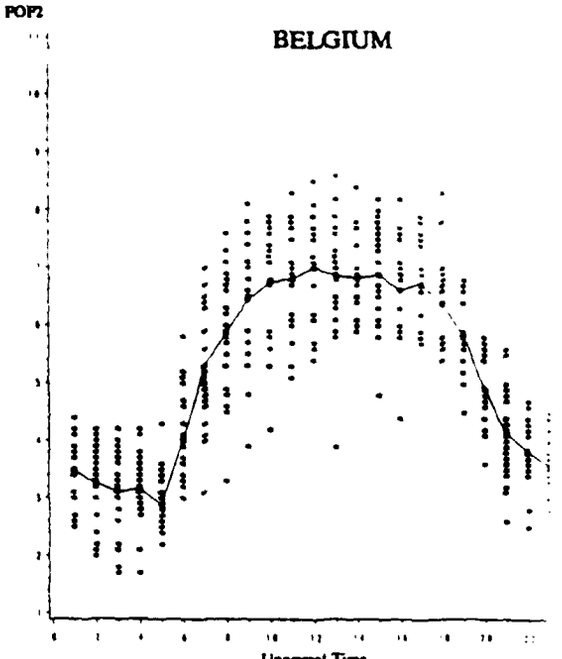


Figure 1.a

March

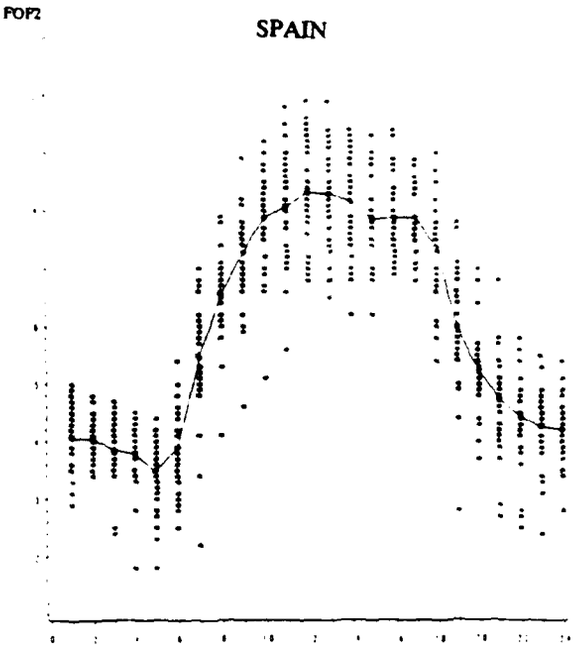


Figure 1.b

March

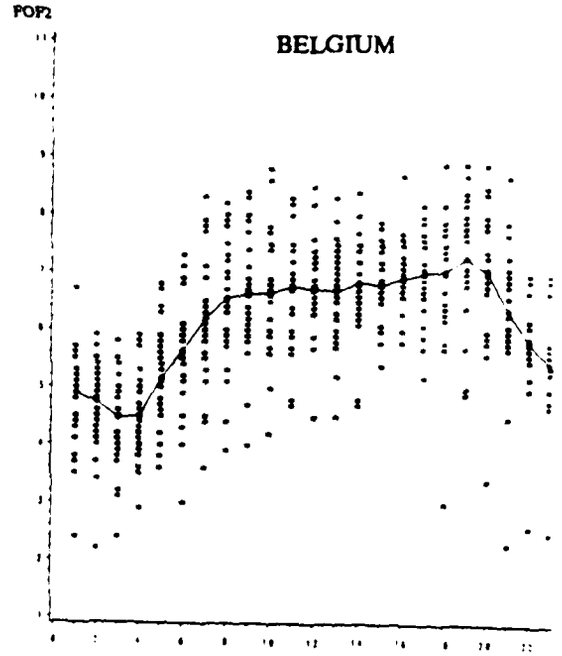


Figure 2.a

May

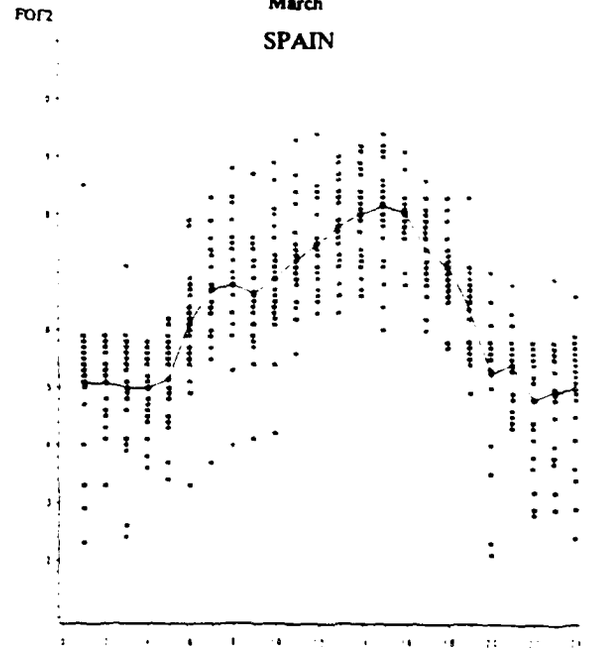


Figure 2.b

May

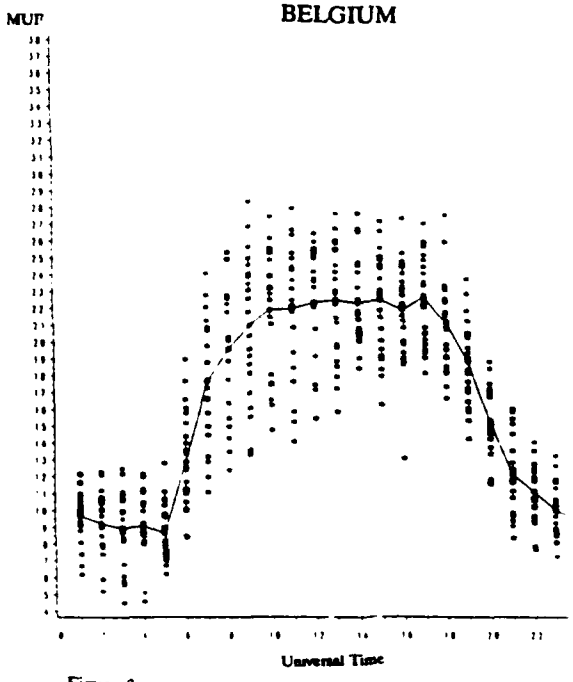


Figure 3.a

March

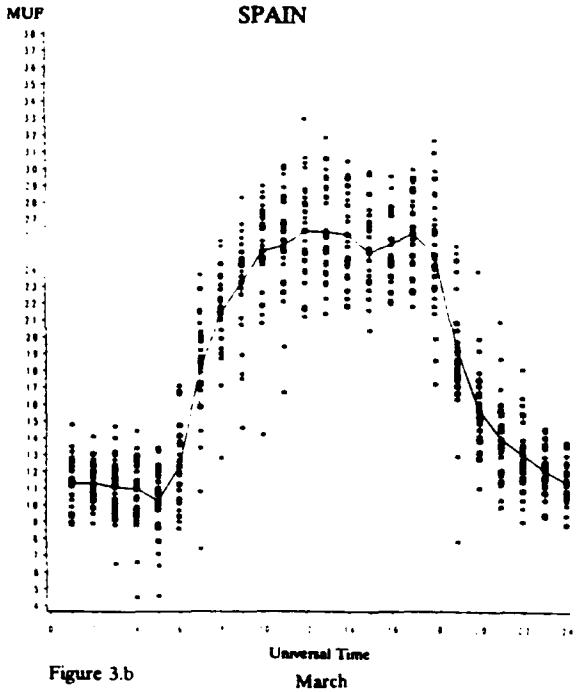


Figure 3.b

March

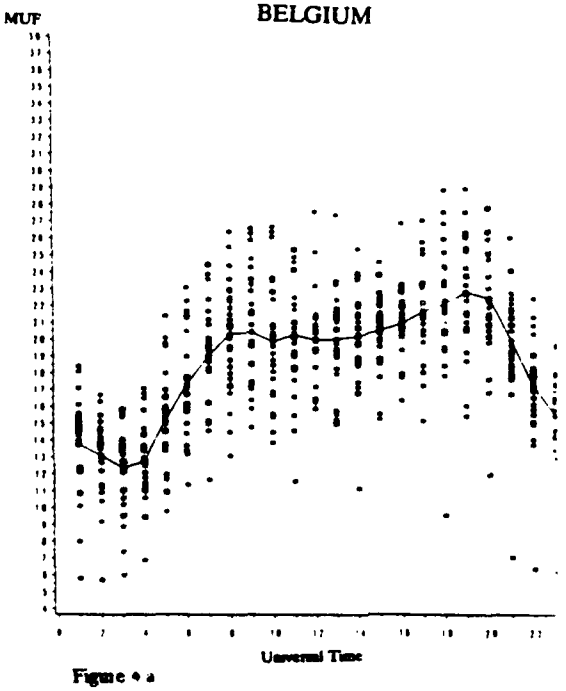


Figure 4.a

May

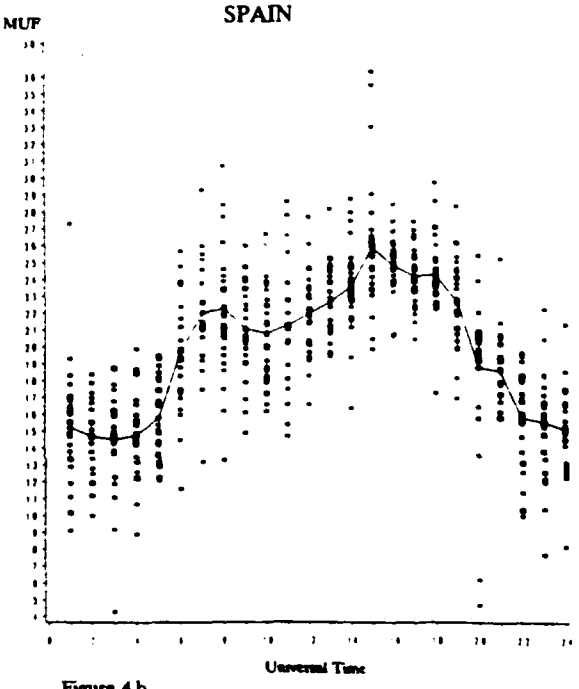


Figure 4.b

May

The hourly values of h'F (minimum F layer virtual height) in Belgium and Spain for the month of March 1988 are shown in Figures 5a and 5b, respectively, and for May 1988 in Figures 6a and 6b, respectively. While lowest mean height was about the same for the two locales, the maximum mean height was higher at Belgium by about 10 km for both months. The standard deviation from monthly means varied at both sites ~10 km to ~50 km for both months.

The hourly f_{min} values (the lowest frequency at which echo traces are observed) in Belgium and Spain for the month of March 1988 are shown in Figures 7a and 7b, respectively; and for May 1988 in Figures 8a and 8b, respectively. The mean values for both stations are not highly variable and show no particular diurnal trend. The standard deviation varied between ~.1 and ~.2 MHz.

CORRELATION ANALYSIS: Correlation analysis for the selected ionospheric parameters was carried out. A least-squares regression line was fitted to the two data sets with the time parameter eliminated between them. The regression lines for f_oF_2, $MUF(3000)F_2$, h'F and f_{min} for 11 March 1988 (day 70), are shown in Figures 9a, b, c and d, respectively. The regression analysis for the specific day shows excellent correlation for f_oF_2, MUF, h'F, and a lesser one for the parameter f_{min}. It appears that at any one day, once a reference point has been established, the ionospheric parameters may be ascertained at the one locale from actual measurements at the other.

The coefficient of correlation, r, is defined as:

$$r = \pm\left(\frac{\sum (Yest-\overline{Y})^2}{\sum (Y-\overline{Y})^2}\right)^{\frac{1}{2}}$$

where Y_{est} represents the value of the parameter at Belgium for a given value of the parameter at Spain estimated from the fitted regression line; $\overline{Y}$ in the mean of Y (the actual Belgium value).

The correlation coefficients for the various parameters w re calculated in daily intervals, i.e., the data throughout a specific day was correlated at the same time for corresponding parameters. Since local time differences (~18 minutes) are within data resolution (hourly) the data sets were not shifted in time for adjustment. Correlation coefficients for f_oF_2, h'F and f_{min} for March 1988 are shown in Figures 10a, b and c, respectively; and for May 1988 in Figures 11a, b and c, respectively.

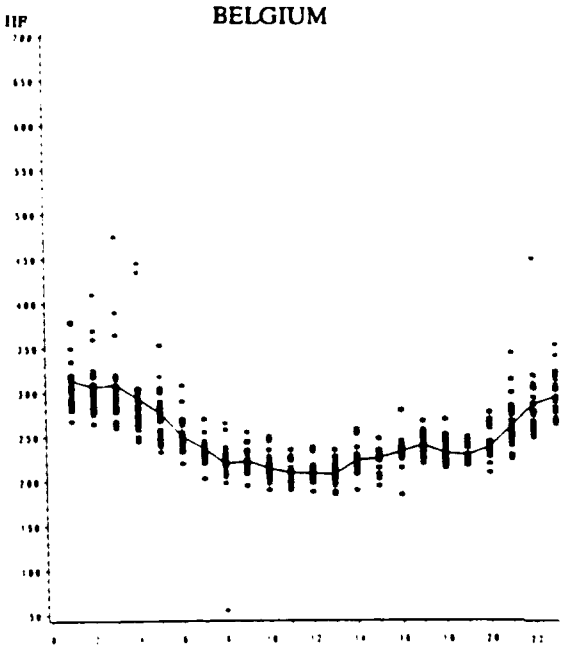


Figure 5.a

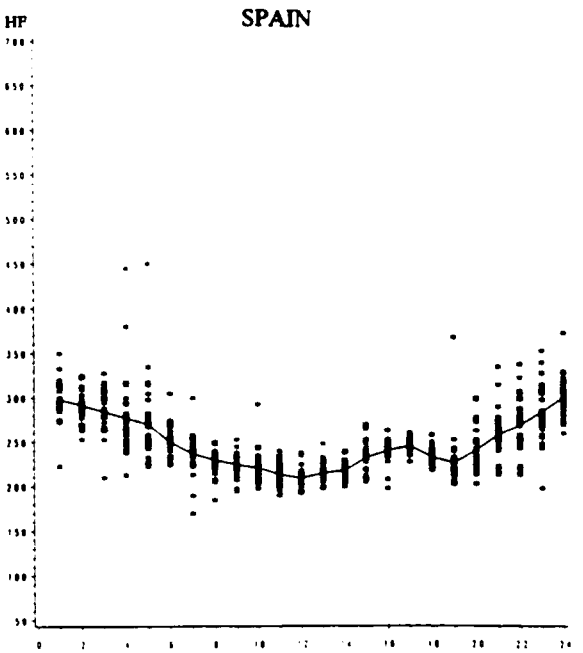


Figure 5.b

March

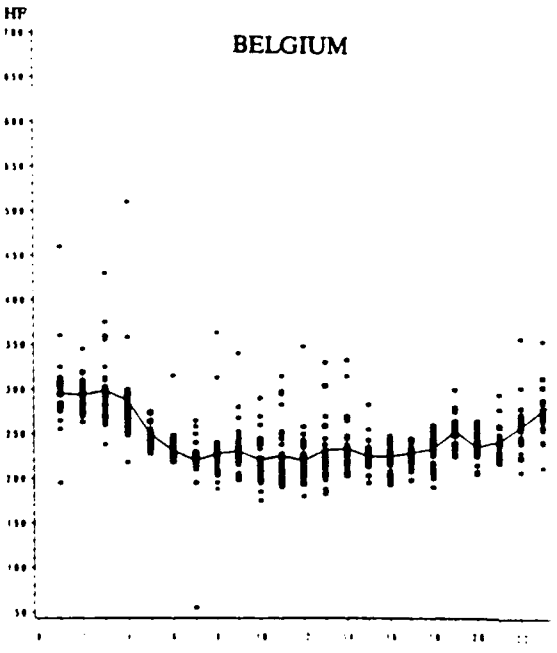


Figure 6.a

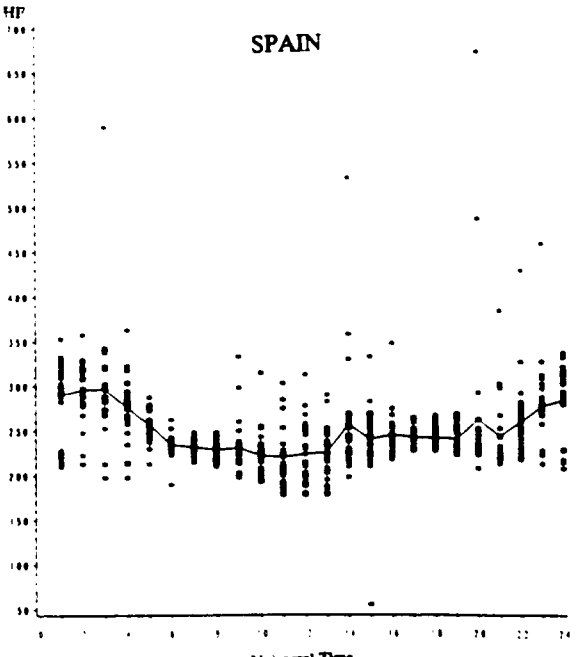


Figure 6.b

May

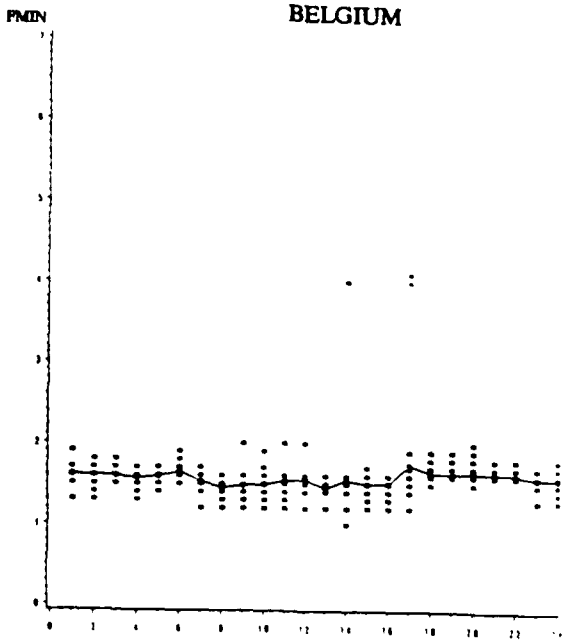


Figure 7.a

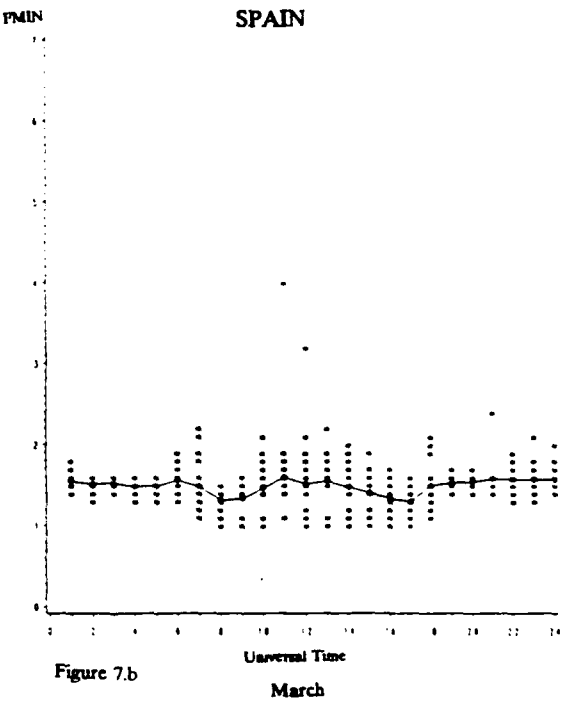


Figure 7.b

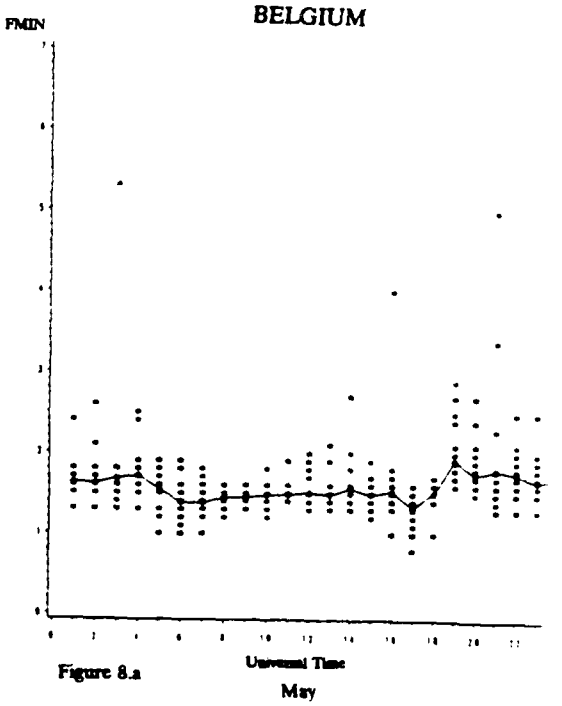


Figure 8.a

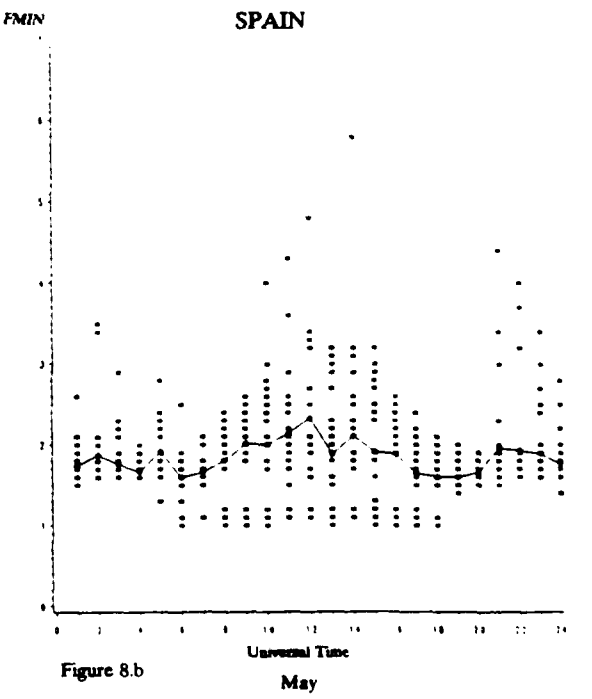


Figure 8.b

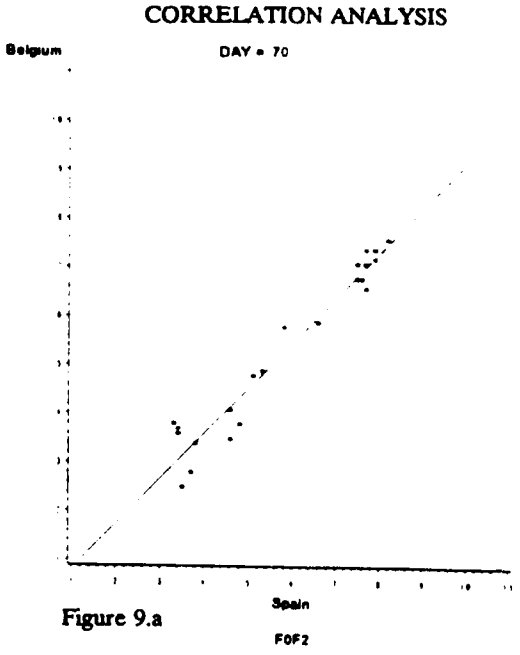


Figure 9.a

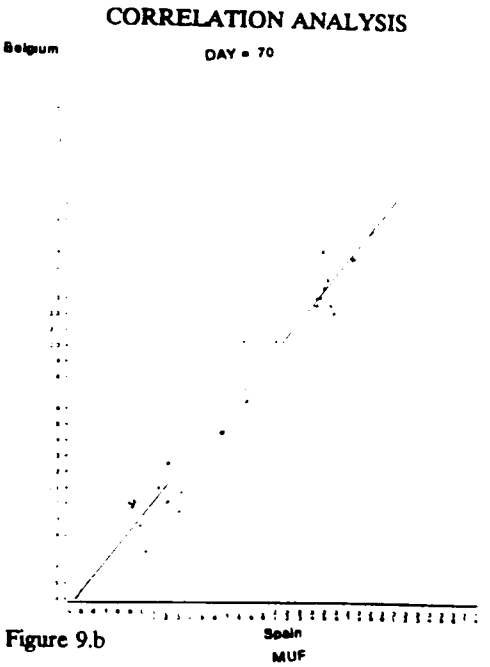


Figure 9.b

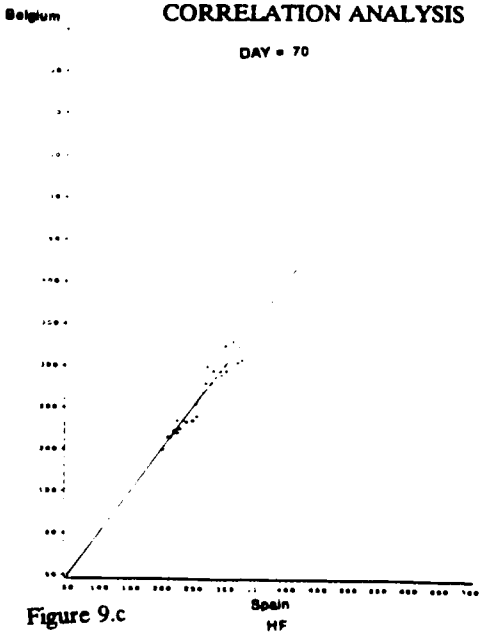


Figure 9.c

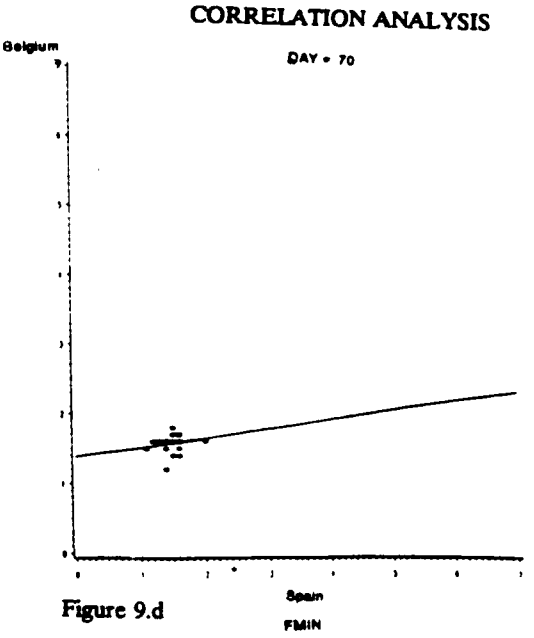


Figure 9.d

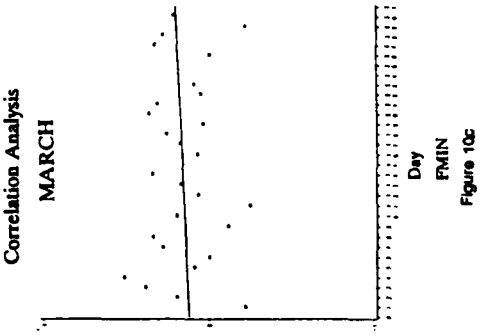


Figure 10c

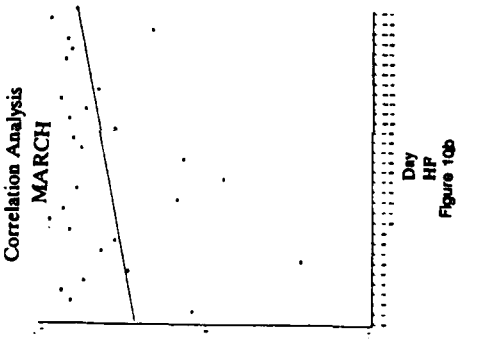


Figure 10b

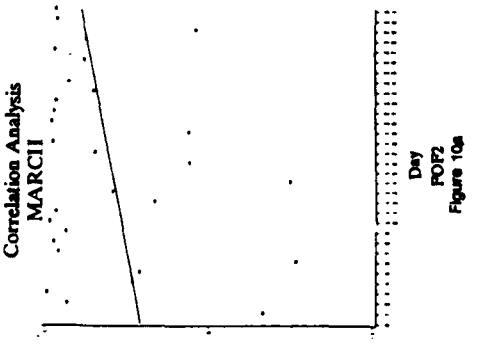


Figure 10a

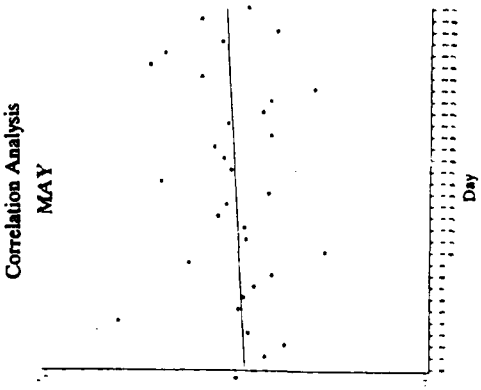


Figure 11c

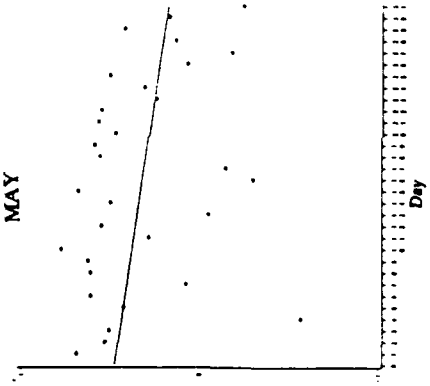


Figure 11b

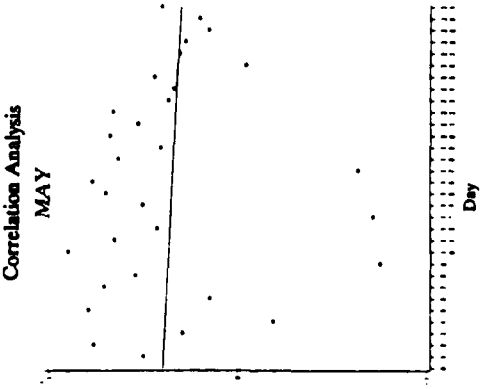


Figure 11a

Correlation coefficients for the entire month of March 1988 indicated values > .6 for most days, with values > .8 for about 50% of all daysfor f_oF_2; while values > .5 for most days for h'F were indicated, with little correlation indicated for f_{min}. For the entire month of May the correlation coefficients are similar but less in value than in March, thus showing a seasonal variation.

It appears that on most days f_oF_2 (and MUF) may be ascertained at any time at one locale from a known current variation of these parameters at the other locale. No such claim may be made for f_{min}. However, f_{min} is not highly variable quantity and its usefulness for communications applications is limited.

ACKNOWLEDGEMENTS: The data at Belgium was taken under the guidance of Dr. J.C. Jodogne, while at Spain it was taken under the guidance of Dr. Luis Alberca of the Ebro Observatory.

REFERENCES:

[1]REINISCH, B.W. [1986] *New Techniques in Ground-Based Ionospheric Sounding and Studies*, RADIO SCIENCE, **Vol. 21**, 331-341.

[2]REINISCH, B.W and HUANG, X. [1982] *Automatic Calculation of Electron Density Profiles from Digital Ionograms, 3. Processing of Bottomside Ionograms*, RADIO SCIENCE, **Vol. 17**, 837-844.

[3]REINISCH, B.W. and HUANG, X. [1982] *Automatic Calculation of Electron Density Profiles from Digital Ionograms, 1. Automatic O and X Trace Identification for Topside Ionograms*, RADIO SCIENCE, **Vol. 17**, 421-434.

FIGURES:

1a,b: Superposed diurnal variation of f_oF_2 at hourly intervals at Belgium (a) and Spain (b), March 1988. Monthly mean values are also indicated.

2a,b: Superposed diurnal variation of f_oF_2 at hourly intervals at Belgium (a) and Spain (b), May 1988. Monthly mean values are also indicated.

3a,b: Superposed diurnal variation of $M(3000)F_2$ at hourly intervals at Belgium (a) and Spain (b), March 1988. Monthly means are also indicated.

4a,b: Superposed diurnal variations of $M(3000)F_2$ at hourly intervals at Belgium (a) and Spain (b), May 1988. Monthly means are also indicated.

5a,b: Superposed diurnal variation of h'F at hourly intervals at Belgium (a) and Spain (b), March 1988. Monthly mean values are also indicated.

6a,b: Superposed diurnal variation of h'F at hourly intervals at Belgium (a) and Spain (b), May 1988. Monthly mean values are also indicated.

7a,b: Superposed diurnal variation of f_{min} at hourly intervals at Belgium (a) and Spain (b), March 1988. Monthly mean values are also indicated.

8a,b: Superposed diurnal variation of f_{min} at hourly intervals at Belgium (a) and Spain (b), May 1988. Monthly mean values are also indicated.

9a,b,c,d: Regression lines for hourly values of f_oF_2 (a), MUF(3000)F_2 (b), h'F (c) and f_{min} (d) at Belgium (ordinate) and Spain (abscissa) for 11 March 1988 (day 70).

10a,b,c: Correlation coefficients versus day of month for hourly values of f_oF_2 (a), h'F (b), and f_{min} (c) at Belgium and Spain for March 1988.

11a,b,c: Correlation coefficients versus day of month for hourly values of f_oF_2 (a), h'F (b) and f_{min} (c) at Belgium and Spain for May 1988.

Chem-Biodetection: Potentiometric and Fiber Optic Biosensor Test Beds

Roy G. Thompson*, Darrel E. Menking, and James J. Valdes, Ph.D.
U.S. Army Chemical Research, Development and Engineering Center
Aberdeen Proving Ground, MD 21010-5423

Moyhee D. Eldefrawi, Ph.D.
Department of Pharmacology and Experimental Therapeutics
University of Maryland School of Medicine
Baltimore, MD 21201-1559

Kim R. Rogers, Ph.D.
U.S. Environmental Protection Agency
Environmental Monitoring Systems Laboratory
Las Vegas, NV 89193-3478

INTRODUCTION

Medical, environmental and biotechnological demands for obtaining dedicated selective chemical analysis for specific chemical/biological reactions in real-time has spurred the research and development of new analytical instruments commonly referred to as "biosensors." Biosensors are typically defined by the joining of a specific biological recognition molecule with a signal-generating element that is capable of transducing molecular interactions occurring at the biosurface interface into an electronic signal. Current research and development of biosensor design draws strongly on the creative union of basic principles from the fields of chemistry, physics, materials science, molecular biology and engineering. By their design, biosensors emphasize a "target-defined" strategy of detecting interactions between molecules based on their native biological significance and function as opposed to the traditional elemental analysis of chemical properties or atomic structure.

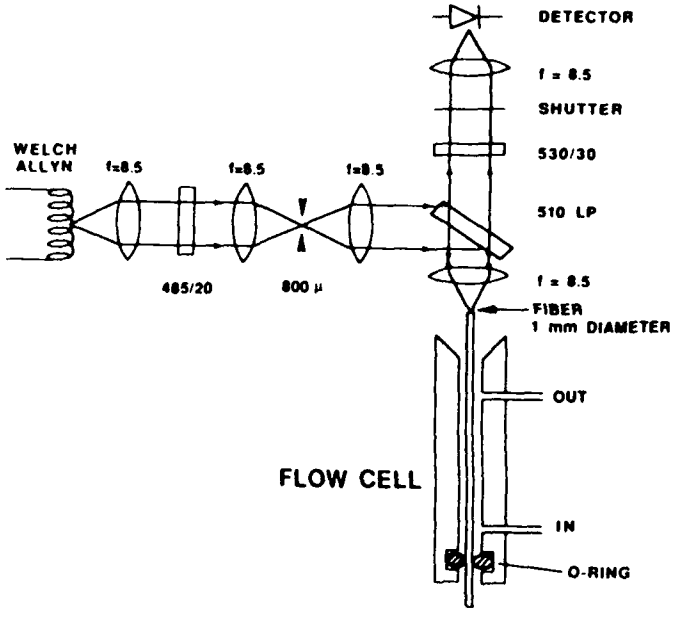
A biological-based "target-defined" strategy for chemical and biological detection offers considerable opportunities for expanding the military's current capabilities and adopting to future exigencies. By selecting certain specific biological

recognition molecules, such as antibodies, sensors can be designed to detect very specific biological agents and/or organisms. By selecting other biological recognition molecules, such as receptors and enzymes, the sensors response can be broadened to detect classes of compounds and/or entirely different classes of compounds that act at the same biological site but by different mechanisms. And by combining a limited number of both specific antibodies and receptors in a single sensor format, one gains the prospect of a semi-generic detector capable of responding to a broad spectrum of chemical and biological threat agents.

This paper will present original research results obtained from research and development of two different biosensor test bed configurations and their response to a variety of chemical and biological toxins when using receptors, enzymes or antibodies as the molecular recognition element. The two sensor test designs are the light-addressable potentiometric sensor (LAPS) and a fiber-optic waveguide-based sensor (FOWG).

I. Fiber Optic Waveguide - Methods and Results.

General. Experimental work with the FOWG was performed on a fluorimeter designed and built by Ord, Inc. (North Salem, NH). The basic components of this system include a quartz halogen lamp, bandpass filters and lenses, and a silicon detector (Fig. 1).



The quartz optical fibers, 60 mm in length and 1 mm diameter with polished ends were also obtained from Ord, Inc. Test solutions are perfused into the flow cell (46 ul vol) holding the quartz fiber by means of a peristaltic pump and was completely exchanged every 14 sec. The FOWG operates on the principle of the evanescent wave effect by exciting a fluorophore just outside the waveguide boundary. A portion of the resulting fluorophore emission then becomes trapped in the waveguide and is transmitted back up the fiber, through the appropriate filters and detected.

Acetylcholine Receptor-Based Sensor. Nicotinic acetylcholine receptor protein (nAChR) purified from Torpedo electric organ was immobilized on the quartz fibers by passive adsorption and ligand interaction with the receptor monitored by the inhibition of fluorescein isothiocyanate labelled α-najatoxin (FITC-α-NT) binding. Specific binding of the FITC- -NT probe to the immobilized receptor is dose-dependent, measured either as the initial rate of change in the fluorescent signal or the maximum signal obtained at apparent equilibrium (Fig. 2).

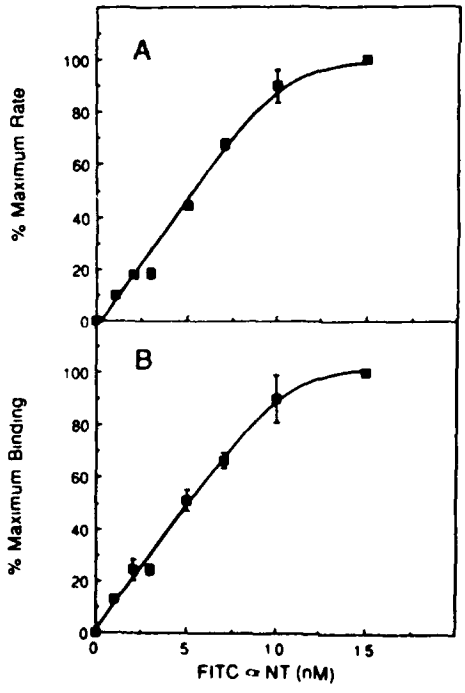


Figure 2. The effect of FITC-α-NT concentration on (A) the initial rate of fluorescence increase or (B) the maximum signal at 10 min. Fibers were perfused with buffered saline and 0.1 mg/ml bovine serum albumin before FITC-α-NT was added at the indicated concentrations. Symbols and bars are means of triplicate measures ± SEM using different fibers for each determination slopes of the linear portion of each curve varied by less than 3%.

The nAChR-coated optical sensor's sensitivity to nicotinic cholinergic drugs and relative insensitivity to non-cholinergic

compounds is summarized in Table 1. For those non-nicotinic compounds which do show inhibition of FITC-α-NT binding - most notably haloperidol - it is significant to note that the nAChR protein carries binding sites associated with its ion channel in addition to the ACh recognition site and this site is known to bind non-competitive inhibitors of receptor function: haloperidol is one such non-competitive blocker.[2]

Table 1. Effects of drugs on the specific binding of FITC-α-NT to nAChR-coated fibers.

Drug[a]	Percentage of inhibition of initial rate of fluorescence change[b]
Nicotinic cholinergic ligands	
Acetylcholine	30 ± 3
Nicotine	25 ± 8
d-TC	58 ± 7
Muscarinic cholinergic ligands	
(+)cis-Dioxolane	20 ± 10
Atropine	0
GABAergic ligands	
GABA	0
Picrotoxin	0
Glutamatergic ligands	
l-Glutamate	0
2-Amino-5-phosphonovalerate	0
Serotonergic ligands	
5-HT	23 ± 8
ICs 205-903	0
Adrenergic	
Norepinephrine	0
Propanalol	0
Dopaminergic	
Dopamine	0
Haloperidol	56 ± 8

[a] The nAChR-coated fibers were pretreated for 5 min with the indicated drugs (1 mM) in PBS + BSA (0.1 mg ml^{-1}) followed by addition of 5nM FITC-α-NT.

[b] Inhibition values are the means ± standard error of the mean of three fibers calculated relative to the rate in the absence of any drug. A value of zero represents non-significant difference from the mean.

The effect of non-competitive inhibitors on nACh receptor binding is also an important indicator of the conformational freedom maintained by the immobilized receptor and offers the possibility of using the receptor complex to detect a wide range of antipsychotic and psychotropic drugs known to affect his site[2].

Acetylcholinesterase Enzyme-Based Sensor. The enzyme acetylcholinesterase (AChE) is well known to be the primary target for the traditional nerve agents as well as a broad spectrum of chemicals used in the pesticide control industry. Coupling of AChE to the fiber optic sensor did not require the use of extrinsic fluorophore to produce a fluorescent signal and measure a compounds inhibition of that signal. Instead, FITC was coupled directly to the enzyme and changes in enzyme activity were measured by the pH-dependent quenching of the FITC label when the enzyme was exposed to its ACh substrate and protons were released by the enzymatic hydrolysis of ACh into choline and acetic acid. By switching the perfusion pump off at set intervals, proton accumulation quenched the intrinsic FITC-AChE fluorescence and turning the pump back on removed excess protons and baseline fluorescence was quickly reestablished (Fig. 3).

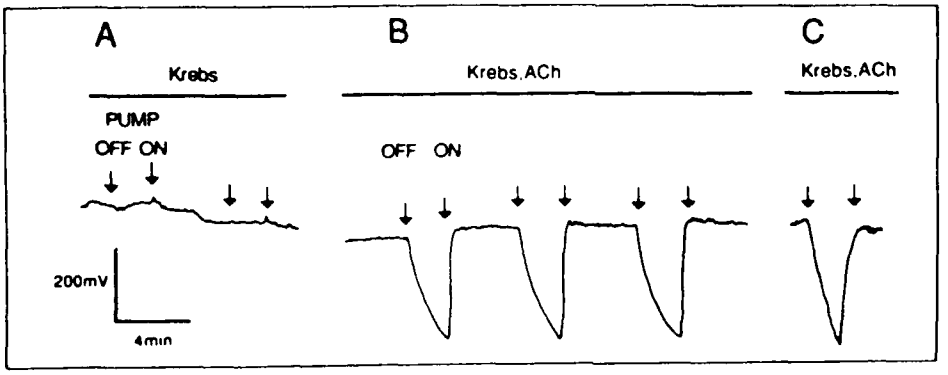


Figure 3. The change in fluorescence was a result of AChE activity (A) Steady-state fluorescence in the absence of ACh was unaffected by interruption in the buffer flow (B) In the presence of 1mM ACh fluorescence was quenched when the pump was turned off and protons accumulated. The baseline fluorescence was quickly reestablished when the pump was turned on an excess protons were removed by the perfusing substrate solution (C) Enzyme activity was measured by the amplitude of signal quenching after 2 min.

Substrate specificity of the AChE-based sensor is in good agreement with other methods for measuring substrate hydrolysis (Table 2).

Table 2. Effect of AChE substrates on the response of the AChE fiber-optic biosensor.

	Relative rate of hydrolysis	
	Fiber-optic sensor[a]	Manometric (Warburg)[b]
Acetylcholine	100 ± 0	100 (0.006 M)
Acetyl-β-methylcholine	63 ± 11	50 (0.03 M)
Benzyl acetate	22 ± 7	34 (0.03 M)
Butyrlcholine	21 ± 6	3 (0.03 M)
Benzoylcholine	<1	<1
n-Butyl acetate	<1	24
Carbamylcholine	<1	NA

[a] The AChE biosensor response expressed as relative to that of ACh hydrolysis rate. The concentration of substrates was 1 mM.

[b] Rate of hydrolysis of substrates by red blood cell cholinesterase expressed as percentages of the rate of hydrolysis of ACh. Concentration of substrates was 0.1 M unless otherwise indicated in parenthesis (Adams, 1949).

Concentration-dependent inhibition of AChE-based sensor for a variety of anticholinesterase compounds is shown in Fig. 4. Worth noting is the sensors sensitivity in detecting as little as 5 ppb of the organophosphate echothiopate and almost complete absence of sensitivity in detecting parathion and malathion. The AChE biosensor also detects the irreversible inhibition of AChE by organophosphates and reactivation by treatment with pralidoxime (2-PAM). Fig. 5 shows the results of this treatment using echothiopate as the irreversible phosphorylating agent. The sensor shows almost complete inhibition of AChE activity after treatment with echothiopate and near full recovery of activity after 2-PAM treatment.

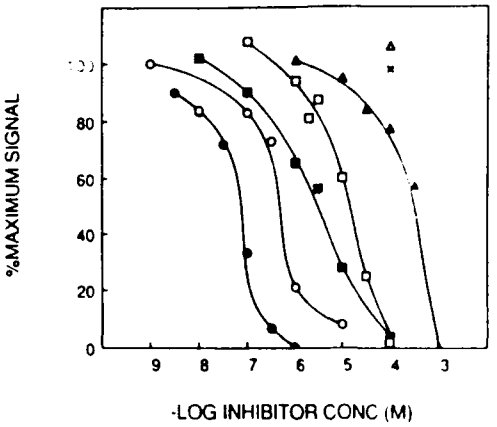


Figure 4. Concentration-dependent inhibition of the AChE biosensor by echothiopate (●), paraoxon (o), bendiocarb (■), methomyl (□), dichrotophos (▲), parathion (x), and malathion (Δ). The AChE biosensor was treated for 10 min with the indicated compound prior to introducing ACh and subsequent assay of activity. The symbols represent means of at least three measurements.

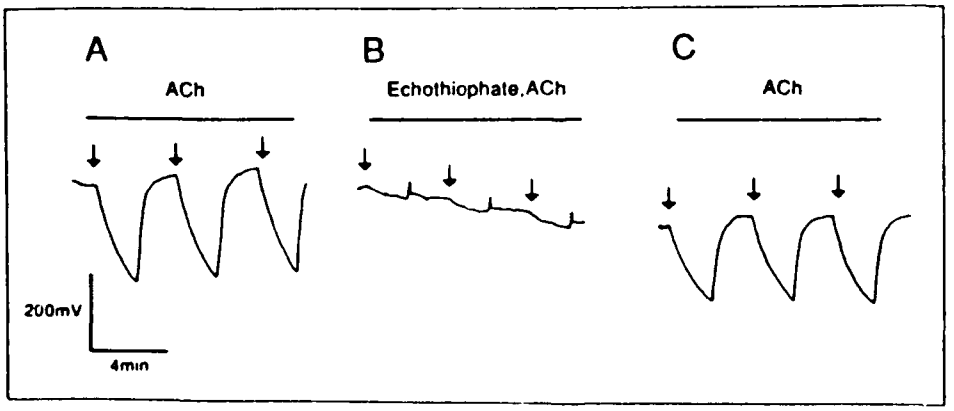


Figure 5. Inhibition of the AChE biosensor by echothiopate. (A) Control responses of the sensor to ACh (1 mM). Echothiopate (0.1 mM) was then added to the ACh-krebs solution and perfused for 10 min before the signal was recorded in (B). Echothiopate was replaced with 1 mM 2-PAM in the perfusion medium and after 10 min 2-PAM was removed and the signal in (C) recorded.

Antibody-Based Sensor Configuration. As noted above,, the AChE based sensor showed very poor sensitivity in detecting the thiophosphate compound, parathion but was extremely sensitive in detecting its oxidized metabolite, paraoxon. In order to detect parathion an antibody-based detection scheme was employed in which a casein-parathion conjugate was immobilized on quartz fibers and FITC-goat antirabbit IgG bound to polyclonal antiparathion rabbit antibodies was used to generate the fluorescent signal. In this design, the presence of free parathion in the perfusate would be predicted to interfere with the binding of the rabbit antiparathion antibody/FITC-goat antirabbit IgG complex and thus decrease the fluorescent signal. The reduction in the optical signal when treated fiber was exposed to free parathion is shown in Fig. 6.

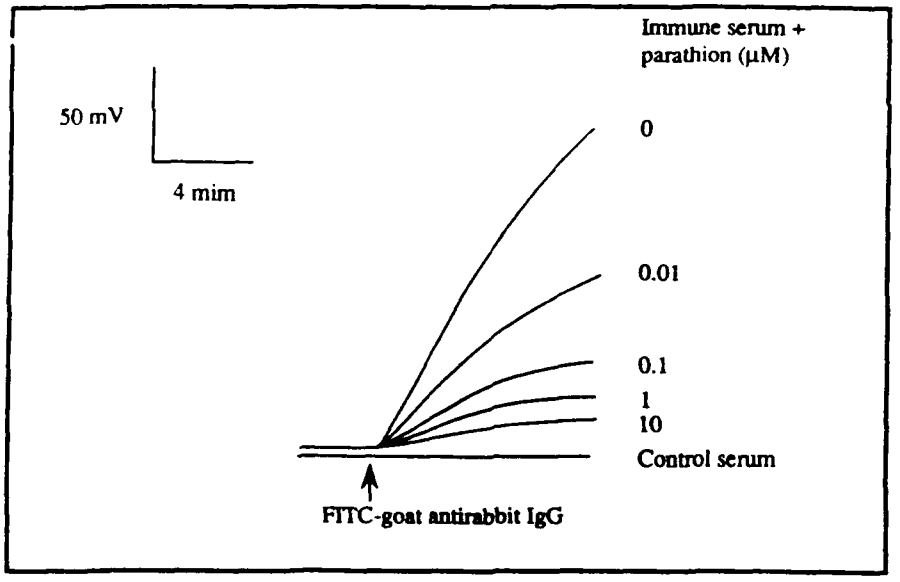


Figure 6. Reduction of the optical signal generated by binding of FITC-goat antirabbit IgG/rabbit antiparathion antibody to casein-parathion conjugates immobilized on the fiber and subsequently perfused in rabbit serum 1:500 with different concentrations of free parathion.

The sensor configuration detects as little as 10 nM parathion which equals 0.3 ppb as opposed to >100 ppm detection afforded by the AChE-based sensor. The antibody based sensor also detects paraoxon but at 100 fold higher concentration than parathion (data not shown).

II. Light-Addressable Potentiometric Sensors (LAPS) - Methods and Results.

General. Experiments with the LAPs technology were performed on the Threshold Instrument (Molecular Devices Corp., Menlo Park, CA) which uses an enzyme-coupled biomolecular complex to generate free H^+ or OH^- ions which are then detected by a pH sensitive silicon nitride surface. In theory, any enzyme which produces H^+ or OH^- ions on reaction with its substrate can be used with this system. The present study used the urease/urea,enzyme/substrate reaction scheme unless otherwise noted.

ACh Receptor-Based LAPS. nAChR was conjugated to carboxy-fluorescein (CF-nAChR) and α-bungarotoxin labelled with biotin (B-α-BGT) according to protocols provided by Molecular Devices Corp. CF-nAChR and B-α-BGT were co-incubated to permit receptor-ligand complex formation and immobilized an biotinylated cellulose nitrate filter via a streptavidin bridge. The immobilized complex was then exposed to anti-fluorescein-urease conjugate for 30 min before quantifying the receptor complex activity in the presence of 100 mM urea. To measure the effect of various compounds on nAChR function. The compounds were incubated with CF-nAChR before incubating with B-α-BGT, immobilization on the membrane and subsequent assay of urease-urea activity. The general immobilization and detection scheme for the nAChR-based LAPS sensor is shown in Fig. 7. The number of B-α-BGT/CF-nAChR/anti-CF-urease complexes immobilized on the membrane was linearly proportional to the amount of CF-nAChR added to the reaction mixture Fig. 8.

Addition of the receptor agonists ACh, carbamylcholine and nicotine reduces the LAPS response in a dose-dependent manner (data not shown) with inhibitory rate constants ranging from 4.4×10^{-6}M for nicotine to 2.9×10^{-8}M for ACh. The receptor antagonists d-tubocurare, α-BGT and α-NT also inhibit the sensor signal in a dose-dependent manner (data not shown) with inhibition constants ranging from 2.7×10^{-7} for dTC to 3.5×10^{-9}M for α-BGT. The ability of agonists but not antagonists to increase receptor binding affinity for subsequent agonist binding is also observed (data not shown).

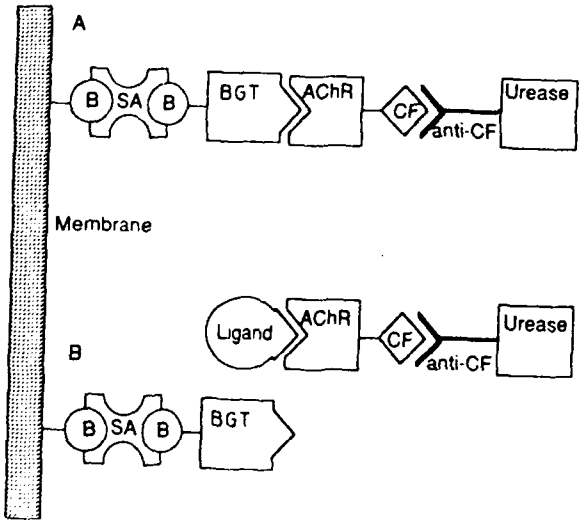


Figure 7. Immobilization and detection scheme for the nAChR-based LAS sensor, B, biotin; SA, Streptavidin; B-BGT, biotinylated -bungarotoxin; CF-AChR, carboxylfluorescein labelled acetylcholine receptor; anti-CF-urease, anticarboxy-fluorescein-urease conjugate.

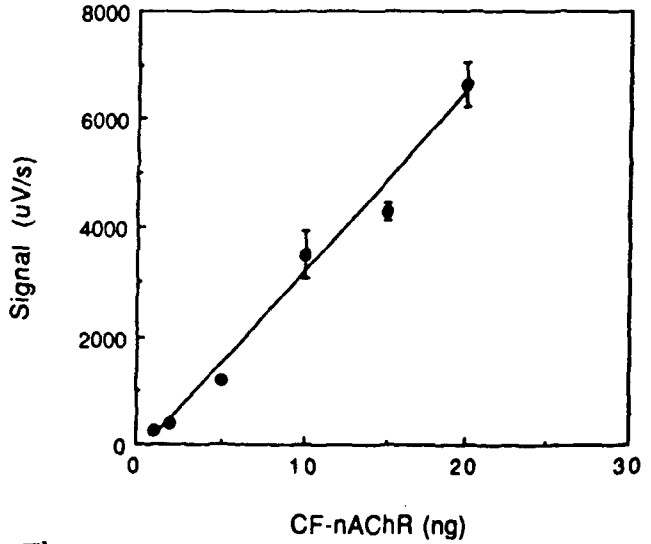


Figure 8. The sensor response as a function of CF-nAChR added to reaction mixture prior to complex immobilization. The nAChR-linked urease activity results in a change in pH when exposed to urea which is measured in units of uV/Sec. Symbols represent menu $\pm$ SEM, n=3.

AChE Based LAPS. As stated previously the LAPS-based sensor format is compatible for detecting any enzyme-based reaction which generates H^+ or OH^- ions and the production of protons by AChE hydrolysis of ACh greatly simplifies the reaction format (Fig. 9).

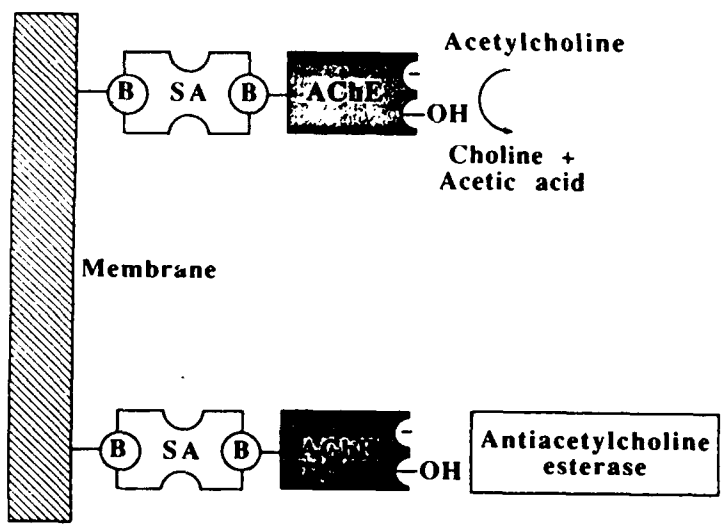


Figure 9. Biotinylated AChE was immobilized onto a biotinylated cellulose nitrate membrane using streptavidin as a bridge. Hydrolysis of the substrate ACh releases protons which generate the potentiometric signal.

Immobilized AChE activity shows a linear correlation with the amount of B-AChE added over a wide range (1-500 ng) and shows autoinhibition at substrate concentrations >100 mM (data not shown. Inhibition of enzyme activity by a variety of anti-cholinesterase compounds is both time and concentration-dependent (Fig. 10).

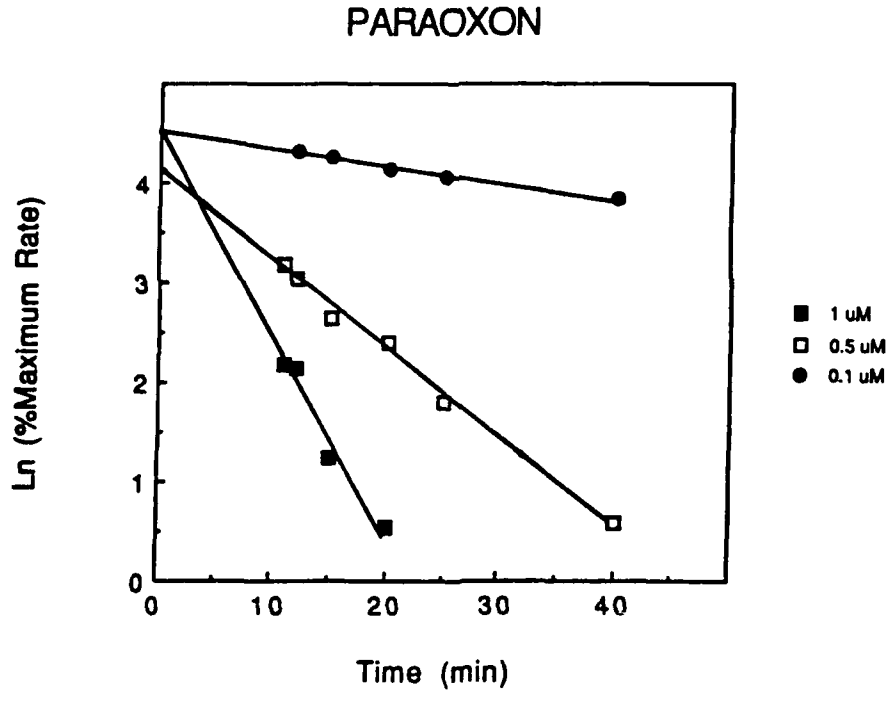


Figure 10. Inhibition of AChE by paraoxon as a function of both concentration and time of exposure.

Reactivation of organophosphate-inhibited AChE is also observed after incubation of the inhibited enzyme with 2-PAM for 10 min. Completely inhibited AChE after echothiopate exposure regained 91% of its original activity after 10 min exposure to 2-PAM.

SUMMARY

Summaries of the LAPS and FOWG-AChR sensors' sensitivities for detecting cholinergic agonists and antagonists are presented Table 3 along with values obtained using radioisotope binding assays in bulk solution. Table 4 presents a comparison between the LAPS and FOWG sensors' ability to measure inhibition of AChE by various anticholinesterase compounds.

Table 3. Comparison of apparent binding affinities of cholinergic ligands to the nACh receptor measured by different methods.

	LAPS biosensor[a] K_i (M)	Fiber-Optic biosensor K_i (M)	Radioisotope Binding Assays K_d (M)
Agonist			
Acetylcholine	2.9×10^{-8}	1.2×10^{-5}	2.4×10^{-8} [b] 4.8×10^{-7} [c]
Carbamylcholine	4.6×10^{-7}	4.5×10^{-3}	6.7×10^{-6} [c] 5.0×10^{-5} [d]
Nicotine	4.4×10^{-6}	6.0×10^{-5}	2.5×10^{-5} [e]
Suberyldicholine	3.4×10^{-8}	NA	NA
Succinylcholine	9.0×10^{-7}	2.4×10^{-5}	6.7×10^{-7} [e]
Antagonists			
d-Tubocurarine	2.1×10^{-7}	8.8×10^{-6}	3.8×10^{-7} [b] 5.0×10^{-6} [d]
Bungarotoxin	3.5×10^{-9}	6.0×10^{-9}	1.0×10^{-10} [c]
Naja toxin	1.2×10^{-8}	1.0×10^{-8}	4.0×10^{-9} [f]

[a] IC_{50} values were determined by Log-Probit analysis (see Figs. 3 and 4) and K_i values calculated using the relationship $K_i = IC_{50}/(1 + [L]/K_d)$.
[b] Data from reference (3).
[c] Data from reference (4).
[d] Data from reference (5).
[e] Data from reference (6).
[f] Data from reference (7).
NA data not available

Table 4. Comparative inhibition of AChE using LAPS, fiber-optic and colorimetric assays.

Compound	LAPS biosensor assay IC_{50} (M)	Fiber-Optic biosensor assay IC_{50} (M)	Colorimetric assay IC_{50} (M)
Echothiophate	$3.1x10^{-8}$	$3.8x10^{-8}$	$3.5x10^{-8}$
Bendiocarb	$9.0x10^{-8}$	$2.2x10^{-6}$	$1.7x10^{-7}$
Methomyl	$1.8x10^{-7}$	$9.0x10^{-6}$	$2.1x10^{-7}$
Paraoxon	$1.5x10^{-7}$	$3.7x10^{-7}$	$4.0x10^{-7}$
Dicrotophos	$1.9x10^{-4}$	$3.3x10^{-4}$	$1.1x10^{-4}$

CONCLUSION

Both the FOWG and LAPS-based sensor configurations show versatility in detecting a broad spectrum of chemical compounds depending on the biological recognition molecule chosen as the sensing element. Using antibodies as the recognition element provides very sensitive (ppb) detection of a specific target compound from closely related molecular forms. The choice of receptors and/or enzymes as the recognition element affords simultaneous detection capabilities for a much broader range of compounds. From the data reported here, it has been shown that antibodies, receptors and enzymes retain their native molecular recognition functions when stripped of their normal biological microenvironments and subject to artificial immobilization and molecular conjugation procedures. Comparison between the FOWG and LAPS-based sensor formats reveal certain advantages of one over the other. The advantages of the FOWG include its rapid response time (<1 min), nominal sample preparation requirements and near continuous on-line monitoring of sample content. The LAPS-based sensor shows an overall greater sensitivity for detecting compounds that act as receptor/enzyme agonists/ substrates. Continued research and development of biosensor technology holds unlimited potential for meeting the chemical and biological threat detection needs of present and future military requirements.

REFERENCES

1. M.E. Eldefrawi and A.T. Eldefrawi, Arch. Biochem. Biophys., 159, 363 (1973).

2. A.T. Eldefrawi, E.R. Miller, D.L. Murphy, and M.E. Eldefrawi, Mol. Pharmacol., 22, 72 (1982).

3. G. Weiland and P. Taylor, Mol. Pharmacol., 15, 197 (1979).

4. D. Colquhoun and H.P. Rang, Mol. Pharmacol., 12, 519 (1976).

5. T. Moody, J. Schmidt, and M.A. Raftery, Biochem. Biophys. Res. Commun., 53, 761 (1973).

6. A.T. Eldefrawi, R. Miller, and M.E. Eldefrawi, Biochem. Pharmacol., 31, 1819 (1982).

7. M. Weber and J.P. Changuex, Mol. Pharmacol., 10, 1 (1974).

Electric Field Effects on the Optical Properties of Layered Semiconductor Structures

Richard L. Tober,* Ph.D., and Garnett W. Bryant, Ph.D.
Harry Diamond Laboratories, U.S. Army Adelphi Laboratory Center
2800 Powder Mill Rd., Adelphi, Maryland 20783-1197

W. Q. Li, Ph.D., and P. K. Bhattacharya, Ph.D.
Department of Computer Science and Electrical Engineering
University of Michigan, Ann Arbor, Michigan 48109-2122

1. INTRODUCTION

The development of novel optoelectronic devices for Army needs depends on the ability to understand and control basic properties of semiconductor structures. Devices used for advanced signal processing, for example, often use artificially layered structures with dimensions such that quantum confinement effects can either enhance or hinder desired optical properties. Also, dopants in semiconductors necessarily produce electric fields that can confine carriers to small regions of space. The optical properties associated with electric–field–induced carrier confinement can be significant enough to affect the overall performance of critical devices.

Quantum well structures have potential profiles that confine states in the conduction and valence bands in such a way that produces sharp exciton features in room–temperature optical spectra. In contrast, modulation-doped GaAs/AlGaAs heterojunction structures can have confined states in either the conduction or the valence band, but not both simultaneously, and therefore do not enhance exciton features. Efforts to use sensitive modulation spectroscopies to probe these structures have had marginal success in that only broad oscillations above the GaAs band gap could be linked to the two–dimensional electron gas that occurs at the heterojunction.[1] Recent calculations of modulated absorption spectra showed no direct evidence of transitions between extended valence band states and electron sub-band states, only oscillations due to the overlap between confined–state wavefunctions and the extended–state Airy functions.[2]

The band bending that results from surface Fermi level pinning can create triangular quantum wells in undoped heterojunction samples not only at the surface but also near an GaAs/AlGaAs interface that is within the surface depletion region. Confinement effects

associated with these undoped structures have been inferred from subtle features in modulated reflectance spectra, but, as with the modulation–doped structures, no definitive evidence was reported.[3] Oscillatory structure has frequently been observed in electric–field–modulated spectra (i.e., photoreflectance as well as electroreflectance) of heterojunction samples. These Franz-Keldysh-like oscillations had atypical periods for which either no speculations[3,4] or only cursory explanations in terms of nonuniform electric fields[5,6] were given.

The experimental and theoretical results presented below show a correlation between carrier confinement near heterojunctions and features observed in modulated spectra. Overlap matrix elements calculated for heterojunction sample geometries have oscillations with periods that depend on whether or not the photon energy corresponds to transitions to electron states in the energy range between the bottom and the top of the conduction band barrier. For energies below and above the barrier, modulated spectra have the normal Franz-Keldysh oscillations (FKO), where the energy between consecutive extrema decreases. In the intermediate range the oscillations spread out in energy with increasing photon energy. The number of oscillations in the energy range between the bottom and top of the barrier depends on the state from which the transitions occur. These trends explain several anomalous features in electric–field–modulated spectra obtained from heterojunction samples.

2. THEORETICAL CONSIDERATIONS

2.1. Franz–Keldysh Analysis

The experimentally measured quantity in electroreflectance experiments, $\Delta R/R$, is represented as a linear combination of Airy functions. When the arguments of the Airy functions are such that the asymptotic limit can be used, $\Delta R/R$ can be expressed as a cosine function. In this case, consecutive extrema in modulated reflectance spectra lie along a straight line of the form[7]

$$(E_n - E_g)^{3/2} = \frac{3}{2}\pi\left[\frac{h\Omega}{2\pi}\right]^{3/2}(n) - \frac{3}{2}\left[\frac{h\Omega}{2\pi}\right]^{3/2}\phi \quad , \qquad \text{(EQ. 1)}$$

where the slope is proportional to the electric field and inversely proportional to the square root of the interband reduced mass through the relation

$$\frac{h\Omega}{2\pi} = \left[\frac{\pi e^2 F^2}{\mu h}\right]^{3/2} ,$$

μ is the reduced mass, E_n is the energy of an extremum, E_g is the band gap energy, F is the electric field strength, and ϕ is a phase factor that depends on the dimensionality of the critical point. Deviations from this linear behavior have been previously reported, but only in connection with samples having both GaAs and AlGaAs layers.[3-6] This prompted us to

investigate the effects that band profiles have on the optical properties of heterojunction samples.

2.2. Heterojunction Effects on Optical Properties

Initial calculations of the effects that scattering from heterojunctions has on Franz-Keldysh oscillations in electroabsorption spectra have been done. We have considered structures with a single heterojunction with a field applied to confine valence states at the heterojunction to obtain qualitative insight. The effective mass equation has been solved to determine conduction band states and trapped valence states. Band-to-band overlap matrix elements have been calculated to estimate the strengths of interband transitions. Detailed results will be presented elsewhere.

Figure 1 shows the strengths for transitions from the n = 5 trapped valence band state to the conduction band of a single GaAs/$Al_{0.3}Ga_{0.7}As$ heterojunction (with a 300-meV conduction band offset) in a 20-kV/cm applied field. For comparison, Fig. 2 shows the overlap matrix elements calculated with conduction band states determined without the barrier in the conduction band. The barrier has little effect on transitions to conduction states with energies below the onset of the barrier because these states never reach the barrier. However, strengths of transitions from the n =5 valence state to states with energies between the bottom and the top of the conduction band barrier (1.58 and 1.88 eV, respectively) are very different for the two cases. Rapid oscillations of the transition strengths occur, as the energy is increased, when the effect of the barrier on the conduction band states is ignored (Fig. 2). Oscillations occur for each additional half-wavelength of the conduction band state that can overlap the trapped valence state. The frequency of these oscillations increases with increasing energy and corresponds to normal FKO. When the effect of the barrier on the conduction band states is included, these oscillations spread out with increasing energy, because the barrier hinders the addition of half-wavelengths of the conduction band state to the region of the trapped valence state. Five oscillations in the transition strength occur in the energy region up to 1.88 eV, where conduction states have energies below the top of the barrier. The number of oscillations in the transition strength in this energy region is determined by the number of oscillations in the valence state. Rapid oscillations occur again when the conduction band states have energies above the barrier at the heterojunction.

The barrier height determines the width of the energy region where this slowdown in the oscillations occurs. Smaller barriers will provide a weaker distortion of the Franz–Keldysh oscillations. The low energy transitions have the largest transition strengths and should make the most important contribution to the absorption spectra. Increasing the electric field increases the number of oscillations that occur at energies below the onset of the barrier and are insensitive to the barrier; thus, bulk Franz–Keldysh oscillations should be recovered for large applied fields.

To correctly model the experiment, calculations should include a sum over all the valence-band states that contribute. This will be done after completing a detailed analysis of the valence band occupancy. Nevertheless, we expect no changes in the qualitative results

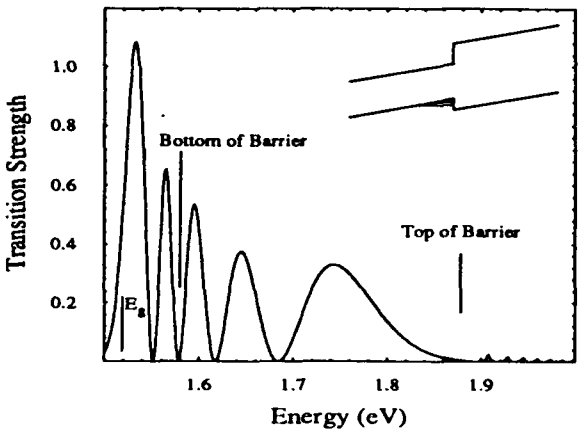


FIG. 1. Band-to-band overlap matrix element for transitions from the $n = 5$ trapped valence state to the conduction band as a function of the energy of the conduction band state for a structure with a single heterojunction. The inset depicts the potential profile considered.

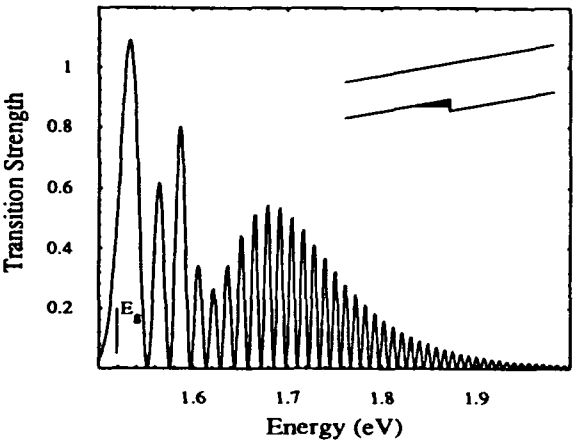


FIG. 2. Band-to-band overlap matrix element for transitions from the $n = 5$ trapped valence state to a conduction band having no barrier, as indicated in the inset.

presented above, only a superposition of contributions such as shown in Fig. 1, for each contributing valence state.

3. EXPERIMENTAL CONSIDERATIONS

3.1. Sample Growth and Preparation

Two of the three samples grown for these studies had structures specifically designed to create triangular quantum wells at GaAs/AlGaAs heterojunctions. The first, the heterojunction sample, had a 3000 Å GaAs buffer grown on a semi-insulating GaAs substrate, followed by a 2000 Å layer of $Al_{0.3}Ga_{0.7}As$, and capped with a 2000 Å layer of GaAs. The second, the diode structure, had a 1.3 mm layer of $Al_{0.3}Ga_{0.7}As$, a 1500 Å layer of GaAs, and a 500 Å layer of $Al_{0.3}Ga_{0.7}As$ grown within the intrinsic region of a p-i-n diode structure. The final structure, the epilayer sample, had an 8000 Å layer of undoped GaAs on an n+ GaAs substrate. Both the GaAs epilayer sample and the heterojunction sample had 1000 Å gold electrodes sputtered on their back surfaces and 60 Å semitransparent nickel layers evaporated on the front surfaces. Electrodes were connected using silver epoxy. The diode sample had 450 mm diameter cylindrical mesas and ring contacts with 250 mm inner diameters to facilitate optical measurements. The samples were epoxied to gold carriers, electrodes were wire bonded to the carrier and to the gold rings. Figure 3 shows schematics of the sample potential profiles produced by the growth geometries and processing. They were intended only to illustrate trends, not the actual potential profiles. The Fermi level pinning that occurs at the surface in the heterojunction sample produces triangular wells in the valence band at the GaAs/AlGaAs interface and in the conduction band as shown in Fig. 3(a). Similarly, the epilayer sample (not

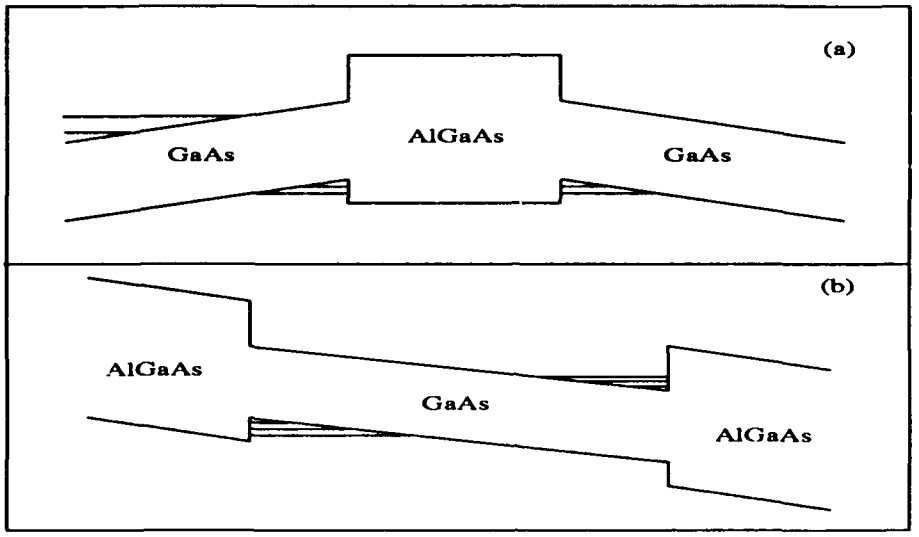


FIG. 3. Schematics of (a) the heterojunction sample and (b) the p-i-n diode sample showing confined states not only at the GaAs/AlGaAs heterointerfaces but at the sample surface as well.

shown here) has a single well in the conduction band at the surface. Figure 3(b) shows that the p-i-n diode sample has quantum wells near heterojunctions in both the valence and conduction bands.

3.2. Optical Measurements

Radiation from a tungsten-halogen lamp was dispersed using a 2/3 meter monochromator, and spatially filtered so that a 250 μm diameter spot could be focused onto the samples. The amplitude of the 400 Hz square–wave modulating voltages used for all three samples produced a 3 kV/cm electric field in the regions probed. The dc bias was varied independently. All samples were mounted in a liquid transfer cryostat so that spectra could be obtained at 4.5 ± 0.5 K. Reflected radiation was measured with a silicon p-i-n diode detector and lock-in amplifier. The ac signal was then digitally normalized to the dc signal (measured at the lock-in auxiliary analog-to-digital converter) to obtain the normalized modulated reflectance, $\Delta R/R$.

4. RESULTS

Figure 4 shows electroreflectance (ER) spectra of the heterojunction sample and the GaAs epilayer sample. They were obtained with applied biases that produced comparable fields in each sample, that is, 13 and 16 kV/cm, respectively. Both spectra have typical Franz-Keldysh oscillations. Their extrema, when plotted according to Eq. 1, fit straight lines with slopes commensurate with the corresponding electric fields (see Fig. 5).

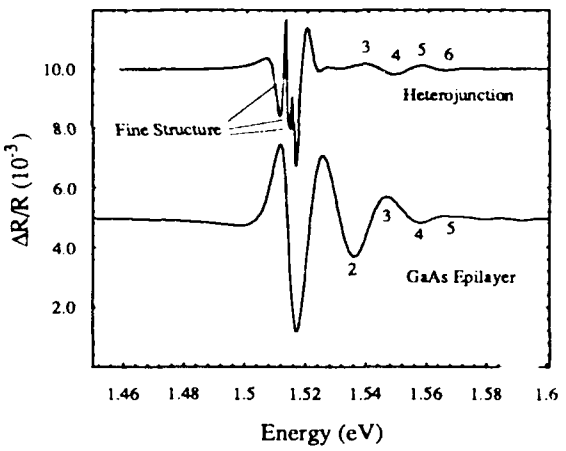


FIG. 4. Electroreflectance spectra, obtained at 5K, of the heterojunction sample and the GaAs epilayer sample. The magnitudes of the fields, as determined from the FKO, are 13 and 16 kV/cm, respectively.

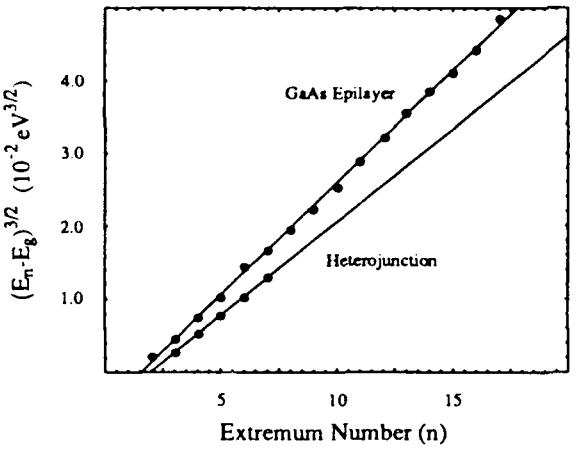


FIG. 5. Linear FKO plots of the extrema shown in both spectra of Fig. 4.

In a lower field regime the periodicity of the oscillations does not change significantly with increasing field strength. This is evidenced by the lines in Fig. 6 connecting extrema of the same index number in spectra obtained from the heterojunction sample at two different field strengths. The low field oscillations decrease in magnitude with increasing fields, reach a minimum, and then evolve into high field oscillations whose periodicity does depend on the field strength. Moreover, the high field FKO plots are linear whereas those for the low–field data are not (see Fig. 7).

All the trends exhibited in Figs. 6 and 7 can be explained in terms of the overlap matrix elements between trapped valence states and the extended states in the conduction band near the heterojunction. The electron wavefunctions slow down near the barrier so that the oscillations spread out in energy (see the discussion in section 3.1). This effect gives rise to the curvature in the low field FKO plot of Fig. 7. As the electric field increases, the relative magnitude of the wavefunctions (i.e., Airy functions) near the barrier decreases so that the transition strength of the oscillations also decreases. Changes in the field strength do not affect the wavefunctions striking the barrier as readily as those below the barrier. Therefore, the period of the low–field oscillations does not change significantly with the applied field. The energy range over which the valence state wavefunctions overlap with the below barrier electron wavefunctions increases with the applied field so that the normal Franz-Keldysh oscillations emerge and extend to greater energies as the low field oscillations diminish. The high–frequency oscillations in Fig. 3 were calculated assuming a field strength of 20 kV/cm. Their periodicity is comparable to that of the high–field (23 kV/cm) experimental results in Fig. 6 and correspond to the normal FKO. Since the spectrum calculated for Fig. 2 accounts for transitions only from the fifth valence state, it has but five oscillations. The low–field spectra in Fig. 6 had contributions from lower as well as higher order states and therefore have more oscillations.

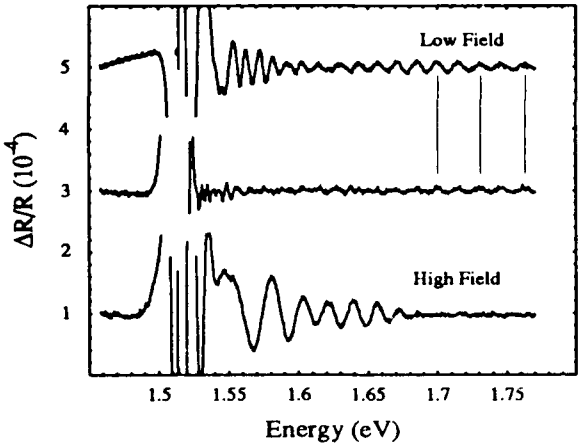


FIG. 6. ER spectra of the heterojunction sample obtained with dc biases of +5.0 (low field), +1.0, and -5.0 V (high field). The high frequency oscillations diminish and give way to the normal FKO as the electric field increases.

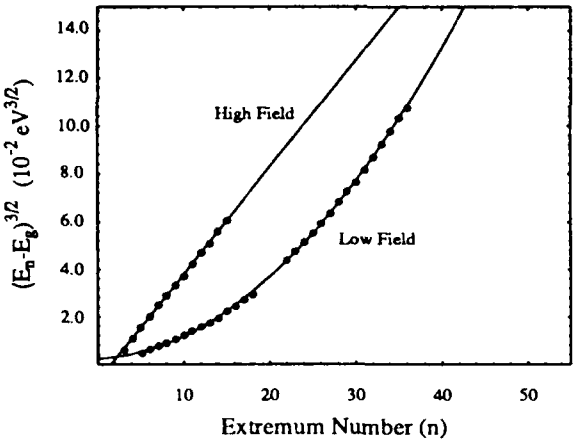


FIG. 7. FKO plots for the low and high field spectra of the heterojunction sample shown in Fig. 6. The slope of the high field plot corresponds to a 23 kV/cm electric field. The slope of a line drawn through the first ten extrema in the low field plot corresponds to about 8 kV/cm.

Similar series of spectra obtained from both the p-i-n diode sample and the epilayer sample have the low field oscillations that diminish and give way to the Franz-Keldysh oscillations. Also, we found that spectra of the epilayer sample did not have the fine structure near the GaAs band gap energy that the other two samples had (see the fine structure in Fig. 4 and the encircled features in Fig. 8). Similar structure has been observed in spectra of GaAs and attributed to bulk exciton effects[7,8] but its absence in the spectra of the epilayer sample suggests other origins. We offer two possibilities not yet verified. First, contributions to the oscillatory structure associated with the many confined valence and conduction states could have phase relations such that they destructively interfere at low energies to produce the fine structure but constructively interfere at higher photon energies to produce smoother oscillations. Subtle indications of continued interference in the high–photon–energy oscillations support this contention. Alternatively, the fine structure could result from transitions between some of the confined electron and valence states.

5. SUMMARY AND CONCLUSIONS

We have observed electric field induced anomolies in optical spectra of artificially layered samples. These irregularities can be used to characterize heterointerfaces in optoelectronic devices such as optical modulators and high–speed amplifiers amplifiers so critical to Army systems. Barriers, either at heterojunctions or at the surface, modify the phase and amplitude of overlap matrix elements between confined valence (or conduction) states and extended conduction (or valence) states. This effect produces irregularities in the Franz-Keldysh oscillations observed in electric–field–modulated spectra. Two field strength regimes exist, each having oscillations with distinct character. In the low field regime barriers impede

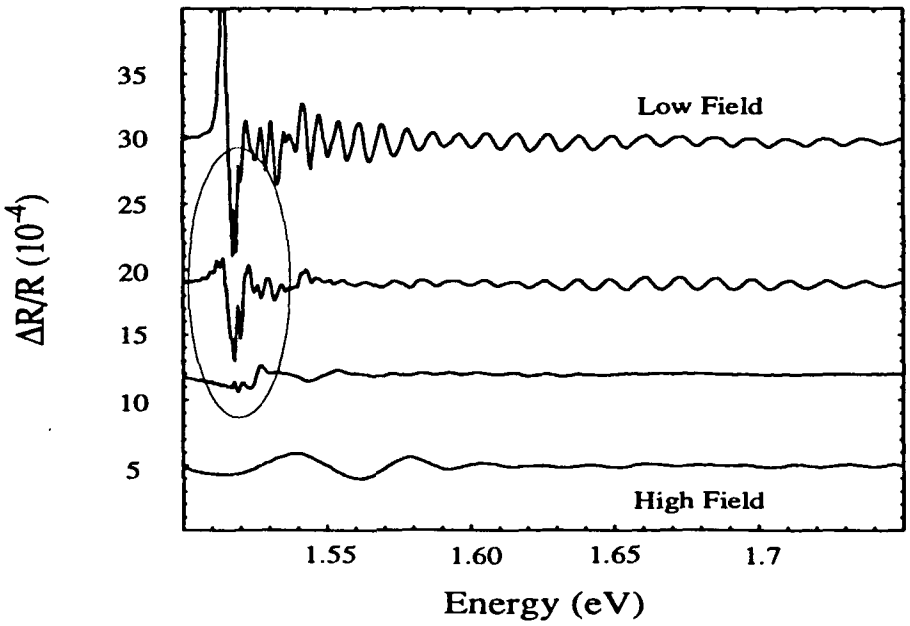


FIG. 8. Electroreflectance spectra of the p-i-n diode sample. Not only does the geometry of this sample support the high frequency oscillations at high field strengths but it enhances the low energy fine structure as well.

wave functions so that oscillations in optical spectra spread out in energy, and diminish with increasing field strength. Corresponding FKO plots ($[E_n - E_g]^{3/2}$ vs. oscillation extrema) have curvature. In the high field regime the normal FKO emerge with increasing field strength. Normal FKO will emerge as a barrier ceases to slow down wavefunctions, that is, when wavefunctions either extend beyond barriers or diminish before reaching them. This latter point suggests that information regarding barrier height and width can be determined from the character of the oscillations. Though no optical transitions between states confined in the triangular well near the heterojunctions could be identified, subtle effects linking the number of oscillations in low–field electroreflectance spectra with the number of oscillations in the confined state wavefunctions were.

6. REFERENCES

1. O. J. Glembocki, B. V. Shanabrook, N. Bottka, W. T. Beard, and J. Comas, Appl. Phys. Lett. 46 (10), 970 (1985).

2. S. Snow, O. J. Glembocki, and B. V. Shanabrook, Phys. Rev. B, 38 (17), 12483 (1988).

3. R. L. Tober, and J. D. Bruno, "Piezoreflectance Investigations of Narrow Barrier Single Quantum Wells," International Conference on Modulation Spectroscopy, Vol. 1286, pp. 291-299, SPIE-The International Society for Optical Engineering, Bellingham, 1990.

4. R. L. Tober, Arthur L. Smirl, Thomas F. Boggess, and J. N. Schulman, J. Appl. Phys., 64(9), 4678 (1988)

5. Michael Sydor, James Angelo, Jerome J. Wilson, W. C. Mitchel, M. Y. Yen, Phys. Rev. B, 40(12), 8473 (1989).

6. N. Pan, X. L. Zheng, H. Hendricks, and J. Carter, J. Appl. Phys. 68(5), 2355 (1990).

7. D. E. Aspens and A. A. Studna, Phys. Rev. B, 7(10), 4605 (1973.

8. F. Evangelisti, A. Frova, and J.U. Fischbach, Phys. Rev. Lett., 29 (15), 1001 (1972).

Simultaneous Electron and Hole Tunneling in Coupled Quantum Wells

Mary S. Tobin, Dr. *
Garnett W. Bryant, Dr. †
Richard P. Leavitt, Dr.
John L. Bradshaw, Dr.
John Pham, Mr.
Harry Diamond Laboratories
Adelphi, MD 20783–1197

1 Introduction

Ultrafast optoelectronic semiconductor device components to be used as optical modulators and bistable optical switches will be needed to satisfy future Army system requirements in optical signal processing, automatic target recognition, and advanced sensor technology. The current semiconductor structures being researched, which have the potential for device application, include quantum well device structures made from systems of uncoupled quantum wells, coupled double-well systems, and superlattices of coupled quantum wells. The optoelectronic response in coupled double-well structures is distinctly different from the response of single wells or superlattices. The response of coupled double-well structures is determined not only by carrier generation and recombination, but also by the dynamics of charge transfer between the coupled wells, which in turn affects carrier generation and recombination.[1–3] The interwell charge transfer and optical response of coupled double-well structures can be controlled by use of either optical pumping or applied bias. However, coupled double-well structures provide restricted flexibility for controlling charge transfer because the electrons and holes must transfer across the same barrier between the same two wells.

In this paper we study more complex coupled well structures that should provide

greater flexibility for control of charge transfer. The simplest new structures that we study are triple quantum well (TQW) structures in which electron-hole generation occurs in the central well and charge transfer occurs to the wells on either side of the central well. The TQW structures can be designed so that at the same applied bias the electrons transfer resonantly to one well while the holes transfer resonantly to the other well. Achieving simultaneous electron and hole resonant transfer is possible in TQW structures because each transfer process involves a different barrier and can be tailored independently. Such control is not possible in double-well systems. Much faster switching is expected if electrons and holes both transfer resonantly. Tailoring the barriers determines not only the tunneling but also the electron-hole polarizability and recombination rate.

We have developed an effective-mass model for the electronic structure of TQW structures. We use the model to design TQW GaAs/AlGaAs structures that provide simultaneous resonant transfer of electrons and holes at a particular applied bias V_r. The designed TQW structures are grown by molecular beam epitaxy (MBE). TQW structures are investigated by photocurrent spectroscopy and time-resolved photoluminescence to determine the dynamics of the electron and hole transfer and the optical response of these structures. For comparison, we investigate the dynamical response of TQW structures in which only electron or hole transfer is resonant at V_r. The time scale of the dynamical response is determined by whether electrons, holes, or both transfer resonantly.

2 Experimental Details

The samples were grown by MBE on an n^+ GaAs [001] substrate. A 500-nm n-type GaAs buffer layer was followed by two short period superlattice-smoothing sequences (an n-type AlGaAs/GaAs superlattice followed by an intrinsic AlGaAs/GaAs superlattice). This was followed by the TQW structure: well-1 was GaAs with thickness d_1, followed by a 2.5-nm AlGaAs barrier, followed by well-2 of thickness d_2, followed by a 1.5-nm AlGaAs barrier, followed by GaAs well-3 of thickness d_3. This in turn was followed by two more short period superlattices (an intrinsic AlGaAs/GaAs superlattice followed by a p-type AlGaAs/GaAs superlattice). The structure was capped by p-type GaAs. The doping for the n- and p-type was Si

and Be, respectively, at 2×10^{18} cm^{-3}. For the $Al_xGa_{1-x}As$ layers, $x = 0.3$. The design thicknesses of the quantum wells for the three samples are listed in Table 1. Samples 1, 2, and 3 have the electron and hole, electron only, and hole only, resonant, respectively. The resonance field of the designed structures is 50 kV/cm.

Table 1.
Nominal thicknesses for each GaAs well of the TQW structure

Sample	d_1 (nm)	d_2 (nm)	d_3 (nm)
1	5	10	3
2	5	10	1.5
3	2	10	3

Each sample was processed into an array of 200-μm diodes and electrical contacts made to the p-i-n structures. For the photocurrent measurements, each sample was mounted in a liquid nitrogen immersion dewar. Chopped light from a tungsten-filament lamp was dispersed by a monochromator and focused onto one of the diodes. A current-sensitive preamplifier and a lock-in amplifier were used to detect the photocurrent generated in the diode as a function of the photon energy and the bias on the sample. Time-resolved photoluminescence was obtained using a time-correlated photon-counting system.[4] For this experiment a Nd:YAG synchronously pumped Rhodamine 6G laser was used for photoexcitation of the sample. The fluorescent light from the sample was passed through a double monochromator and monitored by a photomultiplier tube. The system response time was limited by the response of the photomultiplier tube to about 0.5 ns.

3 Photocurrent Spectra

The photocurrent $I(\lambda)$ is related to the absorption $\alpha(\lambda)$ by the following expression[5]

$$I(\lambda) = \frac{e\lambda P(\lambda)}{hc}(1 - R)\eta \left(1 - \exp(-\alpha(\lambda)W)\right), \qquad (1)$$

where $I(\lambda)$ is the photocurrent, $P(\lambda)$ is the incident optical power at wavelength λ, R is the front-surface reflectivity of the sample, η is the quantum efficiency (the number of electron-holes collected per absorbed photon), $\alpha(\lambda)$ is the absorption

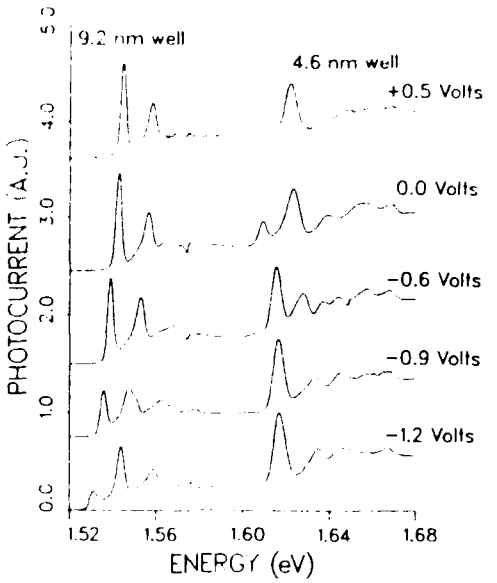


Figure 1: Photocurrent spectra for sample 1 for various values of reverse bias; sample temperature = 77K. The spectra have been vertically offset for clarity.

coefficient within the intrinsic region, and W is the thickness of the intrinsic region of the diode.

In Fig. 1 the photocurrent measurements for sample 1 are given as a function of energy for five values of electric field. Based on the energies of the transitions, we determine the actual values of d_1, d_2, and d_3 to be 4.6, 9.2, and 2.8 nm for sample 1. The lowest energy features (about 1.52 to 1.56 eV) can be attributed to intrawell transitions (1,1H) and (1,1L) of the widest well (9.2 nm); i.e. the lowest electronic level ($n = 1$) combines with the $n = 1$ heavy and light holes, respectively, whereas similar features for the intermediate-size well (4.6 nm) occur at higher energy (about 1.60 to 1.64 eV) as indicated on the figure.

Numerous spectra such as these were taken for the three samples. The peaks were tentatively identified and the transition energy dependence on applied electric field was plotted for the three samples (1, 2, and 3) in Fig. 2 (a, b and c, respectively). The $n = 1$ intrawell transitions for the wells are indicated on the graphs.

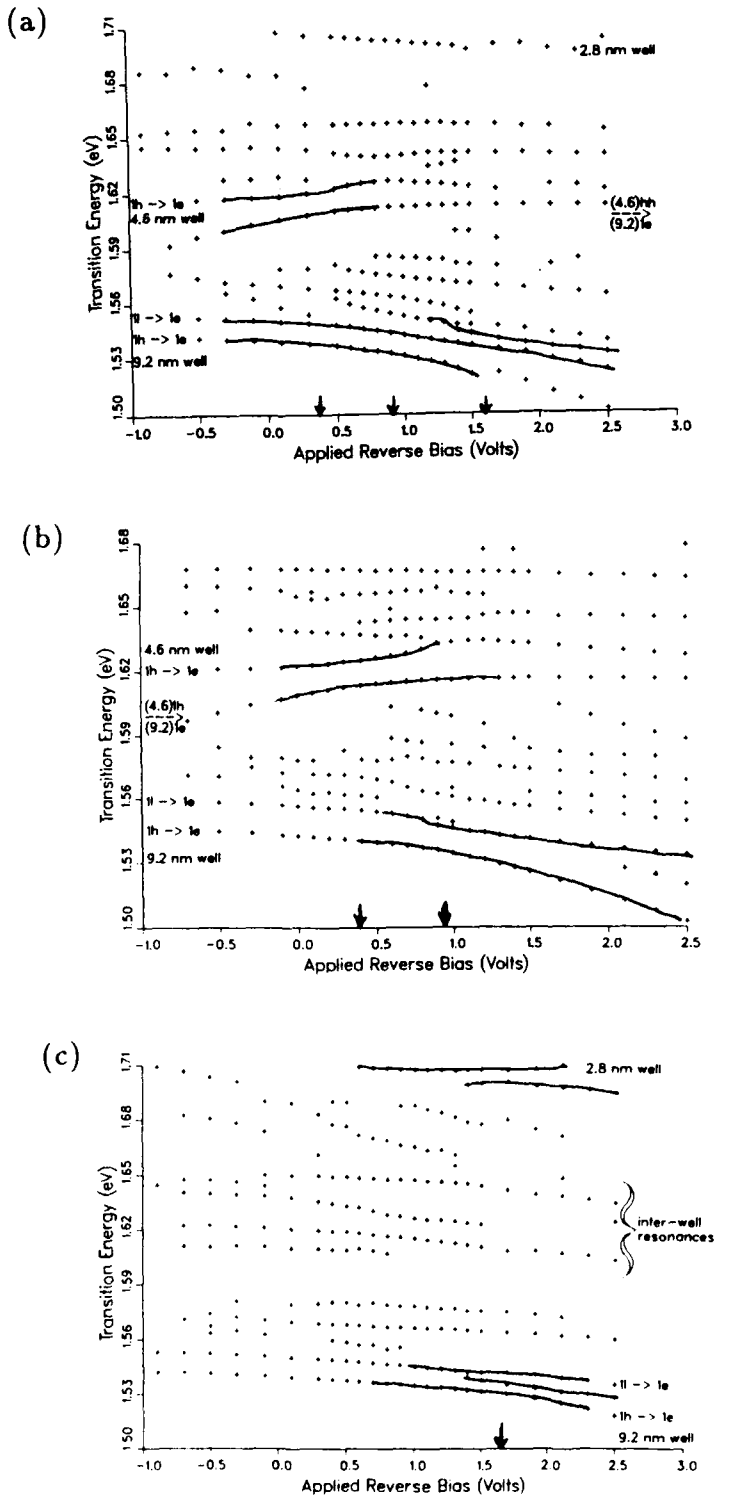


Figure 2: Transition energies deduced from photocurrent spectra taken at 77K for (a) sample 1, (b) sample 2, and (c) sample 3.

The interwell (indirect) transitions are also identified; in this case, the transition occurs between the electron of one well with the hole of one of the other wells. By tracking the transition energies as a function of applied field, crossovers between direct and indirect transitions can be identified. For example, if one follows the indirect transition between the electron from the 9.2-nm well and the hole of the 4.6-nm well as a function of applied electric field, one sees a crossover with the direct transition of the 4.6-nm well at approximately 0.4 volt reverse bias. Similarly, a crossover occurs between the direct transition for the 9.2-nm well and an indirect transition at approximately 0.9 V. These energy crossings are associated with the electron resonance between the $n = 1$ levels of the 9.2- and 4.6-nm wells, which occurs between 0.4 and 0.9 V. The hole resonance occurs near 1.6 V, which is identified by the crossover of the direct transition of the 9.2-nm well and the indirect transition between the 9.2 and 2.8-nm wells. Photoluminescence measurements and time-resolved photoluminescence have also been taken. There was no indication of a simultaneous resonanance from these data. Samples 1 and 2 both have the designed electron resonance. Samples 1 and 3 both have the hole resonance. However, these resonances do not occur at the same applied bias in sample 1.

4 Conclusions

This work represents an initial effort to grow and investigate TQW's. Three samples containing TQW's in a p-i-n structure, which were designed to have a simultaneous electron and hole resonance, an electron-only resonance, and a hole-only resonance, have been grown by MBE. The dependence of the resonances on reverse bias has been investigated by photocurrent spectroscopy. From these results, it has been determined that samples 2 and 3 do have an electron-only and a hole-only resonance, respectively, However, sample 1, which was designed to have a simultaneous electron and hole resonance, was found to have an electron resonance between 0.4 and 0.9 V, and a hole resonance near 1.6 V. Currently, we are determining why the effective-mass theory failed to correctly model the electronic structures of the samples. The results will be used to design and grow new samples with the desired simultaneous resonance.

References

† National Research Council Senior Research Associate at Harry Diamond Laboratories

1. J. E. Golub, K. Kash, J. P. Harbison, and L. T. Florez, "*Long-lived spatially indirect excitons in coupled $GaAs/Al_xGa_{1-x}As$ quantum wells*," Phys. Rev. B **41**, 8564 (1990).

2. R. Sauer, K. Thonke, and W. T. Tsang, "*Photoinduced space-charge buildup by asymmetric electron and hole tunneling in coupled quantum wells*," Phys. Rev. Lett. **61**, 609 (1988).

3. R. Strobel, R. Eccleston, J. Kuhl, and K. Kohler, "*Measurement of the exciton-formation time and the electron- and hole-tunneling times in a double-quantum-well structure*," Phys. Rev. B **43**, 12564 (1991).

4. See for example, D. V. O'Connor and D. Phillips, **Time-Correlated Single Photon Counting** (Academic Press Inc, 1984).

5. R. P. Leavitt and J. L. Bradshaw, "*Transit-time effects on photocurrent spectra of multiple quantum well diodes*," Appl. Phys. Lett **59**, 2433(1991).

Detection of Picogram Quantities of Botulinum Toxin-B (BoTX) Using the Light Addressable Potentiometric Sensor

Dr. James J. Valdes*
U.S. Army Chemical Research, Development and Engineering Center
Aberdeen Proving Ground, Maryland 21010-5423

Dr. James P. Chambers, Dr. James T. Colston, Dr. Pramod Kumar
University of Texas at San Antonio
San Antonio, Texas 78249-0662

Dr. Eppie D. Rael
University of Texas at El Paso
El Paso, Texas 79968-0519

INTRODUCTION

The extremely potent exotoxins produced by *Clostridium botulinum* cause fatal neurotoxicity in both animals and humans. Seven serologically distinct *botulinum* toxins have been isolated from this organism with each sharing similar structural characteristics. A 145,000-150,000 M_r protoxin is enzymatically cleaved into the active toxin consisting of two peptides of approximately 50,000 and 100,000 M_r linked by a disulfide bridge.[1] The active toxin binds specifically to gangliosides and prevents presynaptic release of acetylcholine by an unknown mechanism.[2]

Sensitive and rapid detection of BoTX is invaluable for testing contamination of foodstuffs. In this report, we describe an immunoligand assay using a Light Addressable Potentiometric Sensor (LAPS) to detect picogram quantities of BoTX. The LAPS is a recently developed, sensitive and reliable technology utilizing nitrocellulose filters onto which fluorescein-labeled immunocomplexes are immobilized and labeled with an anti-fluorescein urease conjugated antibody.[3,4,5] Quantitation of toxin is accomplished by insertion of immobilized, urease-labeled immunocomplex into a reader containing a solution of urea. Enzymatic breakdown of urea produces a potentiometric shift proportional to the amount of urease, hence toxin, which is detected rapidly with high sensitivity and reproducibility. Biotin- and fluorescein-labeled anti-BoTX polyclonal antibodies were used in this study to detect picogram quantities of toxin on the LAPS.

MATERIALS AND METHODS

Goat polyclonal antiserum against *botulinum* toxin-B was obtained from Biodesign International. Antibodies were biotinylated and fluoresceinated using a labeling kit which contained the N-hydroxysuccinimide esters of DNP-biotin and carboxyfluorescein. Labeling reagents were reconstituted with dimethylformamide (DMF). Samples were concentrated using Centricon-30 microconcentrators. All general laboratory reagents were of the highest quality. Venoms from *Heloderma horridum*, *Naja nigricollis*, *Agkistrodon acutus*, *Apis mellifera*, *Leiurus quinquestriatus*, *Bitis nasicornus*, *Naja atra*, *Crotalus viridis helleri*, and *Bufo americanus* were purchased from Sigma Chemical Co.

INSTRUMENTATION

All assays were performed on the LAPS purchased from Molecular Devices. Data were collected and stored digitally on a microcomputer.

Purification and Labeling of IgG:

IgG was purified from serum by passage over a Protein A column (Promab, JRH Biosciences, Lenexa, KS) following procedures described by the manufacturer. Antibodies were concentrated using a Centricon-30 microconcentrator (20,000 x g for 20 min) and washed 2 times with PBS until a volume of 1 ml remained. The final concentration of purified IgG was 1.36 mg/ml as determined by the Bradford method using BSA as a standard.[6]

Antibody preparations were separated into two aliquots. One aliquot was biotinylated (DNP biotin labeling reagent for 2 hr at room temperature using a molar coupling ratio, MCR, of 30:1). The remaining aliquot was fluoresceinated (fluorescein labeling reagent for 2 hr at room temperature using a MCR of 30:1). We have determined (data not shown) that a MCR of 30:1 provides adequate and efficient labeling of IgG antibodies. Unreacted haptens were removed by dialysis following six, 1 liter changes of PBS. Dialyzed samples were stored at 4°C.

Assay Procedure:

For each test, BoTX and venoms were dissolved in 100 μl assay buffer (Molecular Devices). Labeled antibodies were diluted in assay buffer to a final concentration of 2 ng/μl. Prior to assaying, biotinylated and fluoresceinated antibodies were combined in a 1:1 ratio such that each 100 μl contained 100 ng of each labeled antibody. We have determined this ratio and concentration of antibodies to be optimal for sensitive detection of BoTX (data not shown).

Primary immunocomplexes were formed by adding 100 μl antibody solution to the reaction mixture containing toxin. After gently vortexing, assay mixtures were incubated at 37°C for 45 min. Following incubation, 1 ml capture reagent (Molecular Devices) containing streptavidin was added to each reaction mixture followed by brief mixing. Each assay mixture was transferred immediately to separate test wells and vacuum filtered through a biotinylated nitrocellulose filter mounted on a test stick (Molecular Devices). Immunocomplexes were captured (immobilized) by formation of a nitrocellulose biotin-streptavidin-biotin conjugate. Unbound material was removed by adding 2 ml wash buffer (Molecular Devices) to each test well followed by vacuum filtration.

Urease conjugated, anti-fluorescein antibody solution (Molecular Devices) was added to each test well and filtered. A final wash with 2 ml wash buffer was performed after which time the nitrocellulose test sticks were removed and placed in a beaker containing wash buffer until read. Each nitrocellulose stick was inserted into the LAPS reader and respective immunoconjugates monitored via hydrolysis of urea.

RESULTS

Figure 1 shows is the effect of increasing amounts of BoTX on output signal. A linear response ($r = 0.98$) over the lowest range of toxin from 100 to 1000 pg was observed, (Frame A). A lower detection limit of 250 pg (78.7 ± 1.5 μV/s) is attainable under these conditions. Shown in Frame B is a dose response curve of increasing amounts of BoTX (1 - 30 ng) indicating saturation to occur at approximately 25 ng with a corresponding mean output signal of 4219.4 ± 301.4 μV/s. Shown in Frame C is the effect of increasing BoTX over an extended concentration range (10^{-10} - 10^{-6} g). A sharp decrease in output signal is observed when toxin is in excess of 30 ng. Shown in Figure 2 is the effect of various venoms on output signal and results indicate very low, assay nonspecificity (< 15% of control).

DISCUSSION

Active *Botulinum* toxin-B consists of a heavy (M_r approximately 100,000) and light chain (M_r approximately 50,000) linked by a disulfide bridge together with several *botulinum* associated proteins yielding a complex with a molecular weight of nearly 500,000 datlons.[7] Toxins produced from nonproteolytic cultures require exogenous protease treatment for protoxin activation.[8] However, in this study *Botulinum* toxin-B was not treated with exogenous protease. Previous studies in our laboratory indicated the toxin complex to be extensively nicked (5 pieces) which upon Western blotting revealed approximately 70% of the immunoreactivity to reside in 2 peptides (M_r = 37,480 and 50,470).[9]

Detection of picogram quantities of BoTX was achieved using labeled polyclonal IgG antibodies. Figure 1A indicates the lowest range of detection under these assay conditions to be between 100 and 250 pg. This amount of toxin approaches the minimal lethal dose in the mouse.[10,11] Although the output signals observed when detecting picogram quantities of BoTX are rather low (<200 μV/s), they are very reproducible as indicated by the fact that the standard error bars are not visibly distinct from the data points (Figure 1A).

Output signal produced by toxin in the ng range (1 - 30 ng, Figure 1B) indicated saturation to occur in the presence of approximately 25 ng toxin. BoTX in excess of 30 ng results in a sharp decrease in signal (Figure 1C). This is most likely due to dilution of antibody by excess toxin. It is essential that antibody concentration be in relatively large excess compared to that of toxin for reliable detection. Detection of toxin in excess of 30 ng is possible but larger amounts of labeled-antibody is required, resulting in a reduction of sensitivity.

It was important to determine the extent of biosensor nonspecificity. To accomplish this, the effect of nine different challenge venoms on signal output was determined. No attempt was made to isolate and purify individual proteins from the venoms for the purpose of exposing the assay to the maximum amount of toxic venom component(s). When compared to an identical assay using BoTX only (Control), output signals were all below 15% (< 600 μV/s) of control. It should be pointed out that the non-specific effect of various venoms was determined using added protein in the nanogram, saturating range. Although nonspecific signal output obtained from this experiment is greater than output values obtained using 100 - 1000 pg BoTX (Figure 1A), the comparison at low and high assay protein can not be made because of the dependence of nonspecificity upon added protein. Thus, when challenged with diverse and complex proteinacious mixtures (venoms), the very low output values suggest high assay specificity.

Results of this study indicate that picogram quantities of BoTX can be detected using the LAPS. Assay conditions were optimized giving the highest sensitivity possible while maintaining the largest useful range for quantitating toxin. Data described here indicate a useful range equivalent to two orders of magnitude (0.25 - 25 ng) was achieved. In conclusion, we describe an immunoligand assay using LAPS which detects picogram quantities of *Botulinum* toxin-B.

FIGURE LEGENDS

Figure 1: Detection of BoTX: Primary immunocomplexes were formed by incubation of labeled antibodies with picogram (Frame A) and nanogram (Frame B) quantities of BoTX as described under "Materials and Methods". Shown in Frame C is detection of BoTX over an extended range (10^{-10}-10^{-6} g) of toxin. Data points represent the mean ± SEM for 3 determinations. In some cases Standard Error bars are not distinguishable from data points.

Figure 2: Nonspecific Output Signal Produced by Nine Different Venoms. Assays were all performed in duplicate using 25 ng of respective venom protein. Mean output is presented as percentage of control (determined using 25 ng BoTX in the absence of added venom, output signal= 4200 μV/s). Assays: 1, Control (BoTX); 2, *Helloderma horridum*; 3, *Naja nigricollis nigricollis*; 4, *Agkistrodon acutus*; 5, *Apis mellifera*; 6, *Leiurus quinquestriatus*; 7, *Bitis nasicornus*; 8, *Naja naja atra*; 9, *Crotalus viridis helleri*; 10, *Bufo americanus*.

REFERENCES

1. B.R. DasGupta, and H. Sugiyama, Biochem. Biophys. Res. Commun. 48, 108 (1972).

2. A.S.V. Burgen, F. Dickens, and L.J. Eatman, J. Physiol. 109,10 (1949).

3. J. Briggs, Nature 329,565 (1987).

4. D. Hafeman, J. Parce, and H. McConnell, Science 240,1182 (1988).

5. J. Parce, J. Owicki, K. Kercso, G. Sigal, H. Wada, V. Muir, L. Bousse, K. Ross, B. Sikic, and H. McConnell, Science 246,243 (1989).

6. M.A. Bradford, Anal. Biochem. 72,248 (1976).

7. J. Wagman, and J.B. Bateman, Arch. Biochem. Biophys. 31,424 (1951).

8. B.R. DasGupta, J. Physiol. (Paris) 84,220 (1990).

9. P. Kumar, J.T. Colston, E.D. Rael, J.J. Valdes, and J.P. Chambers. In press, (1992).

10. J.P. Hanig, and C. Lamanna, J. Theor. Biol. 77,107 (1979).

11. E.J. Schantz, and E.A. Johnson, Microbiol. Rev. 56,80 (1992).

FIGURE 1: DETECTION OF BoTX

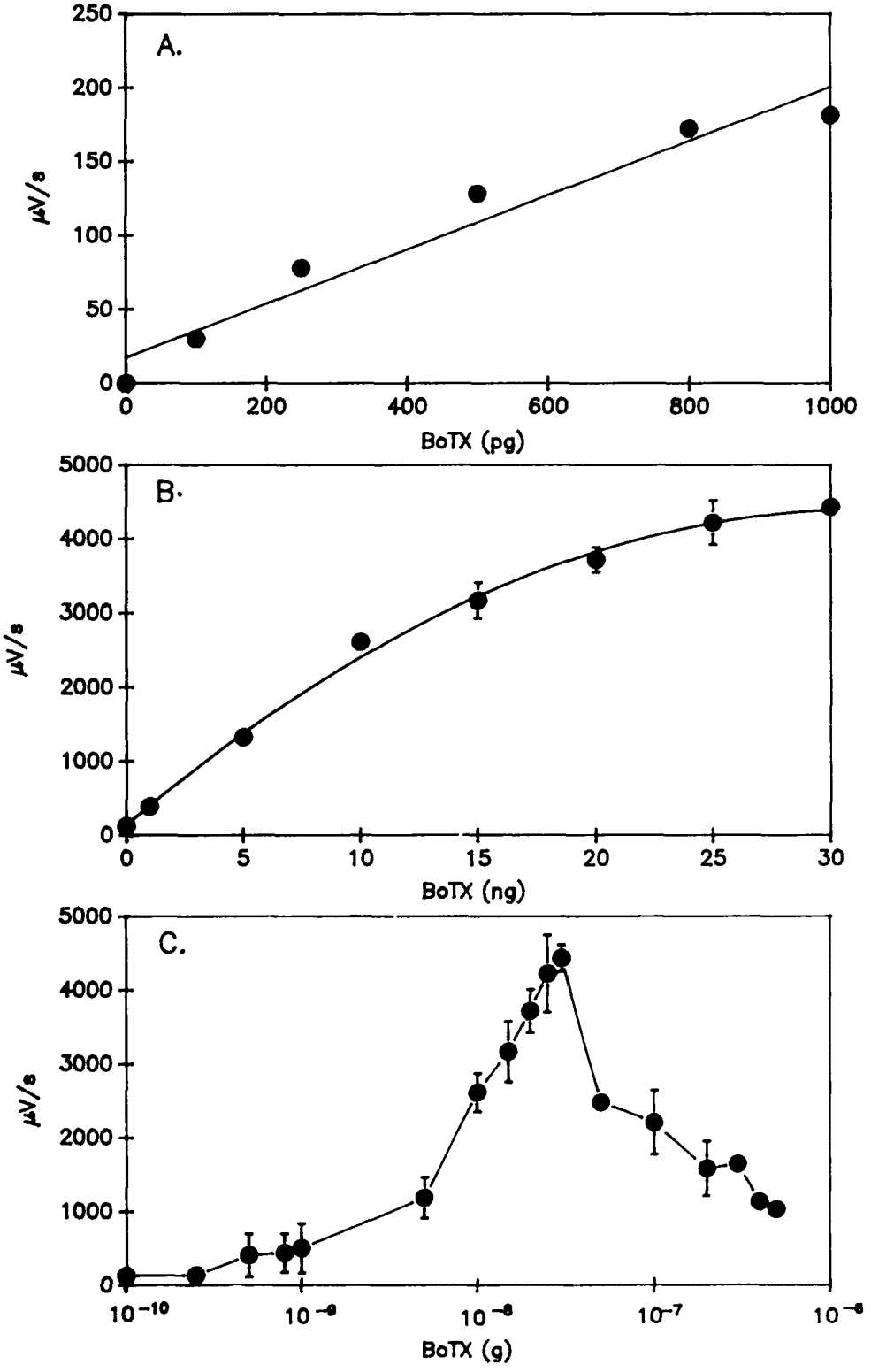
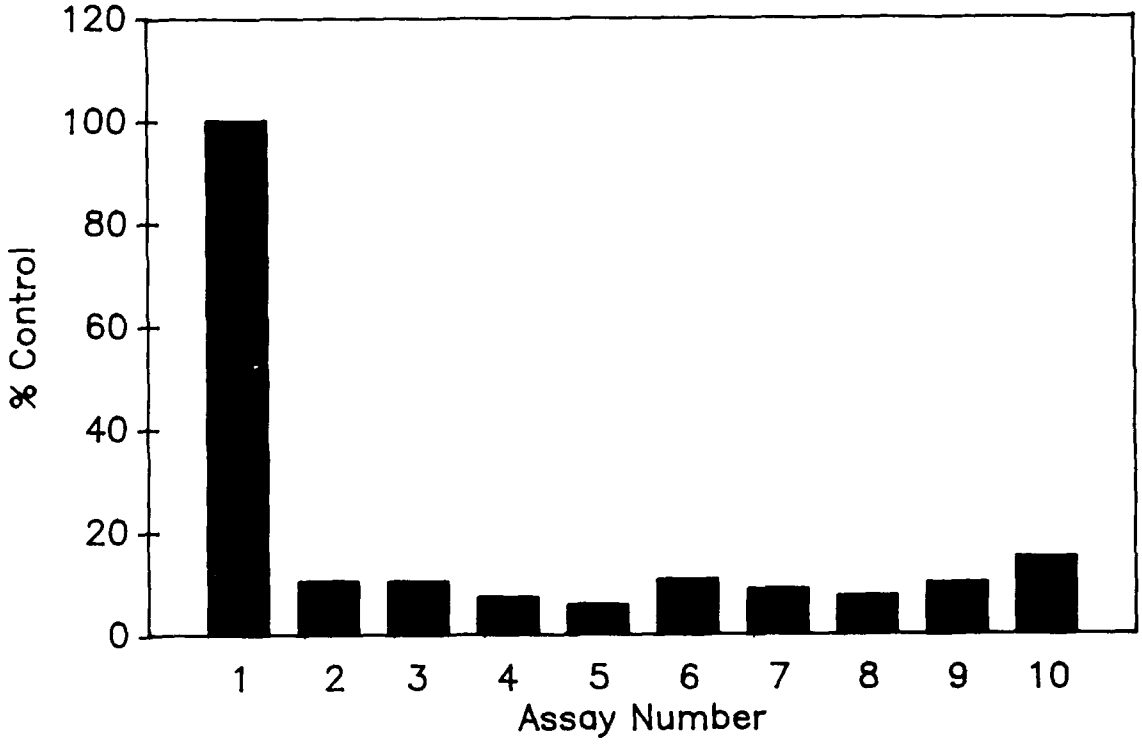


FIGURE 2: NONSPECIFIC OUTPUT SIGNAL PRODUCED BY NINE DIFFERENT VENOMS



Solid-State ^{31}P MAS NMR Study of G-Agent Simulants Adsorbed on Synthetic Resin Catalysts

George W. Wagner, Dr. (1), William T. Beaudry, Mr. (2)*
and J. Richard Ward, Dr. (2)
(1) NRC-CRDEC Research Associate, (2) Research Directorate
U.S. Army Chemical Research, Development & Engineering Center
Aberdeen Proving Ground, MD 21010-5423

INTRODUCTION

Characterizing chemical reactions on synthetic resins is of current interest in our laboratory, since such resins may be tailored to adsorb and react with toxic organophosphorus esters which inhibit acetylcholinesterase activity.[1] In particular, macroreticular resins possess a porous network which should enhance adsorption and provide sites to attach reactive functional groups.[2,3] The U.S. Army recently adopted a kit for skin decontamination using macroreticular resins[4] in lieu of natural clays such as "Fuller's earth." The resin employed, termed XE-555, is actually a mixture of three components: a high surface area, carbonaceous macroreticular styrene/divinylbenzene resin and two macroreticular styrene/divinylbenzene ion-exchange resins. One ion-exchange resin is acidic and the other is basic. The strong acid resin contains sulfonic acid groups, while the strong base resin has quaternary ammonium groups. In solution the toxic, acetylcholinesterase-inhibitors, such as sarin or soman, are rapidly hydrolyzed in strong base and also exhibit acid-catalyzed hydrolysis.[5]

In order for the XE-555 resin mixture to function catalytically, molecular exchange must exist between the sorptive and reactive sites and the reactive sites must be able to effect a heterogeneous reaction with the adsorbed liquid. Previous studies in our laboratory with O,O'-diisopropyl phosphorofluoridate (DFP) have suggested partitioning in the resin mixture.[6] Evidence for this partitioning is the apparent low reactivity observed for DFP adsorbed on the resin, despite the apparent high reactivity with the individual reactive resin components. The nature of this partitioning must be understood in order to design effective systems which both adsorb and react with liquid contaminants.

WAGNER, BEAUDRY,* WARD

This study employs ^{31}P Magic Angle Spinning (MAS) NMR to probe the adsorption and reaction of organophosphorus esters at the molecular level. ^{31}P MAS NMR is perhaps uniquely suited for this purpose, since it can distinguish various molecular adsorption sites[7-10] as well as determine the occurrence and rate of molecular exchange between these sites.[11-15]

EXPERIMENTAL

Dimethyl methylphosphonate (DMMP), purchased from Aldrich, was used without further purification. Para-nitrophenyl diphenyl phosphate (PNDP), provided by Dr. F. R. Longo, CRDEC, was found to be greater than 99% pure by ^{31}P NMR. Ambergard^{R} XE-555 Resin, Amberlite^{R} IRA-900 Strong Base Resin, Amberlyst^{R} XN-1010 Strong Acid Resin and Ambersorb^{R} XE-348F Carbonaceous Synthetic Adsorbent were obtained from Rohm & Haas Company. The latter three materials are similar to the components in the XE-555 resin mixture. The IRA-900, XN-1010 and XE-348F resins, manufactured in the form of small beads, were crushed with a ball mill to yield fine powders. The strong base IRA-900 resin, initially in the chloride form, was exchanged with an aqueous 5% NaOH solution to yield the hydroxide form. The resin was subsequently washed with deionized water to remove NaCl and excess NaOH and then filtered and dried in air at room temperature.

Solution ^{31}P NMR spectra were obtained at 81 MHz on a Varian XL-200 NMR spectrometer. ^{31}P MAS NMR spectra were obtained at 101 MHz using high-power proton decoupling on a home-built 6T FT-NMR spectrometer equipped with a Doty Scientific variable temperature 5mm double-tuned MAS probe. ^{31}P spin-lattice relaxation times (T_1) of less than 2 s are observed for adsorbed DMMP and PNDP which are greatly reduced from their solution values. The short T_1 values alleviate the need for the cross-polarization (CP) technique and, therefore, the more quantitative conventional FT experiment was used. All chemical shifts were referenced to an external sample of 85% H_3PO_4. Positive chemical shifts are at higher frequency (lower field) than the reference. Various reaction products were solvent extracted from the resins and their structures were determined using multinuclear and 2-dimensional ^{1}H-^{1}H COSY NMR techniques.

^{31}P MAS NMR magnetization-transfer (MT) experiments were performed at 81 MHz on the XL-200 using a Doty Scientific high-speed 7mm double-tuned MAS probe without proton decoupling. The MT spectra were recorded for DMMP vapor loaded on the XE-555 resin using the method described by Ford, et al.[13] A FORTRAN program, MT-NMR,[13] obtained from Dr. H. O. Spivey, Oklahoma State University, was used to fit the MT data to yield the rate constants for DMMP molecular exchange.

RESULTS FOR DIMETHYL METHYLPHOSPHONATE (DMMP)

The first phase of the study involves identifying the various adsorption sites on the XE-555 resin. DMMP is an ideal candidate for this purpose since it is a small, simple molecule which is stable except at very high or low pH. Therefore, the molecule is not expected to react with the acid or base groups on the resin. This surface probe permits NMR spectra to be obtained which are not complicated by interfering reaction products and allows the straightforward observation of chemically different adsorption sites.

^{31}P NMR spectra of DMMP vapor adsorbed on the XE-555 resin are shown in Figure 1. Spectrum 1(a) was obtained under normal solution NMR conditions and reveals two overlapping peaks at 38 and 29 ppm. The relative narrowness of these resonances immediately indicates that DMMP is undergoing fast isotropic "liquid-like" motions on the resin which averages the broad chemical

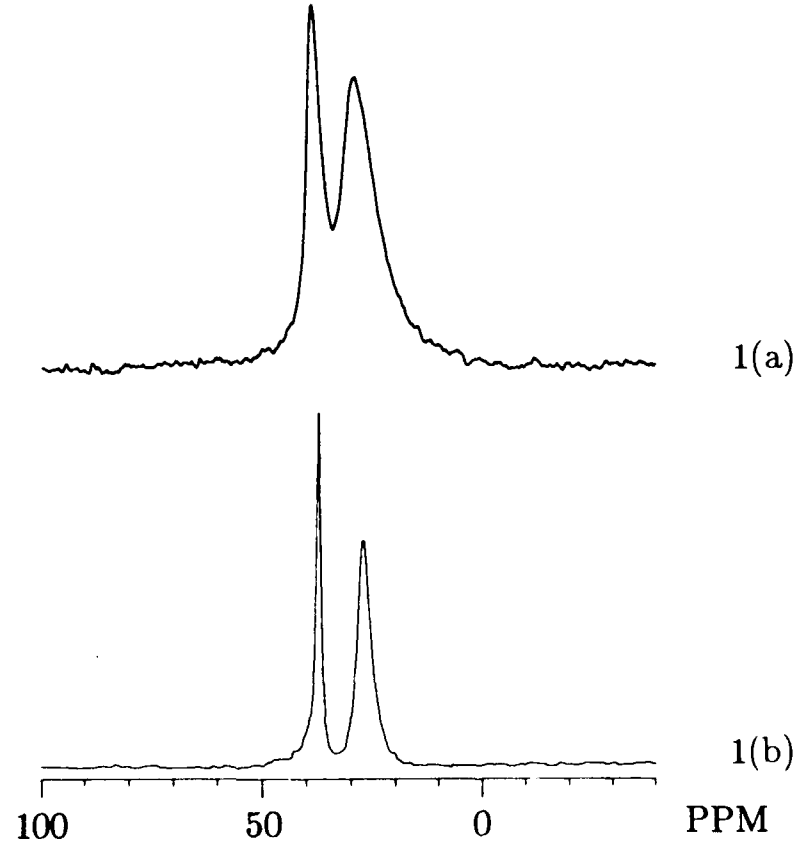


Figure 1. ^{31}P NMR spectra of DMMP vapor loaded on the XE-555 resin obtained using (a) normal solution NMR conditions and (b) MAS with high-power proton decoupling.

shift anisotropy (CSA) powder pattern expected for static organophosphorus esters.[16-19] Spectrum 1(b) was obtained using MAS with high-power proton decoupling. This technique greatly enhances the resolution of the twin peaks as compared to the spectrum obtained without MAS. The narrow downfield resonance occurs at 36.2 ppm which is the same value observed for DMMP in methanol solution. Methanol or 2N HCl (80:20 methanol/H_2O) extractions of the DMMP from the resin result in only the presence of DMMP in the filtrate; thus DMMP does not react with the resin but, in fact, resides in two different adsorption sites. The spectra further show that any molecular exchange occurring between the two sites is slow on the NMR time scale by virtue of the presence of both individual resonances.[20] The width of the peaks in the lower MAS spectrum (ca. 100 Hz) places an upper limit on the exchange rate of 300 s^{-1} at room temperature.

In order to identify the nature of the two adsorption sites, ^{31}P MAS NMR spectra were obtained for DMMP vapor adsorbed on the three individual resins similar to the components in the XE-555 resin. These spectra are shown in Figure 2. In the case of the carbonaceous macroreticular adsorbent, 2(a), a rather broad resonance is detected at 33 ppm (slightly shifted from the solution value of 36 ppm) along with CSA spinning sidebands indicating that the molecules are rather rigidly held. Conversely, DMMP remains highly mobile on the macroreticular strong base (2(b)) and strong acid (2(c)) ion-exchange resins yielding sharp resonances at 36.9 and 36.3 ppm, respectively. Only in the case of the strong base resin does a second signal occur at 25.5 ppm. As was seen for the XE-555 resin, extraction showed only DMMP in the sample. Therefore, the second signal is assigned to DMMP adsorbed at the quaternary ammonium hydroxide base sites rather than to a reaction product. A second signal is not detected for DMMP adsorbed on the strong acid resin, but this does not rule out the possibility of DMMP adsorption at sulfonic acid sites. This resonance may be coincident with the sorptive signal. A complete listing of ^{31}P MAS NMR assignments for DMMP adsorbed on the various resins is given in Table 1.

Table 1. ^{31}P MAS NMR Assignments for Organophosphorus Esters Adsorbed on Various Resins

Ester	XE-555	XE-348F	IRA-900(OH^-)	XN-1010
DMMP	36.2[a]	33[a]	36.9[a]	36.3[a]
	26.3[b]		25.5[b]	
PNDP	-17.9	-17.8	-17.1	-17.9
DPP	-10.1	not observed	-9.6	not observed
DPPA	not observed	not observed	not observed	-10.7
PNPPA	not observed	not observed	not observed	-11.7

[a]Macroreticular region. [b]Quaternary ammonium ion-exchange sites.

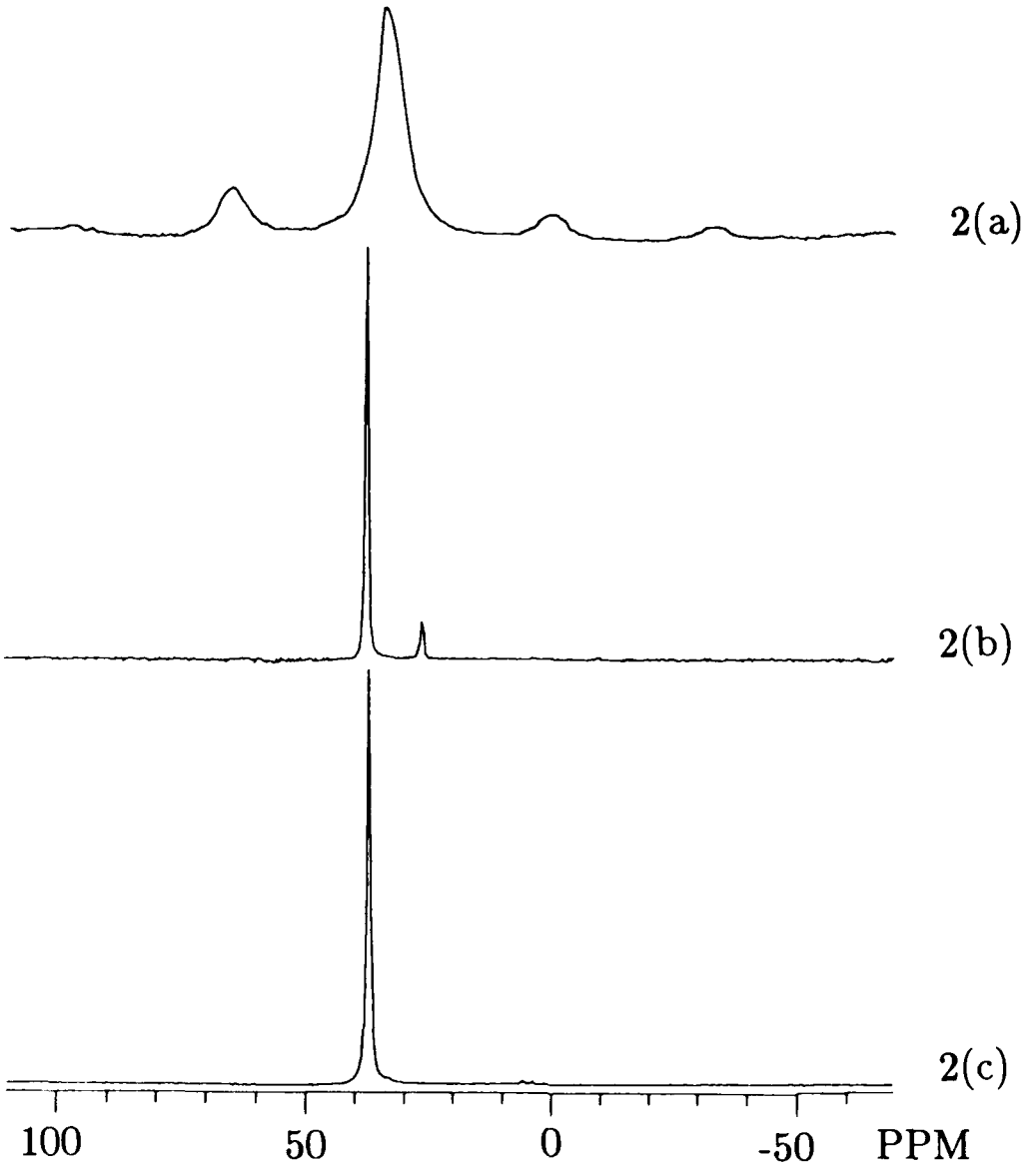


Figure 2. ^{31}P MAS NMR spectra of DMMP vapor loaded on
(a) carbonaceous macroreticular adsorbent, 15 weight %;
(b) macroreticular strong base resin, ca. 3 weight %;
(c) macroreticular strong acid resin, 44 weight %.

^{31}P MAS NMR spectra obtained from XE-555 samples with various DMMP weight % loadings (spectra not shown) indicate that adsorption at the base sites is favored during small weight % vapor loadings. Conversely, the macroreticular region is allowed to fill at large weight % liquid loadings. The relative ratio of the two resonances in these spectra did not change over a several month period indicating that the DMMP partition achieved during the loading process is stable at room temperature. The relative capacities of the two sites is in keeping with the expected greater capacity of the macroreticular adsorption sites, yet DMMP in the small weight % vapor loaded samples does not redistribute to maintain the same constant ratio between the sites.

A reasonable explanation for the adsorption behavior of the XE-555 resin is that the reactive sites facilitate the adsorption of DMMP vapor into the macroreticular region. This can be rationalized by the fact that DMMP is very soluble in water which is expected to be predominately adsorbed at the ion-exchange sites. The reported water contents of the strong base and acid resins are ca. 70% and 50%, respectively. It is interesting to note that in contrast to the macroreticular acid and base ion-exchange resins, DMMP vapor adsorbed on the carbonaceous macroreticular resin (which only contains about 2% water) is rather rigidly held. All three resins rapidly assimilate DMMP liquid to yield adsorbed DMMP which is highly mobile and liquid-like.

The possibility of chemical exchange between the macroreticular sorptive region and the reactive base sites was further investigated by variable temperature ^{31}P MAS NMR. Variable temperature spectra recorded up to 50 °C for DMMP adsorbed on both the XE-555 resin and the strong base ion-exchange resin (not shown) yield no features attributable to molecular exchange. On the contrary, at the higher temperatures the lines actually sharpen due to the increased rate of motion of the molecules within their respective adsorption environments. If molecular exchange were occurring between the sites at a rate approaching the width of the resonances (ca. 100 Hz), then a significant line broadening contribution would be evident as the temperature is raised. Therefore, molecular exchange is very slow (less than 300 s^{-1}) even at 50 °C. Site exchange might be observable by NMR above this temperature, however the quaternary ammonium hydroxide groups begin to decompose at ca. 60 °C. The variable temperature spectra were reversible within the temperature range studied.

Preliminary magnetization-transfer experiments[13,14] are consistent with DMMP exchange occurring between the sorptive region and the strong base sites at rates of ca. 0.5 s^{-1} at room temperature, although we are experiencing difficulty in obtaining exact fits of experimental data to a simple two-site exchange model. It is perhaps not surprising that our system is more complex given the very heterogeneous nature of the XE-555

resin. The observed exchange rate increases with increasing DMMP weight % loading in the 2 to 14% range initially examined. We are continuing these experiments in order to further characterize the exchange process.

RESULTS FOR P-NITROPHENYL DIPHENYL PHOSPHATE (PNDP)

Unlike the DMMP surface probe, PNDP is expected to be more unstable with respect to hydrolysis[21-31] and, thereby, to react with the acid and base groups in the XE-555 resin. In basic solution, the major hydrolysis products are diphenyl phosphate (DPP) and p-nitrophenoxide (a yellow chromophore), although significant amounts of p-nitrophenyl phenyl phosphate (PNPP) and phenoxide also form.[21,24] For acid catalysis, diphenyl phosphoric acid (DPPA) and p-nitrophenol are the sole reported products.[31] PNDP is a white solid at room temperature which necessitates the use of a solvent in order to load the compound onto the resin.

^{31}P MAS NMR spectra of PNDP adsorbed on the XE-555 resin from a 95% (v/v) ethanol/water solution are shown in Figure 3. Two peaks are present at -10.1 and -17.9 ppm in the spectrum of a 1 day old sample (3(a)). The large upfield peak at -17.9 ppm, which has the same chemical shift observed for PNDP in solution,[21] is assigned to PNDP adsorbed in the macroreticular region of the resin. The spectrum observed for adsorbed PNDP is different than the spectrum of neat PNDP solid[32] indicating that the ester does not merely precipitate as the ethanol evaporates, but is rather highly dispersed on the resin. The lack of spinning sidebands for the two peaks indicates that both adsorbed PNDP and the other adsorbed species are highly mobile and "liquid-like." The small downfield peak at -10.1 ppm is due to the hydrolysis product, DPP, as confirmed by extraction and analysis by solution NMR;[21] however, no PNPP product is detected. Thus, the observed DPP hydrolysis product demonstrates that the reactive sites are accessible to PNDP, although no distinct signal is observed for PNDP adsorbed at the acid/base sites.

Solvents play a major role in polymer-supported solution-phase catalysis,[2,34] and ordinarily the XE-555 resin would only contain adsorbed water. Therefore, the effect of coadsorbed solvent on the reaction was investigated in the following experiment. A PNDP/ethanol loaded sample was allowed to dry in air just to the point where no visible solvent remained, at which time a portion of the sample was removed and placed in a capped 5mm NMR tube. The rest of the sample remained exposed to air in order to allow further unrestricted loss of ethanol and equilibration with ambient moisture. The portions were allowed to stand for 23 days before their ^{31}P MAS NMR spectra were recorded. These spectra are also shown in Figure 3. For the capped sample (3(b)), the conversion of PNDP to DPP is nearly complete with only a small signal remaining for unreacted PNDP. Also, note that the signal for DPP lacks CSA spinning sidebands indicating that the

product is still very liquid-like. However, the spectrum of the air-exposed portion (3(c)) reveals a reduced conversion along with a DPP product which is being very rigidly held by the resin.

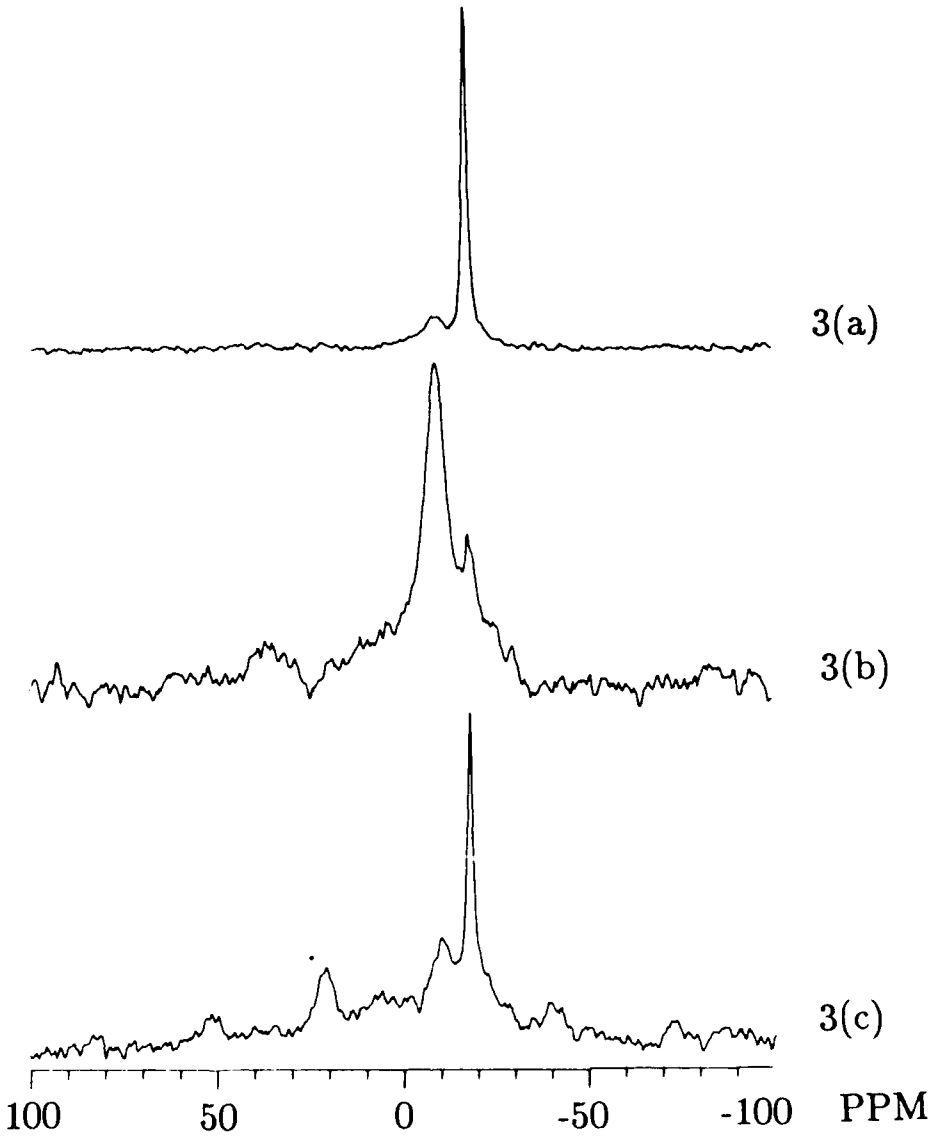


Figure 3. ^{31}P MAS NMR spectra of PNDP adsorbed on the XE-555 resin from ethanol solution. (a) 1 day old sample, capped to retain solvent. (b) 23 day old sample, capped to retain solvent. (c) Same as (b) except that the sample remained exposed to ambient humidity.

^{1}H MAS NMR indicates that no significant amount of highly mobile ethanol remains at the end of the drying period; however, less mobile ethanol would not be detected as a result of severe linebroadening due to ^{1}H-^{1}H dipolar interactions. The capped sample does contain considerably more mobile water than the air-exposed sample, but water is ruled out as capable of enhancing hydrolysis since partially dried samples re-wet with water do not yield increased PNDP conversions. Therefore, these results suggest that coadsorbed ethanol, although of limited mobility, is responsible for promoting PNDP hydrolysis, most likely by promoting PNDP migration rather than allowing release of DPP to regenerate "used" sites. This interpretation is supported by the fact that no turnover is observed when excess PNDP is allowed to react with the XE-555 resin in ethanol solution even though smaller amounts react completely within 4 days. Also, DPP product is not extracted by ethanol or water (acidification is necessary in order to dislodge the product from the resin).

To determine the reactivity of PNDP with the various components of the XE-555 resin, ^{31}P MAS spectra were also recorded for PNDP adsorbed on the carbonaceous macroreticular adsorbent and the strong acid and strong base macroreticular ion-exchange resins. These spectra are shown in Figure 4. PNDP adsorbed on the carbonaceous macroreticular adsorbent (4(a)) yields a single peak at -17.8 ppm, which confirms the previous assignment of this resonance to PNDP adsorbed in the macroreticular region. As expected, no reaction products are observed on this purely sorptive component. PNDP adsorbed on the strong base resin (4(b)) hydrolyzes in the same manner as on the XE-555 resin as indicated by a peak at -9.6 ppm, which is similarly assigned to the hydrolysis product DPP. In support of this assignment, the initially tan strong base resin rapidly becomes yellow after being slurried with the PNDP/ethanol solution, consistent with the release of the p-nitrophenoxide chromophore. The strong base resin further parallels the XE-555 resin behavior in that the spectrum of an air-exposed, 15 day old sample (4(c)) shows a poor conversion along with a large CSA for the immobilized DPP product, whereas the capped control sample (4(d)) yielded a much better conversion along with a highly mobile, liquid-like DPP product. Again, these findings are consistent with the DPP anion remaining bound to the quaternary ammonium ion-exchange sites, and enhanced PNDP hydrolysis due to its greater mobility in the presence of coadsorbed ethanol.

Spectra obtained for PNDP adsorbed on the strong acid resin, Figure 5, are quite different than the spectra obtained for both the XE-555 and strong base resins. Again, the peak at -17.9 ppm in the spectrum of a 1 day old sample (5(a)) is due to PNDP adsorbed in the macroreticular region. The peaks at -10.7 ppm and -11.7 ppm are assigned to the acid catalyzed hydrolysis products, diphenyl phosphoric acid (DPPA) and p-nitrophenyl phenyl phosphoric acid (PNPPA), based on NMR analysis of these species in solution. The spectrum taken after the sample had been kept for 40 days in

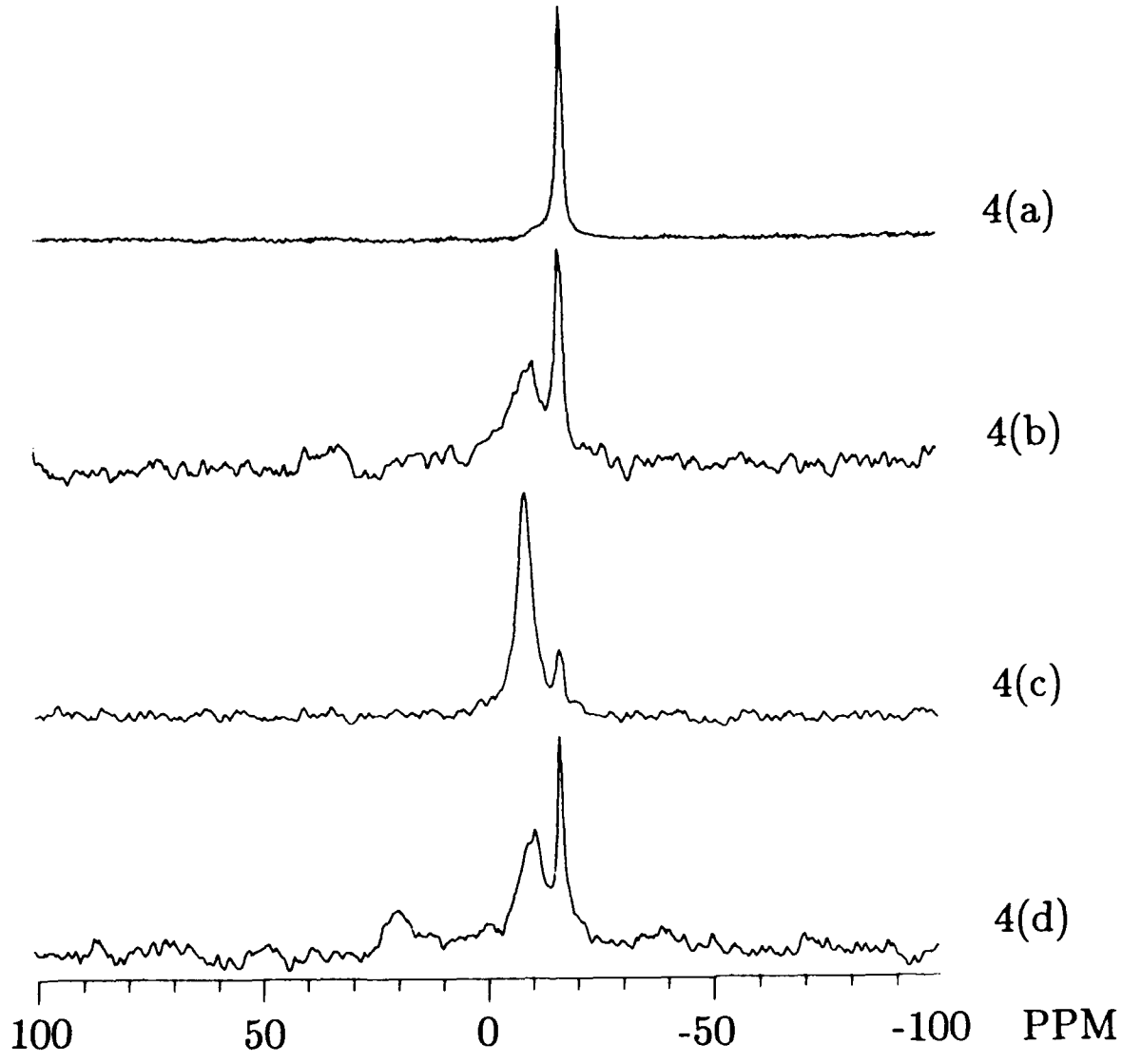


Figure 4. ^{31}P MAS NMR spectra of PNDP adsorbed on (a) carbonaceous macroreticular adsorbent, 1 day, capped to retain solvent; (b) macroreticular strong base resin, 1 day, capped to retain solvent; (c) macroreticular strong base resin, 15 days, capped to retain solvent; (d) macroreticular strong base resin, 15 days, exposed to ambient humidity.

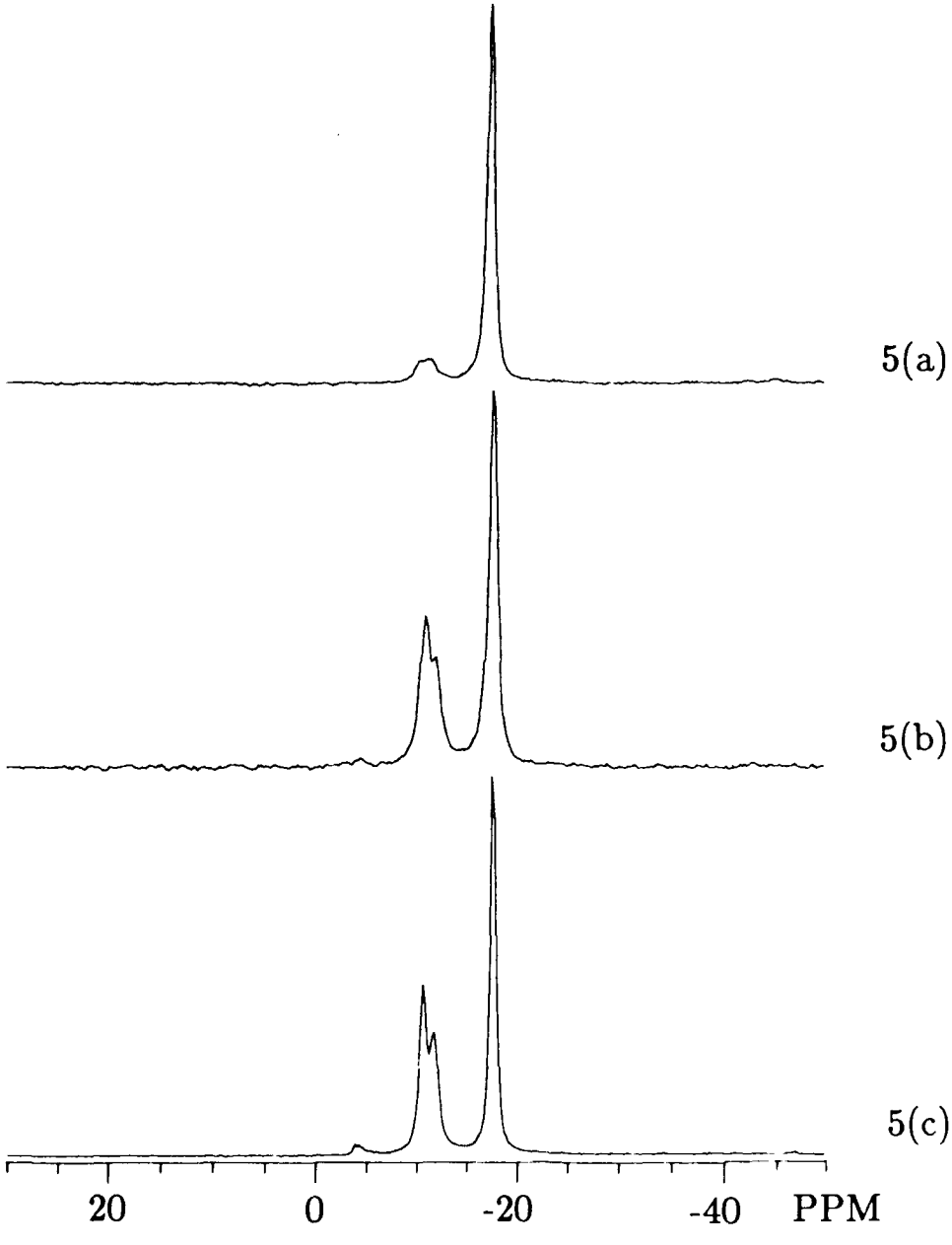


Figure 5. ^{31}P MAS NMR spectra of PNDP adsorbed on the macroreticular strong acid resin. (a) 1 day old, stored in a capped 5mm NMR tube. (b) 40 days old, stored in a capped 5mm NMR tube. (c) Sample exposed to air for 6 additional days.

a capped 5mm NMR tube (5(b)) shows further hydrolysis of PNDP with DPPA apparently being the favored product. Solvent extraction of the strong acid resin yields the same relative amounts of PNDP, DPPA and PNPPA in the filtrate indicating that the products are easily removed and not tightly bound to the sulfonic acid ion-exchange sites. Spectrum 5(c) was obtained from the same sample after it had been exposed to air for 6 days and shows further DPPA and PNPPA formation along with the appearance of a third peak at -4.4 ppm. This signal can probably be attributed to phenyl phosphoric acid formed from further hydrolysis of the PNPPA product,[21] although insufficient amounts were available for spectroscopic identification. Even after long-term air exposure, the neutral acid products on the strong acid resin remain highly mobile and liquid-like in the absence of coadsorbed ethanol, indicating an increased mobility of the neutral species as compared with the anionic base catalyzed products.

CONCLUSIONS

^{31}P MAS NMR spectra of DMMP adsorbed on the XE-555 resin provide direct evidence for the partitioning of adsorbed organophosphorus esters between two different adsorption sites. Comparison of spectra obtained for DMMP adsorbed on strong acid and strong base ion-exchange resins and a carbonaceous macroreticular adsorbent (the three components of the XE-555 resin) allow the assignment of the two adsorption sites to the macroreticular region and the quaternary ammonium hydroxide ion-exchange sites. No resonance is detected for DMMP adsorbed at sulfonic acid ion-exchange sites. Variable temperature spectra obtained for DMMP adsorbed on both the XE-555 resin and the strong base ion-exchange resin indicate that DMMP exchange between the macroreticular region and the quaternary ammonium sites cannot be occurring at a rate greater than 300 s^{-1}. In agreement with this limit, initial magnetization-transfer experiments indicate that the exchange rate for DMMP is ca. 0.5 s^{-1} at room temperature. These experiments are continuing in order to develop a better understanding of the exchange process.

The results obtained for the hydrolysis of PNDP adsorbed on the XE-555 resin (from ethanol solution) reveal that the anionic DPP product appears to be tightly bound to the quaternary ammonium ion-exchange sites rendering the reaction stoichiometric. PNDP hydrolysis is enhanced by the presence of coadsorbed ethanol consistent with an increased rate of diffusion to available reactive sites. Product immobilization does not occur for acid catalyzed hydrolysis on the strong acid resin where the neutral DPPA and PNPPA products remain highly mobile, even in the absence of coadsorbed ethanol. However, the analogous acid catalyzed process does not occur on the XE-555 resin. The reason for the apparent lack of reactivity of the XE-555 sulfonic acid groups remains unclear.

ACKNOWLEDGMENTS

We acknowledge Dr. D. Ralph Leslie (Defence Science and Technology Organization, Melbourne, Australia) for initiating this project as a visiting scientist in our laboratory. The ^{31}P MAS NMR experiments were conducted at the National Institutes of Health, Bethesda, MD, and we are grateful to Dr. D. Torchia for the use of the 6T NMR spectrometer and Dr. H. Cole for her helpful assistance with the operation of the instrument. This work was done while one of the authors (G.W.W.) held a National Research Council-CRDEC Research Associateship.

REFERENCES

1. T. C. Bruice and S. J. Benkovic, *Bioorganic Mechanisms*, Volume II (W. A. Benjamin, New York, 1966), Chapter 6, pp. 100-159.
2. A. Guyot, *Pure & Appl. Chem.*, **60**, 365 (1988).
3. D. C. Sherrington, *Pure & Appl. Chem.*, **60**, 401 (1988).
4. *Ambergard* ^{R} *XE-555 Resin Material Data Safety Sheet* (Rohm and Haas Company, 1989).
5. J. R. Cox, Jr. and O. B. Ramsay, *Chem. Rev.*, **64**, 317 (1964).
6. D. R. Leslie, W. T. Beaudry and L. L. Szafraniec, *Sorption and Reaction of Chemical Agents by a Mixed Sorptive/Reactive Resin*, CRDEC-TR-292, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, October 1991, *UNCLASSIFIED Report*.
7. L. Baltusis, J. S. Frye and G. E. Maciel, *J. Am. Chem. Soc.*, **109**, 40 (1987).
8. L. Baltusis, J. S. Frye and G. E. Maciel, *J. Am. Chem. Soc.*, **108**, 7119 (1986).
9. J. H. Lunsford, W. P. Rothwell and W. Shen, *J. Am. Chem. Soc.*, **107**, 1540 (1985).
10. W. P. Rothwell, W. X. Shen and J. H. Lunsford, *J. Am. Chem. Soc.*, **106**, 2452 (1984).
11. H.-H. Limbach, B. Wehrle, M. Schlabach, R. Kendrick and C. S. Yannoni, *J. Magn. Reson.*, **77**, 84 (1988).
12. A. Schmidt and S. Vega, *J. Chem. Phys.*, **87**, 6895 (1987).
13. W. T. Ford, M. Periyasamy, H. O. Spivey and J. P. Chandler, *J. Magn. Reson.*, **63**, 298 (1985).
14. W. T. Ford, M. Periyasamy and H. O. Spivey, *Macromolecules*, **17**, 2881 (1984).
15. D. Suwelack, W. P. Rothwell and J. S. Waugh, *J. Chem. Phys.*, **73**, 2559 (1980).
16. N. E. Burlinson, B. A. Dunell and J. A. Ripmeester, *J. Magn. Reson.*, **67**, 217 (1986).
17. T. M. Duncan and D. C. Douglass, *Chem. Phys.*, **87**, 339 (1984).
18. A.-R. Grimmer and U. Haubenreisser, *Chem. Phys. Lett.*, **99**, 487 (1983).

19. J. Herzfeld, R. G. Griffin and R. A. Haberkorn, *Biochem.*, **17**, 2711 (1978).
20. R. K. Harris, *Nuclear Magnetic Resonance Spectroscopy* (Pitman, London, 1983), Chapter 5.
21. B. A. Burnside, L. L. Szafraniec, B. L. Knier, H. D. Durst, R. A. Mackay and F. R. Longo, *J. Org. Chem.*, **53**, 2009 (1988).
22. B. L. Knier, H. D. Durst, B. A. Burnside, R. A. Mackay and F. R. Longo, *J. Solution Chem.*, **17**, 77 (1988).
23. R. A. Mackay, F. R. Longo, B. L. Knier and H. D. Durst, *J. Phys. Chem.*, **91**, 861 (1987).
24. S. H. Gellman, R. Petter and R. Breslow, *J. Am. Chem. Soc.*, **108**, 2388 (1986).
25. R. A. Moss, K. Y. Kim and S. Swarup, *J. Am. Chem. Soc.*, **108**, 788 (1986).
26. D. A. Jaeger and D. Bolikal, *J. Org. Chem.*, **50**, 4635 (1985).
27. D. R. Jones, L. F. Lindoy and A. M. Sargeson, *J. Am. Chem. Soc.*, **106**, 7807 (1984).
28. R. A. Moss, K. W. Alwis and J.-S. Shin, *J. Am. Chem. Soc.*, **106**, 2651 (1984).
29. C. A. Bunton, F. de Buzzaccarini and F. H. Hamed, *J. Org. Chem.*, **48**, 2457 (1983).
30. C. A. Bunton and L. Robinson, *J. Org. Chem.*, **34**, 733 (1969).
31. C. A. Bunton, S. J. Farber and E. J. Fendler, *J. Org. Chem.*, **33**, 29 (1968).
32. The room temperature ^{31}P MAS NMR spectrum of neat PNDP solid yields two resonances at -17.9 and -12.6 ppm which also display completely isotropic CSA's indicative of rapid isotropic molecular tumbling within the crystal lattice of the solid. The relative intensity of the two signals vary smoothly as the temperature is raised with the downfield resonance nearly disappearing just prior to the melting point at ca. 50 °C. As the temperature is lowered, a rather abrupt onset of an observable CSA occurs at ca. -45 °C where the two resonances are of equal intensity. This infers the presence of two equally populated crystal lattice sites in the rigid phase formed at this temperature. Similar behavior has recently been observed for crystalline phosphoric acid, although only a single resonance is detected in the low temperature rigid phase of this solid.[33]
33. A. J. Vila, C. M. Lagier, G. W. Wagner and A. C. Olivieri, *J. Chem. Soc., Chem. Commun.*, 683 (1991).
34. F. Svec, *Pure & Appl. Chem.*, **60**, 377 (1988).

A Taxonomy for the Vulnerability/Lethality Analysis Process

Dr. James N. Walbert*
Dr. Michael W. Starks
Dr. J. Terrence Klopcic
US Army Ballistic Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

1. Introduction[1]

An important aspect of the formulation of any scientific process is the definition of the mathematical framework within which that process is considered. This mathematical basis defines the limitations of the process, provides the means for conducting analyses, and helps ensure uniformity and consistency of results. It is quite common for analytical processes to evolve over a long period of time before the underlying mathematics is fully understood and illuminated. This is the case with Vulnerability/Lethality (V/L) analyses, long considered more an art than a science. In this section, the mathematical framework for V/L analyses is defined, demonstrating how each part of a well-known process fits into this framework, and identifying parts of this analytical process which are in error and/or require additional research. It will be shown that this taxonomy allows a rational scientific approach to the V/L analysis process.

2. Background

Traditionally, the V/L analysis process has been that of inferring some loss of effectiveness, or combat utility, from damage inflicted by a munition on a military system. The association of remaining utility with damage has been accomplished by a wide variety of means, from intuitive inference to empirical correlations to Monte Carlo simulations on computers. Among the numerous difficulties with this process is defining "effectiveness" or "utility," since these terms tend to be related to particular mission or combat

scenarios. Conclaves of experts in military science have been convened for the purpose of providing estimates of utility, generally expressed as a percent, given the loss of certain combinations of components or subsystems on a particular vehicle.[2] Such estimates, or Damage Assessment Lists (DALs), use intuitive inference to link component damage to loss of combat utility. These estimates represent a kind of average over all possible missions for the vehicle, and are therefore devoid of detail about specific system capabilities. The most common interpretation of these estimates (an incorrect one, as emphasized by Starks[3]) is as a probability of complete "kill," in either mobility or firepower or both.

In the late 1950s, a series of tests at the Canadian Armament Research and Development Establishment (CARDE) represented the first modern attempt at extensive collection of empirical data to relate damage to loss of function.[4] From these CARDE Trials came a number of correlation curves relating hole sizes in armor to loss of capability. Extensions of these curves are still used today, even though in many cases the combat systems to which they are applied bear little resemblance to those tested at CARDE. The unfortunate effects of this extrapolation are pervasive, even infecting computer codes written thirty years after the tests.

In the cases of aircraft and ships, although current analytical practices are different, many of the same shortcomings apply. For aircraft, vulnerability analyses have long included performance-oriented measures of effectiveness (MOE); examples of these include "Forced Landing", "Time-dependent Crash Landings", etc. However, along with additional mission-oriented MOEs such as "Mission Abort", aircraft V/L analyses traditionally suffer from similar logical disconnects between weapon effects and target response. In the case of ship analyses, several shortcomings apply. Therefore, although the language of this report is cast in terms of armored vehicles, the applicability of the concepts is universal.

Computer models which have evolved to assist in this analysis process are a reflection of the level of understanding the analysts have of the various physical and engineering phenomena involved. The lumped-parameter model known as the Compartment Model, for example, assumes each system consists of "black box" compartments such as ammunition, crew, and engine.[5] A perforation by a munition anywhere into one of these compartments results in a standard type of loss of function; that is, all components in the compartment are "lumped" into a single group for analysis. Another class of models, Point Burst, include more extensive component descriptions and attempt to distinguish between different shot lines by tracing the lines through a detailed target description.

In about 1985, the task of making pre-shot predictions for the Abrams tank live fire tests

underscored the widely-known fact that deterministic models fail to represent adequately the uncertainties of projectile impact attitude, armor and component fracture mechanics, spall production, fragment ricochet and numerous other factors involved in damaging a combat vehicle. Stochastic or Monte Carlo techniques were introduced in an attempt to provide more realistic estimates of damage to vehicles. For an historical perspective on vulnerability testing and modeling, the reader is directed to reference 6. In that paper, Deitz and Ozolins recognized the need to understand more rigorously the analytical processes and relationships by introducing the concept of spaces for V/L analysis. These spaces and the mappings between them have been used in a number of papers over the past several years.[7,8,9,10] Although there is no question that the notion of spaces has been heuristically useful, there have been both changes in usage and a lack of mathematical precision in the ongoing dialogue. It is thus important to make rigorous the complete taxonomy with its definitions, assumptions, and limitations. What follows is a description of the spaces and mappings, as well as a detailed discussion of the application of this taxonomy to V/L analyses.

It is anticipated that the framework presented in this paper will pervade much of the future work in vulnerability and lethality, both theoretical and empirical. In fact, the terminology discussed herein has already become part of the working vocabulary in the community.

3. Vulnerability within Survivability and Lethality

In the context of this paper, the term "vulnerability analysis" refers to the evaluation of the effects of a warhead on a target. This excludes consideration of such factors as acquisition, munition flight, etc. Similarly excluded are mission/scenario factors that describe the effect of the resulting damage upon combat utility. This relationship is illustrated in Figure 1.

4. V/L Spaces and Mappings

a. V/L Levels and Spaces The basis for the taxonomy of V/L Spaces comes from the recognition that V/L analyses pass through distinct levels of information in a precise order. These levels are:

1. Threat-Target Interaction - Initial Configurations (including Initial Conditions),

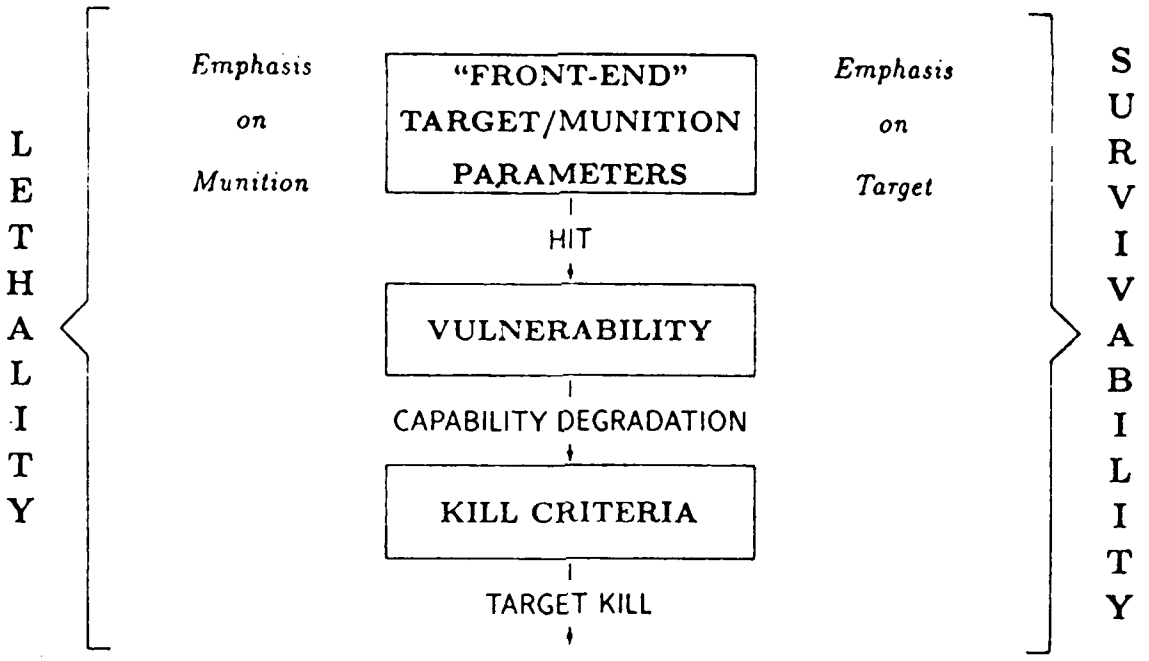


Figure 1: The Relationships Between Vulnerability, Lethality and Survivability

2. Target Component Damage States, and
3. Target Capability States.

From the Target Capability States can be derived the various mission-oriented losses of function such as "Firepower Kill" and "Mobility Kill".

The mappings by which one passes from one level to the next are dependent on different kinds of information at each level. For example, going from Level 1 to Level 2 (threat-target initial configuration to target damage) essentially involves physics; going from Level 2 to Level 3 (target damage to degraded capability) requires engineering measurement. The process can be shown pictorially as in Figure 2.

It is important at the outset to differentiate between "Levels", which are composed only of states of existence, and the "mappings", operators – with the data and algorithms to which they have access – which relate a state at one level to a state at another.

A *Level* contains all the information required to define the state of the system at the associated stage of a V/L analysis/experiment. At each level, one can define a space of

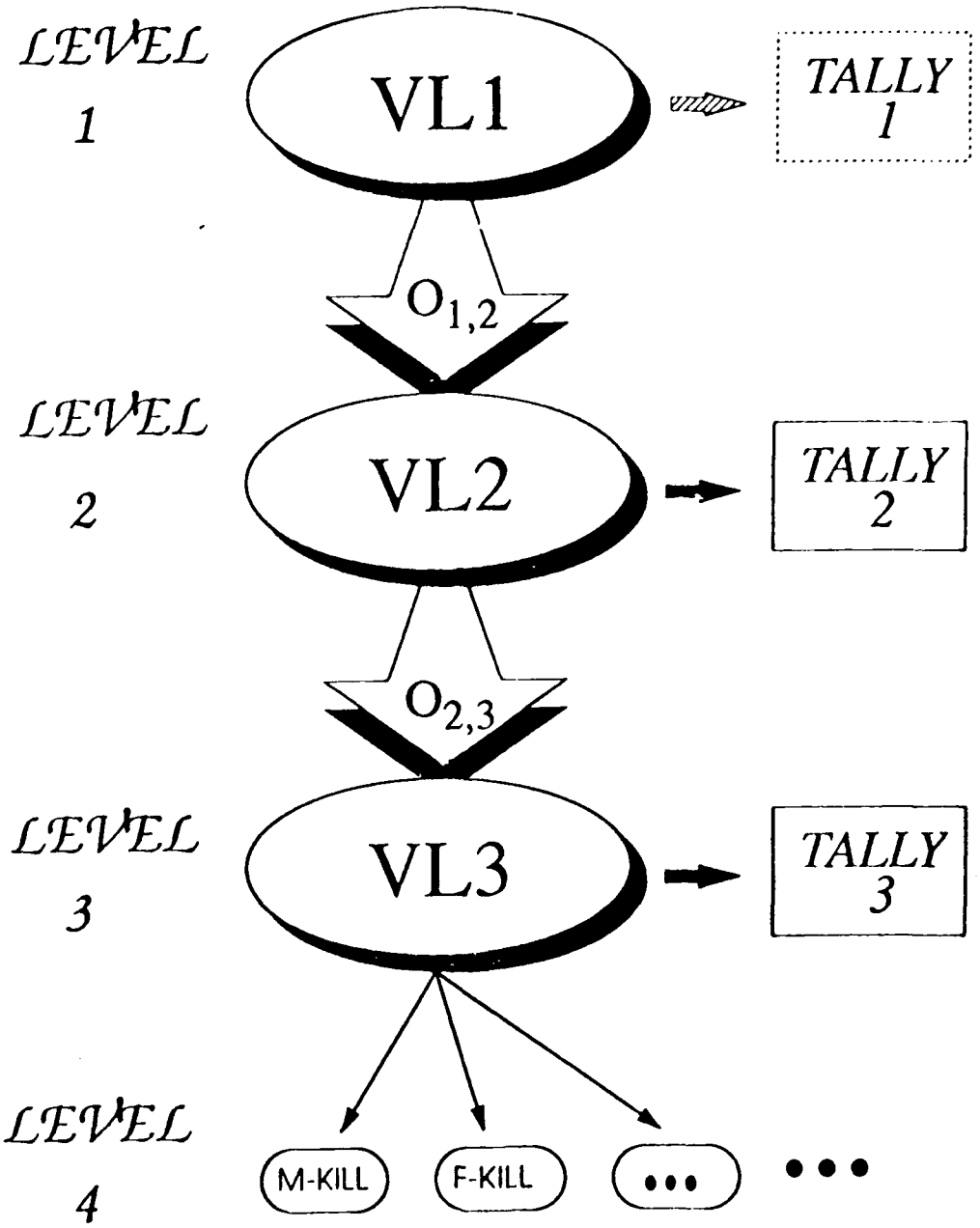


Figure 2: The Vulnerability Analysis Process

points, each point being a vector whose elements correspond to the status of a particular entity related to the target. For example, in Space 2 (Damage States), each element may refer to the status of a particular component/sub-system. The spaces thus defined are the "V/L Spaces", and represent, at each level, the state of the target system.

A *Mapping* represents all of the information (physics, engineering, etc.), known or unknown, required to associate a point in a space at one level with a point in a space at the next level. Mappings have access to information such as: fundamental data (penetration parameters [level 1 to level 2], leakage rates [level 2 to level 3], etc.); intermediate data generated by the mapping (line-of-sight thicknesses [1 to 2], temperature rise in an uncooled engine [2 to 3]); algorithms (depth of penetration [1 to 2], fault trees [2 to 3]).

The V/L experimental and analytical processes can then be expressed as a series of mappings which relate a state vector in one space (the domain) to a resultant state vector in a next higher-level space (the range).

Note that at each transition to the next level some detail about the target system may be lost: a broken bolt in level 2 may be the cause of degraded mobility influencing mission effectiveness, but at level 3 the bolt is no longer recognized as an entity. It is now widely acknowledged that skipping over levels (such as inferring remaining combat utility directly from the size of the hole in the armor) loses so significant an amount of information that continuity and auditability are lost.

b. Axioms and Definitions In order to provide a consistent structure to the topology about to be defined, the following axioms (assumptions) are given:

1. There are 3 levels of information making up the vulnerability analysis universe; spaces can be defined at each of these levels.
2. The points in each of the spaces are, in principle, observable and/or measurable.
3. The points in each space are vectors, consisting of one or more elements.
4. There exist mappings from each level to the next, specifically from a point in a space at each level to a corresponding point in a space at the next level.

With these axioms in place,

Definition 1:

1. V/L Space 1, or VL1, is a set of possible initial configurations for target/munition interaction.
2. V/L Space 2, or VL2, is the set of all possible damage vectors which can result from the initial configurations contained in VL1. The elements of the vectors in VL2 indicate the status of all critical components/subsystems.
3. V/L Space 3, or VL3, is the set of all possible system capability degradation vectors resulting from the damage states in VL2. The elements of the VL3 vectors indicate degrees of capability (for movement, communication, firepower - or, at a finer level of resolution, speed, acceleration, etc.)

Definition 2: The dimension of a space is the number of elements in a vector (point) in that space.

Definition 3: The cardinality of a space is the number of vectors (points) in that space.

Definition 4: The mapping from VL1 to VL2 is denoted by O12; similarly, the mapping from VL2 to VL3 is denoted by O23.

It is important to recognize that it is possible to construct many different spaces at any particular level. For example, note that the number of elements in a point (vector) in a space may depend upon the granularity of the target description to be used. This appears to be practically unavoidable: a human enumerates the different elements which will be evaluated in deciding what state exists after a single shot. Thus, there may be any number of spaces which could be created to describe the post-shot evaluation. Yet, they could all be "Spaces", as defined in this section. (That is, they can be closed, possess an identity element, be amenable to the defined operators, etc.)

A potentially fruittul area for future study is the relationship between different spaces at the same level that differ only in their degrees of granularity. This naturally leads to the concept of an "ultimate" space at each level. For example, consider a sequence of spaces of damage vectors (level 2), each succeeding space having more elements in its vectors. Since each element of a damage vector refers to the status of a particular part of the target, such a sequence could result from a progressively finer dissection of the target into successively smaller parts. The endpoint of this sequence is a construct whose discrete elements are replaced by continuously varying ones which detail the infinitesimal, point-by-point status of the target. Ideally, this "endpoint" will also form a space under

the same definitions that are listed herein, with the necessary replacement of discrete entities with continuous ones.

The relationship between this "ultimately fine space" and a space of coarser granularity may be of more than academic interest. For example, suppose a predictive result is expressed in terms of a set of damage vectors assembled into space VL2a. An experiment is independently conducted, with results expressed in terms of the damage vectors in VL2b. Consider the task of determining how close the prediction is to the experimental result. Since the spaces are different, comparison of the damage vectors, in a mathematical sense, may be unfounded. However, if both results can be related to their associated points in VL2* (the "ultimately fine space"), a comparison between the points can be made.

c. Relationships Between Spaces at Different Levels

i. Mappings Next, consider the mappings, the association of points from a space at one level with those in a space at another level. As described above, the points in the spaces are *determined by system design*, construction, and intended application; specifically excluded are the physical and engineering factors that relate points in one space - for example, a set of initial conditions - to points at the next level - for example, the resultant damage. Rather, such factors are incorporated in the mappings, either actually, if the mapping is accomplished by a field experiment, or algorithmically, if the mapping is accomplished by analysis or simulation. Analytical mappings are characterized by empirical or theoretical relationships such as penetration algorithms, fracture mechanics, etc., in the case of the mapping from VL1 to VL2. When going from VL2 to VL3, an analytical mapping may consist of a series of fault trees derived from an engineering model. Thus, in this taxonomy, knowledge gaps are quite clearly linked to the ability to construct a mapping from one space to another.

There may also be a certain variability inherent in the processes of penetration, fracture mechanics, and so on. If it exists, such variability would be a characteristic of the mapping function; that is, two applications of a mapping function to the same point in its domain could result in two different image points in its range.

ii. Repeated Mappings and Probability Distributions Consider the following procedure: Construct spaces at Levels 1 and 2 (VL1 and VL2) as described above. Also

construct a "scorecard" at Level 2 which allows one to count how many times each damage state point in VL2 is reached. Then select only one set of initial conditions (a fixed point in VL1) and iterate the mapping O12, counting the number of times each point in VL2 is reached. It is clear that, following a large number of mappings, the information in the scorecard provides an indication of the likelihood that a certain damage state point in VL2 will occur from a given set of threat-target initial conditions in VL1. In fact, it is a straightforward process to infer from the scorecard information a probability distribution associated with the mapping and the initial conditions. The boxes marked "TALLY" in Figure 2 represent such scorecards.

In principle, the process could be repeated for several sets of initial conditions. In this way, one can arrive at an understanding of the stochastic nature of the physics or engineering underlying the O12 and O23 mappings.

Once the spaces are defined at each level and the mappings (O12, O23) are known for a particular vulnerability problem, then the analysis process can proceed. Selecting a set of initial conditions for threat-target interaction, one applies the O12 mapping to determine a damage state vector in VL2. Using that damage state vector as the domain point, one then applies the O23 mapping to determine a loss-of-capability vector in VL3. By repeated application of the O12 mapping from the same initial conditions, one can infer the likelihood of occurrence of each of the damage state vectors. Similarly, by repeated application of the O23 mapping to the same damage state vector, one can infer the likelihood of occurrence of each loss of capability.

It is essential to appreciate two points:

1. These likelihoods, or probabilities, are functions of the mappings, and not of the spaces; if the mappings are changed, the probabilities which they associate with the vectors in the spaces will change.
2. The mappings have their domains and ranges in the V/L spaces, not in the sets of probabilities. (A forthcoming paper will make clear some of the problems which can result from not properly maintaining the distinction.)

iii. Non-invertibility It is also important to realize that the mappings O12 and O23 are not, in general, invertible. That is, given a capability state vector in VL3, it is not possible to determine which damage state vector in VL2 was mapped into it by O23. In fact, there will generally be numerous damage state vectors which could produce, under O23, a given capability state vector.

A capability state vector in VL3 could be (M=0,F=1,C=1), indicating full firepower and crew capability, but no mobility. Notice that this piece of information by itself reveals nothing about why there is no mobility. Since the information does not survive in the capability state vector, the mapping O23 is clearly not invertible. Stated another way, the O23 mapping is "many-to-one" (or "many-to-many").

Similarly, O12 is not invertible; it, too, is many-to-one or many-to-many.

The consequences of this non-invertibility can be significant, particularly impacting the development process for military equipment (the subject of a forthcoming paper).

d. Impact of Cardinality It has been implied by the previous paragraphs that a combination of testing and modeling can be used to characterize system performance. As was noted in the previous section, the cardinalities of VL1, VL2, and VL3 could be finite or infinite. If the cardinality of VL2 is finite, it may be possible (though very expensive) to examine, through testing and simulation, the full spectrum of images of VL2 in VL3 under the O23 mapping. If the cardinality of VL2 is infinite, this is simply not possible.

Similarly, if the cardinality of VL1 is infinite (which is quite certainly the case in the real world of continuously-varying coordinate systems, masses and velocities), then it is clearly impossible to analyze every aspect of O12. Thus, the best which can be hoped for is to identify a reasonable approximation to O12, and to O23 as well. This is where the introduction of stochastic simulation into the process can pay off handsomely.

5. Relevance of the Taxonomy to the Vulnerability Analysis Process

a. Initial "Set-up" of Problem Given a completely defined threat and target, how does one construct the appropriate V/L spaces for a given problem and determine and/or approximate the mappings between them? The first step in any analysis process is to determine the required precision. This requirement will dictate the level of detail required in computer target descriptions, the level of precision needed in test instrumentation and data reduction, the number of components identified as critical, and the level of performance capability testing to be done. (It seems reasonable that precision in the results of an analysis depends upon the level of detail in the target description. However, the quantification of this dependence is most difficult in practice.)

Having determined the required precision, the analyst quantifies and enumerates all rele-

b. Extraction of Results for End-Uses To this point, the discussion has been limited to the vulnerability analysis process the results of which are specific, quantifiable, observable measures. However, for various analytical purposes one must also be equipped to make more general assertions concerning the military utility of actual or hypothesized weapon systems. For example, to use vulnerability data in all its fine detail may be prohibitive. The remedy is to coalesce several data into a few numbers, weighting each datum based upon tactical/mission related probabilities.

Figure 3 provides a notional scheme which shows how such a statistical aggregation is formed. The battlefield utility of the target is broken into its potential mission components. Each component is further broken down until, ultimately, each branch ends on a specific target capability degradation (VL3 point). Since each branch has been weighted, it is then a straightforward task to "roll-up" an overall weighted loss of battlefield utility.

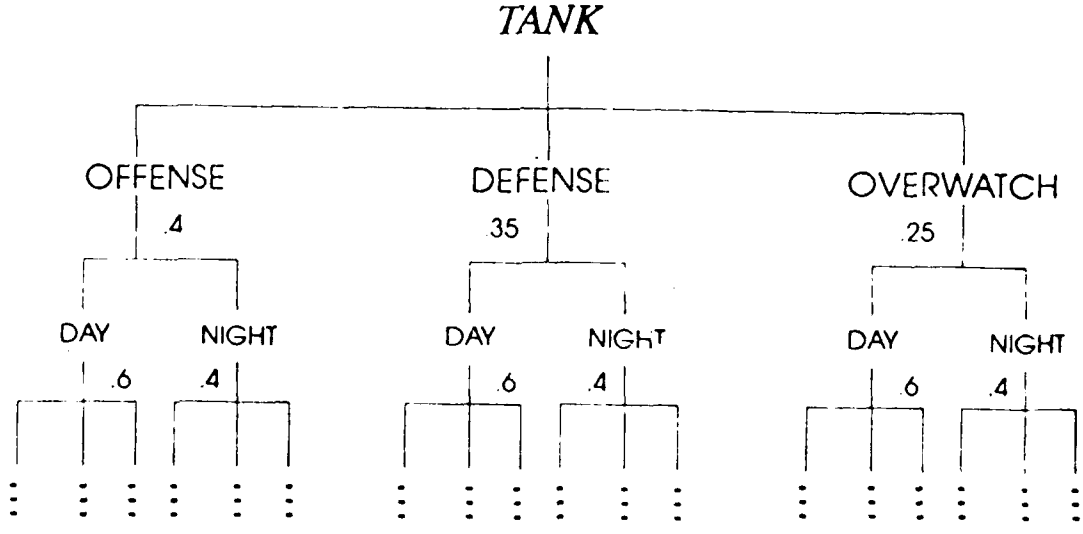


Figure 3: Aggregation to Level 4

6. Conclusions and Recommendations

A taxonomy has been developed for the vulnerability analysis process. It has been shown that this taxonomy represents an appropriate and internally consistent mathematical

b. Extraction of Results for End-Uses To this point, the discussion has been limited to the vulnerability analysis process the results of which are specific, quantifiable, observable measures. However, for various analytical purposes one must also be equipped to make more general assertions concerning the military utility of actual or hypothesized weapon systems. For example, to use vulnerability data in all its fine detail may be prohibitive. The remedy is to coalesce several data into a few numbers, weighting each datum based upon tactical/mission related probabilities.

Figure 3 provides a notional scheme which shows how such a statistical aggregation is formed. The battlefield utility of the target is broken into its potential mission components. Each component is further broken down until, ultimately, each branch ends on a specific target capability degradation (VL3 point). Since each branch has been weighted, it is then a straightforward task to "roll-up" an overall weighted loss of battlefield utility.

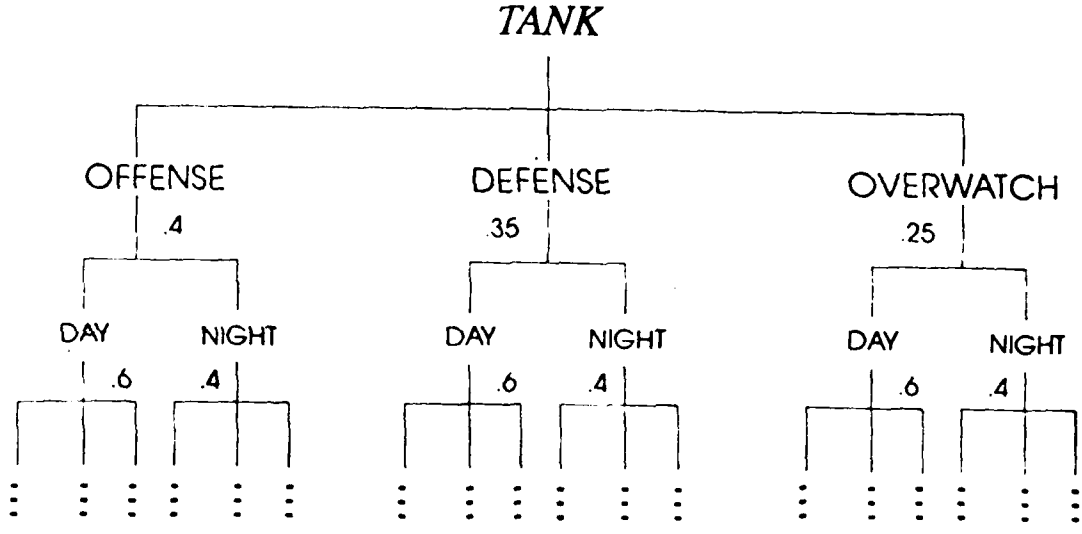


Figure 3: Aggregation to Level 4

6. Conclusions and Recommendations

A taxonomy has been developed for the vulnerability analysis process. It has been shown that this taxonomy represents an appropriate and internally consistent mathematical

foundation for vulnerability science, providing a framework for analytical processes and a means for identifying knowledge gaps.

The astute reader will have surmised that there is likely an algebra of the vectors in the spaces which can be defined, with a "norm" or distance, and a binary operator for combining two vectors. This algebra is being pursued, and is the subject of a report now in progress.

At this time it is impossible to predict the degree to which such a formal algebra, with rigorously defined entities, will be able to be fitted to the mathematical structures described above. However, if nothing else ever comes out of the exercise, the taxonomy and associated vocabulary that have been developed have already proven to be extremely useful to those who now routinely use it. The Levels of results, the meaning of a Space of state vectors, the constituents of a mapping - these concepts have considerably sharpened information and idea exchange in a wide range of vulnerability applications, including managerial and operational issues, as well as methodological developments.

References

1. J.T.Klopcic,M.W.Starks and J.N.Walbert, *A Taxonomy for the Vulnerability/Lethality Analysis Process*, BRL Memorandum Report, BRL-MR-3972, May 1992

2. G.A. Zeller and B.F. Armendt, *Volume X: Vulnerability Models; Part 1A: Update of the Standard Damage Assessment List for Tanks: Underlying Philosophy and Final Results.*, ASI Systems International, HQS Armament Division Report No. AD-TR-87-65, November, 1987

3. Michael W. Starks, *New Foundations for Tank Vulnerability Analysis,* **The Proceedings of the Tenth Annual Symposium on Survivability and Vulnerability of the American Defense Preparedness Association,** Naval Ocean Systems Center, San Diego, CA, 10-12 May 1988

4. *Tripartite Anti-Tank Trials and Lethality Evaluation, Part I,* **Canadian Armament Research and Development Establishment,** November 1959

5. Bradshaw F. Armendt, Jr., *Methods of Assessing Anti-Armor Weapons Lethality,* Working Paper 51 of Subpanel 3 of NATO AC/225, July 1974

6. P.H. Deitz and A. Ozolins, *Computer Simulations of the Abrams Live-Fire Field Testing,* **Proceedings of the XXVII Annual Meeting of the Army Operations Research Symposium,** 12-13 October 1988, Ft. Lee, VA; also, Ballistic Research Laboratory Memorandum Report BRL-MR-3755, May 1989

7. Michael W. Starks, *Assessing the Accuracy of Vulnerability Models by Comparison with Vulnerability Experiments,* Ballistic Research Laboratory Technical Report BRL-TR-3018, July 1989

8. P.H. Deitz, M.W. Starks,J.H. Smith and A. Ozolins, *Current Simulation Methods in Military Systems Vulnerability Assessment,* Ballistic Research Laboratory Memorandum Report BRL-MR-3880, November 1990

9. Michael W. Starks, *Improved Metrics for Personnel Vulnerability Analysis,* Ballistic Research Laboratory Memorandum Report BRL-MR-3908, May 1991

10. Paul H. Deitz *et al.*, Current Simulation Methods in Military Systems Vulnerability Assessment, **Proceedings of the XXIX Annual Meeting of the Army Operations Research Symposium,** October 1990, Ft. Lee, VA

Camouflage Flicker: A Highly Observable Low Observable

*Mr. Wendell R. Watkins, Mr. Samuel B. Crow, Mr. Fernando R. Palacios, and Mr. Daniel R. Billingsley
U.S. Army Atmospheric Sciences Laboratory; SLCAS-AM-T
White Sands Missile Range, New Mexico 88002-5501

INTRODUCTION

Nature has given the chameleon an excellent defense against its would-be predators by providing it with built-in visible camouflage. However, if the predators had thermal vision capabilities, the chameleon's skin would need a better design. Camouflage netting used to conceal the United States Army's vital assets needs the same care in its design too. Unfortunately, in camouflage design, the suite of sensors that can be used to detect targets spans the visible through the millimeter-wave spectral regions; yet, trying to blend anthropogenic objects into natural backgrounds is no easy chore. This paper highlights several issues in the area of target acquisition and camouflage design in the 8- to 12-μm or far-infrared (IR) regions that have not been adequately understood or addressed.

Among the issues that are atmospherically or meteorologically impacted is that of solar loading induced contrast changes. The energy budget that affects surface temperatures has a fairly consistent diurnal cycle, but rapid heating and cooling of vegetative backgrounds occur during partly cloudy conditions. This effect has been under investigation at the U. S. Army Atmospheric Sciences Laboratory (ASL) for several years with an excellent example (showing how dramatic target contrast changes due to solar loading variability can be) recorded at the Smoke Week 5B Test held at Fort Huachuca, AZ, in 1983.[1] In 3 min the average background radiance apparent temperature cooled 2 °C due to a cloud shading the target region. An uncamouflaged vehicle in this scene became very highly contrasted as a result. The radiance from camouflage, if used on this vehicle, would have to track these rapid changes in the background radiance level to be effective. As will be shown, the prevailing windspeed is an extremely important parameter to consider in assessing how well camouflage works.

To tackle this problem, the authors developed and perfected (over a 5-yr period) a new measurement tool that provides the proper method for characterizing and isolating the parameters that cause target or camouflage

contrast changes. This tool is called the Mobile Imaging Spectroscopy (MISL) with the imaging portion termed the Target Contrast Characterizer (TCC) and the spectral characterization performed by using a Fourier transform spectrometer (FTS).[2] The TCC of the MISL utilizes comparison of optically matched closeup or near-field (NF) images with engagement range or far-field (FF) images in three spectral bands (visible, 3 to 5 μm, and 8 to 12 μm). Comparison of NF and FF images allows scene changes to be separated into inherent and propagated components. The MISL has been recently used to measure image distortion due to optical turbulence in the far-IR for horizontal paths,[3,4] contrast transmission in the far-IR including the degrading effects of path radiance,[5] and almost by accident the phenomenon that will be referred to here as camouflage flicker.[6]

CAMOUFLAGE FLICKER

During 9 days of August 1990, the authors collected a set of MISL measurements on vehicular targets, target boards, and camouflage netting. Upon reviewing hundreds of hours of analog imagery data tapes in the fast forward mode, the authors noted that the camouflage netting at times would begin to act almost like a beacon. The netting would change between exhibiting more thermal contrast than an exposed tank and then disappearing into the background in rapid succession. Having seen the dramatic contrast changes that solar loading variability produced on backgrounds during the Smoke Week 5B Test, the authors isolated the times of partly cloudy conditions and viewed them in more detail. At times, the camouflage flicker was present; at others, it was not. Correlation between camouflage flicker and meteorological data was needed to unravel this mysterious behavior.

Plots of prevailing solar irradiance, windspeed, relative humidity, and temperature for the 9 days are shown in figure 1. These data were recorded at 5-min intervals (typical of the usual temporal frequency of meteorological data collection) and were sufficient to isolate the windspeed as the driving parameter behind camouflage flicker. The first 2 days of testing, which show a cloud-free rise and fall of solar irradiance during the daylight hours, had midday moderate-to-high windspeed (4 to 5 m/s) and low windspeed (1 to 2 m/s), respectively. During the course of these days, traces of the camouflage flicker phenomenon were present when the windspeed fluctuated rapidly; but, generally, the camouflage tracked the background very well. These days were hot and dry and were followed by a weekend rain shower on the fourth day. Here there is an overall increase in relative humidity and a corresponding decrease in temperatures. This is followed on days 5 and 6 by partly cloudy conditions as indicated by the rapid fluctuations in the solar irradiance plots. What is significant here is that Mother Nature was exceedingly cooperative. On the fifth day the midday windspeed was high (5 to 6 m/s) followed on the sixth day with low midday windspeeds (2 m/s), whereas the other meteorological conditions were nearly the same, as can be seen by comparing the daily meteorological plots in figures 2a and 2b. The midday far-IR imagery from these 2 days showed a

striking difference in camouflage effectiveness caused by solar loading changes coupled with the prevailing windspeed.

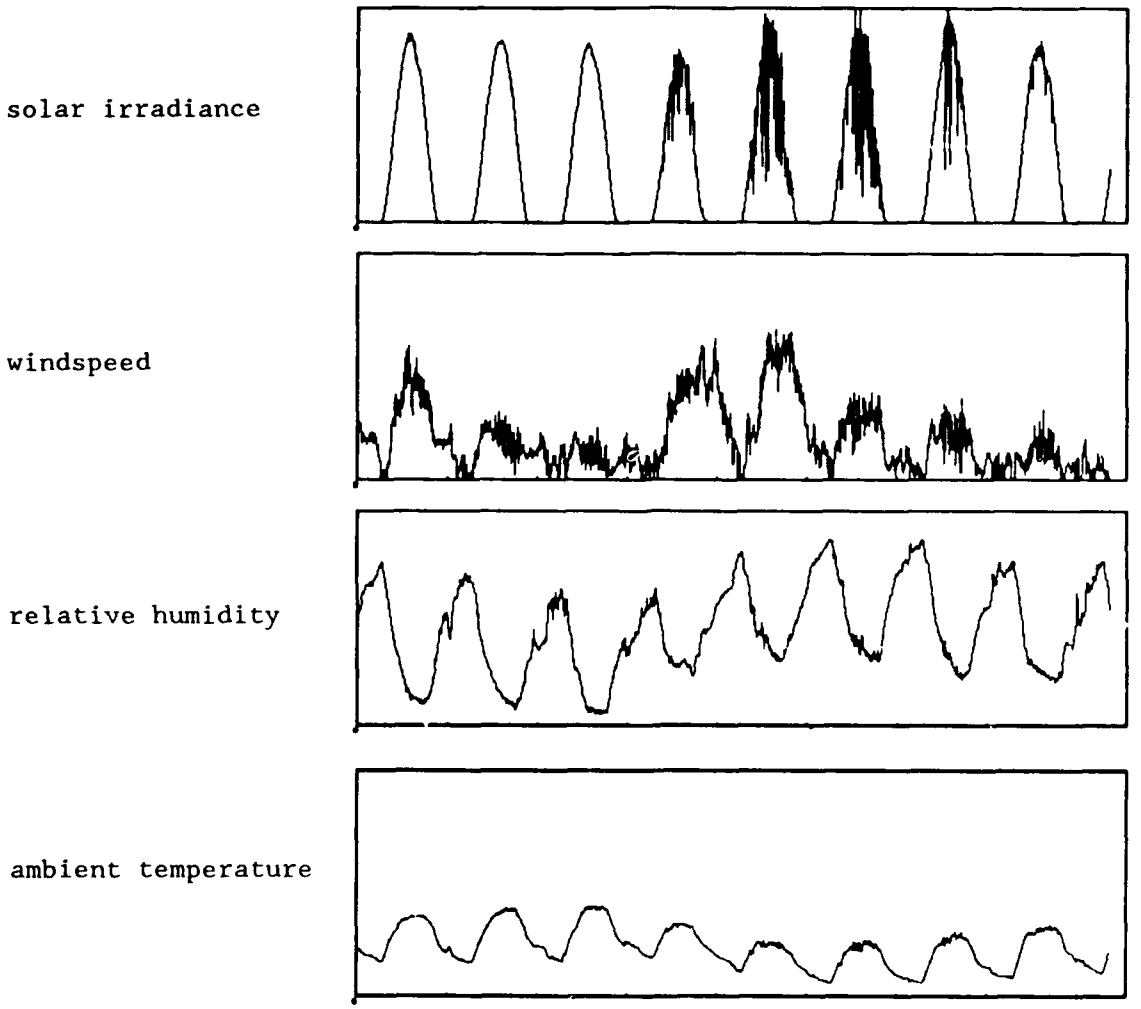
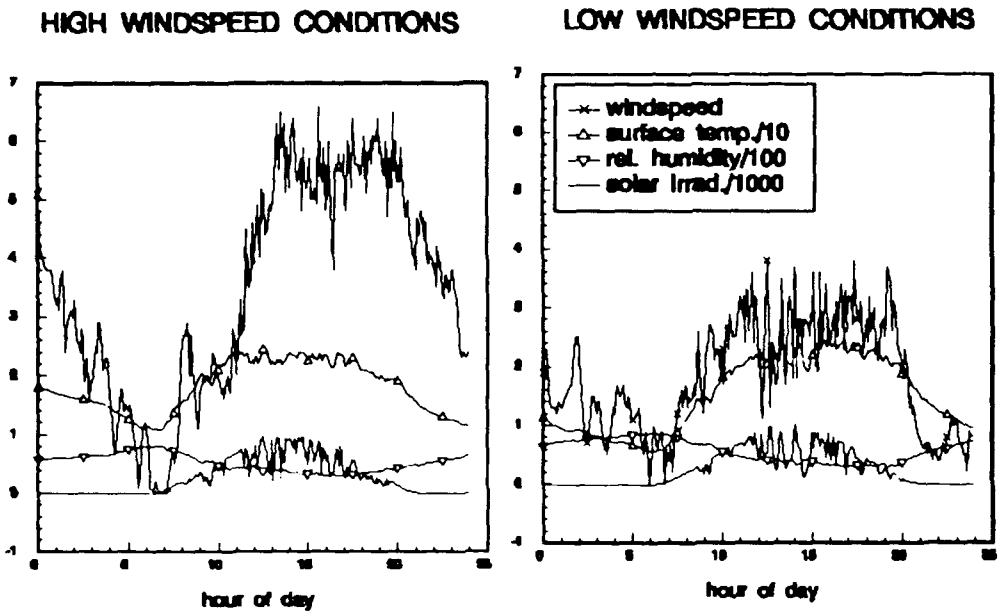


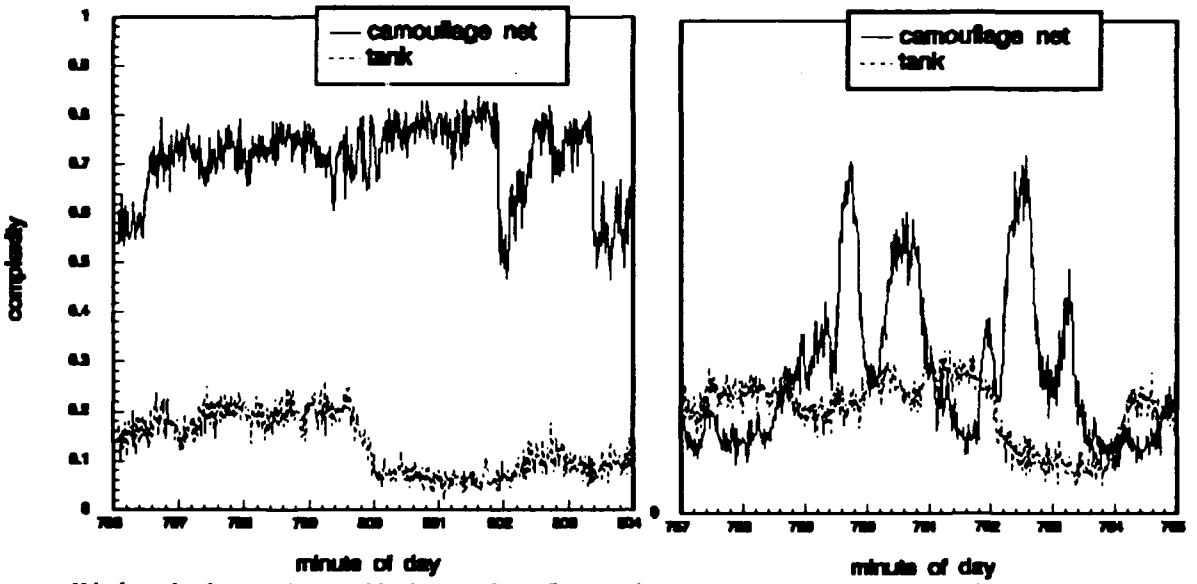
Figure 1. Eight-plus day plots of 5-min interval meteorological data. Top plot solar irradiance in watts/meter^2 from 0 to 1000. Second plot windspeed in meters/second from 0 to 10. Third plot relative humidity percent from 0 to 100. Last plot ambient temperature from 0 to 100 °C.



a. Fifth day with high windspeeds. b. Sixth day with low windspeeds.

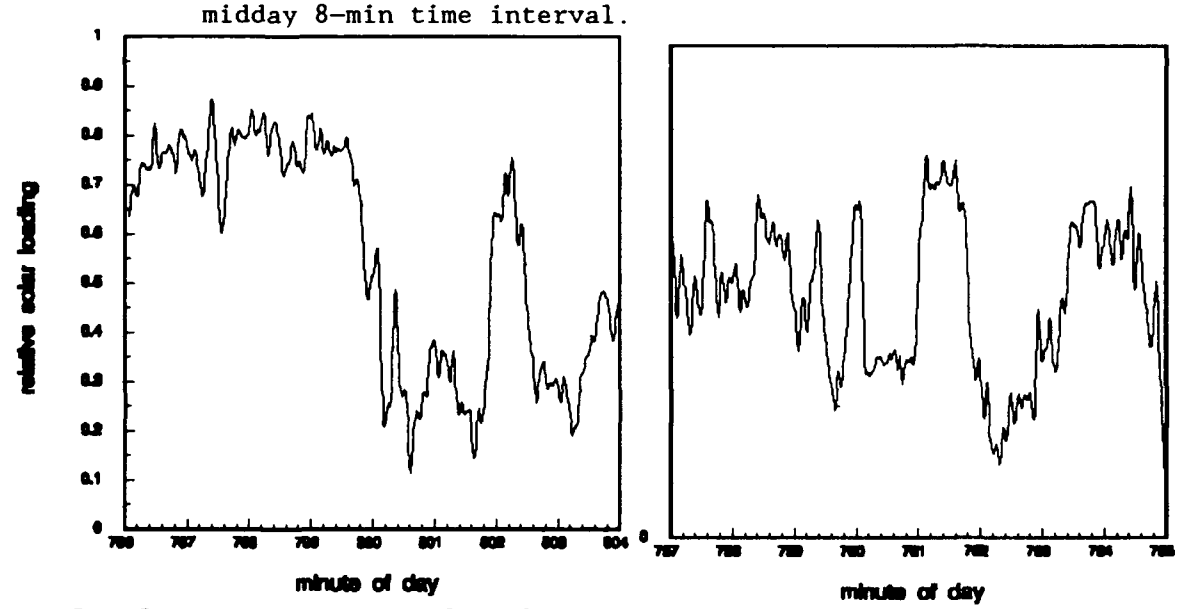
Figure 2. Composite meteorological plots from fig. 1.

The disparity in camouflage effectiveness between two comparable time intervals of 8 min duration on these 2 days in terms of camouflage netting and exposed tank contrast can be assessed in terms of a grayscale histogram metric called the "complexity measure."[7] The complexity measure is essentially the overlap of equal area target (camouflage netting/exposed tank) and background intensity distributions obtained from digitized analog images. If the target blends into the background, the intensity distributions are nearly equal and the complexity will be close to 1.0. If the target is highly contrasted, the distributions will be nearly disjoint and the complexity will be close to 0.0. Plots of complexity measure for the camouflage netting-to-background and exposed tank-to-background are shown in figures 3a and 3b. Trying to isolate the tracking of rapid solar loading changes by using 5-min interval meteorological data is not practical, as figures 2a and 2b illustrate for high-wind and low-wind cases, respectively. Hence, the complexity measure was applied to the visible imagery to determine the rapid changes in solar loading in the background of the camouflage net and tank by comparing it at 1-s intervals with a portion of the scene that was sunny as shown in figures 4a and 4b. These solar loading increases and decreases and, respectively, decreases and increases in the complexity measure of the camouflage net compared to the background for low windspeed conditions are shown in figures 3b and 4b.



a. High windspeed conditions day 5. b. Low windspeed conditions day 6.

Figure 3. Complexity measure plots of exposed tank to background (dashed curve) and camouflage net-to-background (solid curve) over a midday 8-min time interval.



a. Day 5—same time as in fig. 3a. b. Day 6—same time as in fig. 3b.

Figure 4. Solar loading plot of 1-s intervals derived from visible imagery complexity measure analysis for the same time intervals as fig. 3 where high values represent sunny and low values represent cloud conditions.

What is the reason for camouflage flicker? The vegetative background has cooling processes that the camouflage net does not. About one-third of the incident solar radiation is absorbed by chloroplast through the basically nonheating process of photosynthesis; and, additionally, evaporative cooling significantly reduces the solar loading heating process. Even so, leafy or grassy vegetation has low thermal mass and does change dramatically with solar loading changes, as observed at Smoke Week 5B. Considering these dynamics it is amazing that the camouflage netting, which attempts to match the background through the use of low emissive paints with only windspeed as a convective cooling process, works at all. Yet, the camouflage netting blends into the background quite well during high windspeed conditions in spite of the solar loading changes, as shown in figure 3a. Note the contrast changes of the exposed tank shown by complexity measure drops in figure 3a as the target area changed from sunny to shady. On the other hand, it is just as striking that a change from 5–6 m/s to 2 m/s can cause the camouflage netting to act like a beacon fluctuating between background radiance level to more than an exposed tank radiance level in a matter of seconds. As a direct result of these measurements, the MISL meteorological data acquisition instrumentation has been augmented to include 1-s interval data at the target area for signature change characterization.

MISL CHARACTERIZATION

At this point a diversion is necessary before resuming the discussion of the camouflage netting design and testing. Proper characterization of signature changes requires more than meteorological data. The MISL was designed to address many of these issues. Paramount to the success of the MISL is a unique large area (1.8 m by 1.8 m) target board (figure 5). Using this target board in conjunction with the TCC NF–FF comparison technique, the authors used the MISL to characterize optical turbulence degradation for IR imaging systems over horizontal paths.[3,4] Previously, optical turbulence was thought not to appreciably affect far-IR devices, but since Desert Storm a reassessment has been made.[8] Figure 6 shows 16 consecutive FF images from a 30-Hz Inframetrics, Incorporated, far-IR imager where the distortion of 3.5-cycle target board bar pattern is clearly visible for the 1.6-km horizontal path range. As will be addressed later, the MISL data can be used to characterize the temporal behavior of far-IR isoplanatic patches for active imaging use.

In addition to spatial degradation, the intensity of the scene radiation is altered through absorption, scattering, and path radiance. Spectral characterization is needed to measure these quantities, and the MISL has a fast Fourier transform spectrometer to accomplish this. The spectral response of the MISL imagers has been measured and special cold finger filters installed to reduce the effects of path radiance.[9] The MISL far-IR imagers were used to measure the net contrast reduction due to

propagation over a 1.6-km path by using NF and FF images of the large area target board and comparing the difference in radiance between this resolvable temperature and a uniform vegetative background.[5] The resulting values of contrast transmission agreed to about 2 percent with LOWTRAN7 predictions.[10] A typical example for a hot, dry day yielded a value of 0.80 compared to the LOWTRAN7 prediction of 0.82. If the 3.5-cycle bar pattern is used and the hot and cold bars that have 0.15 mrad spacing are compared, the temperatures are no longer resolvable at 1.6 km range. The corresponding contrast transmission value during daytime optical turbulence was 0.13, a factor of 6 lower than the actual intensity loss.

Figure 5. Visible image of the MISL large area target board with the 3.5-cycle vertical bar pattern installed.

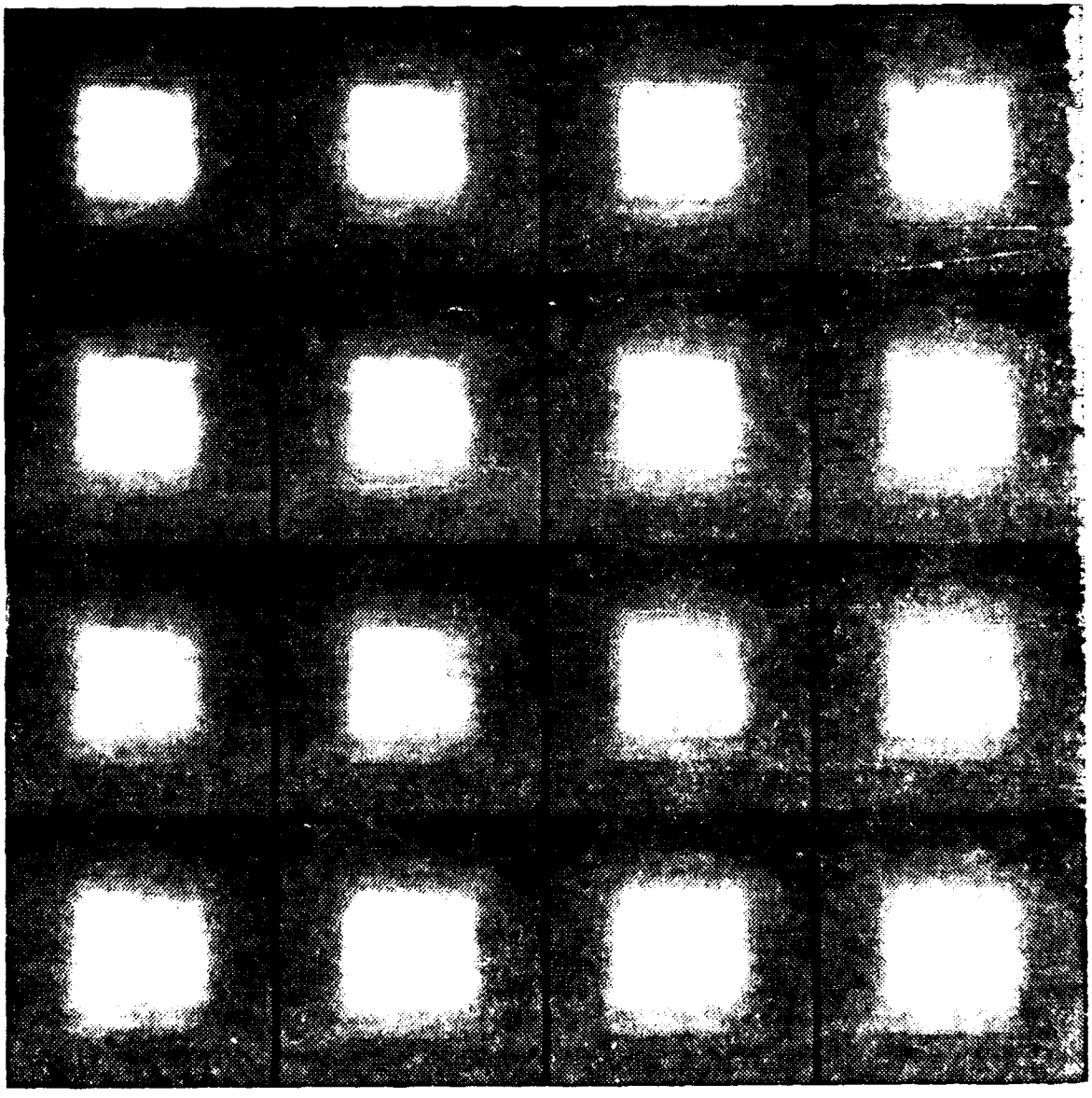


Figure 6. A set of 16 consecutive far-IR frames (proceeding left to right and top to bottom) from the TCC 30-Hz FF imager of the MISI large area target board from 1.6-km range during afternoon high optical turbulence. The distortions of the vertical bar pattern are easily distinguishable with individual distortion and isoplanatic patches persisting from frame to frame.

For the case where a resolvable temperature difference is present, the contrast transmission, $\tau_c(R)$, for a range, R, can be expressed as [11]

$$\tau_c(R) = \frac{1}{1 + \dfrac{\int_{\nu_1}^{\nu_2} L_p(\nu)\, F(\nu)\, d\nu}{\int_{\nu_1}^{\nu_2} L_b(R,\nu)\, \tau(R,\nu)\, F(\nu)\, d\nu}} \qquad (1)$$

where $L_p(\nu)$ is the path radiance at frequency, ν; L_b is the spectral background radiance from the target area range, R; $\tau(R,\nu)$ is the spectral atmospheric transmission over the range, R; and F(v) is the net imager spectral response including a spectral cold finger filter to reduce path radiance covering frequencies, ν_1 to ν_2. The MISL FTS is required for characterizing imager spectral response function, $F(\nu)$, and determining the optimizing spectral filter bandpass.[9] Now for many cases τ_c can be very closely approximated by simply ratioing the FF to NF temperature differences between the large area target board and the uniform background. This works because the propagated target and background radiances retain nearly the same spectral content as the blackbody radiance used to calibrate the thermal images. This is expressed for the target as follows:[11]

$$\frac{\int_{\nu_1}^{\nu_2} {}_{ap}L_t(R,\nu)\, F(\nu)\, d\nu}{\int_{\nu_1}^{\nu_2} F(\nu)\, d\nu} = \frac{\int_{\nu_1}^{\nu_2} L_t(0,\nu)\, \tau(R,\nu)\, F(\nu)\, d\nu}{\int_{\nu_1}^{\nu_2} F(\nu)\, d\nu} + \frac{\int_{\nu_1}^{\nu_2} {}_{ac}L_{amb}(\nu)\, [1-\tau(R,\nu)]\, F(\nu)\, d\nu}{\int_{\nu_1}^{\nu_2} F(\nu)\, d\nu} \qquad (2)$$

where ${}_{ap}L_t(R,\nu)$ is the propagated apparent radiance, $L_t(0,\nu)$ is the actual inherent target radiance, and ${}_{ac}L_{amb}(\nu)$ is the actual ambient temperature path radiance. Figure 7 shows a comparison between the difference in estimated (FF imager measured) propagated apparent temperature spectral radiance (left side of equation (2)) and the actual propagated apparent spectral radiance (from actual target and ambient temperature radiances) for the imager (right side of equation (2)) for dry conditions with a 311 K target and humid conditions with a 318 K target over a 1.6-km path. The resulting apparent temperatures for the resulting radiances were identical for both the dry and wet cases. An example will be shown later that will illustrate how effective the MISL spectral response and filter optimization are.

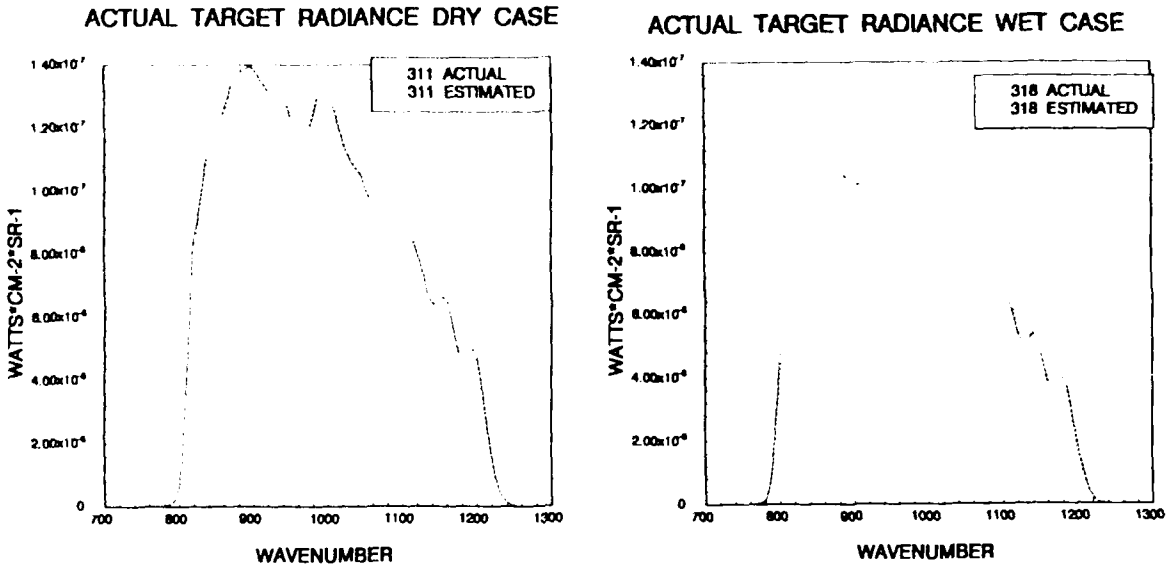


Figure 7. Comparison of actual and imager blackbody estimated propagated target board spectral radiance for dry and humid conditions.

ACTIVE IMAGING

Thus far only the emissive component of camouflage netting has been considered. The low emissive paints that are used for vehicular surfaces and camouflage nets can be applied in textured patterns to better blend into backgrounds, but there is another consideration. To obtain a low emission (and absorption) coefficient, these opaque surfaces that do not exhibit a photosynthesis process must have a high reflection coefficient because the sum of the emission and reflection coefficients equals unity. How does this reflective surface affect target contrast? During the day the reflection of cold sky radiation in the far-IR helps to reduce the overall radiance from the camouflage. At sunrise and sunset, however, there is a problem of solar glint from the side surfaces of the net. Also at night, especially when no clouds are present, the top surface of the camouflage net reflects the cold sky (typically -40 °C) and is easily distinguished from vegetative backgrounds that do not have highly reflective properties since they are naturally designed to be highly absorbing. The day-to-night differences for the camouflage net can be seen in figure 8.

Now consider what will happen if a 2,273 K flare or other explosive source is detonated, especially an airburst, at a range more distant than the camouflage netting. If the highly reflective top surface of the camouflage net, during a hot day with ambient temperature of 38 °C (311 K), reflects effectively 10 percent of the flare temperature radiance level, then the net will appear to the imager to have a blackbody temperature of

654 K. The spectral composition of the actual radiance from the net is quite different from the imager inferred 654 K blackbody being skewed to higher frequency or lower wavelength where the 2,273 K flare radiance peaks. How does this impact the temperature difference approach to contrast transmission expressed in equation (2) and figure 7? The two spectral radiances (the effective blackbody and the actual flare plus ambient) are put in equation (2) as the $L_t(0,\nu)$ and the apparent temperature spectral temperature spectral radiance calculated for each for both the dry and humid conditions considered in figure 7 for a 1.6-km path. These radiance curves are shown in figure 9. The apparent temperatures derived by the far-IR imager would be 576.4K and 518.2 K for the actual source radiance for dry and humid conditions and 576.7 K and 518.5 K for the effective blackbody source radiance, respectively, for dry and humid conditions. These differences were significantly reduced by the MISL imager spectral response and filter function, $F(\nu)$, optimization that is tailored to reduce the effect of path radiance. But not to forget the original goal of this example, in either case the highly reflective surface of the camouflage net can be exploited in an active source mode to light up its location like a Christmas tree in the far- as well as mid-IR spectral regions.

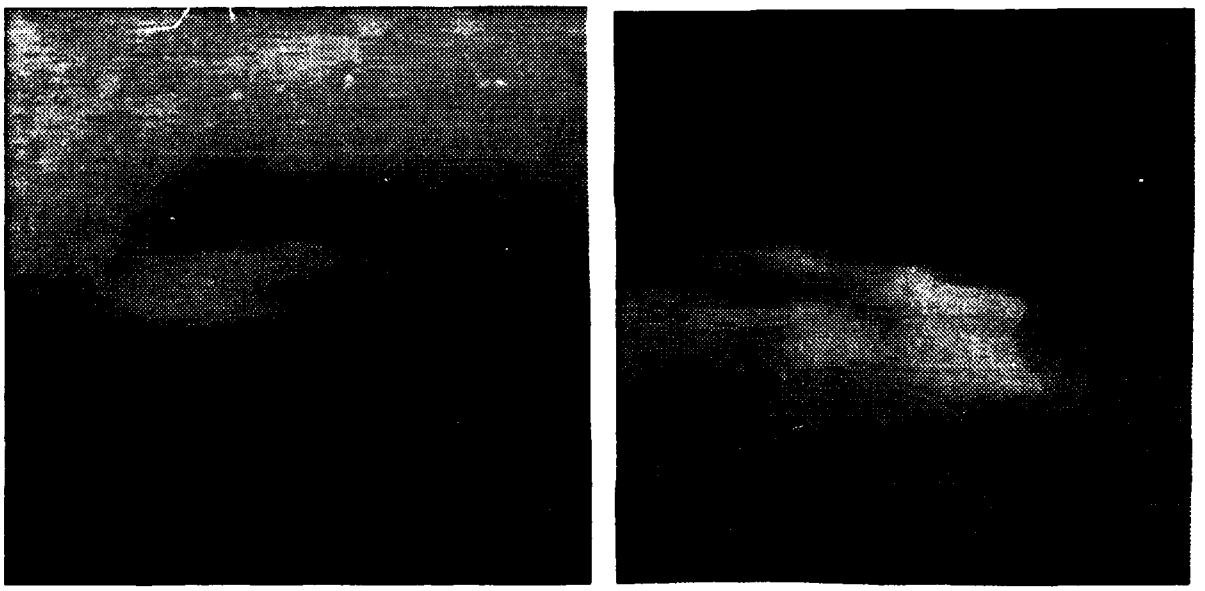


Figure 8. Daytime versus nighttime far-IR thermal image regions taken with an Inframetrics, Incorporated, model 610 imager with a 17-power lens at a range of 1.6 km. The cold sky reflections from the top of the camouflage net can be easily seen in the nighttime scene. On the left is a nighttime scene and on the right is a daytime scene of the camouflage net.

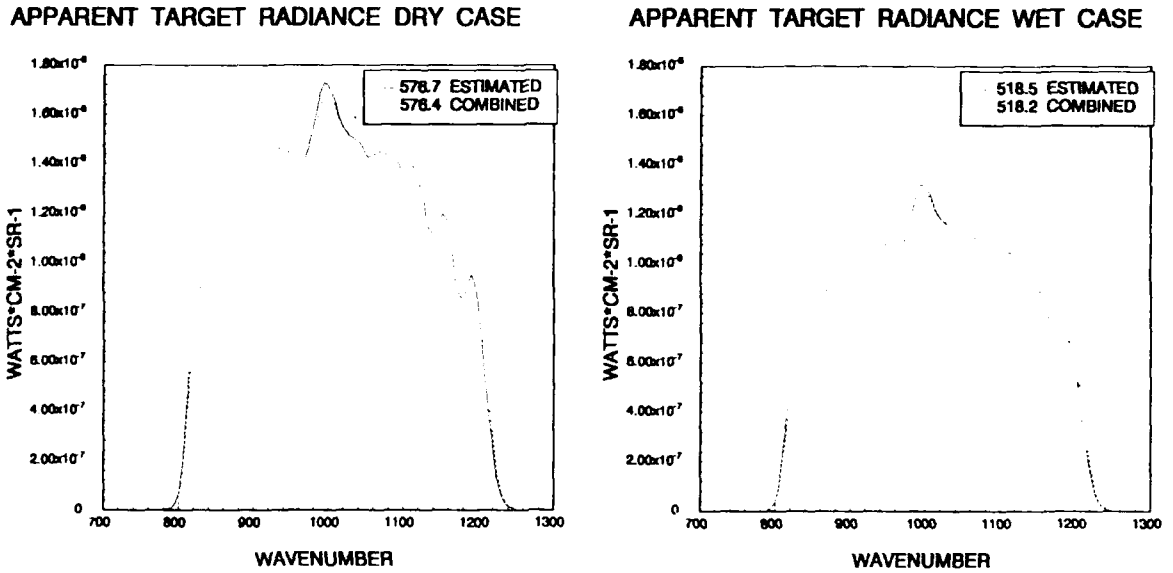


Figure 9. Comparison of actual and imager blackbody estimated propagated flare and camouflage netting spectral radiance for dry and humid conditions.

As a final example, the use of an illuminating polarized carbon dioxide laser will be discussed even though its application on the battlefield may be limited because laser users often expose their positions by using a probing laser. There is the potential to obviate some of the far-IR optical turbulence distortion along horizontal observation paths using laser-based active imaging. The key to this partial image restoration is the enhanced backscattering, especially from metal and dielectric surfaces by polarized laser radiation.[12,13,14] The enhancement occurs because of the constructive interference of the backscattered polarized beam and its time reversed partner returning along the outgoing laser beam path. This constructive interference is degraded by the optical turbulent distortions, but not in the unperturbed or isoplanatic patches within the target area. In the visible these unperturbed patches are quite small for severe horizontal path optical turbulence. In the far-IR the patches observed are on the order of 1-to-2 target board bar widths and persist between fields or frames of the 30-Hz imagers.

To illustrate this persistence of the perturbed and unperturbed regions in the IR scene, we use the set of 16 consecutive FF frames shown in figure 6. These images show the optical turbulence perturbed scenes of the 3.5-cycle bar target from along a 1.6-km path where the middle third of the path was from 1 to 3 m above the ground. The prevailing wind was 2 m/s and was nearly parallel to the viewing direction. What should be noted is that focusing and defocusing turbule lens regions appear and disappear in the dark and light bars of the vertical pattern that has no turbulent distortion in the corresponding NF images. These dark and light distortion patches

persist from frame to frame but then disappear. Note the dark patch in the lower right in frames 3 and 4 of figure 6 and the light patch in the lower left in frames 9 and 10. In similar fashion, but not as readily apparent, the isoplanatic patches of comparable size persist between frames as well.

The method to exploit the characteristics of the isoplanatic patch size and persistence in the far-IR is illustrated in figure 10. The imager and illuminating carbon dioxide laser must be coaligned through the use of a reflecting mirror on the imager telescope secondary or the use of a zinc-selenide (ZnSe) or barium-fluoride (BaF_2) beam splitter. For image restoration the CO_2 laser is chopped at 30 or 60 Hz and synchronized to the imager scanning frequency so as to illuminate every other field or frame. The backscatter enhancement in the target region and perhaps over the entire scene will be brightened in the illuminated field or frame, allowing a binary mask to be formed to define the unperturbed regions in the scene. This mask can be applied to the subsequent nonilluminated field or frame to isolate unperturbed portions of the IR image. The summation of these unperturbed partial images over several frames will allow reconstruction of the target as the isoplanatic patches move about to cover the entire scene. If the target contour can be defined in each of these partial scenes, the reconstruction process can include overlap of these contours and significantly reduce the image blurring due to large turbule-induced wander of the target in the scene as well. If successful, this far-IR image restoration technique may be extrapolated in some fashion to the less tractable problem of visible image distortions due to optical turbulence. Use of low emissive paints will accentuate the differences in enhanced backscatter returns from camouflage netting versus background. The summation of just the binary masks may segment the camouflage net without using the unilluminated fields or frames that may show essentially no contrast temperature difference and all this in spite of severe optical turbulence!

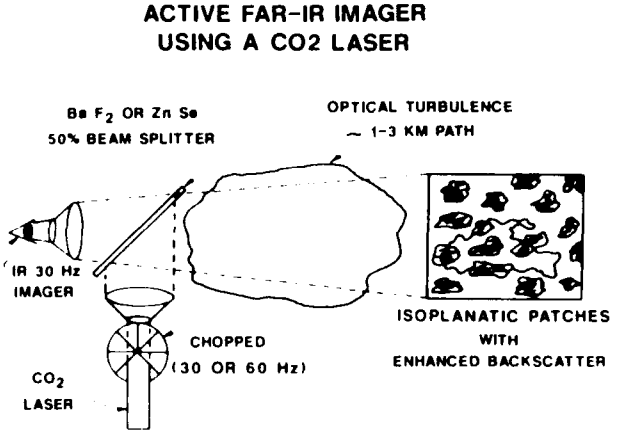


Figure 10. Experimental setup for target and camouflage detection through horizontal path optical turbulence, including partial image restoration by using a chopped carbon dioxide (CO^2) laser.

ANCILLARY OR ALTERNATE APPROACH

Considering the problems that have been shown herein that are related to the design and testing of low-emission, high-reflection surfaces used in camouflage netting, one might question the current philosophy in camouflage design. Totally reflecting surfaces may be a better approach. If a target were enclosed in a multisided tent the sides of which were essentially broadband reflector venetian blinds with the outward slat normals pointed down from the horizontal, and if this surface were viewed from horizontal or elevated positions, then the background immediately in front of the side would be reflected toward the observer or imager. The detection of this camouflage could be exceedingly difficult in the visible and IR because there is no longer any emissive component to cause camouflage flicker and the reflective component cannot be exploited by active imaging. The top of the camouflage tent would require careful design, but a mixture of totally reflective and textured surfaces could be used. Care would also have to be taken to reduce the millimeter-wave signature through texturing of the sides as well. However, considering the choice of remedying existing camouflage design and the proposed new approach, the new approach looks promising indeed. At least, it does not suffer from camouflage flicker.

CONCLUSIONS

Measurements of camouflage netting as a low observable have been presented that illustrate the real-world problem of trying to blend a man-made object into a natural background. The combination of low windspeed and solar loading changes due to partly cloudy conditions created an almost beacon-like characteristic to low emissive painted camouflage netting that otherwise performed very well. This camouflage flicker was characterized by using ASL's MISL that could isolate the driving meteorological parameters that caused the radiance variations. Because of the low emissive nature of current camouflage design, the highly reflective surfaces can also be exploited through the use of high temperature or laser source active imaging. Finally, a new total reflection design may be used to mitigate some of the problems of the current camouflage design.

REFERENCES

1. W. R. Watkins, F. T. Kantrowitz, and S. B. Crow, in *SPIE Proceedings* (SPIE-The International Society for Optical Engineering, 1988), **926**, p. 69.

2. W. R. Watkins, S. B. Crow., and F. T. Kantrowitz, *Opt. Eng.*, **30(10)**, 1563 (1991).

3. W. R. Watkins and J. B. Jordan; in *ICO Proceedings, Atmospheric, Volume and Surface Scattering and Propagation* (Florence, Italy, 1991), p. 315.

4. W. R. Watkins et al., in *SPIE Proceedings*, **1486** (SPIE–The International Society for Optical Engineering, 1991), 17.

5. W. R. Watkins, in *AGARD Conference Proceedings* (Izmir, Turkey, 1991), **502**, 26–1.

6. S. B. Crow et al., in *SPIE Proceedings* (SPIE–The International Society for Optical Engineering, 1991), **1486**, p. 333.

7. J. J. Carlson, J. B. Jordan, and G. M. Flachs, in *SPIE Proceedings* (SPIE–The International Society for Optical Engineering, 1988), **901**, p. 35.

8. B. D. Guenther, R. Buser, and W. Morrow, *Opt. & Photonics News*, 2(11), 8 (1991).

9. F. T. Kantrowitz and W. R. Watkins, in *SPIE Proceedings* (SPIE–The International Society for Optical Engineering, 1990), **1311**, p. 180.

10. F. X. Kneizys et al., AFGL–TR–88–0177, 1988.

11. W. R. Watkins, F. T. Kantrowitz, and S. B. Crow, in *SPIE Proceedings* (The International Society for Optical Engineering, 1989), **1115**, p. 179.

12. A. A. Maradudin, in *ICO Proceedings*, Atmospheric Volume and Surface Scattering and Propagation (Florence, Italy, 1991), p. 91.

13. E. Jakeman and D. L. Jordan, in *ICO Proceedings*, Atmospheric Volume and Surface Scattering and Propagation (Florence, Italy, 1991), p. 95.

14. K. A. O'Donnell and M. E. Knotts, *ICO Proceedings*, Atmospheric Volume and Surface Scattering and Propagation (Florence, Italy, 1991), p. 231.

Propellant Response to Shaped Charge Jet Impacts

Mr. Jerry L. Watson*
Ballistic Research Laboratory
Aberdeen Proving Ground, MD. 21005-5066

INTRODUCTION

With the increased interest in vehicle and crew survivability in today's fleet of fighting vehicles, a good understanding of how the ammunition is responding to a given threat is essential. Work is ongoing at the Ballistic Research Laboratory to quantify these responses so that vulnerability predictions can be made. This paper deals with the response of propellants to impacts from shaped charge jets. To do this the use of a ballistic pendulum has been employed[1]. This device is calibrated so that the propellant response can be equated to an equivalent amount of TNT which gives the same impulse load. This does not mean that the propellant response given in terms of TNT equivalence gives the same peak pressure or loading time, but that the integral of pressure during the loading time are the same. In addition to the ballistic pendulum effort, x-rays and novel gaging techniques have been used to give insight into the reactions that are occurring within the propellant bed. Through these latter techniques different types of reactions have been observed. Two types of reactions have been identified. Each type of reaction will be discussed below. The parametric relationships developed to date for each kind of propellant response to the jet parameters that impact it are also discussed in detail.

TYPE I PROPELLANT RESPONSE

The Type I propellant response is a propellant burn based upon the localized pressure applied to the propellant bed from the impacting jet. Figure 1 is an x-ray of a shaped charge jet impacting a propellant bed. As the jet enters the bed the jet parts the material forming a crater just like it would in an inert material. The material will fracture as it is impacted and the amount of fracturing will depend upon the mechanical properties of the propellant material. The material between the crater and the shock front is compressed. As the shock moves into the bed it also decreases in magnitude and slows forming a parabolic shaped shock front which degrades into an acoustic wave. The material which is ignited from the impact of the jet tip burns according to the localized pressure. As the pressure drops so does the burn rate. Large amounts of propellant are recovered from these kinds of impact. Carbon gauges[2] placed in the propellant beds verify this theory.

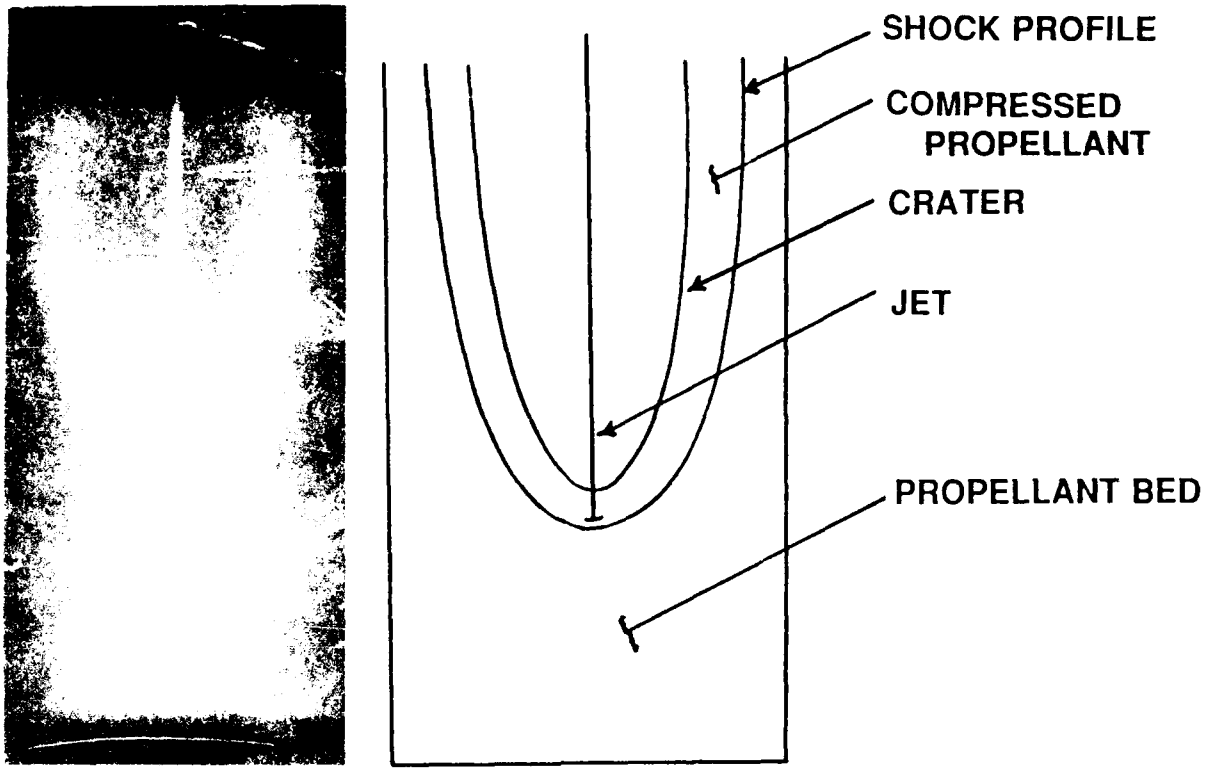


Figure 1. X-Ray of a Type I Propellant Response

A schematic of the carbon gauge test set-up used is shown in Figure 2. The gauges were placed 25.4 mm (one inch) apart with the first one being 25.4 mm (one inch) from where the jet impacts. Figure 3 shows the data curves from a typical Type I propellant response. Calibration work is on-going for these gauges so only data records are presented here. These records show the trends in signal magnitude, shape and time of response and gives a good indication of what is happening in the bed of propellant. For a Type I response the signals get progressively smaller and the time between signals get progressively longer as the reaction moves into the bed. The shape of the signal also changes from a very fast risetime for the first gauge to a longer and longer ramp wave as it moves to the more distant gauges.

TYPE I JET/PROPELLANT PARAMETRICS

All propellants seem to follow the Type I response. However, there are several propellants which do have thresholds where the reactions increase significantly when jet energy is increased. This can be easily seen by firing shaped charges through different amounts of armor to vary the input energy. Figure 4 shows the responses of three different propellants caused by a shaped charge attack into the propellant through varying amounts of armor. The shaped charge used for all these tests was

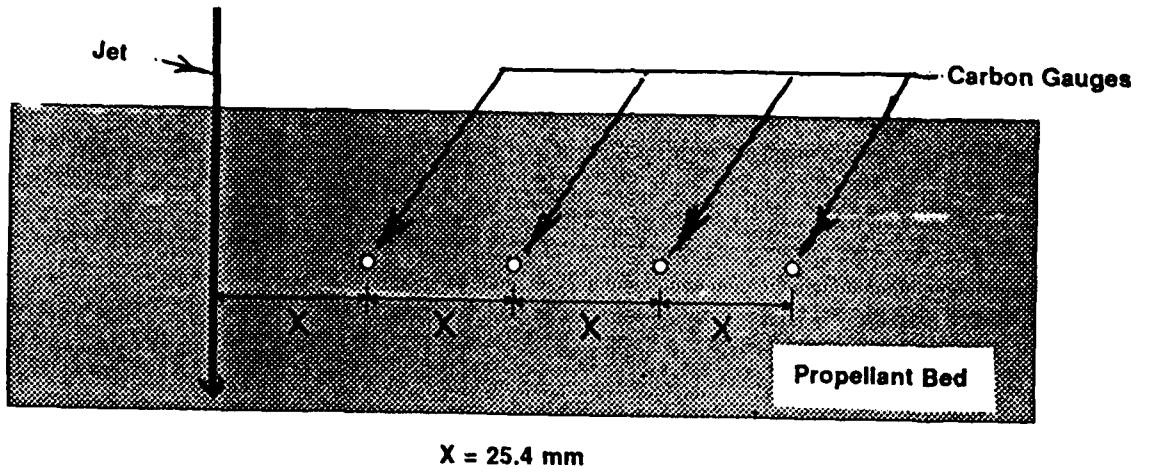


Figure 2. Schematic of Carbon Gauge Test Set-up.

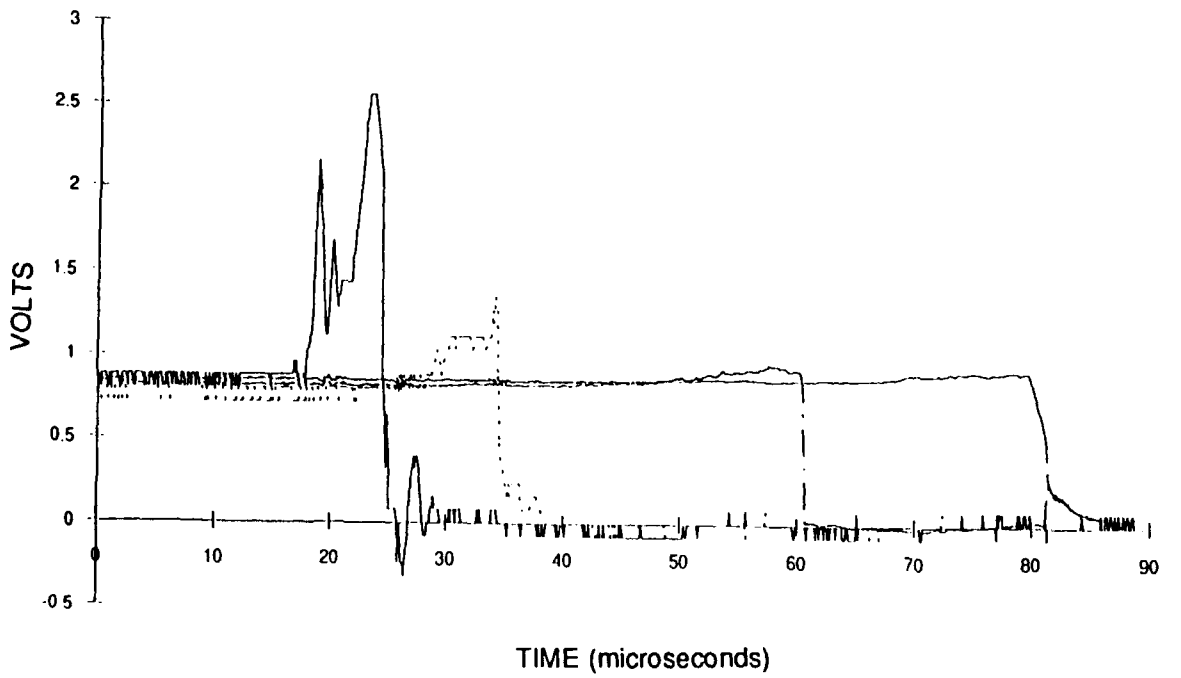


Figure 3. Pressure Records from Carbon Gauges in a Type I Propellant Response.

a 65mm copper cone with a 2mm diameter jet. Reactions are measured using a ballistic pendulum which has been calibrated with TNT charges. Several things become apparent from these reaction curves. All three propellants look alike when the jet is conditioned with 178mm of rolled homogeneous armor (RHA), but look greatly different for less RHA conditioning. Propellant C is obviously detonating for RHA conditioning below 178mm while the other two show lesser reactions. Propellant C has definitely changed from a Type I response to a more energetic reaction or a detonation (Type II response). In addition propellant B shows a much more violent reaction at small RHA conditioning than does the propellant A. This means that the Type I response varies with the kind of propellant. Recording the responses in this manner is not the most convenient or useful method for a universal application of the data. It would be more useful to record the propellant output in terms of the behind the armor jet parameters so that predictions can be made for other shaped charges.

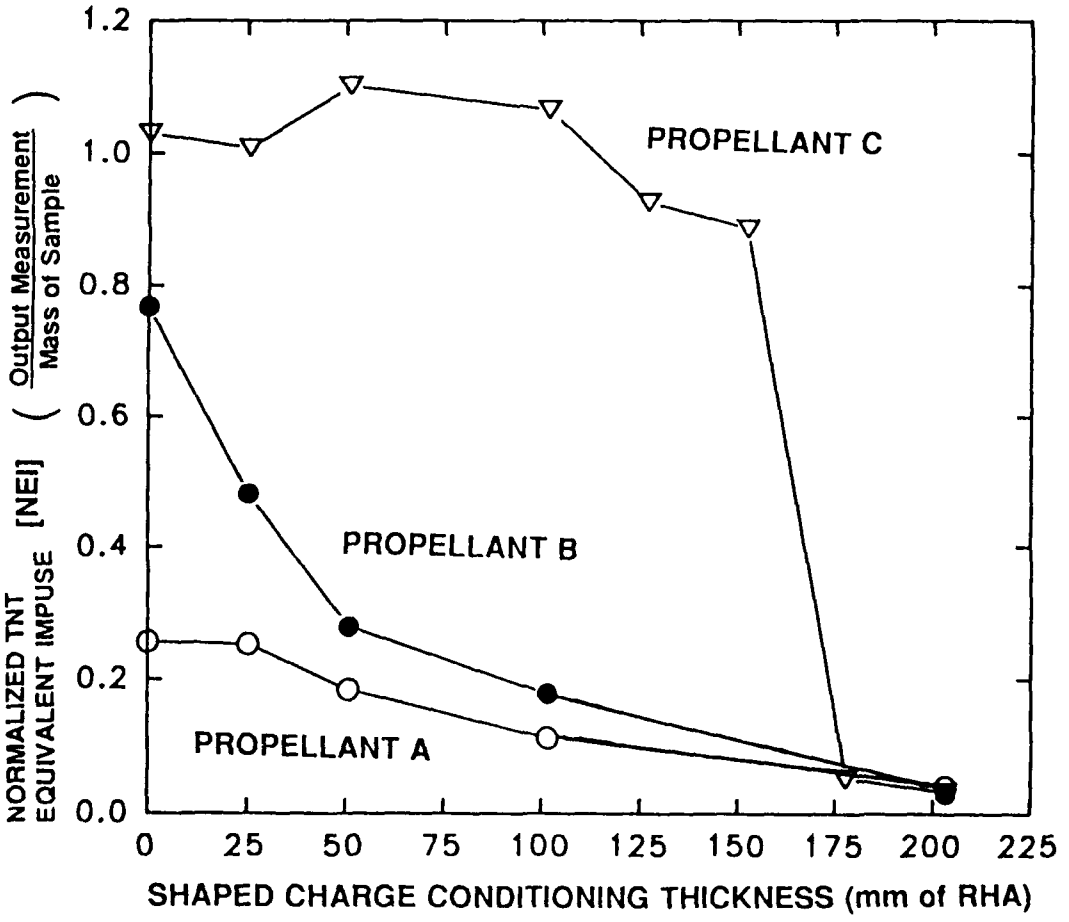


Figure 4. Response of Propellant to a Shaped Charge Attack

For the Type I reactions (deflagration) a study is on-going to determine the relative response of the propellant to various shaped charge attacks. In this study the jet parameters have been shown to directly effect the amount of energy that the propellant will release. Figure 5 shows a plot of the propellant output as a function of jet impact velocity. This data is shown for five different propellants and is plotted on log-log scale so that any functional relationships can be seen. The data form five straight lines, all with the same slope, which is a cubic function. A least squares fit for each of the data sets has been made and the resultant equation plotted on the figure. From this analysis it can be seen that the propellant response is proportional to the cube of the jet impact velocity.

The other jet parameter which is of importance is the jet diameter. This same kind of analysis can be done with it. If we divide the propellant output by the cube of the jet impact velocity, which was previously established, and plot it as a function of jet diameter (at break-up), the relationship between propellant output and jet diameter can be determined. Data has been taken for several different propellants and for several jet sizes. A plot of the data for two propellants and two jet sizes is shown in Figure 6. Again this is done on log-log scale to see if a functional relationship exists. The data plot shows that there appears to be a relationship with the square of the diameter. More data with other jet sizes need to be gathered before this can be definitely proved. The squared relationship may indicate that the output is related to the jet impact area.

Bed depth (length of propellant the jet passes through) should also be important. To study the effect that bed depth has upon the reaction another

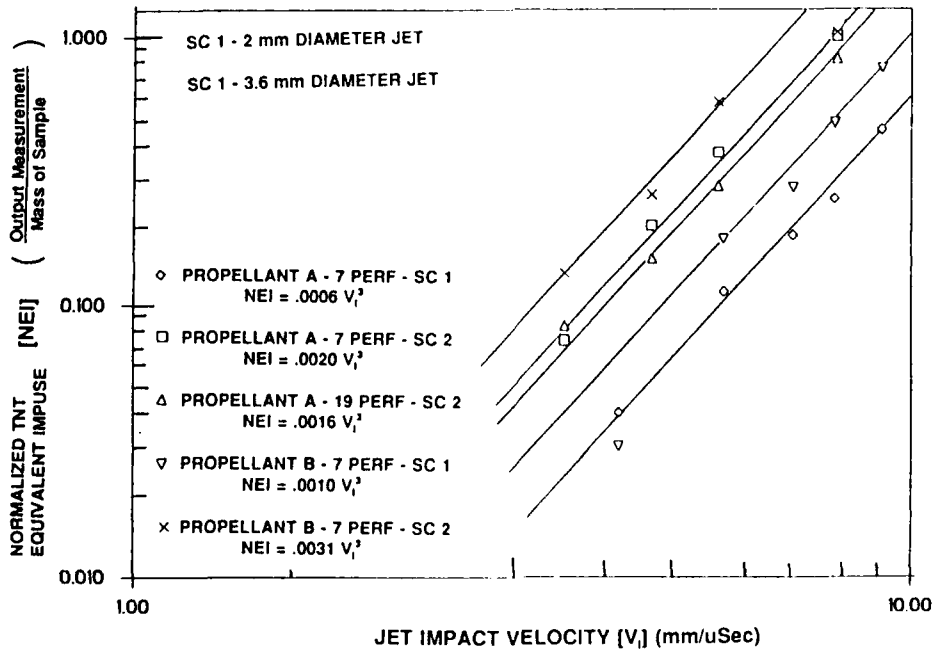


Figure 5. Propellant Response as a function of Jet Impact Velocity

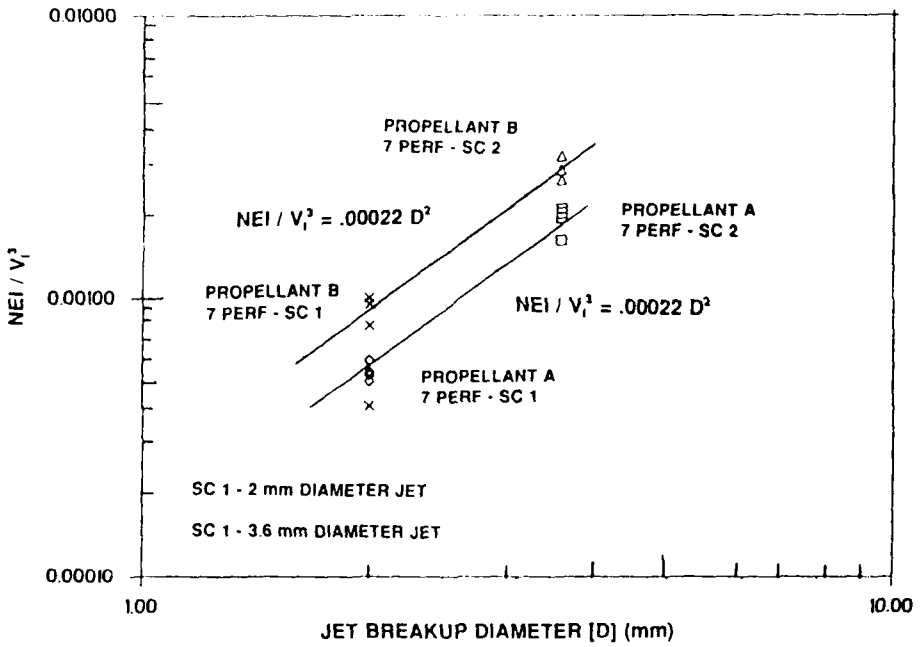


Figure 6. Response of Propellant as a function of Jet Diameter.

experiment was performed. In this experiment shaped charges were fired into beds of different depth so that the jet would pass through different lengths of propellant. Propellant A was used. The bed length varied from 127mm (five inches) to 1016mm (40 inches) with the diameter of the bed being held constant at 1016mm (6 inches). The larger shaped charge (3.6mm diameter jet) was used and was fired down the axis of the bed as in the earlier experiments. The results of these experiments are shown in Figure 7. The response does increase with the increased length but not linearly.

Since Majerus[3] and others have expressed the opinion that the response of propellant to shaped charge attack is related to the kinetic energy deposited into the bed of propellant, a curve of the form $A[1-(S/\{S+T\})^B]$ was fitted through the data, where A is a constant and B is another constant related to the density of the bed and jet. S is the stand-off distant between the shaped charge and the propellant and T is the length of propellant engaged by the jet. This form was used because it expresses the difference in the entering and exiting velocities from the propellant bed[4]. This form does seem to fit the data rather well. The B constant is a function of the square root of the ratio of the propellant bed density to the jet density (γ) which is .33 for our bed. Since the fitted curve yielded an exponent of .99, this means that $B = 3\gamma$. Since this exponent is a cubic function as was our earlier expression for velocity it means that it is really a part of the velocity expression which we had earlier. Therefore the parametric relationship should be a function of $V_i^3D^2[1-(S/\{S+T\})^{3\gamma}]$. It should be noted that for a given test set-up where the shaped charge, T and S are constant the expression reduces to V^3. Figure 8 is a plot of both this data and the

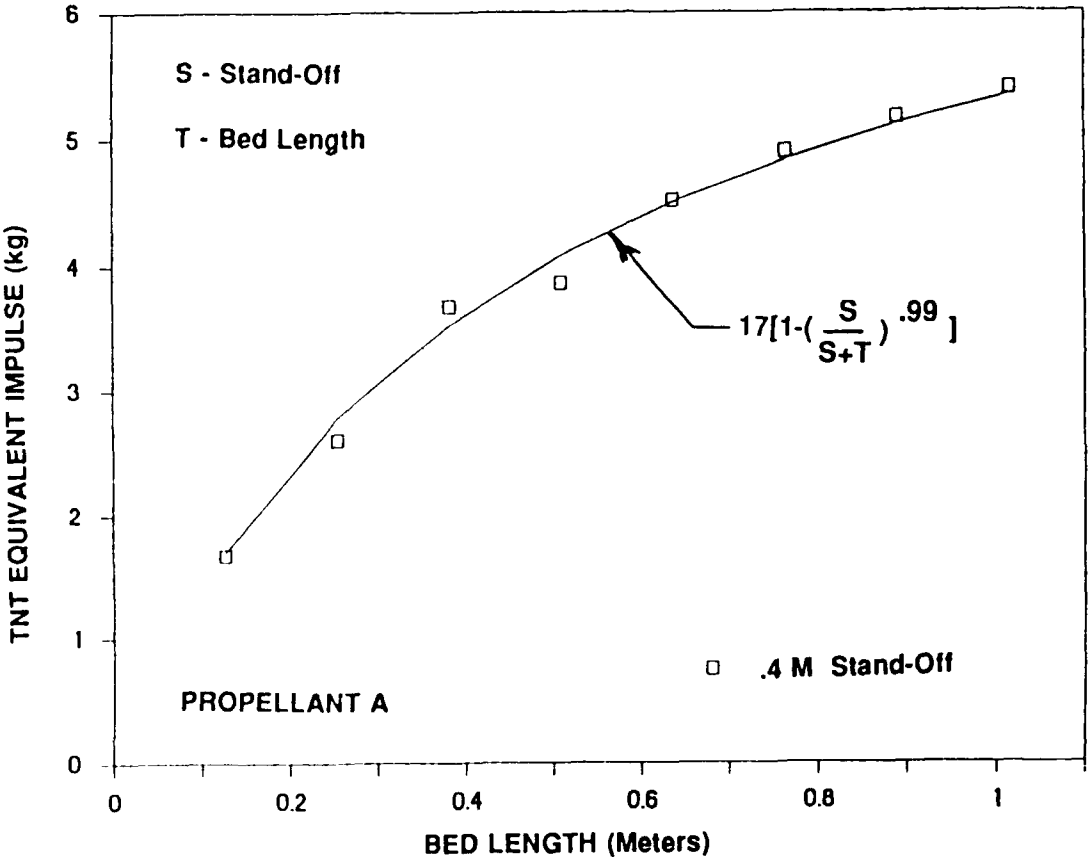


Figure 7. Propellant Response as a function of Target Length of Propellant.

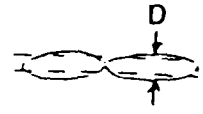
earlier test data as a function of this new parametric relationship. Both sets of data fall on a straight line which gives some assurance that the relationship is correct.

If this parametric expression is examined carefully it can easily be changed into the equation for kinetic energy. If the expression is multiplied by the break-up time of the jet (t_b) and the appropriate constants of ρ , π , .5, .25, .333 (average velocity correction), 4/9 (jet diameter correction) the results will be kinetic energy in terms of Joules of energy. The total expression breaks down in the following manner;

$$\text{Kinetic Energy} = .5\ M\ V^2 = .5\ \rho\ (\text{Volume})\ V^2$$

$$= .5\ \rho\ (\pi\ d^2\ \{1/4\}) \int_{X_1}^{X_E} V^2\ dx \qquad V = x/t \quad dx = t\,dV$$

$$= .5\ \rho\ (\pi\ d^2\ \{1/4\} \int_{V_1}^{V_E} V^2\ t\ dV$$



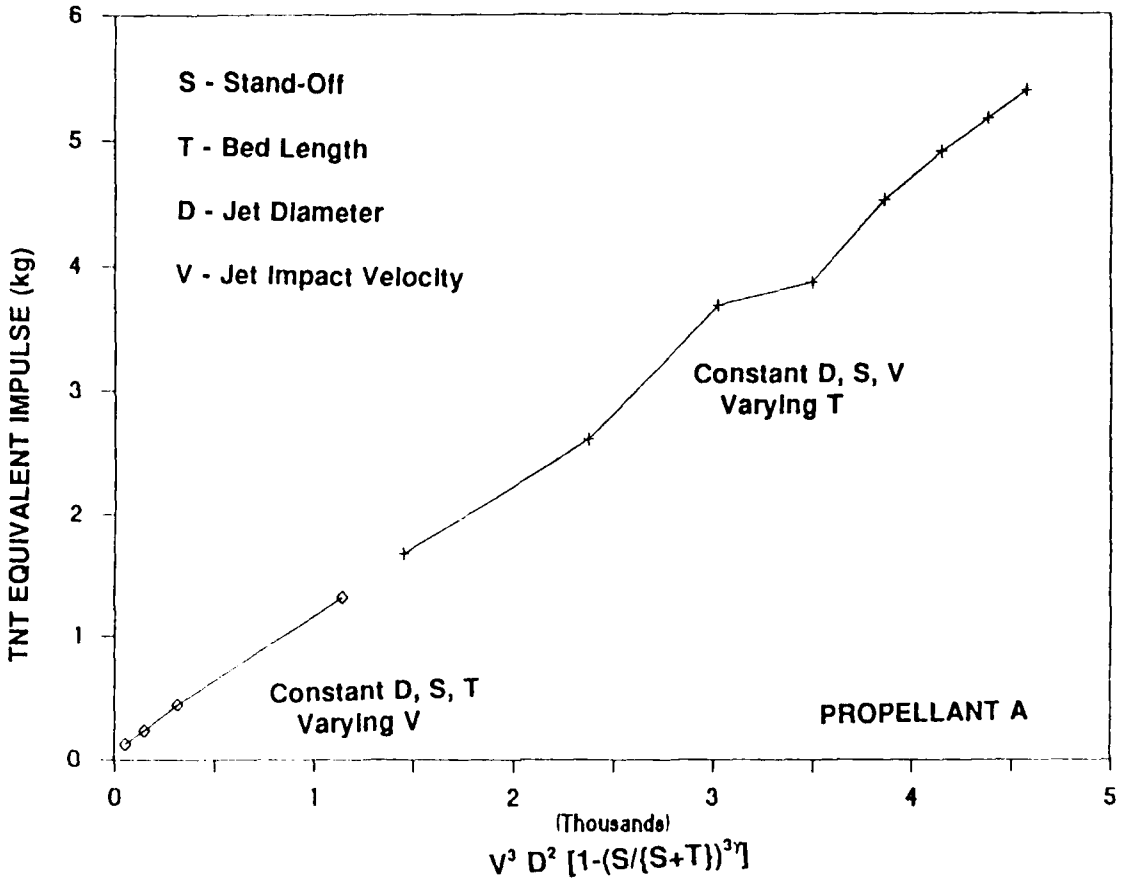


Figure 8. Propellant Response as a function of Jet Velocity, Diameter, Stand-off, and Propellant Target Thickness.

$$= .5\,\rho\,(\pi\, d^2\,\{1/4\}\,t)\,(V^3/3)\Big|_{V_I}^{V_E} \qquad \text{and } d = 2/3\ D \text{ at } t_b$$

$$= .5\,\rho\,(\pi\,\{4/9\}\,D^2\,\{1/4\}\,t_b)\,(V_I^3 - V_E^3)/3 \qquad \text{and } V_E = V_I\,(S/\{S+T\})^{\gamma}$$

$$= .5\,(4/9)\,(1/4)\,(1/3)\,\rho\,\pi\,t_b\,D^2\,V_I^3\,[1-(S/\{S+T\})^{3\eta}]$$

$$= .0185\,\rho\,\pi\,t_b\,D^2\,V_I^3\,[1-(S/\{S+T\})^{3\eta}] \quad \text{(Joules)}$$

It does appear that the propellant output is indeed proportional to the deposited kinetic energy delivered to the bed by the shaped charge. This expression is still being investigated for other jet sizes, bed depths, jet stand-offs and jet velocities. As said earlier this expression does reduce to V^3 for given test set-up.

TYPE II PROPELLANT RESPONSE

The Type II propellant response is a propellant detonation. The detonation starts in the impact zone where the jet tip comes in contact with the propellant. Once the propellant starts to detonate it continues into the bed and slows along the shock front to its steady state detonation velocity for the propellant bed. The penetration velocity of jets for most propellant beds will be above the detonation velocity of the propellant bed. Figure 9 shows carbon gauge records in a detonating bed of propellant. These pressure records are very large compared to the Type I reactions and are of equal magnitude occurring at uniform intervals of time. This indicates that the reaction is travelling at a constant velocity of equal magnitude.

Figure 10 shows two x-rays of a propellant that exhibits a Type II response. These x-rays are quite different from the earlier Type I response x-ray. The shock front is almost cardioid rather than parabolic in shape. The crater area is not well defined and may extend completely to the shock front. There also doesn't appear to be any compaction of the propellant behind the shock front. Since the material is detonating it is reacting in the shock and is being converted to detonation products very quickly. The cardioid shape is caused by two different effects. The first effect is caused by a release wave which slows the propagation at and near the surface. The second is the difference between the penetration rate and the steady state detonation velocity in the bed. This causes the shock front to be slightly conical around the jet axis rather than spherical.

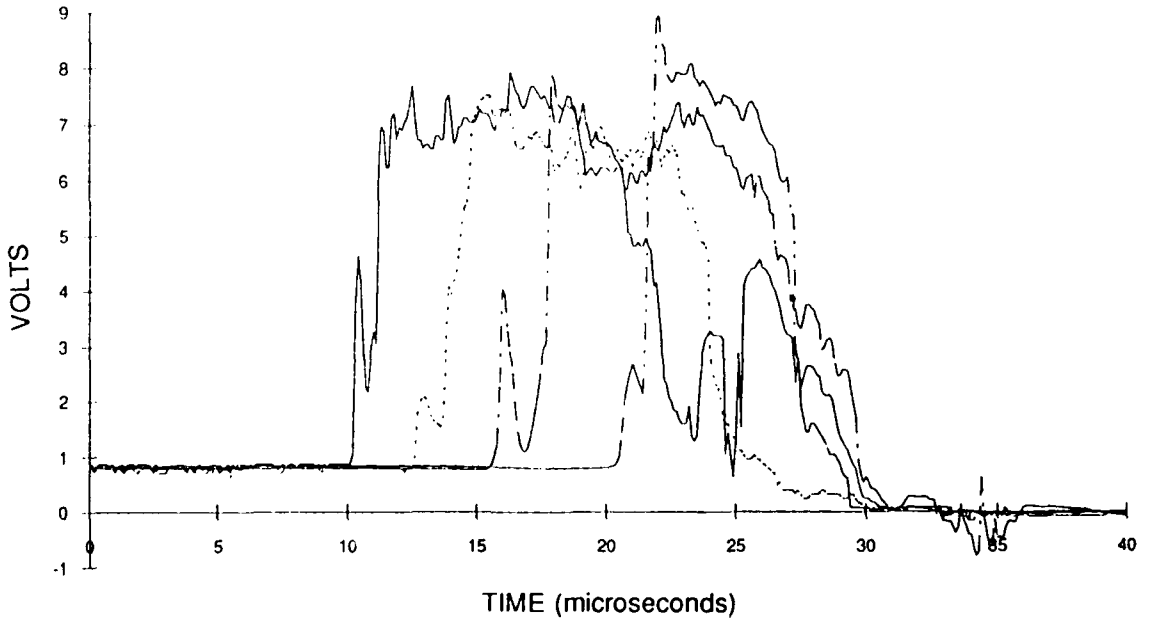
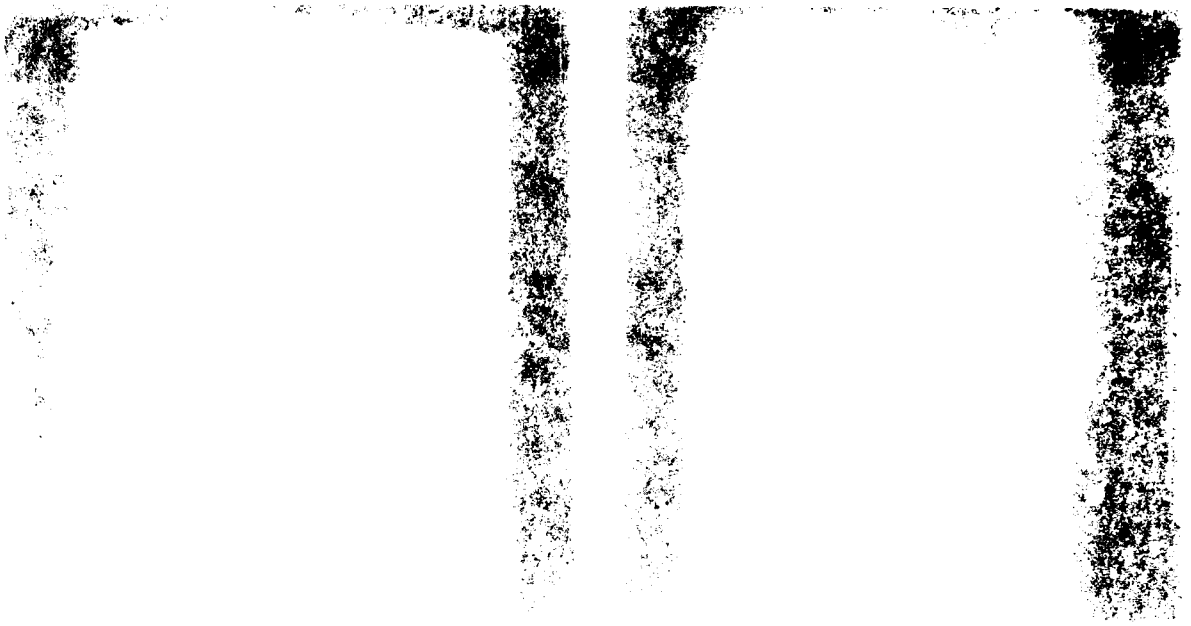


Figure 9. Pressure Records from Carbon Gauges in a Type II Propellant Response.



10 usec after impact 20 usec after impact

Figure 10 Two X-Rays of a Type II Propellant Response.

PROPELLANT RESPONSE

As shown earlier for a given test set-up using the same shaped charge and propellant thickness, the response equation versus deposited kinetic energy for Type I responses reduces to

$$\text{Output} = C_1 V^3$$ where C_1 becomes the assimilated value of all the above constants into a single value.

This means that for a given test set-up, relative comparisons can be made by comparing the output from the propellant to the cube of the jet impact velocity as long as the response is strictly a deflagration. As shown earlier there are other responses seen in propellants with detonative characteristics. Figure 11 shows three propellants that show a detonative characteristic when enough energy is put into the propellant bed by the shaped charge jet. A 65mm (cone diameter) shaped charge with a jet tip velocity of 9.15 mm/usec was used in these experiments. The tip speed was changed by firing through various amounts of RHA. Each of these propellants show a different threshold where they change from a deflagrative reaction to a much more responsive reaction. In the past all these responses were considered to be detonations, but when examined more closely show three different responses.

To study these responses a test set-up was used as shown in Figure 12. This

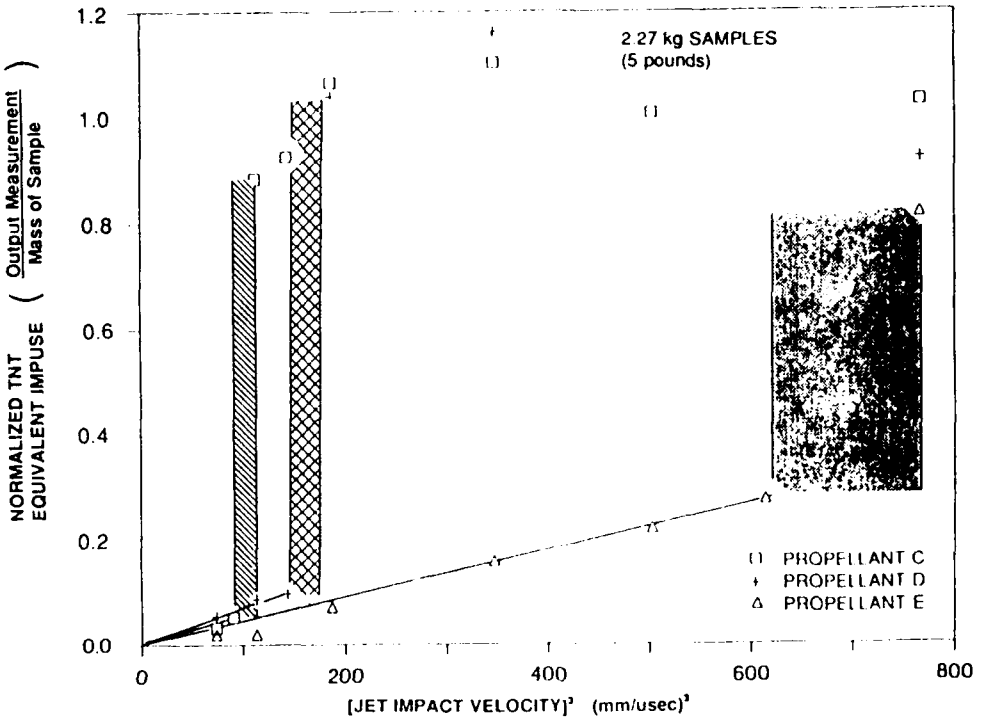


Figure 11. Output Response from Three Propellants.

test set-up used a modified version of the coaxial velocity gauge[5] where foils were used in place of the wire and tube. They are labeled Lyman gages in the figures. By using this type of gauge a continuous record of the detonative reaction within the propellant bed could be obtained. Propellant C was the first propellant to be investigated. Based upon the five pound tests and other even larger tests, this material was felt to be truly detonating when struck with a shaped charge jet. An unconditioned shaped charge was used to initiate the reaction with the propellant and be above the threshold for detonative responses. Data from this test is shown in Figure 13. The reaction along the gauge could be considered to be constant for the 120 mm of gauge length used. Because of the large propellant

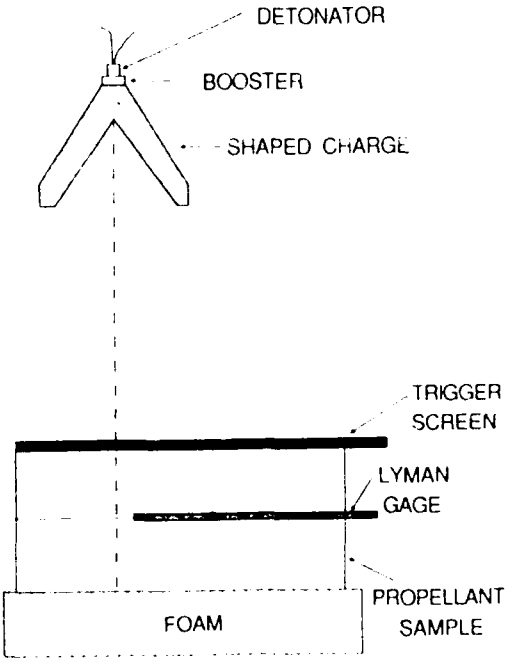


Figure 12. Test Set-up for Measuring Detonative Reactions.

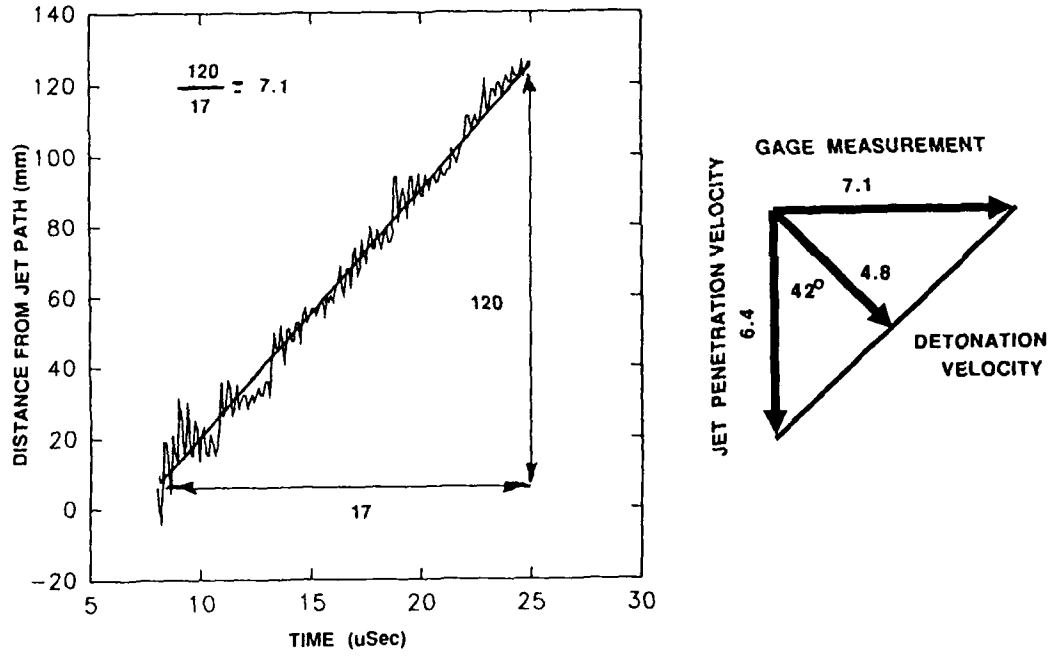


Figure 13. Reaction Velocity Propellant C.

grain size the curve shows rather erratic features, but the reaction does not appear to be slowing down in the propellant in the region that was measured by the gage. The velocity of the reaction front along the gauge was measured to be 7.1 mm/usec. Estimating the penetration velocity of the jet into the propellant then the detonation velocity becomes 4.8 mm/usec moving into the propellant at an angle 42 degrees off of the jet axis. To be truly detonative the bed size must be greater than the critical diameter for this propellant type and size, which is unknown at the moment. However, there is all indications that the jet impact is sufficient to initiate a detonative reaction if the bed is above the critical diameter.

Propellant D was the second propellant to be investigated. Again a unconditioned 65mm shaped charge was used to initiate the reaction which is above the detonative reaction threshold as seen in Figure 11. The test set-up is exactly the same and the data is plotted in Figure 14. There is definitely not a linear response with time as seen with the experiment with Propellant C for the entire length of the gage. However, the first 31.6mm of the gage does shows some degree of linearity with a velocity of 5.4 mm/usec. Doing an analysis similar to that for the Propellant C, the detonation velocity would be 4.1 mm/usec moving into the propellant at an angle of 50 degrees. Again, because of the large grain size the data is rather erratic, but if the material being impacted by the jet is undergoing an over-driven detonation, the material within the reaction front moving faster than 4.0 mm/usec can probably be considered to be detonating, or at least releasing its energy in some degraded

reaction since the velocity and pressures in the shock front are so high. The reaction front is definitely slowing down over the gage length. If the data is now redrawn on semi-log paper, a reasonable straight line can be drawn through the data. Figure 15 is a plot of this data on semi-log paper. This says that the reaction front is decaying exponentially. An equation can be now written to give the location of the reaction front with respect to time for this propellant. The velocity can now be calculated with respect to time or distance into the bed. That would be of the form;

$$x = C' \ln t + C''$$

and $$dx/dt = V_R = C'/ t$$

where V_R *is the reaction velocity along the gage*

or $$V_R = C'/ e^{((x+C'')/C')}$$

for this data set C'= 56.1 and c" = -107 when x is in mm and t is in usec.

This gives a reaction velocity component (along the gage length) of 7.3 mm/usec at the beginning of the gage and a reaction velocity of .88 mm/usec at the end of the gage. This velocity is not the reaction velocity but a component of that velocity. The other component at 90 degrees is needed to determine the magnitude and direction of the reaction front which will be less than either component and changes in magnitude and direction. These two propellants do show similar

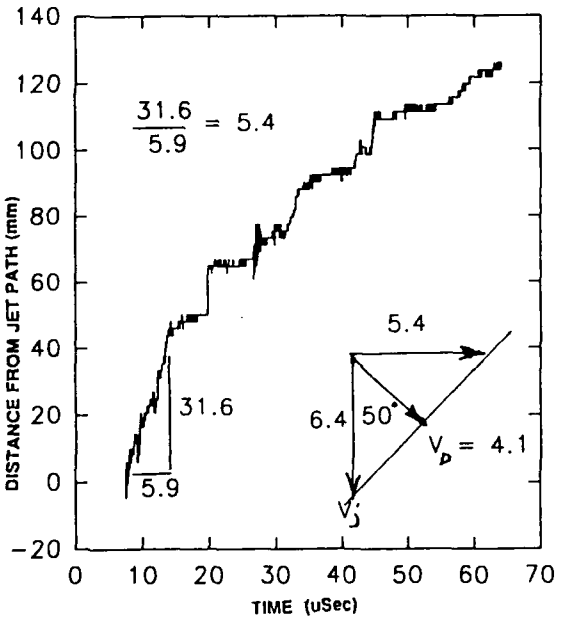


Figure 14. Reaction Rate in Propellant D.

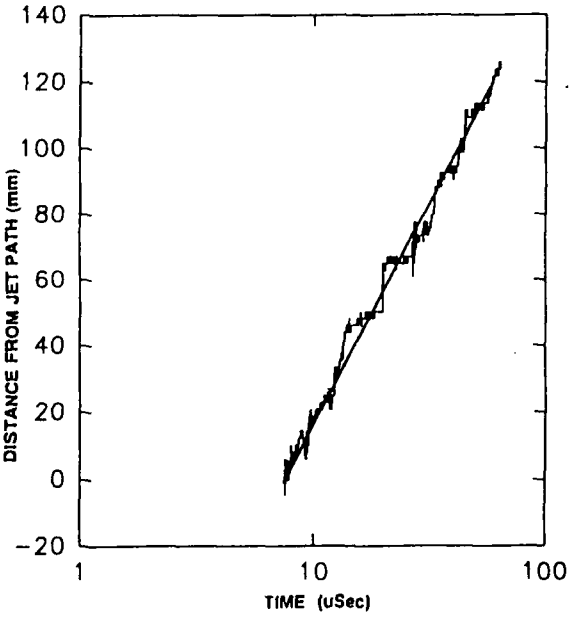


Figure 15. Reaction Rate of Propellant D.

characteristics in that they will detonate under the jet tip, but differ in that the Propellant C will sustain a detonation front into the propellant bed, but the Propellant D does not. This can be easily shown in the velocity formulas where;

for Propellant C
$V_{REACTION} = V_{DETONATION} = Constant$

for Propellant D
$V_{REACTION} = C'/ time$

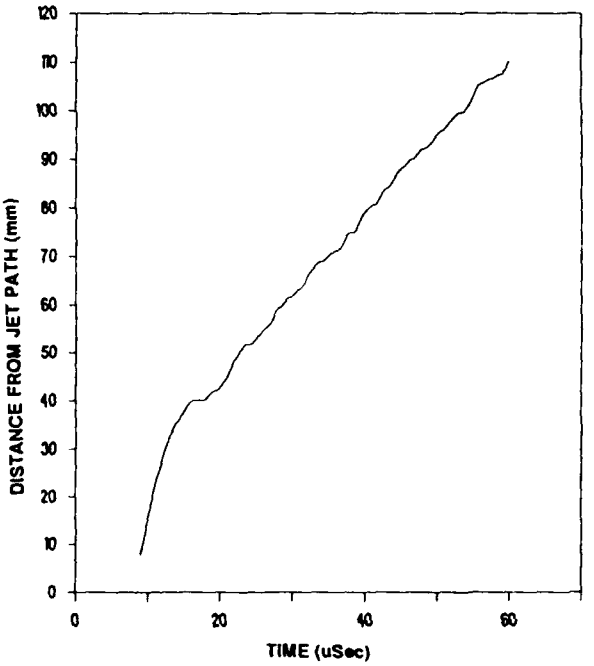


Figure 16. Reaction Rates for Propellant E.

The third propellant tested is a fine ball powder, Propellant E. This material also has a detonation threshold for this shaped charge jet which is much higher than the others. The test was performed in the exact same manner and the results are shown in Figure 16. This material has much smaller grains and more detail can be seen in the data. This data can be divided into two regions. The first region is that of the over-driven detonation which is decaying very fast and oscillates before reaching a fairly constant velocity. The second region is that of a slow moving reaction front in the bed. A plot of the reaction velocities with time is shown in Figure 17. It appears that a steady state condition has been reached and the reaction velocity along the gage is 1.5 mm/usec. This is much different from the response of the other two propellants studied earlier. Figure 18 is a comparison of the reaction velocities with respect to time for the three propellants. For these rather small scale tests, blast measurements were all very close for these samples once the detonation threshold was exceeded, but if the sample increases, the differences in how the reactions are decays will become more and more evident.

CONCLUSIONS

Propellants have a shaped charge velocity threshold above which the propellant detonates under the tip of a shaped charge jet during impact. For impact velocities below this threshold the response is a mild deflagration and is directly related to the shaped charge jet impact parameters. For impacts above this threshold the propellant bed properties become the controlling parameter. The response of the propellant to this over-driven impact detonation can be of several forms. It can detonate the entire sample or it may limit itself to the region around the jet path.

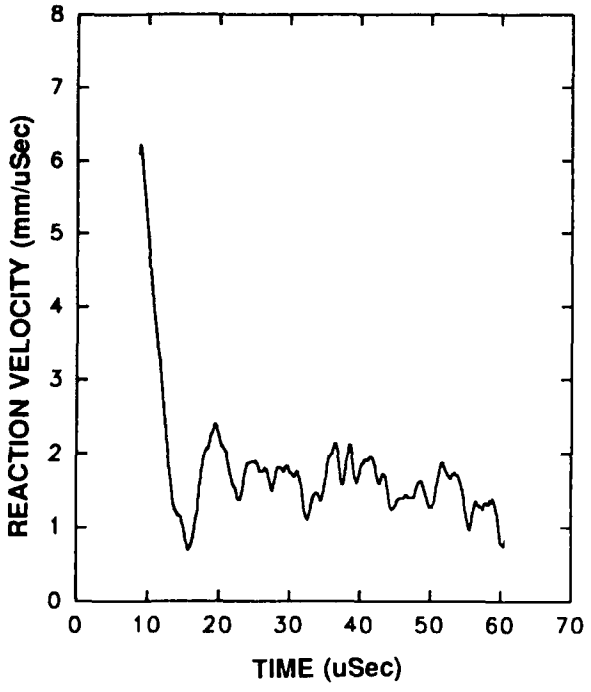


Figure 17. Reaction Velocities in Propellant E.

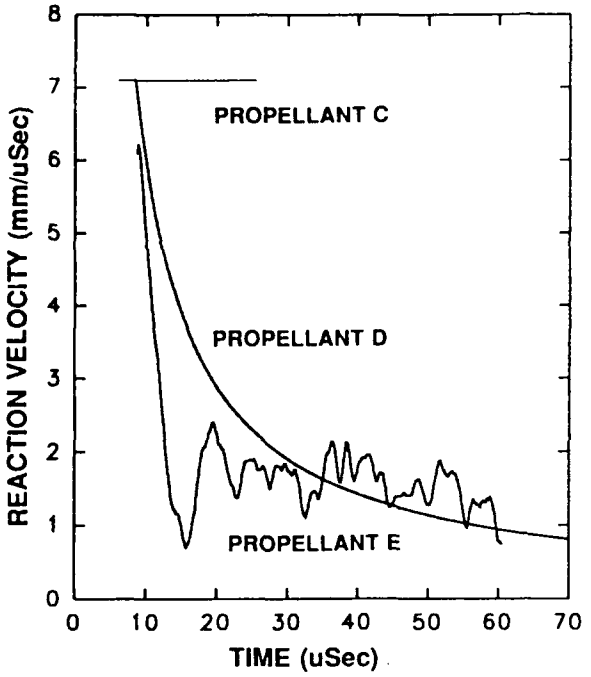


Figure 18. Comparison of Reaction Velocities for Three Propellants.

REFERENCES

1. J. L. Watson, *1990 JANNAF Propulsion Systems Hazards Subcommittee Meeting* **Vol 1**, pp. 25-32, April 1990.

2. J. Stanliewicz and R. L. White, *Rev. Sci. Instr* **42, (7)**, 1067-1068 (July 1971).

3. J. N. Majerus and A. B. Merendino, *Ballistic Research Laboratory, Report No.* **ARBRL-TR-021028**, September 1978.

4. William P. Waters and Jonas A. Zukas, *Fundamentals of Shaped Charges*, (Wiley Interscience, 1989, pp. 137-142

5. Michael J. Ginsberg, Blaine W. Asay, and Eric Ferm, *1991 JANNAF Propulsion Systems Hazards Subcommittee Meeting,* pp. 273-280, 18-22 March 1991.

New Directions in Computational Dynamics at TACOM

Dr. Roger A. Wehage and Mr. James L. Overholt*
System Simulation and Technology Division (AMSTA-RY)
U.S. Army Tank-Automotive Command (TACOM)
Warren, MI 48397-5000

1. INTRODUCTION

TACOM has used high-resolution, computer-based vehicle models for more than ten years to quantify dynamic performance specifications and determine upper bounds on safe operating performance envelopes. Validated model predictions have been used to rapidly answer numerous technical evaluation questions regarding design and product improvements. This process has saved the Army millions of dollars by significantly reducing expensive, time consuming and sometimes dangerous field tests.

Because computer-based dynamic analysis is so important, demand is outstripping our modeling and simulation capability. Effective utilization of dynamic simulation requires (1) accurate characterization of vehicle systems, (2) conversion of data into validated, high resolution computer-based models, and (3) ease of application to many diverse problems. More stringent performance specifications are being placed on vehicles to meet changing battlefield requirements. Emerging technology is pushing prototype vehicle designs toward increased subsystem complexity to meet these demands. The operator's influence on system performance is now more critical than ever, because military vehicle design trends are placing major emphasis on systems that adapt to the human, and his capabilities. Thus, computer-based vehicle models must be able to accurately account for operator inputs when they are significant. While algorithms can be developed to emulate a human's physical response characteristics, no algorithms or computers exist which can represent his reasoning and decision making processes, which are critical to successful vehicle operation.

The most feasible solution is to put the human directly into the dynamic vehicle model. However, several technical challenges must be resolved before this can happen. The computer-based vehicle models have to run in real-time

because the operator thinks and reacts in real-time. The operator must have sufficient sight, sound and touch sensory inputs from the model so he can analyze them and respond accordingly. He must also be given adequate controls to input his response commands into the model. Operator interfaces could range from high resolution graphics workstations to fully developed simulators such as TACOM's Crew Station/Turret Motion Base Simulator (CS/TMBS) facility.

Real-time simulations are now feasible because TACOM has developed a new methodology called Symbolically Optimized Vehicle Analysis System (SOVAS), which can generate the necessary high resolution vehicle models. A second methodology called Dynamic Response—Interactive Vehicle Emulator (DRIVE), that interfaces human operators with SOVAS-generated vehicle models, is also being developed to run in real-time on Iris graphics workstations. DRIVE will eventually be extended to operate with the CS/TMBS to give the operator more feedback and control capability.

Primary research is concentrating on the SOVAS and DRIVE methodologies to achieve real-time simulation capability and rapid model development. Very few real-time multibody simulation programs have been reported in the technical literature, and none allow rapid vehicle model development. SOVAS is based on a new symbolic equation preprocessing technique which reduces run-time operations count by one to two orders of magnitude and eliminates recursive operations that bottleneck parallel processors. When fully operational, SOVAS will convert user-supplied topological and parametric vehicle descriptions directly into highly optimized algorithms which are ready to run on the computer.

Vehicles also have many critical subsystems which must be accurately modeled and optimized to run in real-time along with the equations of motion. These interdisciplinary models become integral parts of the equations of motion, and SOVAS will fuse them directly into the optimized equations.

A brief overview of the recent history and future of computational mechanics is given to better illustrate TACOM's position in the area of computational vehicle dynamics research. The Army's current needs in this general area, along with a number of critical subsystem model development and analysis research efforts being carried out at TACOM, are then reviewed. This will set the stage for a general overview of the SOVAS and DRIVE methodologies. Finally, projections of Army's future vehicle modeling and simulation needs, and how TACOM plans to meet them, are made.

2. ARMY'S NEED FOR HIGH RESOLUTION COMPUTATIONAL DYNAMICS

The driving force behind the development of high resolution computational vehicle dynamics is its cost effectiveness. Low cost, high performance computers and powerful software programs now make real-time, high resolution vehicle system simulations a reality. Representative model predictions should augment or replace costly or dangerous field and laboratory tests. The word "representative" is stressed here because inaccurate predictions made with faulty models can themselves be costly and dangerous.

The aircraft community has enjoyed a greater emphasis on the importance of computer-based modeling and simulation because "you cannot park it" when it fails. Recent experiences in the desert clearly indicate the need for better understanding and quantification of mobility, reliability, stability and safety. Performance specifications are increasing and drivers need better training and instrumentation to effectively operate the vehicles. Billions of dollars have been spent on improving aircraft reliability and safety, and the automobile industry is investing heavily in reliability, stability and safety research. All of these research efforts should carry over to military vehicle design

There is one major difference between military ground vehicle, and aircraft or automobile performance. Off-road ground vehicles are operated closer to the upper limits of their dynamic performance envelopes a greater percentage of the time than aircraft or automobiles. Unfortunately, ground vehicle operators receive little, if any, training in accident recovery or avoidance procedures, either because controlling a vehicle is not perceived as a demanding skill or training is too expensive and dangerous.

Ground vehicles operate in harsh environments and are expected to perform their intended missions. Based on these missions, TACOM is responsible for defining realistic performance standards for new vehicles and insuring that they are met. Traditionally, new performance specifications are derived by proportionally scaling up existing ones with little, if any, understanding of the consequences When new vehicles do not meet these requirements, they are redesigned and/or the specifications are changed until compromises are reached. The Army often ends up with substandard systems because realistic performance specifications cannot be defined.

Now TACOM is zeroing in on a new approach which will allow the definition of more realistic performance specifications by using high resolution concept vehicle simulations. Iterations on specifications and design can be made on the computer before issuing a request for proposal Thus, procurement and

contractors alike can be assured of a greater probability of meeting the new specifications. This will result in considerable savings in design and development, and redesign costs when the contractors have a clearer understanding of the Army's needs.

The major performance areas impacted most by high resolution vehicle simulations are mobility, stability, reliability and safety. Clearly, each of these critical functions depends on many parts of a vehicle system and the driver as well. The following paragraphs summarize the importance of major vehicle subsystems and indicate the level of difficulty associated with ensuring that the models are accurate enough to make the necessary predictions.

The steering, powertrain and suspension systems in most vehicles are tightly coupled together and interact strongly with each other. Sufficient details must be included in models of subsystems whose dynamic interactions significantly affect vehicle performance, otherwise they could result in erroneous predictions. Thus, representative models must be developed for all critical subsystems which are important to vehicle performance.

"Wheel hop" is a major wheeled vehicle problem which can immobilize a vehicle in seconds on soft soils or snow and can cause loss of control on hard surfaces. It is the result of complex internal dynamic oscillations involving the vehicle chassis, powertrain, suspensions, tires and soil. Wheel hop does not exist in a number of vehicles, so it is clearly a function of design. TACOM is using high resolution models to quantify this phenomenon so performance specifications will prevent it from occurring in future vehicle systems.

The Army is investigating Antilock Braking Systems (ABS) to improve safety and mobility in off-road applications. Very little information is available on ABS performance on irregular surfaces or in soft and loose soils. Severe wheel hop may render ABS systems totally ineffective, or ABS may help to reduce it. ABS systems are tied in with engine controls, transmission, steering and brakes. Thus effective prediction of ABS performance will require accurate models of ABS and all systems it interacts with.

Suspension design also affects vehicle ride quality and component reliability. Some vehicles which exhibit no wheel hop have demonstrated superior ability to isolate occupants and cargo from terrain-induced disturbances. Ride quality and component reliability are both influenced by the level of shock and vibration transmitted to the chassis by the suspensions. ABS may also improve ride quality and component reliability, so a study in this area will involve the above ABS models as well.

The Army transports large volumes of fuel and other liquids in tankers and other vehicles to many off-road destinations. These vehicles have relatively high centers of gravity (cg) which can shift considerably due to the liquid's sloshing actions. This combination of elevated cg and transient inertial force loads can induce rollovers at much lower speeds than normal. Thus, TACOM is advocating support for research in the development of liquid sloshing methodologies which can be included in vehicle dynamics models for studying this phenomenon.

High resolution track dynamics is another critical modeling and simulation area that TACOM is investigating. The Army spends hundreds of millions of dollars on track maintenance and procurement each year. At high vehicle speeds, the inertial, internal and external loads on tracks are very high, and standing waves can develop at various speeds and loading conditions. These effects cause tracks to absorb large amounts of energy resulting in premature wear and overheating. Poorly designed tracks can also absorb considerable power from the engine and reduce maximum vehicle range. The interaction dynamics between track, vehicle components and terrain are critical to achieving representative models. However, these models cannot be run in real-time on serial or vector processor-based computers. Therefore, TACOM is developing algorithms to exploit the architecture of parallel computing devices such as the 512 processor, 50 MHz MC68040-based Connection Machine, CM 5[1].

Real-time vehicle models are being developed to support exploratory research efforts as well. One area of interest to the Army is automated risk assessment/accident avoidance. Newer vehicles generally have more power and are capable of achieving higher speeds, but they do not necessarily have improved dynamic stability properties. Thus operators are over driving them and having more accidents. This happens because a driver either fails to comprehend the severity of an impending event, or because he gets confused with a choice of multiple responses. The best solution is to give drivers high risk training so they will avoid these situations, and will be able to make quicker and more informed decisions if they do happen. Considerable inexpensive training could be done with real-time vehicle simulators where the student driver is controlling vehicle steering, throttle, braking, etc. The automated risk assessment/accident avoidance concept would use multiple sensors mounted on the vehicle to give an indication of its orientation relative to the surrounding terrain. Multiple reduced-order, faster-than-real-time, look-ahead simulations of the vehicle using its current state as initial conditions would be running on board. These models would predict ahead in time using various response options. They would provide warnings or recommended corrective actions to the driver. Real-time simulation models will allow these concepts to be designed, tested and verified on the computer before installation in a vehicle.

3. HISTORY OF COMPUTATIONAL DYNAMICS

Computational dynamics, as we know it today, began with the development of analog and digital computers in the 1950s and 1960s[2], but the theory of constrained multibody dynamics was developed in the late 1600's to the 1800s[3,4]. One can easily find more than one hundred different multibody formalisms in the literature and they can be classified roughly into three categories which are briefly described[5].

Newton defined equations for the translational motion of a particle and Euler developed them for the rotational motion of a rigid body. Together, Newton-Euler equations[3] (NEE) give the absolute spatial motion of rigid bodies. Joints are represented by imposing algebraic constraints on the displacements and by appending the second time derivative of these equations to the NEE using Lagrange multipliers. Many modern computer algorithms are based on NEE with appended constraints[5]. The number of variables in models based on tnis method can be quite large and the constrained equations will generally have the factored matrix form

$$\begin{bmatrix} M_a & J_a^T \\ J_a & 0 \end{bmatrix} \begin{bmatrix} a \\ \lambda \end{bmatrix} = \begin{bmatrix} g \\ \gamma \end{bmatrix} \tag{1}$$

Lagrange developed a second approach based on kinetic energy expressed in terms of joint relative velocities. He defined a procedure to expand a kinetic energy expression into Lagrange's equations[3] (LE) and used multipliers to append constraint equations to get

$$\begin{bmatrix} M_p & J_p^T \\ J_p & 0 \end{bmatrix} \begin{bmatrix} \ddot{p} \\ \lambda \end{bmatrix} = \begin{bmatrix} Q_p \\ \gamma \end{bmatrix} \tag{2}$$

Later, others such as Appell, Gibbs[3] and Kane developed procedures to derive equations of motion with no appended constraint equations. Their formulations which are the most strongly coupled and the most difficult to derive also have the simplest matrix form

$$M_q \ddot{q} = Q_q \tag{3}$$

For simplicity, Eq. 3 will be referred to as Kane's equations (KE), although there is much more to his method than implied by this equation[6].

The thousands of research papers in computational multibody dynamics attest to the interests and challenges in this area. Modern day researchers are not primarily concerned with deriving the governing differential equations, but with numerically integrating them so system states can be studied[7]. In the above equations, second derivatives of the state variables are not directly accessible to numerical integrators. Thus, they first have to be isolated, usually by some form of L-U factorization, followed by forward elimination and back substitution. The matrix in Eq. 1 is generally very large and sparsely populated with nonzero entries, so specially designed factorization algorithms, which manipulate only the nonzero entries, have been developed to efficiently handle them. Sparse matrix manipulation algorithms have considerable overhead and are not always reliable[8] Appending constraints as in Eqs. 1 and 2 results in systems of differential-algebraic equations and dependent variables, which creates additional numerical and stability problems. Vehicle simulations based on Eq. 1 are slow and the results may be unpredictable if a user lacks sufficient numerical analysis background.

Computer programs based on Eqs 2 and 3 cover a broad spectrum of solution strategies. Many use full matrix L-U factorization because the coefficient matrix may be densely populated with nonzero entries. More recent algorithms use symbolic procedures to generate computer programs which give the L-U factors directly[9,10,11], and others are being tailored to take advantage of concurrent and parallel processing[12]. However, effective parallelism is limited by the so-called serial recursion operations which are required to evaluate the equations and compute the L-U factors. Vehicle simulations based on Eqs 2 and 3 tend to be more efficient than those based on Eq 1, but the models may be more difficult to set up.

In the 1950s and 1960s, emphasis was placed on developing highly optimized, but oversimplified vehicle models to take advantage of limited computing power. Each model was essentially a hand-developed research project which could take two or more years to complete[13,14]. An automated sparse tableau approach to analyzing electrical circuits[15,16] evolved in the 1960s which allowed complex circuit models to be assembled from simple building blocks using appended constraints. In the 1970s, this approach found its way into mechanical systems analysis[17,18], primarily in the form of Eq. 1. It was first used for large scale military vehicle applications by TACOM in the late 1970s and is still in use today. As noted earlier, vehicle models can be set up quickly in this method, but it is inefficient and may be unreliable

4. REAL-TIME VEHICLE DYNAMICS AT TACOM

TACOM's current modeling and simulation capability is unsuitable for real-time vehicle simulation applications so an alternative approach using Eq. 3 has been developed. However, all known algorithms for isolating the state variables in Eq. 3 are incapable of achieving real-time simulations of the Army's large-scale vehicle models. Thus, a new methodology under development at TACOM, which precomputes most of the quantities used in the equations, will potentially reduce current operations count by three or more orders of magnitude[19]. The main features of this methodology, which is called Symbolically Optimized Vehicle Analysis System or SOVAS, are reviewed.

A large number of rigid bodies, joints and kinematic loops are required to accurately represent the kinematics, dynamics and kinetics of complex vehicle systems. A model could easily require two hundred or more rigid bodies, several hundred joints and a hundred or more kinematic loops. It could also require several hundred state variables, with only a small percentage being independent. Existing solution methods for Eq. 3 waste time solving for the dependent state variables which do not have to appear explicitly in the equations. They also waste time evaluating Eq. 3 using these variables, and uncoupling the equations so the independent variables can be evaluated. The SOVAS methodology does most of this work once in a preprocessor.

A vehicle's equations of kinematics and motion are defined relative to a coordinate frame embedded in its chassis. Most of its bodies are either attached directly to the chassis or are not far from it in the kinematic hierarchy. Nearly all vehicle state variables correspond to coordinates of joints which attach the bodies together. A good percentage of these variables become dependent when the body-joint combinations form closed kinematic loops. This means that the dependent variables can be given as implicit or explicit functions of a selected set of independent variables. This set of independent variables will be relatively small and the dependent to independent variable ratio could approach twenty or more in systems with many closed kinematic loops. The dependent variables are nearly always given implicitly because the constraint equations are highly nonlinear. Thus iterative procedures such as Newton-Raphson iteration are the only known general methods to solve for the dependent variables. However, Newton-Raphson iteration requires L-U factorization of large matrices which require considerable simulation computer time. Because of physical restraints such as jounce and rebound stops, steering stops, etc., within a vehicle, all variables inside the kinematic loops have well defined limits imposed on them. This makes it possible to apply Newton-Raphson iteration in a preprocessor to solve for all dependent variables over their ranges at selected grid spacings of the independent variables

over their domains. The resulting sets of points define smooth manifolds which approximate all dependent variables as explicit functions of the independent ones. These functions are stored and used during a simulation to evaluate all dependent variables, thus avoiding Newton-Raphson iteration entirely.

All coefficients in the dependent velocity and acceleration equations, and in Eq. 3 itself, depend on the joint variables. However, most of these terms depend only on the variables inside the kinematic loops. In addition, the variables inside the loops depend only on the independent ones, so nearly all quantities required to solve for the dependent velocities and accelerations, and to solve Eq. 3, can be given as explicit functions of this small set of independent variables. As described above, each term is precomputed over its range for incremental values of the independent set. This even applies to the L-U factors.

Precomputed quantities can be evaluated en masse on parallel processors to eliminate the recursive bottlenecks of traditional methods that inhibit parallel performance. However, there are other problems which have to be resolved to squeeze out peak performance from the algorithms. First, it is not always obvious which independent variables a given term will depend on. In most cases, each term can be forced to depend on at most, two variables, and other terms can be decomposed into sums where each summand has reduced variable dependency. In addition, the magnitude of many terms will be insignificant so they can be eliminated. Symbolic methods are being developed to help resolve these problems.

Model sensitivity to coefficient variability is another important research topic. All dimensions and parameter values used in a model are only as accurate as the data supplied. However, many parameters such as masses and inertias cannot be measured with high accuracy, or they will have to be estimated when no data is available. Thus it makes no sense to ask for a .01 percent tolerance on data which may be five percent or more in error. In some parts of a model, reduced tolerances could significantly reduce operations count. In other parts, reduced tolerances could cause numerical instabilities and lead to large errors.

A number of complex wheeled vehicle models are currently under development in SOVAS. Some of the more challenging ones are for the U.S. and Soviet heavy equipment transport systems. Each model contains well over one hundred rigid bodies and fifty kinematic loops. All of the models are being benchmarked on a Cray-2 vector processor, a 512 processor Connection Machine, CM 5 and a high-end, eight-processor Iris workstation. It is expected that the large models will easily run in real-time on the CM 5 and small to medium-sized models will run in real-time on all three computers.

SOVAS-generated vehicle models will be used to analyze the design and performance problems discussed earlier. The simulations will rely on the user to provide steering, throttle, braking and other control operations from the terminal of a high-end Iris graphics workstation. His control inputs to the models are required because it is impossible to preprogram complex human decision making and dynamic response characteristics. In addition, each model will have to run in real-time and the user must be able to visualize its interaction with the surrounding environment. He will have to monitor various vehicle states to help in his decision making processes. If a vehicle simulation is running on the Cray 2 or CM 5, a limited amount of control and display data will have to be passed across high speed communication links.

The protocol of interfacing a user to real-time vehicle simulations will be handled through a turnkey graphics oriented program called DRIVE which will run on an Iris workstation. The DRIVE concept will use software packages, which are commercially available, to rapidly define scenarios and vehicle graphical images, and to generate the real-time dynamic displays. These interactive programs contain hundreds of powerful functions which are used to minimize the time and cost to develop a satisfactory visualization and control system. Accurate and high resolution displays are crucial because the operator will be basing his control decisions on what he sees and hears. The DRIVE interface routines are being designed to create a realistic operating environment for the user.

The turnkey DRIVE workstation will contain libraries of user-selectable, predefined terrain profiles and vehicle models. The user will be given many options to configure test scenarios and vehicle system models from the libraries. In order to achieve real-time simulations, DRIVE will store only the minimum amount of control information in high speed memory necessary to rerun a simulation any number of times. A stored simulation can be replayed without user control inputs. During reruns, the user will be able to select virtually any viewing configuration such as zooming in for close inspection of critical components. He will be able to look at a wide range of system states from many different perspectives. He will also be able to run simulations in slow motion or transfer the graphical images to video tape.

5. FUTURE SIMULATION NEEDS

As a result of more than ten years of simulation experience in supporting vehicle acquisition projects, TACOM has identified significant factors which, if incorporated when developing performance specifications and used during the source selection and evaluation activity, would make the modeling and simulation efforts more successful. First, it is essential that the vehicle system proponent (e.g.,

Program Executive Officers), the user, and vehicle simulation specialists, work together to define the vehicle system mission and to develop sets of representative, realistic-use scenarios. Paramount to this process is a sufficient knowledge and understanding of the detailed engineering characteristics and quantitative performance levels of existing fleets which form the basis for developing comparable quantitative performance specifications, and establishing product improvement goals of new systems. Second, with respect to source selections, quantitative performance specifications must be clearly defined in solicitation packages so that prospective bidders can identify and provide the information necessary for useful vehicle simulations. This point cannot be over emphasized Manufacturers, in TACOM's experience, sometimes fail to support their bids with sound design data. Requiring detailed data packages as a part of bids, besides aiding simulation, will help identify those bidders exhibiting competent engineering capabilities. Finally, the use of high-resolution simulation models and realistic vehicle input characteristics ensures that source selection and evaluation boards can perform detailed, discriminating, and objective technical evaluations of proposed systems.

If properly applied, simulation can effectively augment the test planning and validation process. For example, a critical part of vehicle system acquisition is DT and OT testing. Through modeling and simulation during the development phases, test environments and instrumentation requirements for field testing can be determined in advance with better certainty. Also, simulation results can identify potential vehicle problems that may arise during field tests which should be addressed prior to testing Doing this affords the potential for substantially reducing test costs and time. TACOM's RDE Center and the U. S Army Test and Evaluation Command (TECOM) are jointly developing procedures to use physical simulation in lieu of field testing in certain cases. For example, structural integrity testing of truck and trailer frames and components in support of Comparison Production Tests and Production Quality Tests is an area where physical simulation has demonstrated cost and time savings of up to fifty percent. This permits the creation of simulation database libraries, which can be used for future applications. Relying on simulation in test planning and validating simulation results against carefully controlled tests greatly enhances the simulation database libraries and increases confidence in simulation

Using simulations to quantify military ground vehicle system performance characteristics has steadily gained acceptance over the past ten years. TACOM has been demonstrating that simulation and modeling leads to significant time and cost savings compared to traditional "build-test-break-fix" approaches. Army leadership has come to recognize the value of simulation as a tool for reducing the costs and time associated with traditional approaches to vehicle development

Simulation also allows analysis of concepts and scenarios which cannot be replicated economically (or not at all) with test beds

Even with the past and current successes of TACOM's simulation-based vehicle acquisition strategy, it is still necessary to examine the future trends and needs of concept, design and evaluation approaches. Rapidly changing technology (particularly in the field of high performance computing and computer graphics imagery) have had dramatic affects on advancements in simulation methodology. A great deal of the recent efforts have concerned themselves with the theory behind real-time simulation and animation of large models (even with the supercomputers of today, real-time simulation of only the simplest of models is possible.) These efforts have fallen short due to their reliance on simplifying assumptions and/or computing environments that do not exist.

The Army is investing a large amount of resources into the creation of programs dedicated to developing virtual prototyping (VP) and rapid response modeling capabilities. VP is a concept of designing future vehicle systems entirely on computers before "bending metal." Even though these programs do not emphasize the need for real-time simulation and animation, they nonetheless, represent a significant leap in technology with their reliance on intelligent database developments (these programs are being structured based on augmenting the concept and design process by quickly extracting critical vehicle parameters from intelligent computer networks.) The main ground vehicle proponents of the VP approach are TACOM and the Defense Advanced Research Projects Agency (DARPA). TACOM researchers are currently working with engineers and computer scientists from the University of Iowa, General Dynamics Land Systems (GDLS), BMY and FMC in an effort to use the DARPA Initiative for Concurrent Engineering (DICE) database for the design and analysis of new vehicles. It is pointed out that a VP approach, which does not emphasize total model fidelity, critical real-time human decision making processes and control inputs, will be incapable of quantifying critical dynamic performance characteristics, e.g. upper limits on stable vehicle operation.

Besides the emphasis on the creation of new methodologies such as SOVAS and DRIVE, TACOM is also investigating the acquisition of new physical simulators. These new devices (not of the real-time breed) will allow engineers to perform more detailed laboratory tests on a wide variety of vehicle performance characteristics, such as frame flexure, axle reliability and trailer pintle dynamic effects. The data and experience gained from these tests will allow engineers the opportunity to refine and validate vehicle and subsystem mathematical models.

6. CONCLUSIONS

Ongoing dynamics and controls research in TACOM's System Simulation and Technology Division continues to strive toward high resolution, real-time man-in-the-loop simulation capabilities. With the prospect of defense budgets continuing to decrease over the next few years, it is imperative that design and analysis methods be developed that are timely and cost effective, while maintaining a high level of confidence in the resulting product. The future of military vehicle modeling and simulation must emphasize the coordinated use of rapid, automated, high resolution model development capabilities with real-time simulations and animations. Furthermore, these capabilities must be transitioned to engineers and scientists tasked with creating and evaluating new vehicle designs. TACOM's development of the SOVAS and DRIVE methodologies will greatly assist in the efforts to meet the requirements of the 21st century Army.

7. BIBLIOGRAPHY

[1] *The Connection Machine CM5 Technical Summary:* Thinking Machines Corporation, 245 First Street, Cambridge, MA, (1991).

[2] *Electronic Analog and Hybrid Computers:* G.A. Korn and T.M. Korn, (McGraw-Hill, Inc., New York, NY, 1964).

[3] *A Treatise on Analytical Dynamics:* L.A. Pars, (Ox Bow Press, Woodbridge, CT, 1979).

[4] *A History of Mechanics:* R. Dugas, (London, 1957); English Translation of *Historie de Mécanique*, (Neuchatel, 1950).

[5] *Multibody Systems Handbook:* W. Schielen, (Springer-Verlag, New York, NY, 1990).

[6] *Dynamics: Theory and Applications:* T.R. Kane and D. A. Levinson, (McGraw-Hill, New York, NY, 1985).

[7] *Computer Aided Analysis and Optimization of Mechanical System Dynamics:* E.J. Haug, Editor, (Springer-Verlag, New York, NY, 1984).

[8] *Direct Methods for Sparse Matrices:* I.S. Duff, A.M. Erisman and J.K. Reid, (Clarendon Press, Oxford, 1986).

[9] *Symbolic Computer Methods to Automatically Formulate Vehicle Simulation Codes:* M.W. Sayers, Ph.D. dissertation, The University of Michigan, (1990).

[10] *High Performance Muitibody Simulations via Symbolic Equation Manipulation and Kane's Method:* D.E. Rosenthal and M.A. Sherman, *The Journal of the Astronautical Sciences*, Vol. 34, No. 3, (1986), pp. 223-239.

[11] *Symbolic Uncoupling and Efficient Solution of Tree Structured Linear Equation Systems:* R.A. Wehage, *Transactions of the Seventh Army Conference on Applied Mathematics and Computing*, (1990).

[12] *Symbolic Uncoupling of Multibody Equations of Motion, Parts I and II:* R.A. Wehage and M.J. Belczynski, papers presented at the Ninth Army Conference on Applied Mathematics and Computing, June 1991 (unpublished data).

[13] *Simulation and Model Verification of Agricultural Tractor Overturns:* D. Davis, Ph.D. dissertation, Cornell University, (1973).

[14] *Simulation of Motor Scraper Overturns:* S. Okuno, Ph.D. dissertation, Cornell University, (1977).

[15] *A Sparsity Oriented Approach to the Dynamic Analysis and Design of Mechanical Systems, Parts I and II:* N. Orlandea, M.A. Chace and D.A. Calahan, *ASME Journal of Engineering for Industry*, Ser. B, Vol. 99, (1977), pp. 773-784.

[16] *A New Efficient Algorithm for Solving Differential-Algebraic Systems using Implicit-Backward Differential Formulas:* R.K. Brayton, F.G. Gustavson and G.D. Hachtel, *Proceedings of the IEEE*, Vol. 60, No. 1, (1972), pp. 98-108.

[17] *DADS User's Manual, Rev. 6.0:* Computer Aided Design Software, Incorporated, Oakdale, IA, (1989).

[18] *ADAMS User's Manual, Version 5.2.1:* Mechanical Dynamics, Inc., Ann Arbor, MI, (1989).

[19] *High Resolution Vehicle Simulations Using Precomputed Coefficients:* R.A. Wehage and M.J. Belczynski, to be presented at the 1992 ASME Winter Annual Meeting, Anaheim, CA.

Navier-Stokes Predictions of Pitch Damping for Axisymmetric Shell Using Steady Coning Motion

Mr. Paul Weinacht* and Dr. Walter B. Sturek
US Army Ballistic Research Laboratory
Aberdeen Proving Ground, Maryland 21005-5066
and
Dr. Lewis B. Schiff
NASA Ames Research Center
Moffett Field, CA 94035

INTRODUCTION

Prediction of the in-flight motion of projectiles requires the determination of the aerodynamic forces and moments that act on the body. These aerodynamic forces and moments may be determined by experimental or theoretical means, such as Computational Fluid Dynamics. From a computational standpoint, much of the research effort has focused on determining the static aerodynamics, such as drag and pitching moment. Only a limited number of studies have focused on numerical prediction of dynamic aerodynamic derivatives such as pitch damping force and moment. In the current research effort, a parabolized Navier-Stokes technique has been adapted to predict the pitch damping force and moment using steady coning motion.

By applying linear flight mechanics theory such as that developed by Murphy[1], it can be shown that aerodynamic side force and moment coefficients acting on a projectile in steady coning motion can be related to the pitch damping force and moment coefficients. Steady coning motion is defined as the motion performed by a missile flying at a constant angle with respect to the free stream velocity velocity vector and undergoing a rotation at a constant angular velocity about a line parallel to the freestream velocity vector and coincident with the projectile center of gravity. This is shown schematically in Figure 1. Coning motion is, in fact, a specific combination of two orthogonal planar pitching motions. The use of steady coning motion to determine

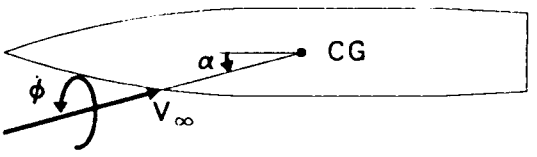


Figure 1. Schematic of coning motion.

the pitch damping aerodynamic coefficients provides an interesting and cost effective approach for determining the aerodynamics which are normally associated with unsteady or time-dependent motions.

Experimental studies of conical bodies undergoing spinning and coning motions were performed by Schiff and Tobak[2]. Their wind tunnel results showed that, at low angles of attack, the slopes of the side force and moment with angle of attack normalized by the coning rate were in good agreement with predictions of the damping-in-pitch force and moment coefficients obtained using linearized theory. They also demonstrated that the Magnus force and moment (variation of side force and moment with spin rate and angle of attack) was negligible, thus the linear pitch damping coefficients could be determined from the side force and moment due to coning alone for these bodies.

Later, Schiff[3] computed the supersonic inviscid flow about a conical body undergoing coning motion. To compute the flow around the body in coning motion, Schiff made use of a rotating coordinate frame. Within the rotating coordinate frame the flow was steady, thus the steady Euler equations could be solved. The governing equations were modified to include the centrifugal and Coriolis force terms. His computed results compared well with experimental results and with estimates of pitch-damping coefficients using a linear theory. Later studies by Agarwal and Rakich[4], and Lin[5] also employed rotating coordinate frames to compute the supersonic viscous flow about conical bodies in coning motion. More recently, Weinacht and Sturek[6] performed computations for finned projectiles in coning motion to determine the pitch damping coefficients.

In each of the previous efforts, the pitch damping coefficients were determined from the side moment due to lunar coning motion which required the Magnus moment to be neglected. The predictions of the pitch damping coefficients presented in this paper were determined from the side moment due to a specific combination of spinning and coning motion which allows the side moment due to this motion to be directly related to the pitch damping force and moment coefficients.

In the current paper, application of this approach has been made to axisymmetric shell. The flow field about these projectiles in steady coning motion has been successfully computed using the parabolized Navier-Stokes computational approach of Schiff and Steger[7]. The computations are performed in a rotating coordinate frame similar to that employed originally by Schiff. From the computed flow field, the side moments due to coning motion, spinning motion, and combined spinning and coning motion are used to determine the pitch-damping coefficient. Computations have been performed for a generic shell configuration, a secant-ogive cylinder boattail. Predictions are made for various length to diameter ratios and supersonic Mach numbers. The parabolized Navier-Stokes predictions are compared with predictions made using an Euler code originally reported by Schiff. Results are also presented for a series of ogive-cylinder configurations

which have been fired through an aerodynamics range located at the Ballistic Research Laboratory. Comparison between PNS results and range data are made.

THEORETICAL BACKGROUND

As was discussed previously, steady coning motion is defined as the motion performed by a missile flying at a constant angle with respect to the free stream velocity vector (angle of attack) and undergoing a rotation at a constant angular velocity about a line parallel to the freestream velocity vector and coincident with the projectile center of gravity. This is shown schematically in Figure 1. With respect to a non-rolling coordinate frame, the vertical and horizontal components of the angle of attack, α and β, can be shown to be sinusoidal, constant amplitude pitching motions as the projectile rotates about the free-stream velocity vector. Because the components of the angle of attack differ in phase by a quarter cycle, the total angle of attack, $\alpha_t = \sqrt{\alpha^2 + \beta^2}$ is constant. It can be seen then, that coning motion is composed of a specific linear combination of two orthogonal planar pitching motions.

Side Moment for Lunar Coning Motion

In the current study, there are two particular types of coning motion of interest. The first motion is described as lunar coning motion. In lunar coning motion, the angular velocity of the body results only from the rotation of the body about the freestream velocity vector. By applying linear flight mechanics theory, the side moment produced by lunar coning motion can be written in terms of other aerodynamic coefficients as shown below.

$$C_n = \delta(\frac{\dot{\phi}l}{V})(C_{n_{pa}} + [C_{m_q} + C_{m_\alpha}]) \tag{1}$$

The side moment coefficient, C_n, represents the net moment which causes rotations of the body out of the plane of the angle of attack. For lunar coning motion, the side moment consists of contributions from the pitch damping moment coefficient, $C_{m_q} + C_{m_\alpha}$, which is a moment proportional to the angular rate, and the Magnus moment coefficient, $C_{n_{pa}}$, which accounts for a side moment due to flow asymmetries from a combination of spin and angle of attack. This relation is identical to that presented by Schiff and Tobak[2] for bodies of revolution. Assuming that the side moment due to lunar coning motion and the Magnus moment can be determined, this relation will allow the pitch damping coefficient to be determined.

Because the projectile body does not rotate with respect to the pitch plane while undergoing lunar coning motion, the flow, when observed in the coning coordinate frame, will be steady for both axisymmetric and non-axisymmetric bodies. If the Magnus moment can be neglected, lunar coning motion can be used to determine the pitch damping

coefficients for a wide class of symmetric flight bodies using steady flow approaches. As an example, pitch damping predictions for finned projectiles have been obtained using lunar coning motion[6].

Side Moment for Combined Spinning and Coning Motion

A second type of coning motion can be formulated which allows the side force and moment to be directly related to the pitch damping force and moment. In this motion, the angular velocity of the body is described by a rotation of the body about the freestream velocity vector as well as a rotation of the body about is own axis (spinning motion). The rate of spin is equal in magnitude to the coning rate, but opposite in sign.

For this type of coning motion, the side moment can be written as shown below.

$$C_n = \delta(\frac{\dot{\phi}l}{V})[C_{m_q} + C_{m_\alpha}] \tag{2}$$

Here, the side moment is directly proportional to the pitch damping moment coefficient. Despite the simplicity of this expression, the Magnus problem has not been entirely removed from the problem, however. This is because the motion in the coning coordinate frame (computational frame) involves coning and spinning motions. Thus, the computational approach must be capable of capturing the Magnus effect, otherwise the pitch damping predictions will be in error by the degree to which the Magnus force and moment are improperly determined. Additionally, because of the combined spinning and coning motions, the fluid motion will only be steady for axisymmetric projectile geometries.

Similar expressions relating side force due to coning to the pitch damping force and Magnus force for the cases of lunar coning and combined spinning and coning can be developed using the same approach as discussed above.

COMPUTATIONAL APPROACH

Computation of the viscous flow field about the axisymmetric shell configurations was accomplished by solving the thin-layer Navier-Stokes equations using a parabolized Navier-Stokes technique (PNS). The computations were performed in a coordinate frame that rotates at the coning rate of the projectile. The fluid flow relative to the rotating coordinate frame does not vary with time, allowing the steady (non-time varying) Navier-Stokes equations to be applied. The solution of the steady Navier-Stokes equations can be performed at a reasonable computational cost. In order to implement the rolling coordinate frame, the governing equations were modified to include the effects of centrifugal and Coriolis forces. The steady thin-layer Navier-Stokes equations are shown

below.

$$\frac{\partial \hat{E}}{\partial \xi} + \frac{\partial \hat{F}}{\partial \eta} + \frac{\partial \hat{G}}{\partial \zeta} + \hat{H}_c + \hat{H} = \frac{1}{Re}\left(\frac{\partial \hat{S}}{\partial \zeta} + \hat{S}_c\right) \tag{3}$$

Here, $\hat{E}$, $\hat{F}$, and $\hat{G}$ are the inviscid flux vectors, $\hat{S}$ is the viscous flux vector, $\hat{H}_c$ and $\hat{S}_c$ are inviscid and viscous source terms due to the cylindrical coordinate formulation, and $\hat{H}$ is the source term containing the Coriolis and centrifugal force terms which result from the rotating coordinate frame. Each of these matrices are functions of the dependent variables represented by the vector $q(\rho, \rho u, \rho v, \rho w, e)$, where ρ and e are the density and the total energy per unit volume, and u, v, and w, are the velocity components in axial, circumferential, and normal directions. The source term containing the Coriolis and centrifugal force terms is also a function of the angle of attack, coning rate, center of gravity position, and spacial coordinates.

The pressure, p, which appears in the flux terms, can be related to the dependent variables by applying the ideal gas law.

$$p = (\gamma - 1)\left[e - \frac{\rho}{2}(u^2 + v^2 + w^2)\right] \tag{4}$$

The turbulent viscosity, μ_t, which appears in the viscous matrices, was computed using the Baldwin-Lomax turbulence model.[9]

The thin-layer equations are solved using the parabolized Navier-Stokes technique of Schiff and Steger.[7] Following the approach of Schiff and Steger, the governing equations, which have been modified here to include the Coriolis and centrifugal force terms, are solved using a conservative, approximately factored, implicit finite-difference numerical algorithm as formulated by Beam and Warming.[10]

The computations presented here were performed using a shock fitting procedure reported by Rai and Chaussee.[11] This procedure solves the five Rankine-Hugoniot jump conditions, two geometric shock propagation conditions, and one compatibility equation to determine the values of the five dependent variables immediately behind the shock, as well as the position of the shock. By including the implicit part of the source term due to the rotating coordinate frame in the circumferential inversion, the shock fitting procedure of Rai and Chaussee can be used without modification, as long as the free-stream conditions are modified to account for the rotation due to the coning motion.

At the body surface, no-slip, constant wall temperature boundary conditions were applied. For the cases with spin, the tangential velocity, v, was set equal to the local velocity of the body surface due to solid body rotation.

The computational results presented here were obtained using a grid that consisted of 60 points between the body and the shock. In the circumferential direction, gridding

was performed over a 360 degree sector due to the lack of symmetry from the combination of angle of attack, spin, and coning motion. Thirty six grid points were used in the circumferential direction. Grid resolution studies showed virtually no difference in the computed aerodynamic coefficients as the number of grid points were increased from 36 to 72 circumferential points using cylindrical coordinates. The computations were performed using a Cray X-MP supercomputer and typically required less than 20 minutes of CPU time for complete calculation over a single configuration.

RESULTS

Computations were performed to determine the aerodynamics of several axisymmetric shell configurations in steady coning motion. The first set of predictions is for a generic shell configuration; a secant ogive cylinder boattail (SOCBT). Calculations were performed over a range of Mach numbers and body lengths. The PNS results were compared with Euler code results. A second set of results was obtained for a series of ogive-cylinder configurations (Army-Navy spinner rocket, ANSR) which were fired through an aerodynamics range located at the Ballistic Research Laboratory. The computational results are compared with the aerodynamics determined from the in-flight motion of the projectile.

Results for the SOCBT

A schematic of the secant ogive cylinder boattail configuration (SOCBT) is shown in Figure 2. The parabolized Navier-Stokes results were compared with pitch damping results obtained using the inviscid code developed by Schiff[3]. Results were obtained for three body lengths (5, 6, and 7 calibers) and a range of supersonic Mach numbers. The center of gravity position used in the calculations was located at sixty percent of the body length from the nose. The PNS (viscous) results were obtained using the combined spinning and coning motion approach, while the Euler (inviscid) results were obtained using lunar coning motion. The use of lunar coning motion in the inviscid approach, instead of combined spinning and coning motion, is a consequence of the viscous nature of the spin boundary condition.

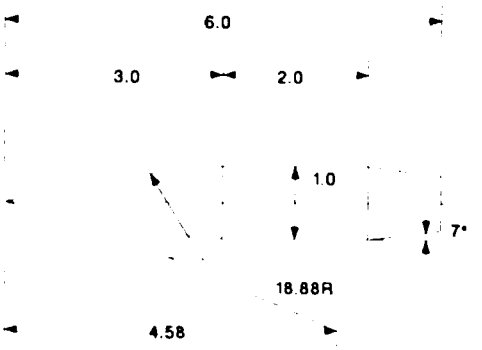


Figure 2. Schematic of SOCBT configuration.

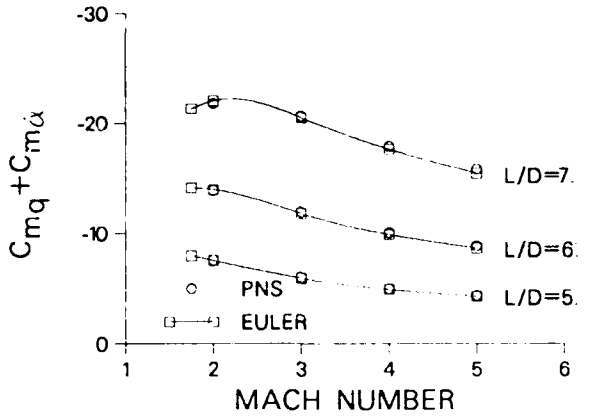


Figure 3. Pitch damping moment coefficient versus Mach number for various body lengths, SOCBT configuration

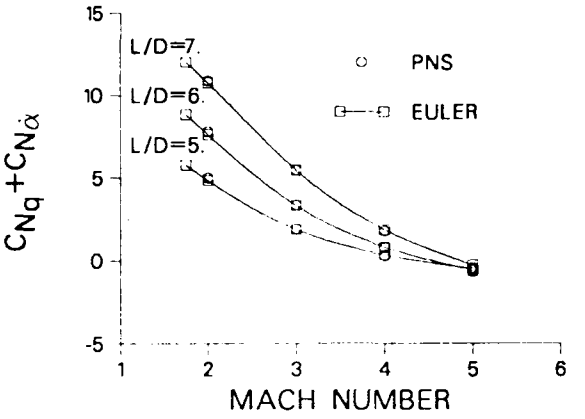


Figure 4. Pitch damping force coefficient versus Mach number for various body lengths, SOCBT configuration

Figure 3 shows the variation of the pitch damping moment coefficient as a function of Mach number for the three body lengths. In the supersonic regime, the pitch damping moment shows a decreasing trend with increasing Mach number. The inviscid results at $L/D = 7$ show a maximum near Mach 2. The results also show a significant increase in the pitch damping moment with increasing body length. Though not shown here, the computed results show that the boattail reduces the pitch damping moment by 20 to 30 percent compared with a completely cylindrical afterbody.

The comparison of the pitch damping moment predictions from the PNS approach and the Euler approach shows very good agreement across the range of parameters considered here. This is not surprising since the pitch damping appears to be primarily an inviscid phenomenon. However, the degree of agreement between the PNS and inviscid code is not an absolute indicator of the magnitude of viscous effects, since there does appear to be some code-to-code variation of the predicted pitch damping which is on the order of the viscous effect. The relevance of these differences is probably not significant especially in light of difficulties in measuring this coefficient experimentally.

The predicted variation of the pitch damping force coefficient with Mach number for the SOCBT configuration is shown in Figure 4. These results were obtained from the side force due to combined spinning and coning motion. The results show a strong decrease in the coefficient with increasing Mach number across the range of parameters examined here. The results also show an increase in the coefficient with increasing body length. The PNS results are in good agreement with Euler results.

Results for the Army Navy Spinner Rocket

Computations were performed for the ANSR series of projectiles. These projectiles were fired in an aerodynamics range and the aerodynamics determined from the projectile motion[12]. The projectiles consisted of a two caliber ogive nose with several different length cylindrical bodies as shown in Figure 5. The total body lengths were five, seven, and nine calibers. For each body length, projectiles were fabricated and fired with three different center of gravity (CG) locations. This allowed the aerodynamic forces to be determined from the variation of the aerodynamic moments with CG location.

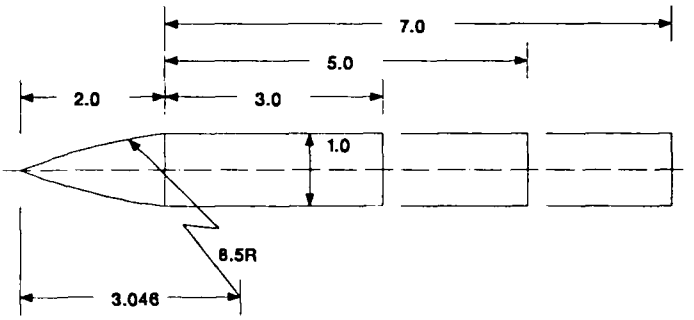


Figure 5. Schematic of the Army Navy spinner rocket (ANSR)

Figure 6 shows the variation of the pitch damping moment coefficient with CG location for the five, seven, and nine caliber bodies at Mach 1.8. Figure 7 shows a similar plot at Mach 2.5. In both of the figures, the computed results are compared with the experimental measurements. The computational results are in excellent agreement with the experimental data. The computational predictions were made by computing the flow field for each of the CG locations. This was performed because the coning motion produces a rotation about the CG, producing a different flow field in each case. These results are shown by the open triangular symbols. However, once the aerodynamics of a given configuration are determined, the CG translation relations[1] can be applied

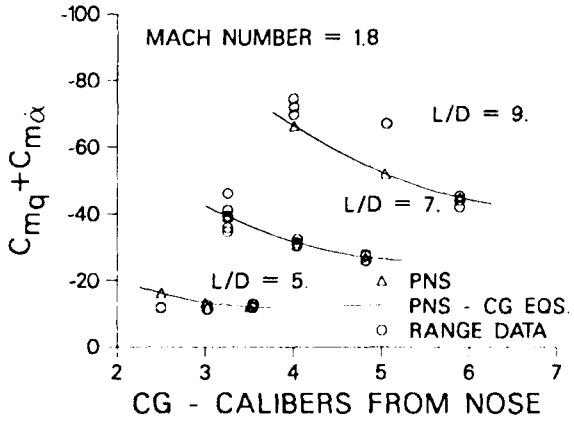


Figure 6. Pitch damping moment coefficient versus CG position, Mach 1.8, ANSR

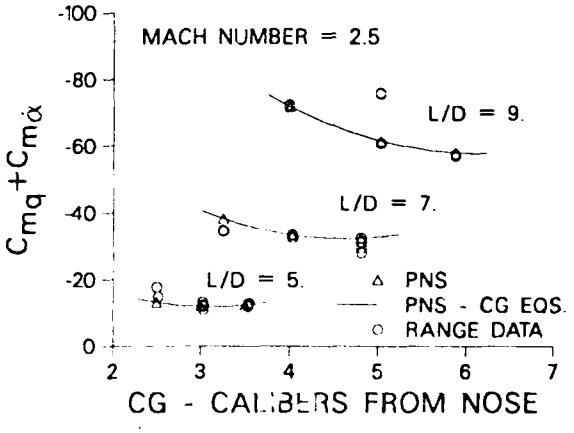


Figure 7. Pitch damping moment coefficient versus CG position, Mach 2.5, ANSR

to predict the aerodynamic coefficients for the same configuration with a different CG location. Several of these relations are shown below.

$$\begin{aligned} \hat{C}_{N_\alpha} &= C_{N_\alpha} \\ \hat{C}_{m_\alpha} &= C_{m_\alpha} - s_{cg} C_{N_\alpha} \\ \hat{C}_{N_q} + \hat{C}_{N_\alpha} &= C_{N_q} + C_{N_\alpha} + s_{cg} C_{N_\alpha} \\ \hat{C}_{m_q} + \hat{C}_{m_\alpha} &= C_{m_q} + C_{m_\alpha} - s_{cg}(C_{N_q} + C_{N_\alpha}) \\ &\quad + s_{cg} C_{m_\alpha} - s_{cg}^2 C_{N_\alpha} \end{aligned} \tag{5}$$

The aerodynamic coefficients for the modified configuration are denoted by the "^", while the aerodynamic coefficients for the baseline configuration are shown on the right-hand side of the equations. The center of gravity shift, s_{cg}, is in calibers and is positive for a CG shift towards the nose. Using these relations and the predicted aerodynamic coefficients for the middle CG position, the variation of the pitch damping moment coefficient with CG location was determined. This variation is shown in Figures 6 and 7 by the solid line. The difference between the pitch damping moment coefficients predicted from the CG translation relations and the pitch damping moment as determined from the direct computations is less than 0.1 % . This serves as a consistency check for the computational approach.

It is noted that at both Mach numbers, for the middle CG position of nine caliber body, there are several experimental data points which deviate from the trend shown by the predictions and the apparent trend shown by the experimental data. The cause of this is unknown.

As seen above from the center of gravity translation relations, the pitching moment coefficient, C_{m_α}, varies linearly with the CG shift. The slope of the variation is the normal force coefficient slope, C_{N_α}. By firing projectiles with the same external shape but with different center of gravity positions, the normal force can thus be determined from the variation of the pitching moment with CG location. A similar approach can be used to determine the pitch damping force coefficient from the variation of the pitch damping moment coefficient with CG location. Because the pitch damping moment varies in a nonlinear fashion with the CG shift, a modified damping moment is defined as shown below. When the left hand side of this equation is plotted as a function of CG shift, the results should be a line with slope equal to the pitch damping force coefficient.

$$\hat{C}_{m_q} + \hat{C}_{m_\alpha} - s_{cg}\hat{C}_{m_\alpha} = C_{m_q} + C_{m_\alpha} - s_{cg}(C_{N_q} + C_{N_\alpha}) \tag{6}$$

It should be noted that for all practical purposes, determination of the pitch damping force from range firings can only be obtained from the variation of the pitch damping moment with CG locations because the pitch damping force coefficient contributes very little to the in-flight motion of the projectile. Hence, direct determination is not practical.

From the CG variation of the pitch damping moment, the pitch damping force coefficient was determined from the experimental measurements. Figure 8 shows the variation of the pitch damping force coefficient with body length for the middle center of gravity location. Note that, unlike the normal force coefficient, the pitch damping force varies with CG position. The agreement between the computational predictions and experimental results are within the experimental accuracy and show the correct variation with body length and Mach number.

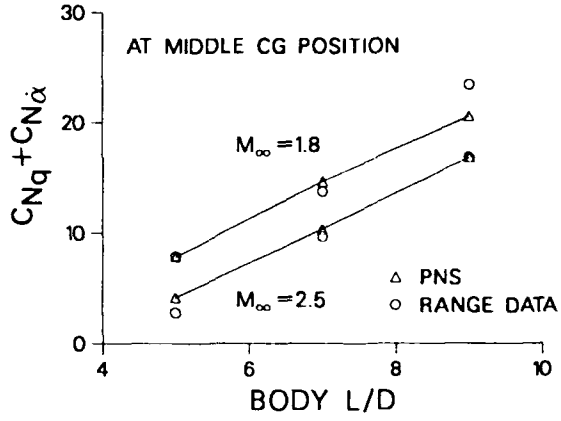


Figure 8. Pitch damping force coefficient versus body length, middle CG position, ANSR

As mentioned previously, the pitch damping predictions were obtained using the combined spinning and coning motion which allows the pitch damping force and moment to be determined directly from the side force and moment. The expected differences between applying combined spinning and coning motions, and lunar coning motion is reflected in the Magnus moment coefficient. In the current effort, the Magnus force and moment have been computed for the ANSR configuration and comparison made with range data obtained from the same series of firings as shown in Figures 9 and 10. The computed results were obtained for a fully turbulent boundary layer. (There is some evidence from the experimental program to indicate laminar flow over a portion of the body. particularly near the nose.) The computational results are in reasonable agreement with the experimental data. The predictions show that determining the pitch damping

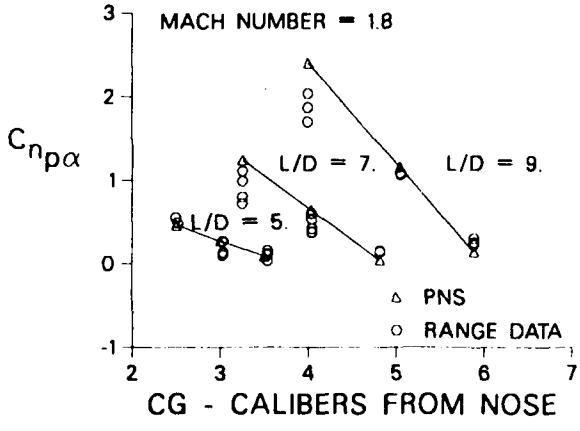


Figure 9. Magnus moment coefficient versus CG position, Mach 1.8, ANSR

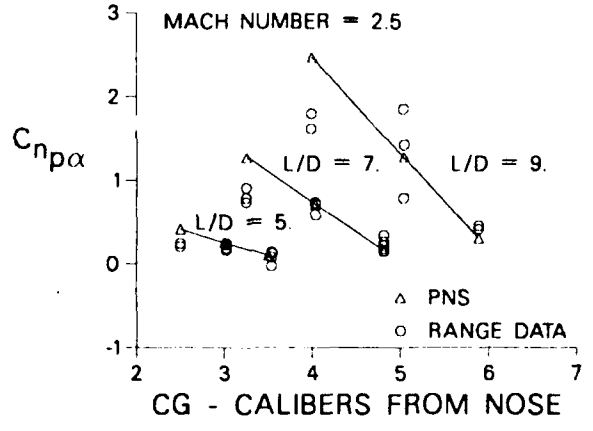


Figure 10. Magnus moment coefficient versus CG position, Mach 2.5, ANSR

coefficient from the side moment due to lunar coning motion and completely ignoring the Magnus moment will result in errors of less than 5 percent for this configuration.

Results were also obtained to determine the pitch damping coefficient from separate calculations involving lunar coning and spin. The results, not shown here, are in excellent agreement with the results obtained using a single calculation utilizing combined spinning and coning motion. The maximum difference between the two approaches was less that 0.1 % . This demonstrates the lack of coupling between the spinning and coning motions over the range of coning rates, spin rates, and angles of attack considered here. (It interesting to note that at higher angles of attack, a non-linear variation of Magnus moment with angle of attack was predicted. Even at these angles of attack, no coupling between the coning and spinning motions was observed.)

Predicted In-flight Motion of the ANSR

Prediction of individual aerodynamic coefficients, such as pitching moment, lift, or pitch damping moment coefficients, can provide useful information about some of the aerodynamic characteristics of a given projectile. However, if the in-flight motion of the projectile is desired, a minimum of eight aerodynamic coefficients must be determined. Various researchers have provided capabilities for predicting many of these coefficients. With the development of a capability to predict the pitch damping force and moment coefficients, it is now possible to determine all of the necessary coefficients required to predict the in-flight motion of axisymmetric shell using Computational Fluid Dynamics (CFD) techniques. To date, this capability has been demonstrated for supersonic, small angle of attack flight. Extension of the capability to predict pitch damping in the transonic and subsonic flow regimes appears possible and is the subject of on-going research.

As a demonstration of this capability, the in-flight motion of the ANSR configuration was predicted using coefficients determined using the computational approach. The predicted motion was compared with the motion obtained from the range firings. Figure 11 shows the angular motion of the five caliber long projectile launched at Mach 1.8. This three-dimensional plot displays the vertical and horizontal components of angle of attack, α and β, as a function of the distance downrange, Z. The thickened curve displays the three-dimensional motion. Two-dimensional projections of the motion in the vertical and horizontal planes are also shown. It is useful to regard the curves shown here as the path traversed by the nose of the projectile as it flies downrange. The motion predicted using the CFD derived aerodynamic coefficients is denoted by the solid lines, while the motion obtained from the range firings is shown by the circular symbols. The agreement between the two motions is excellent.

Figure 12 shows a similar result for the in-flight motion of a seven caliber long version of the ANSR launched at Mach 2.5. Again, excellent agreement is observed.

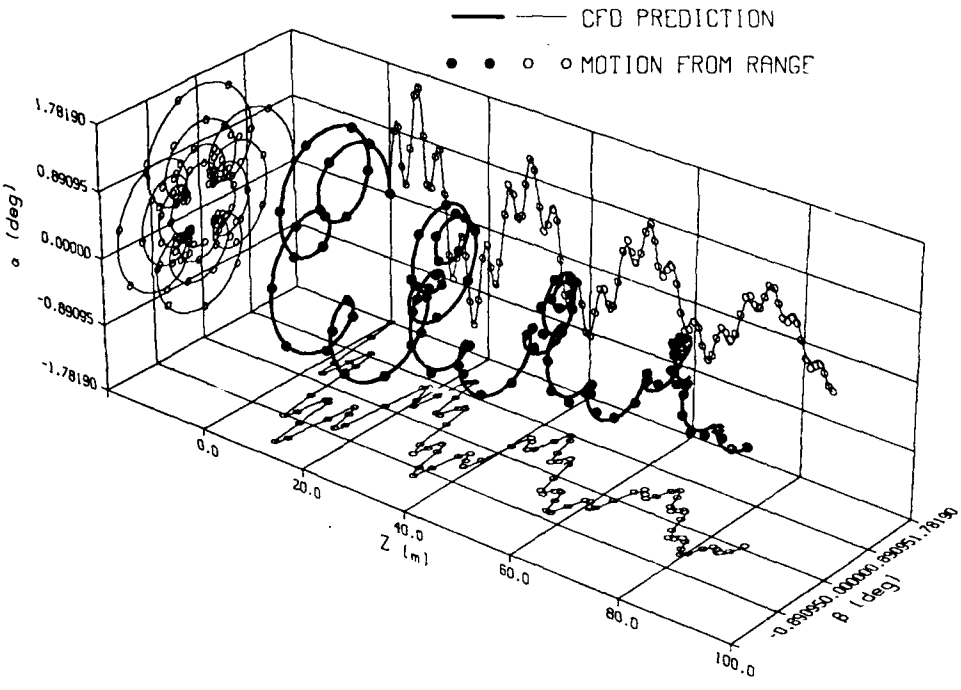


Figure 11. In-flight Yawing Motion for ANSR projectile, $L/D = 5$, Mach 1.8

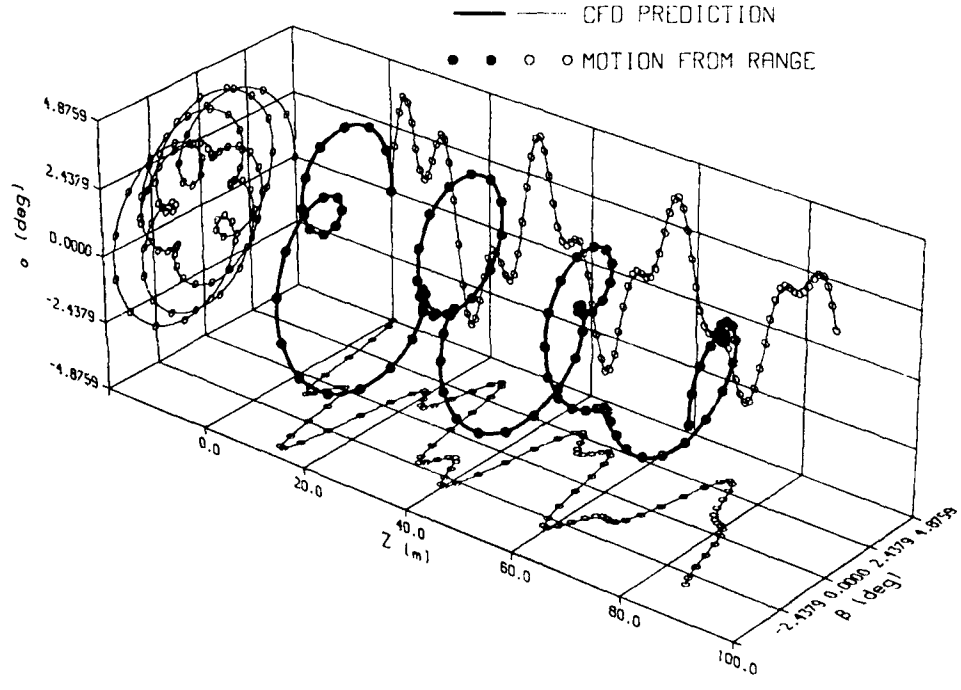


Figure 12. In-flight Yawing Motion for ANSR projectile, $L/D = 7$, Mach 2.5

CONCLUSION

A computational approach for predicting the pitch damping coefficients using steady coning motion has been successfully applied to several axisymmetric shell configurations. Through the use of a combined spinning and coning motion, the pitch damping force and moment have been obtained directly from the side force and moment using a single calculation. This approach does not require that the Magnus force or moment be ignored or determined from an auxiliary calculation, as in the case of lunar coning motion.

The computational predictions for the SOCBT configuration showed good agreement with results obtained with a previously published inviscid code. The results showed an increasing trend in the pitch damping coefficient with increasing length-to-diameter ratio and a decreasing trend in the coefficient with increasing Mach number. The predictions of both the pitch damping force and moment coefficients for the the Army Navy spinner rocket were seen to be in excellent agreement with the data obtained from aerodynamic range testing. The computational results predicted the correct variation in the pitch damping moment coefficient with changing center of gravity location, body length, and Mach number. For each of the configurations examined here, the effect of viscosity on the pitch damping coefficients was small.

The capability for predicting individual aerodynamic coefficients, such as the pitch damping force and moment coefficients, is significant because it allows particular aspects of the projectile aerodynamic performance to be evaluated. The ability to predict the pitch damping coefficients has additional significance because a complete capability based on numerical solution of the Navier-Stokes equation now exists for determining the basic set of aerodynamic coefficients required to predict the supersonic, small angle of attack, in-flight motion of axisymmetric projectiles. For the Army Navy spinner rocket, the predicted in-flight motion obtained from the computed aerodynamic coefficients showed excellent agreement with the inflight motion obtained from range firings.

The use of the combined spinning and coning motion is currently being applied to time marching codes for the prediction of pitch damping at subsonic through low supersonic velocities. In this velocity regime, it is expected that viscous effects will be of greater importance than in the supersonic regime examined here.

References

1. Murphy, C.H., "Free Flight Motion of Symmetric Missiles," U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, Report No. 1216, July 1963. (AD A442757)

2. Schiff, L.B., and Tobak, M., "Results from a New Wind-Tunnel Apparatus for Study-

ing Coning and Spinning Motions of Bodies of Revolution", AIAA Journal, Vol. 8, No. 11, November 1970, pp. 1953-1957.

3. Schiff, L.B., "Nonlinear Aerodynamics of Bodies in Coning Motion", AIAA Journal, Vol. 10, No. 11, November 1972, pp. 1517-1522.

4. Agarwal, R., and Rakich, J.V., "Computation of Supersonic Laminar Viscous Flow Past a Pointed Cone at Angle of Attack in Spinning and Coning Motion", AIAA Paper 78-1211, AIAA 11th Fluid and Plasma Dynamics Conference, Seattle, WA, July 1978.

5. Lin, T.C., "A Numerical Study of the Aerodynamics of a Reentry Vehicle in Steady Coning Motion", AIAA Paper 78-1358, AIAA Atmospheric Flight Mechanics Conference, Palo Alto, CA, August 1978.

6. Weinacht, P., and Sturek, W.B., "Navier-Stokes Predictions of Pitch Damping for Finned Projectiles Using Steady Coning Motion," AIAA Paper 90-3088, AIAA 8th Applied Aerodynamics Conference, Portland, OR, August 20-22, 1990.

7. Schiff, L.B., and Steger, J.L., "Numerical Simulation of Steady Supersonic Viscous Flow," AIAA Journal, Vol. 18, No. 12, December 1980, pp. 1421-1430.

8. Levy, L.L., and Tobak, M., "Nonlinear Aerodynamics of Bodies of Revolution in Free Flight", AIAA Journal, Vol. 8, No. 12, December 1970, pp. 2168-2171.

9. Baldwin, B.S., and Lomax, H., "Thin Layer Approximation and Algebraic Model for Separated Turbulent Flows," AIAA Paper 78-257, 16th Aerospace Sciences Meeting, January 1978.

10. Beam, R., and Warming, R.F., "An Implicit Factored Scheme for the Compressible Navier-Stokes Equations," AIAA Journal, Vol. 16, No. 4, 1978, pp. 85-129.

11. Rai, M.M., and Chaussee, D.S., "New Implicit Boundary Procedure: Theory and Applications," AIAA Paper 83-0123, Reno, Nevada, January 1983.

12. Murphy, C.H., and Schmidt, L.E., "The Effect of Length on the Aerodynamics Characteristics of Bodies of Revolution in Supersonic Flight," U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, Report No. 876, August 1953.

REVIEW OF ENVIRONMENTAL RESEARCH SPECIFIC TO SMART WEAPONS OPERABILITY ENHANCEMENT FOR THE BATTLEFIELD ENVIRONMENT

*Dr. James P. Welsh, U.S. Army Corps of Engineers, Cold Regions Research & Engineering Laboratory, Smart Weapons Operability Enhancement, 72 Lyme Road, Hanover, NH 03755-1290

Mr. Mike Hardaway, U.S. Army Corps of Engineers, Topographic Engineering Center, Fort Belvoir, VA 22060-5546

Mr. Wade West, U.S. Army Corps of Engineers, Waterways Experimentation Station, 3909 Halls Ferry Road, Vicksburg, MS 39180-6199

SWOE Scene Generation Process

The SWOE scene generation process is an integration of measurement, information bases, models and rendering capabilities. The Corps of Engineers (CoE) Tech Base effort, relevant to SWOE, is focused in four areas: measurements, information bases, modeling, and rendering.

Measurement and Information Bases

What, where, when, and how to measure, to obtain representative samples, and infer energy interactions between surface and volume scatters (including spatial scales of less than 0.5 meters). Construction of information

bases for a global variety of battlefield environment conditions, particularly to develop site-to-site and area-to-area comparisons methods. Figure 1 summarizes the CoE effort for measurement and information bases.

Figure 1. Tech Base Measurement & Information Bases

△ WES □ TEC ○ CRREL

TASK			
Measurements	Classification & Regression Based Winter Parameters ○	95 GHz Measurements Temperate & Desert ○△	Winter Scene Sampling ○
	Standardize Snow Microstructure Methods ○	Forest Edge MMW Temperate ○△	Surface Energy Exchange in Winter ○
	3-D Scene Parameters Temperate ○△	Slant Range IR Temperate & Desert ○△	Surface Scatter Polarimetrics ○△
	35 GHz Measurements Temperate ○△	Forest Edge IR Temperate ○△	
Evaluation	IR Metrics Ground to Ground ○△	MMW Metrics Air to Ground ○△	
		IR Metrics Air to Ground ○△	MMW Metrics Ground to Ground ○△
		Correlation to Scenes ○△	
Info Bases	Data Compression △	Microstructure Info Base Snow Covered Area ○	MMW Signature Data Base ○△
	Interface GIS & DBMS △	Prototype 3-D Temperate, Desert & Tropic Info Bases ○△	IR Signature Data Base ○△
	Texture Reference Data △	Transient Atmospheric Phenomena Info Base ○	
		Enhanced Data Base Procedures △	Enhanced Data Compression △

An integral part of the 3-D scene generation procedure, the information base contains all spatial and attribute data required to define the total landscape environment (land, water, atmosphere, sky, etc.). The content and

structure of the information base are driven largely by the various requirements of the numerical models and the scene generation software.

The information base contains three kinds of data: digital terrain data (e.g., topography, soil types, vegetation types); physical, thermal, and spectral terrain attribute data (e.g., moisture content, emissivity, reflectance); and meteorological (weather and atmospheric) data (e.g., air temperature, visibility).

Digital terrain data are representations of portions of the earth's surface stored in computer-compatible formats. These data depict characteristics such as elevation, vegetation types, soil types, and other relevant environmental information. Digital terrain data used in the scene generation procedure are stored in raster and vector formats and managed by a geographic information system (GIS).

The physics-based thermal signature prediction models used in the scene generation procedure require as inputs complete, quantitative descriptions of the physical, thermal, and spectral attributes of each landscape feature. These data are most efficiently stored and retrieved in tubular format in a relational database management system (RDBMS). The RDBMS associates each stored numerical value with the corresponding landscape feature depicted in the GIS.

Meteorological data are required throughout the scene generation process and have particular importance to the radiation field prediction models. Both surface weather and upper atmospheric profile data are required. These data are also stored in tabular format in the RDBMS.

Models

The goal of the SWOE modeling effort is to assemble and integrate 3-D fundamental physics models of important environment phenomena and objects (natural & manmade) for the IR and MMW spectral bands. Specific objects are:

trees, with & without leaves, buildings, vehicles, roads, bridges, etc. The boundary regions between adjacent objects, tree canopies, row crops, forest edges, and other textured surfaces are also important. Models of the energy budget are significantly effected by heterogeneity in the 3-D distribution of energy emitters and scatters. Figure 2 summarizes the CoE effort relevant to modeling.

Figure 2. Tech Base Modeling

△ WES □ TEC ○ CRREL

Task			
Energy Budget — 3-D Models	Geometric Mesh & Mechanics For 3-D Multiphase Calculations ○ Radiosity & Ray Tracing Methods For 3-D △ Low Resolution Mesoscale Weather Model △	3-D Multiphase Snow & Soil Model ○ Intermediate Resolution Mesoscale Weather Model △	3-D Model With Water Transport ○ Total Energy Flux Model △ High Resolution Mesoscale Weather Model △
Energy Budget — IR	First Generation Laser Scattering Model △	IR Scene Composit Models (Buildings, Cultural Features Etc.) ○ △	
Energy Budget — MMW	Evaluation Of Lang 35GHz Snow & Vegetation Models ○ Active & Passive Signature Modeling Concepts ○ △	Extension Of Lang Model To 95 GHz For Snow ○ First Gen Coupled 35 GHz Surface & Volume Scattering Models (Shi) ○ Interim Single Band Signature Model △	Extension Of Lang Model To 95 GHz For Vegetation ○ Integrate 35 GHz Lang & Coupled Models (Lang / Shi) ○ Multi-band Signature Model ○ △
Texture — IR	Statistical IR Snow Relations ○ Second Generation Synthetic IR Texture Procedure △	Statistical IR Mixed Snow & Ground Model ○ First Gen Physics Based Texture Models ○ △	Hybrid IR Snow & Frozen Ground ○ Refined Physics Based Texture Models △
Texture — MMW	First Generation Synthetic MMW Texture Concepts ○ △	Statistical MMW Snow Texture Model ○ First Generation MMW Vegetation & Soils Synthetic Texture Model △	Statistical MMW Mixed Snow & Ground Model ○ Enhanced MMW Vegetation & Soils Texture Model △
Boundary Layer		Model For Turbulent Heat Flux Exchange ○	Mesoscale Transient Boundary Layer Model ○

The SWOE Interim Thermal Model is used to calculate the surface temperatures for a wide variety of surfaces, including vegetated and non-vegetated surfaces, bodies of water, and snow/ice-covered surfaces. The model results are valid for all seasons.

The thermal models, in the package, are driven by conventional weather data, such as standard surface weather observations and radiosonde data. Default databases of seasonally dependent thermal properties are provided to cover a set of standard surfaces that are commonly encountered in scene simulation.

The SWOE thermal models package accommodates various vegetation effects. The effects of simple vegetation, such as grasses and crops, and forests can be included in the 1-D heat balance of soils. A separate 3-D model of the thermal balance for individual trees is included. Two geometric representations of trees, based on measurements of actual trees, have been included to calculate the temperature fields for trees.

The SWOE thermal models are driven by the radiation fields from the atmosphere. The atmospheric radiation budget is calculated using a modified version of LOWTRAN7, which is the standard atmosphere radiance and transmission code used by the DoD community.

The SWOE Radiance models software system contains two parallel computational paths, one for terrain and one for 3-D objects (currently individual trees and two military targets). The terrain radiance path is built around a new Fortran model, called IBRM ("Improved Background Radiance Model").

Radiance's for 3-D objects are computed with the Hardbody module of the SPIRITS code, a U.S. Government standard for aircraft. Both radiance models utilize the same basic algorithms and phenomena, which include:

- radiance's computed spectrally at 2 to 20 cm-1 resolution, and bandpass integrated (with optional filter function) only after atmospheric effects are added;
- radiance sources of thermal emission, the sun, the sky, and surrounding terrain;
- sky emission from broken clouds;
- solar shadowing;
- spectral directional emissivities for each material;
- a spectral bidirectional reflectivity for each material;
- spectral atmospheric transmission and radiance (thermal and solar scatter) along all paths connecting the terrain, sun, sky, and sensor, utilizing the Air Force MODTRAN model (an upgrade to LOWTRAN7).

A separate model, SHADOW, automatically generates faceted shadows of the 3-D objects for inclusion within the scene.

The terrain is modeled with a set of textured polygons which overlay the topography grid. The polygon definitions and geometry are determined as part of the information base effort. The radiance models software computes a list of in-band radiance's for each polygon, based in part on temperatures computed by the SWOE thermal models.

Trees and targets are described with a triangular geometry, typically with 3000 to 20,000 triangles per object. Tree geometry's are based on trunk and branch measurements taken from sets of real tree measurements. Faceted leaves are generated using a fractal technique. The resulting geometry, plus a file with a separate temperatures for each triangle, are input to Hardbody for the tree radiance computations.

Targets require that the user have or prepare ahead a set of computed target temperatures for the scene specified conditions; Hardbody then computes the radiance's. Utility software is provided to convert thermal computations to the Hardbody format. Hardbody computations include facet-to-facet reflections, in addition to the those listed above. It also computes images for the objects as individuals.

Clouds are one of the more important modulators of the surface energy balance. The SWOE model package considers the influence of clouds for solar and infrared downwelling flux. During the thermal loading phase, also known as the model spin-up phase, a simple model is used to modulate the broadband downwelling flux in both spectral regions based on the geographical location of the scene, time, surface characteristics (slope and albedo), atmospheric conditions, cloud amount, and cloud type. This approach does not provide the radiant field information required at the time of the scene simulation. At scene simulation time a modified version of LOWTRAN is used to calculate the spectrally dependent solar direct and diffuse, and infrared downwelling flux. This information is used in the computation of reflections off of and between scene elements; and absorption and scattering by atmospheric gases, aerosols, and clouds. Cloud shadows at the time of the scene simulation are generated by the Cloud Scene Simulation Model (CSSM). The CSSM uses a Successive Random Additions (SRA) fractal algorithm to generate the horizontal distribution of the clouds based on the cloud amount and type (stratiform, cirriform, or cumuliform). 1-D SRA and 2-D SRA algorithms are used to generate the upper and lower surface of the cloud while a 3-D SRA algorithm is used to modulate

the liquid water density (LWD) information at each cloud grid point in the cloud volume. The mean LWD information as a function of cloud type and altitude has been obtained from an extensive cloud database. In the future, the LWD information will be used to determine the cloud microphysical and optical properties for use in a model that will calculate the full 3-D cloud radiative interactions. The scene generated cloud characteristics are controlled by the Hurst and Lacunarity parameters in the SRA algorithm. Model default values controlled by the cloud type are used in the cloud simulation, but the user can modify these parameters. Cloud shadows are determined using a ray tracing technique, the 3-D cloud spatial distribution, and the solar azimuth and zenith angle or scene location and time of year and day.

Rendering

The rendering effort of the CoE laboratories is summarized in Figure 3. The SWOE rendering software provides the capability necessary to create 2-D visualization of 3-D objects and backgrounds. The software uses the depth buffer approach to resolve hidden surfaces. The output is a projection of the data contained in the information base onto a 2-D image or pixel file. The input files contain physics models, initialization information, haze and lighting, and viewpoint . These inputs are used to generate the pixel data to create an image using the following processes:

- Viewpoint manager - processes input initialization data and viewpoint information to generate a sun vector, ambient and diffuse lighting parameters, bounding planes, and the world space to a viewpoint space transformation matrix, which is referred to as "viewpoint data".

- World manager - uses viewpoint data and bounding planes to select root nodes in the data base for the terrain region. A node is defined as a subset

of the information base which contains position data, information for level of detail (LOD) and field of view (FOV), materials properties, etc. and pointers to the items associated with each node.

- Node processor - traverses the node tree and uses the bounding planes to determine which nodes are in the FOV. The FOV test creates a list of active nodes, loads texture maps and allocates nodes to the correct LOD.

Figure 3. Tech Base Scene Rendering

△ WES □ TEC ○ CRREL

	TASK			
Object Rendering	2-D Features	Standard Data Structures □	Non- Standard Data Structures □	
	3-D Solids	Initial Model Library □ Solid Model Insertion □	Expanded Model Library □ Solids Model Interactions □ Textured Ellipsoid □	Model Library Editing Functions □ Dynamic Models □
	Amorphous Objects	COMBIC Smokes □ Multiple Point Sources □	Other Obscurrents □ Clouds / Fog / Haze / Humidity □	Terrain / Wind Dynamics □
Techniques	Texture	Synthetic IR Texture Mapping □	Pixel Map Texturing □	Physics Based Texture □
	Rendering	Terrain Shadowing □ Anti-Aliasing □ Filtering / Smoothing □	Attribute Capability □ Integrated Z-Buffering / Ray Tracing / Textured Ellipsoid □ Solids Objects Shadowing □	Amorphous Object Shadowing □ Tracker Algorithm □ Radiosity Rendering □
	MMW	Technology Assessment □ Preliminary Geometry □	Interface Sensor Models □	Initial Phase / Brightness Soft ware □

- Item processor - transforms sun vectors, calculates triangle face normals, eliminates back faces, determines polygon coloring and shading, converts vertices to viewpoint space, clips to hither plane, and projects polygons to screen space.

- Pretiler - clips polygon to screen space and creates triangles with incremental color, depth, and texture information.

- Tiler - produces pixels for display for each triangle through the graphics processor, and may modify color attributes of textured items.

The result of the rendering is a 2-D pixel space representation of the radiometric energy arriving at the aperture of a sensor for a specified viewing geometry.

Summary and Conclusions

The SWOE concept, as developed, depends on the research products of DoD technology base programs. The CoE technology base contributions to the SWOE program are significant. CoE has taken the lead by commitment to a focused tech base effort required for enhancement of future smart weapon system performance. This commitment has resulted in extremely high levels of coordination and cooperation between CoE laboratories, as well as, serving as a model for unprecedented levels of coordination and cooperation between all the armed forces in the SWOE program.

Novel Techniques to Determine Ecological Effects at U.S. Army Sites

Dr. Randall Wentsel, CRDEC, APG, MD 21010-5423*
Dr. Ronald Checkai, CRDEC, APG, MD 21010-5423
Dr. Robert Parmelee, Rutgers Univ., New Brunswick, NJ 08903

To support U.S. Army Installation Restoration and Base Closure programs, Ecological Risk Assessment (ERA) is required to assist in the determination of the appropriate clean-up levels. The increase in a CERCLA-required clean-up criteria by an order of magnitude can cost millions of dollars. Therefore, technically sound environmental data that can be defended to the scientific and regulatory community are critical to reach informed decisions at U.S. Army sites. Proposed ERA procedures have been developed by the U.S. Environmental Protection Agency (EPA). Important aspects of the ERA procedure are the characterization of ecological effects, and characterization of exposure. The accurate characterization of the fate and effects of contaminants at a site will produce a technically sound ERA.

To support ERA, novel techniques are needed to address critical environmental questions in site clean-up. We have developed two such techniques. The transport and transformation of chemical contaminants (fate) in soil have been studied using intact soil columns. To assess ecological effects, research was conducted to develop a soil fauna microcosm.

The controlled environment soil-core microcosm unit (CESMU) controls environmental parameters that impact on soil chemical, physical, and biological properties. These methods embodied a collection of techniques that began with soil sampling in the field, and continued throughout the laboratory investigation of chemical fate, migration, and degradation in site-specific soils; it was a cost-effective investigative methodology that can be used to screen chemical materials before initiating high-cost environmental field studies. Intact soil cores were collected in the field using a hydraulically controlled probe, delivering intact soil-cores with minimal disturbance directly into high density polyethylene pipe (10.3-cm ID). The inert polyethylene pipe was an effective hydrophobic barrier that remained an integral part of the soil-core column, obviating subsequent transfers of soil. Columns were sealed in the field and transported to the laboratory. In the laboratory, each soil-core was fitted with a porous ceramic plate, and a polyethylene endcap containing fittings for teflon tubing (Fig 1). Intact soil columns were then transferred into the CESMU chamber, a controlled temperature unit with capacity sufficient for maintaining a constant temperature within entire soil columns. Controlled tension

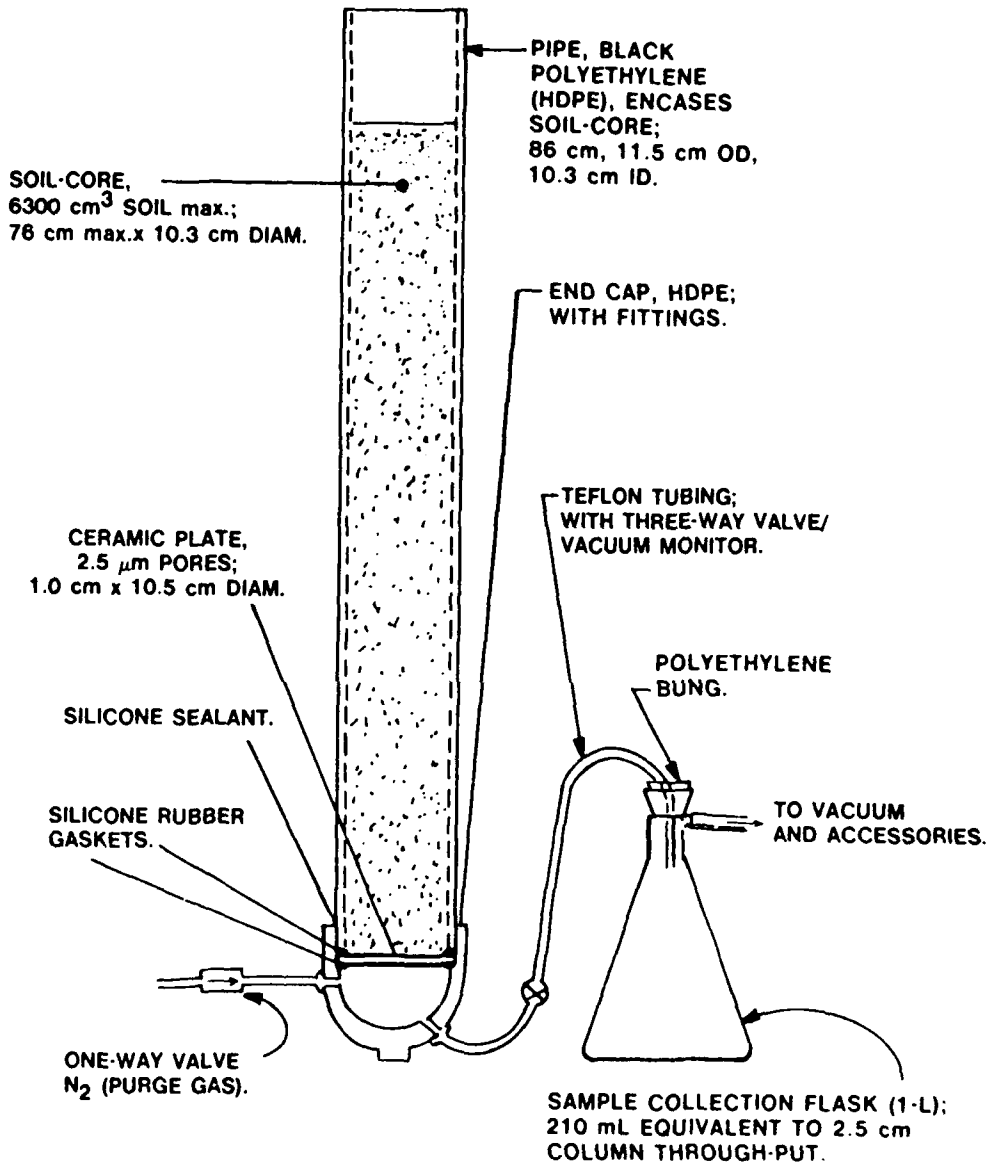


Figure 1. Intact Soil Column Unit

(vacuum) was applied equally at the bottom of each soil column across the controlled-pore ceramic plate, at 30-35 kPa; tension was regulated and monitored. The tension that was applied is comparable to that encountered in the field as a result of combined soil matric and gravitational forces; thus avoided were undue flooding, the buildup of a handing column of water in the lower portion of columns, and artificial changes in soil redox potential in response to steady-state alteration of the soil water content, as can happen when gravitational forces alone are relied upon to promote water flow through soil columns. Soils were saturated with water and equilibrated (48h minimum), allowing them to come to field capacity before initiating studies. Artificial rain was added twice per week by peristaltic pump, at rates simulating rainfall. Leachate solutions were collected under tension via teflon tubing into flasks in darkness, and kept at soil column temperature inside CESMU until harvested for analyses. During the investigations, soil columns were harvested at intervals for sectioning by depth, extraction, and soil analyses. The munitions under study were added atop the study columns within a subsample of the soil.

The delivery and application (input) of artificial rainwater, formulated to represent the major constituents and acidic pH of rainfall in the mid-Atlantic coastal region, was driven by peristaltic pumps and delivery occurred at the top of each soil-core column via capillary tubing, applied at the center of each soil-core at the rate of 7 $\mu m\ s^{-1}$ (1 inch per hour). Artificial rainwater input on each column occurred twice per week on Mondays and Fridays during the course of studies, and simulated semiweekly rainfall. This method did not consider any losses of rainfall due to runoff as may occur in the field situation, and made each rainfall event equal; however, rainfall input was calculated from average native conditions at the sampling sites, and semiweekly additions allowed for wetting and drying cycles that simulated natural field conditions. Solution thru-put (leachate) was collected via teflon tubing into 1-L flasks in darkness, and kept at soil column temperature inside CESMU until removed for analyses. Nitrogen gas was provided for purging the end-cap but only during collection of aqueous samples, eliminating the entry of ambient air into the bottom of columns during leachate solution removal.

Replicate treatment columns were harvested at regular intervals during the course of studies, sealed, and frozen; then the HDPE pipe containing the frozen soil-core was carefully cut length-wise using an electric router and hand guide, allowing the resulting intact soil-core to rest in the lower half of the HDPE pipe. The soil-cores were then slowly thawed in the horizontal position to effectively eliminate longitudinal migration, then the soil was sectioned into 2.5 cm sections by depth.

In these studies the fate, migration, and degradation of munition residues were investigated. Results from one set of studies are presented in Figures 2 and 3. The transport of HMX and RDX in the soil columns indicated that RDX moved rapidly through the soil. After 6.5 weeks RDX had migrated through the soil column (27 inches). HMX was not as mobile as RDX. After 6.5 weeks HMX had migrated to a depth of 9 inches. At 13 weeks the compound was measured at a depth of 15 inches, and by 19 weeks HMX had migrated through the silt loam soil column.

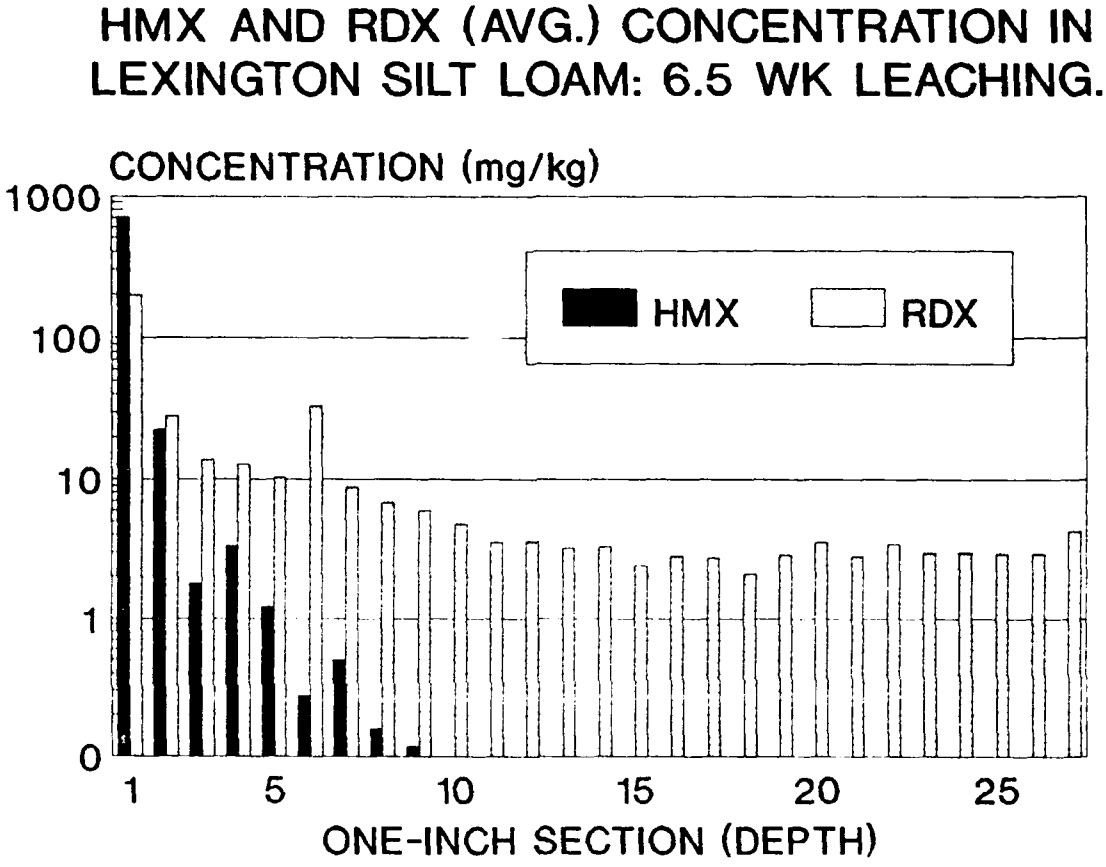


Figure 2. HMX and RDX (Avg.) Concentration in Lexington Silt Loam: 6.5 Wk Leaching

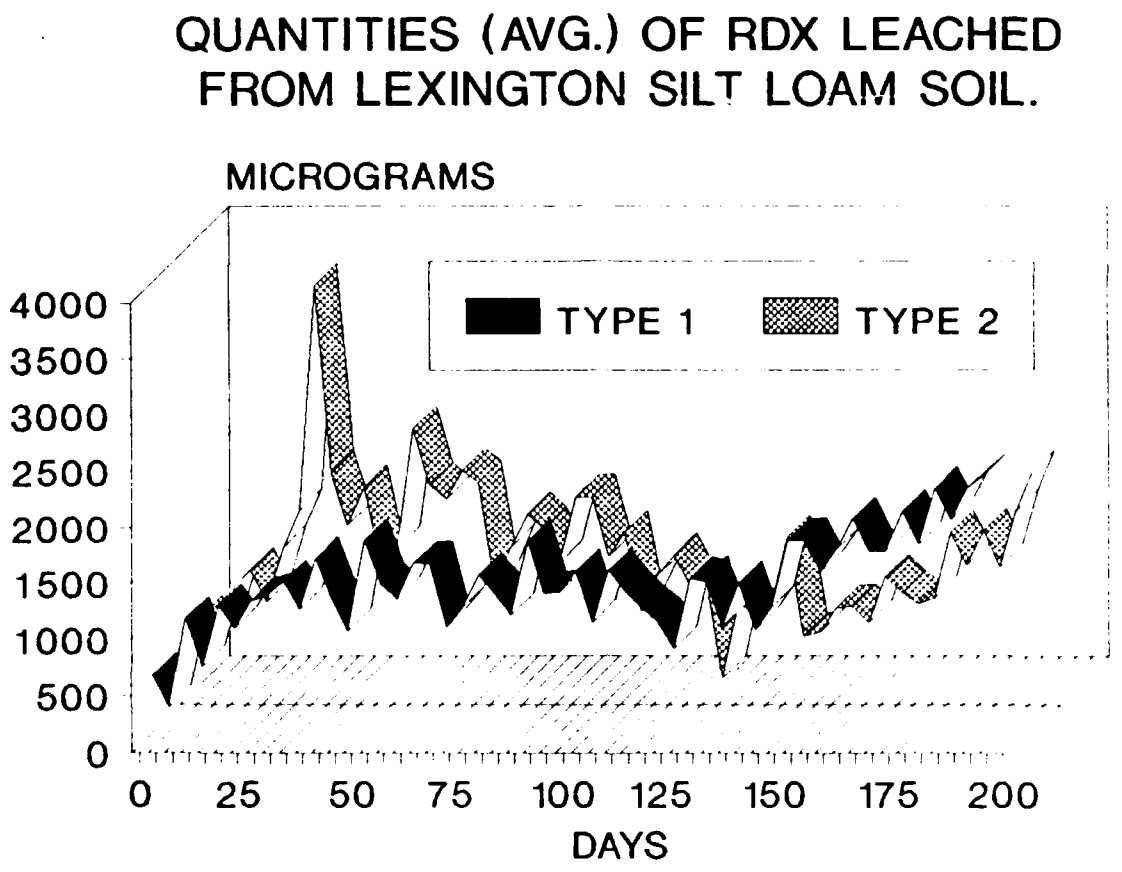


Figure 3. Quantities (Avg.) of RDX Leached From Lexington Silt Loam Soiol.

As was observed in the soil data, RDX moved rapidly through the column and it was observed at day 3 in the leachate samples. Half of the soil columns (type 2) had a breakthrough peak occurring near day 35. The remaining columns did not have a noticeable peak but had a broad and gradual increase in RDX concentrations over time. The mass balance data are presented in Table 1. For RDX 58% remained in the soil, while 12% leached through the soil column. HMX did not migrate to as great a degree with 87% in the soil; primarily in the top 2 inches. The mass balance for HMX and, RDX were 88 and 70%, respectively, indicating good recovery of the compounds in the soil and leachate samples, and thus good fate estimates.

Many of the methods used to characterize ecological effects of chemicals on soil systems are single species or a battery of single species tests. Most do not assess population interactions or community and system levels effects of toxicants on ecosystems. Existing procedures measure the acute effects of chemicals on inhibition of seed germination, and seeding growth, soil microbial biomass, earthworms and isopods.[1]

At superfund sites other methodologies have been developed to assess chemical impacts. These include methods such as: honey bees, earthworms, crickets, and plant mutagenicity tests.[2]

Single species tests are not adequate to assess effects of chemicals on exceedingly complex soil decomposer food webs, and more sophisticated tests are needed to enable scientists to accurately assess ecological damage to terrestrial ecosystems. Tests which can measure community level responses, and structural and functional changes would generate valuable information for ecological risk assessors and risk managers. Analysis of soil microfauna (nematodes and microarthropods) communities and trophic structure in response to chemicals may prove to be particularly useful in achieving this goal.

Community and trophic analysis of field-collected nematode and microarthropod communities has major advantages over ecotoxicological methods currently in use. Soil microfauna communities typically have large numbers of individuals (10^4 to 10^6 m^{-2}), are very diverse, and contain bacterivores, fungivores, herbivores, omnivores and predators. Therefore, they provide a large number of different species and trophic levels on which to test for direct effects of contaminants. Analysis of both nematode and microarthropod communities also can provide information on toxicology in different soil microsites; nematodes inhabit the water film around soil particles and microarthropods inhabit air-filled soil pores.[3] Soil fauna affect ecosystem processes such as decomposition and mineralization largely by regulation of the microbial community. Trophic analysis can then provide evidence for indirect effects of toxicants and disruption of soil food web structure and alteration of ecosystem processes. For example, elimination of microarthropods with insecticides revealed that microarthropods regulate litter decomposition and N-dynamics by feeding on fungi and bacterivore nematodes in deserts,[4] and by grazing on fungi in agroecosystems.[5] Therefore, soil fauna communities are ideal mode assays to test the effects of chemical pollutants because they have a large number of species and different trophic levels to detect direct effects and because trophic analysis can provide indirect evidence for disruption of food web structure and

HMX & RDX RECOVERED (AVG. OF DUPLICATES) IN LEACHATES & SOIL*: 6.5 WK LEACHING.

	HMX	RDX
AMOUNT INITIALLY ADDED (mg)	160	160
RECOVERED IN LEACHATE (mg)	1.0	19.4
PERCENT OF INITIAL	0.6%	12%
RECOVERED IN SOIL (mg)	139	93.1
PERCENT OF INITIAL	87%	58%
TOTAL RECOVERED (mg)	140	112
PERCENT RECOVERY	88%	70%

* LEXINGTON SILT LOAM SOIL

Table 1. HMX & RDX Recovered (Avg. of Duplicates) in Leachates & Soil*: 6.5 Wk Leaching.

function.

Our objective was to develop a simple inexpensive microcosm technique to evaluate the effect of chemical contaminants on field populations of soil nematodes and microarthropods. We used community and trophic level analysis of natural populations to investigate the ecological effects of copper and several other compounds, to demonstrate the utility of a microcosm approach.

Soil was collected from a mature oak-beech forest with a well developed understory on the property of the U.S. Army Aberdeen Proving Ground in Edgewood, Maryland, U.S.A. The organic horizon was removed and from the top 10 cm of A-horizon soil was collected, mixed and sieved through a 1.3 cm mesh screen. The soil was 56% silt, 33% sand and 11% clay with 5.9% organic matter. Soil CEC was 6.2 meq/100 g and pH was 3.8.

For soils treated with copper, copper sulfate was applied by making the appropriate concentration in 250 mL and then sprayed on the soil (approx. 3000 g) in a fine mist while the soil was loosely and continuously mixed. Copper was applied at 0, 100, 200, 400 and 600 $\mu g\ g^{-1}$.

Treated soils were loosely packed in plastic leach tubes (4 cm top i.d. by 20.7 cm long, with a volume of 150 cm^3). Tubes were placed in racks and incubated at room temperature for 7 days. To maintain the initial moisture level, water was periodically added to tubes until initial tube wet weight was obtained. Soil was removed from the individual tubes then gently mixed, and individually sampled for chemical concentration, percent moisture, and soil nematodes and microarthropods.

Omnivore-predator nematodes and mesostigmatid and oribatid mites were the most sensitive components of the soil microfauna community, and were significantly lower than controls at 100 $\mu g\ g^{-1}$ copper (Fig. 4 and 5). Total nematode and microarthropod numbers did not show a significant reduction until concentrations were above 200 $\mu g\ g^{-1}$ copper (Fig. 4 and 5). The decline in numbers of nematodes at 400 and 600 $\mu g\ g^{-1}$ copper occurred because of negative effects on all trophic groups, but copper only had negative effects on mesostigmatid and oribatid mites in the microarthropod community. Nematode abundance was actually greater than controls at lower levels of copper application. Total numbers of nematodes were higher than controls at 100 and 200 $\mu g\ g^{-1}$, but were significantly so only at 200 $\mu g\ g^{-1}$. High populations of herbivore nematodes (>35 g^{-1}) in these treatments accounted for the increase. We observed an increase in nematode abundance at intermediate levels of copper concentration, and trophic analysis suggest that the increase occurred because of negative effects on nematode predators. Both omnivore-predator nematodes and mesostigmatid and oribatid mites were significantly lower than controls at these concentrations. Although many of the feeding habits of the nematodes in the omnivore-predator nematode group are unknown, mesostigmatid and oribatid mites are both known to consume nematodes.[6] Increased abundance of prey bacterivore nematodes and prostigmatid mites has also been observed when predators were eliminated by pesticides in deserts and agroecosystems.[4,6] These novel soil ecosystem techniques will support ecological assessments at U.S.

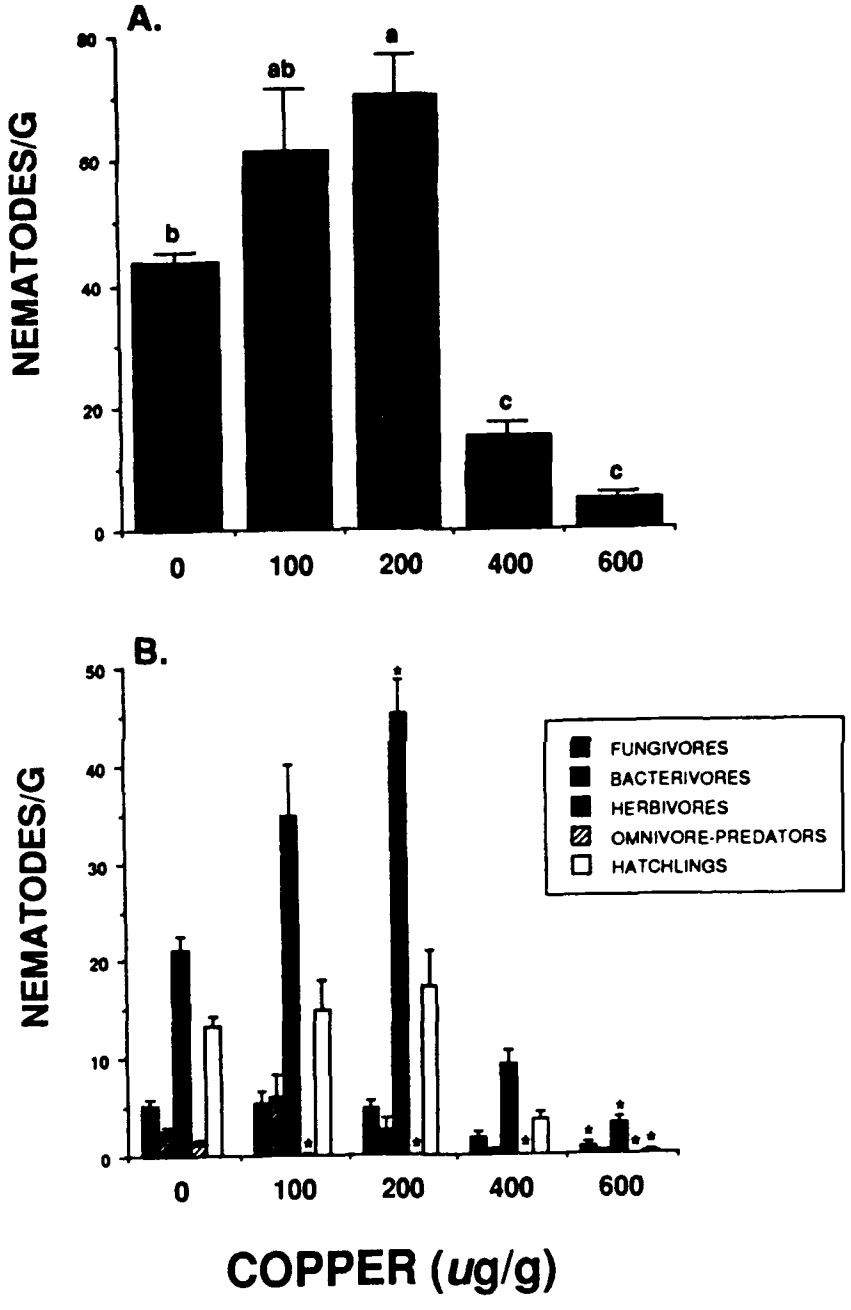


Figure 4. Effects of Copper on Nematode Populations.

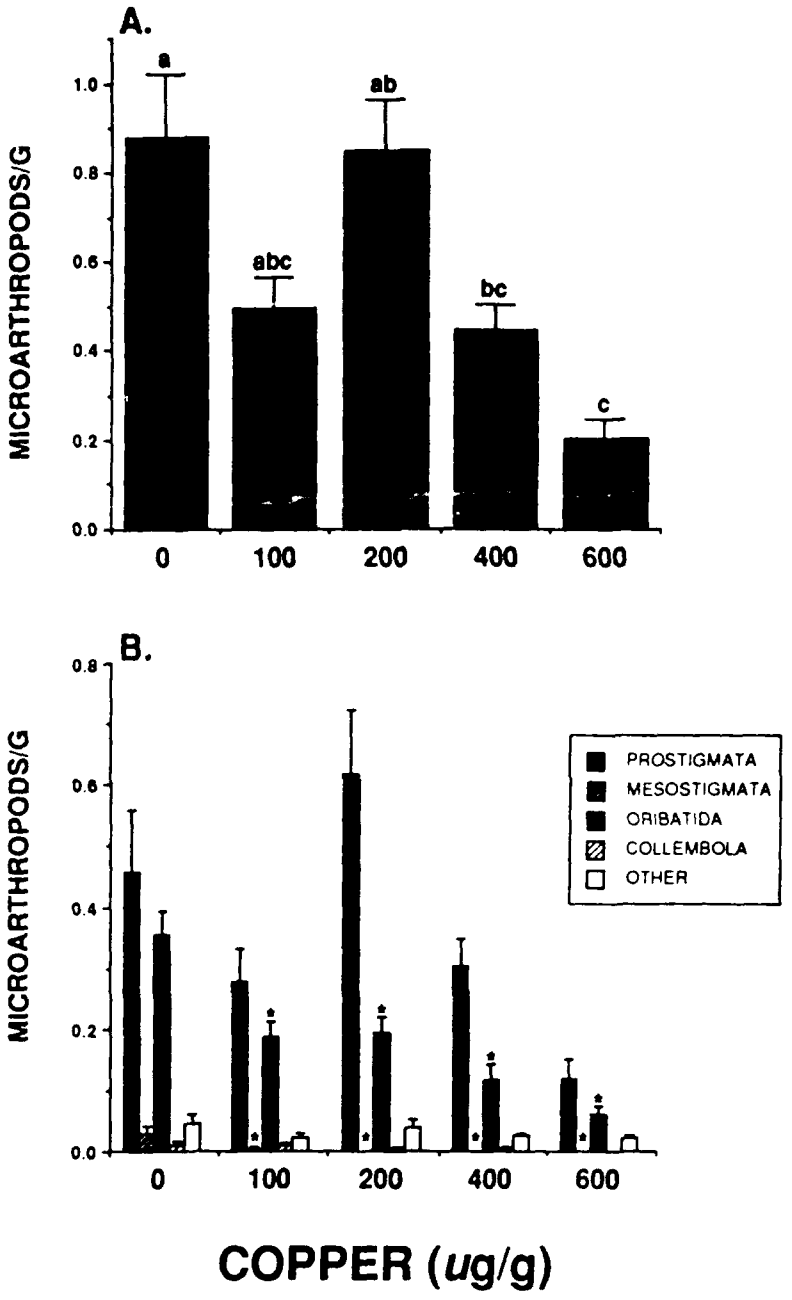


Figure 5. Effects of Copper on Microarthropod Populations.

Army sites.

Our studies demonstrated that trophic structure and community analyses of soil fauna in simple laboratory microcosms were sensitive measures of direct and indirect effects of chemical pollutants on terrestrial ecosystems. We observed higher sensitivity to chemicals for some trophic groups that was not indicated by total counts alone, and demonstrated that soil microfauna were as sensitive as, if not more so, than other components of the biotic community of chemicals. Our method should have wide utility because it can easily and inexpensively detect community and ecosystem level effects in soils with natural populations of nematodes and microarthropods.

The results from this research were the development of intact soil column units to address the dynamics of the transport and transformation of chemicals in soil (fate) and the soil fauna microcosm which will provide soil community and ecosystem level data on soil systems (effects). These techniques are currently being used at DoD sites to address environmental fate and effects questions in the development of clean-up criteria.

REFERENCES:

1. EPA 1982. *Environmental Effects Test Guidelines.* EPA 56016-82-002. Washington, DC.

2. EPA 1989. *Ecological Assessment of Hazardous Waste Sites: A Field and Laboratory Reference Document.* EPA/600/3-89/013. Corallis, OR.

3. Anderson, J.M. 1988. *Spatiotemporal effects of invertebrates on soil processes.* Biol. Fertil. Soils 6:216-227.

4. Parker, L.W., P.F. Santos, J. Phillips and W.G. Whitford. 1984. *Carbon and nitrogen dynamics during the decomposition of litter and roots of a Chihuahuan desert annual, Lepidium Iasiocarpum.* Ecol. Mono. 54:339-360.

5. Beare, M.H., R.W. Parmelee, P.F. Hendrix, W. Cheng, D.C. Coleman and D.A. Crossley, Jr. 1992. *Microbial and faunal interactions and effects on litter nitrogen and decomposition in agroecosystems.* Ecology (In Press).

6. Hendrix, P.F. and R.W. Parmelee. 1985. *Decomposition, nutrient loss and microarthropod densities in herbicide-treated grass litter in a Georgia piedmont agroecosystem.* Soil Biol. Biochem. 17:421-428.

Projectile Tracking Device Using GPS

Mr. George C. Wiles, Mr. John S. Eicke*
Mr. Brian T. Mays, Mr. Andrew P. Ladas
US Army Harry Diamond Laboratories, SLCHD-TA-TS
2800 Powdermill Rd, Adelphi, MD 20783-1197

PROGRAM OVERVIEW

Many efforts are under way to improve artillery accuracy and effectiveness. Besides precise determination of gun and target location, advanced registration techniques employing projectile tracking and trajectory modeling to predict the point of impact can greatly increase artillery effectiveness. Automated methods allow for increased range without requiring the use of a forward observer.

One automated method under consideration is the tracking of rounds using the GPS (see fig. 1). A cooperative effort is under way by LABCOM's Harry Diamond Laboratories, Human Engineering Laboratory, and Ballistic Research Laboratory to develop a device contained in a standard artillery fuze volume.

The four-year program is now in the Proof of Concept phase, in which a GPS translator contained in a large portion of the projectile volume (see fig. 2) is fired and tracked via post-mission processing. The Proof of Principle phase will miniaturize the translator into a standard profile 9 cubic inch fuze volume and track the round in near real-time with on-site data processing.

TRADEOFF ISSUES

An initial study determined that there are two GPS-based options for projectile tracking. The first, and most obvious, is to place a GPS receiver on board the projectile to compute the position and relay it to the ground. The second solution calls for an RF translator to be placed on the projectile. A translator receives the satellite signals and retransmits them at a different frequency, allowing a ground

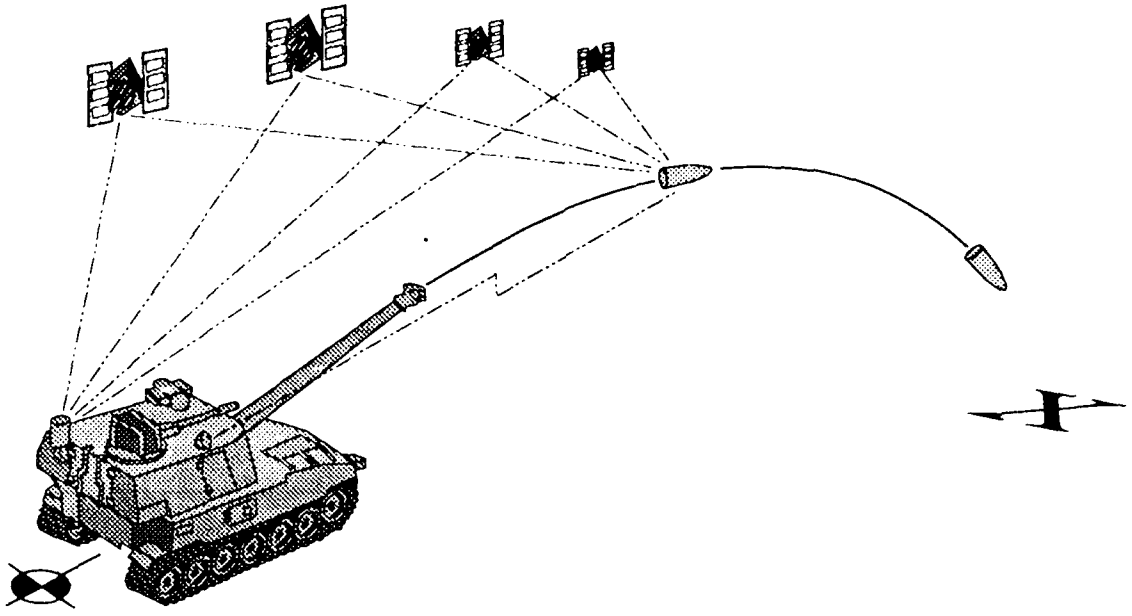


Figure 1: System Concept

receiver to measure the relative time delay of the satellite signals and determine the position of the translator antenna. Any additional time delay due to the translator or downlink path affects all the signals equally, and is absorbed at the receiver as an additional clock error. Translators have been used with great success as range instrumentation in both the Trident and ERIS programs.

A tradeoff study indicated that the translator concept was the best approach for projectile tracking. A translator is a much simpler device, consisting of an amplifier chain, an oscillator, and a mixer. This device would have a lower per unit cost and a much better chance of surviving gun launch. A receiver can require several minutes for a time-to-first-fix (TTFF) from a cold start; this time can be reduced to seconds with the use of a warm start (having previous knowledge of position and precise time) or multiple correlators operating in parallel. While this acquisition problem exists for both the receiver and translator approaches, the translator scheme provides several advantages. The requirements on oscillator stability after gun launch to enable a receiver to maintain lock or perform a warm start are several orders of magnitude more stringent than those of the translator. The translator can broadcast a harmonic of its local oscillator to allow the ground receiver to remove all the frequency uncertainty except the downlink Doppler, which itself may be closely modeled. Also, when the process-

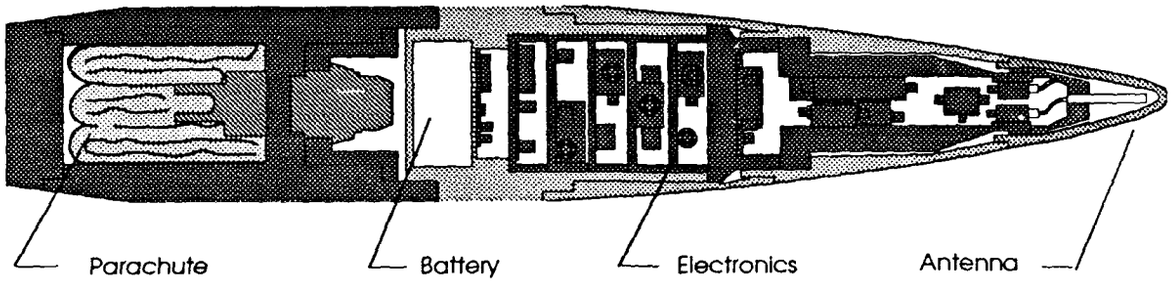


Figure 2: Proof of Concept Test Round

ing is being performed on the ground, the data may be buffered in memory to allow a TTFF that is longer than the projectile flight time.

SPECIAL CONSIDERATIONS

Building a translator into a standard fuze volume suitable for gunfire has some unique technical obstacles. Of particular concern was the antenna design, the master oscillator, and self-interference. A typical GPS antenna has a hemispherical pattern to receive all satellites in view. The projectile requires an isotropic pattern to work over the entire trajectory, and there has to be an axial phase center to minimize the effects of the projectile spin. A linearly polarized receive pattern is required, due to the reversal of polarization (relative to the satellite) that would occur after the projectile reaches apogee. Finally, significant isolation must exist between the receive and transmit ports. These requirements, which would be hard to meet without any size constraints, are particularly difficult when limited to a cone that is less than 4 inches high and 2 inches in diameter. This antenna was identified early in the program as a potential technical barrier.

The translator master oscillator was originally identified as another technical barrier, but became less of a concern after the decision was made to perform a coherent upconversion keyed to a pilot tone. In this scheme, all the mixer local oscillators are generated from a single source, and a harmonic of this frequency is broadcast to the ground. The ground receiver, using the pilot tone frequency, may then determine where to search the frequency spectrum for the translated GPS signals. Then the oscillator requirement after gunfire was to remain sufficiently stable to maintain the intermediate and output frequencies centered within the translator filter and antenna bandwidths.

Another potential problem was that of self-interference. A translator functioning at several kilometers of range must have greater than 120 dB of gain to make up for the spreading loss incurred in the reradiation of a wave from a distant source. To compound the problem, translators that use a pilot tone are designed with harmonics that are within the input and output frequency bands. This opens the way for numerous oscillation and interference problems due to signal connection and crosstalk which were exacerbated by the small volume.

PROOF OF CONCEPT CONFIGURATION

The following subsections describe the hardware that was gun-fired as a Proof of Concept demonstration and the ground equipment used to support the test.

TRANSLATOR

The function of the translator is to receive the GPS signals, move them in frequency, and transmit them to the receiver. The translator acts as an extension of the GPS receiver's RF front end, or like an extremely long antenna cable of indeterminate length. Just like a receiver front end, a translator consists of a series of amplifiers, mixers and filters (see fig. 3). The translator must pass all of the GPS spectrum without distorting it, as well as reject any interfering signals. The translator must also have a low noise figure, because it establishes the noise floor in the receiving process.

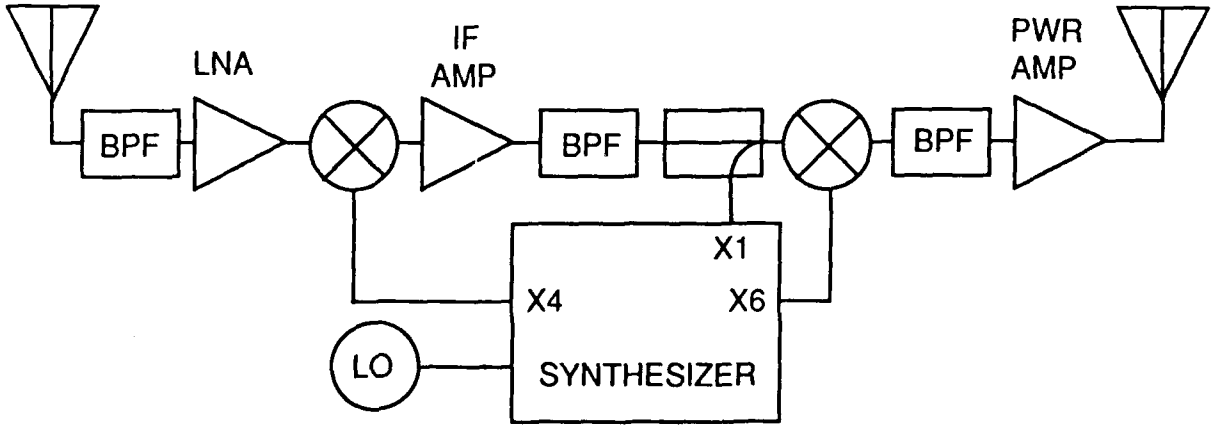


Figure 3: Translator Block Diagram

While the topology of the translator (filter-amp-mixer...) is well determined by its function, the IF and transmit frequencies are left up to the designer. In this instance, the transmit frequency was chosen to be S-band (2207 MHz) for two reasons. One reason is that the frequency has to be well below or above the GPS center frequency of 1575 MHz to avoid self-interference. Since a lower frequency would require a larger antenna element, the S-band was chosen as a good high-frequency compromise between antenna size and path loss. A second reason was one of convenience, since there already exist a number of telemetry band antennas and receivers to support development.

The translator up-conversion takes place in two stages. The intermodule isolation requirements are less severe when the gain is distributed over several frequencies. The IF frequency was chosen well below the GPS frequency to allow for a narrow filter, since filters that are a small percentage of their center frequency are difficult to construct. Besides rejecting image frequencies and spurious signals, the IF filter sets the shape of the output spectrum. Range and missile tracking translators have a strict requirement to not interfere with other translators which may be on the air. The projectile translator, being a tactical device, has no such restraint, and is bandlimited only to conserve transmitter output power.

A translator is unusual in that no carrier is modulated. The spread-spectrum GPS signal at the translator front end is so weak that it is actually below thermal noise within its signal bandwidth. The satellite can be detected only by collapsing the signal by correlating it with a replica of the coded signal. The noise plus signal is what is upconverted and amplified by the translator. The noise density (dBm/Hz) required at the output depends on the downlink range and tracking antenna gain, but the total output power may be minimized by bandlimiting the system noise to the minimum required to pass the GPS C/A code spectrum of 2.046 MHz. The 5-MHz IF bandwidth chosen for the projectile application reflects both the minimum achievable filter bandwidth with acceptable loss and the minimum system bandwidth given the uncertainty of the oscillator frequency after gun launch (see fig. 4).

Although the Proof of Concept test had the luxury of a large portion of the projectile volume, we decided it would be prudent to restrict the antennas to the external portion of the fuze volume, to adequately address the surface area concerns. Since the translator sets the noise figure in the receiving chain, it is important to maximize the GPS receive antenna gain. Many candidate ap-

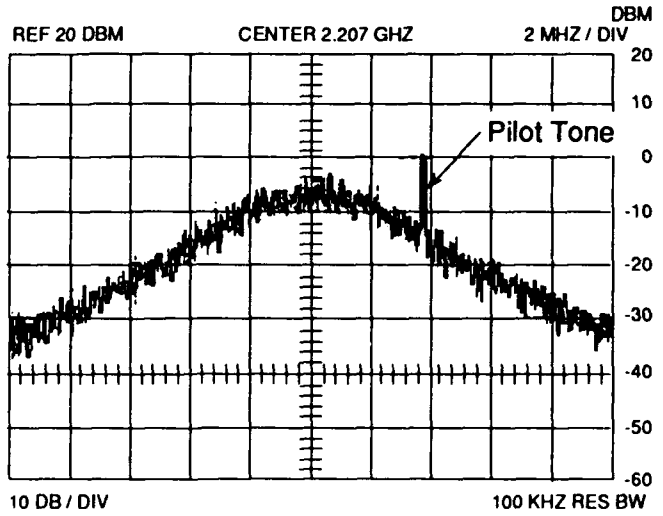


Figure 4: Translator Output Spectrum

proaches for an antenna design were prototyped and measured on a projectile mock up to determine their suitability; in the end the choice was limited to various combinations of two elements. The prototypes revealed that the best choices were a quarter wave monopole and a dual-element conical wrap microstrip patch. Both antennas exhibit a monopole radiation pattern, with no appreciable phase variation in the azimuth (roll) plane. However, as with any low-frequency antenna mounted on a projectile, there was a characteristic scalloping and nulling of the pattern in the elevation plane. This is an unavoidable consequence of the interaction of the projectile body with the antenna. Even though the resulting pattern exhibits a null off the forward axis, studies indicate that this should have a minimal impact when the full GPS satellite constellation is in place.

Three possible configurations of the two candidate elements were considered: a GPS monopole with an S-band microstrip, a GPS microstrip with an S-band monopole, and a tandem microstrip design. A tradeoff study of performance vs. economy of volume showed that the best choice was the tandem design using two microstrip antennas mounted on the outside of the conical surface. However, the GPS monopole has a small gain advantage over the microstrip element. In order to maximize the chances of a successful Proof of Concept test, the voluminous GPS monopole with S-band microstrip was chosen (see fig. 5).

A significant portion of the hardware design was invested in finding a suitable oscillator and IF filter. A surface acoustic wave (SAW) oscillator design was found "off the shelf" that would survive the 6000 g airgun screen with a minimal (3 to 4 kHz) shift. A group of these oscillators were pre stressed by subjecting them to a centrifuge, and then carefully tuned to the translator center frequency. The 6000 g screen level is an overtest for the expected Proof of Concept test flight, which was chosen to be lower than the maximum 16000 g to favor the use of readily available commercial parts.

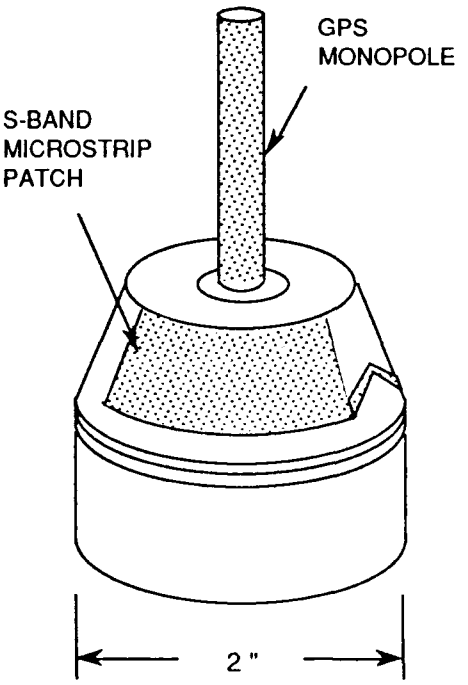


Figure 5: Dual Frequency Projectile Antenna

As the breadboard design was compressed to fit inside the projectile, the importance of point-to-point isolation became apparent. In some cases the required out-of-band frequency rejection of the filter was greater than the isolation between the filter connectors. The filter rejection was distributed among the translator filters as much as possible. Several types of filters were obtained and airgun tested, as well as a lumped element design constructed in-house. A miniature cavity filter filled with foam potting was chosen as the best compromise of size and performance.

The large volume of the Proof of Concept test projectile allowed many of the breadboard parts to be used in the flight test without redesign. Samples of the parts were subjected to spin and airgun shock tests to evaluate survivability. The translator was partially miniaturized by replacing some breadboard components with TO-8 can versions, and mounting them as stripline modules. Assembling the power and RF interconnects for these modules inside the projectile was not unlike building a ship in a bottle.

GROUND EQUIPMENT DESIGN

The use of a translator does not negate the need for complicated signal processing and precise frequency reference; it merely allows it to be performed on the ground in a benign environment. The acquisition process can further be aided by a priori knowledge of the projectile's flight path, and from synchronized GPS timing. This can be supplied in the tactical environment by the ballistic computer and static GPS receiver resident in the howitzer.

In a typical GPS receiver, position is calculated by the simultaneous solution of four equations, carried out in a Kalman filter. These multi state linear estimators typically use information from the correlator tracking loops in their calculations. It was discovered early in the program that the position solution from a dynamic translator system requires a different filter configuration to effectively model the translator's drifting oscillator and translator-to-ground Doppler shift. These shifts and drifts drive the correlator tracking loops of a standard receiver in a manner that causes the solution to diverge, since the dynamics of the loops do not match the expected dynamics of a direct GPS solution. Also, most receivers are ill-equipped to deal with the increasing time delay from a translator moving toward a satellite but away from the ground antenna. Therefore, even if the translated signal was returned to its original frequency, it would not likely produce a position solution in an unmodified commercial receiver.

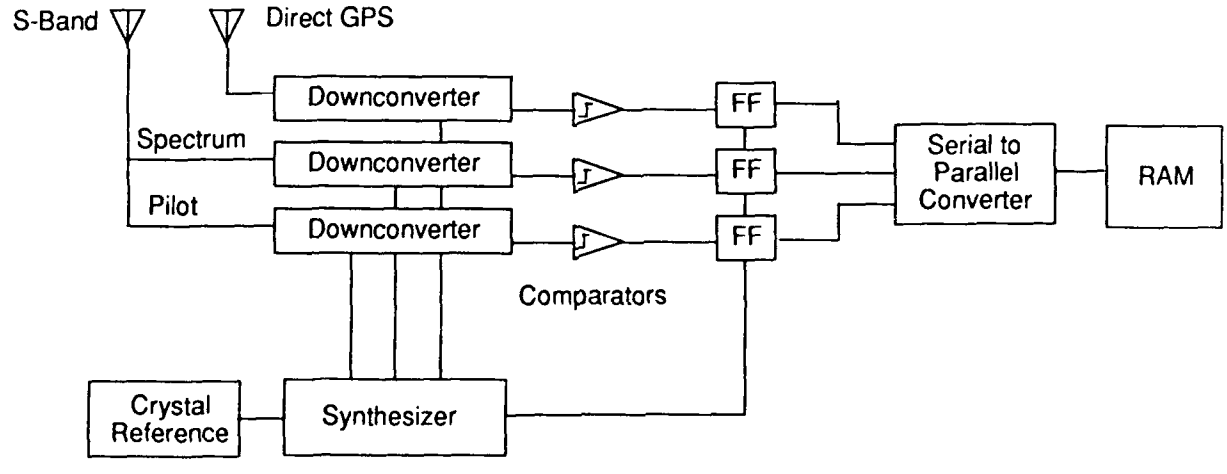


Figure 6: Ground Equipment Block Diagram

The original plan of using a modified standard receiver was abandoned, and a method of data capture and rudimentary processing was implemented for the Proof of Concept test. This method consisted of reducing the translated signal to near baseband, hard limiting it for a one-bit quantization, and recording it for later processing in software (see fig. 6). One bit sampling, while not optimum, is more than adequate, is considerably simpler to implement, and takes less memory space than multilevel sampling. Three channels of data were recorded. An L-band antenna captured the GPS signal at the reference location (which would be the gun in a tactical scenario) for use as a differential baseline. The translated S-band signal was captured, and the pilot tone was captured in a separate channel. The pilot tone was captured separately and at a lower sample rate (1.25 MHz vs 2.5 MHz for the broadband spectrums) to conserve on memory space.

The short time of flight and one-bit quantization allowed a simplified recording scheme to be used. Instead of an expensive, precise wideband tape recorder, the entire flight was captured in 80 MB of RAM. This provided over 100 seconds of recording time. The short-term frequency stability of the system was maintained by deriving all downconversion and sample clock frequencies from a crystal reference.

DATA PROCESSING

Once the data were recorded, the satellite signals were processed using traditional GPS receiver techniques implemented in software (see fig. 7). First the pilot tone channel was examined to find the pilot frequency as it appeared on the ground. Next, an estimate of the translator oscillator was calculated from the measured pilot tone and the expected Doppler shift found from the computer prediction of the flight. A good estimate of the

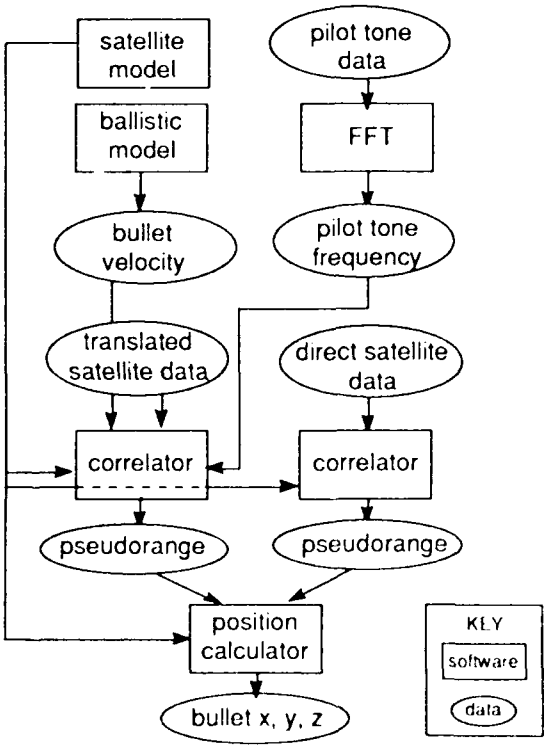


Figure 7: Data Processing Software

translator oscillator, the bullet dynamics, and the GPS time of day provides the starting point for the narrowband search for the satellite signal. A software correlation routine compares the sampled signal to a generated replica and integrates the matches to determine if the signals are correlated. The relative code delay when correlation occurs is saved as the pseudorange. The position may then be calculated point by point by using a least squares solution. These same pseudoranges may be processed in a Kalman filter to reduce the position solution noise by taking advantage of the ballistic nature of the flight path and modeling system noise sources.

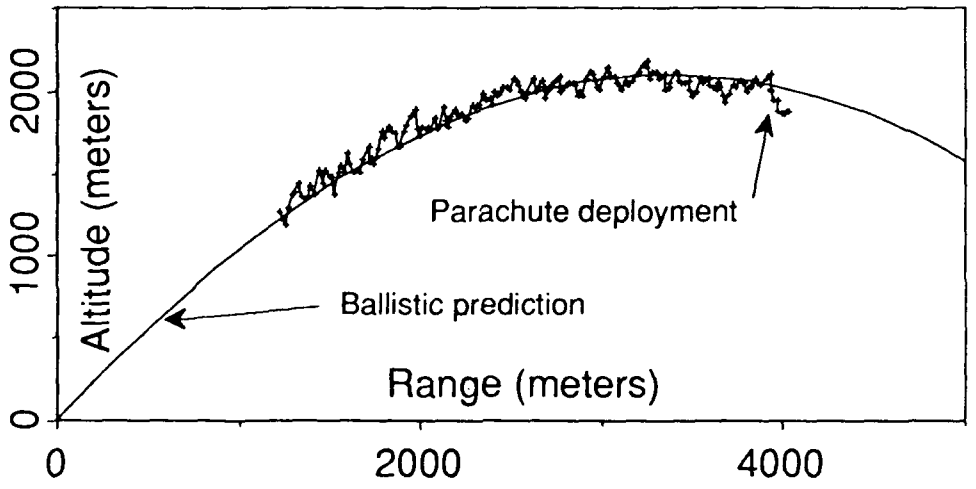


Figure 8: Least Squares Position Solution

TEST RESULTS

The initial Proof of Concept test firings took place on 4 October, 1991 at the U.S. Army Yuma Proving Grounds (YPG) in southwest Arizona. Preliminary test results have shown the test to be a success. Two test rounds were fired at zone 3 conditions out of a 155-mm howitzer and recovered by parachute. These conditions resulted in a 3000 G setback acceleration and initial spin of 100 rps. The spotter rounds and the first test round were tracked by range MPS radars. Each round's launch and impact locations were surveyed in GPS (WGS-84) coordinates. Both units successfully transmitted pilot tone information which was recorded along with the translated GPS information. At the time of this writing, data processing

has recovered the signals from three satellites in the translated data record of the second test round. Calculations have provided the position information shown in figure 8 for two dimensions. Efforts are under way to improve the accuracy of the pseudorange tracking and to Kalman filter the data.

FURTHER WORK

Two more firings are planned, using the recovered Proof of Concept projectile hardware. Efforts to miniaturize the translator hardware will culminate in the Proof of Principle test firing. The focus of the miniaturization effort is a Small Business Innovative Research contract to produce five MMIC modules that will perform the down- and up-conversion, frequency synthesis, and amplification needed in the translator. A small thermal battery power supply is under development by Harry Diamond Laboratories. The signal processing algorithms for correlation, tracking and position solution on the ground are being performed also, with final system integration to be performed by Harry Diamond Labs. The Proof of Principle firing of a miniaturized translator with a near realtime position solution is scheduled for the end of calendar year 1993.

SYNTHESIS OF METALLO-TETRABENZPORPHYRINS POSSESSING HIGH THIRD-ORDER OPTICAL NONLINEARITY FOR MILITARY LASER EYE PROTECTION

*Reginald A. Willingham, Dr., John H. Cornell, Dr.,
David M. Alabran, Mr., John W. Cullen, Mr.,
Frank H. Bissett, Dr., Masato Nakashima, Dr.,
David E. Remy, Dr. and Joseph F. Roach, Mr.
U.S. Army Natick RD&E Center
Natick, MA 01760-5020

INTRODUCTION

A new threat to the soldier in the field has emerged through the potential use of tunable laser weapon systems. Laser range finders and target designators are already in use on the modern battlefield. Coherent radiation from laser sources even at low energy levels is capable of inflicting severe damage to the retina of the eye. One mode of protection against such laser threats is the use of nonlinear optical (NLO) media, which permit the transmission of visible light at low intensities but will reversibly block visible and near infrared radiation at intensities harmful to the eye. To be useful, the NLO media must have ultrafast response times.

The interaction of an electric field with a medium produces a separation of charges, resulting in a polarization in the medium. When the applied fields are weak, the induced polarization is linearly dependent on the applied field and the propagation behavior is determined by the linear susceptibility of the material, Chi(1). At higher intensities, the polarization includes a nonlinear dependence, which becomes increasingly more important, and the susceptibility must be defined by additional terms such as Chi(2) and Chi(3). These terms represent second-order and third-order susceptibilities of the material and are determined by features at the molecular level, including electronic delocalization, molecular structure and overall molecular packing. Chi(3) is a property of a bulk material. In addition, one can extract from Chi(3) an additional parameter, gamma, which is an intrinsic nonlinear property of the molecule.

Enhanced third-order susceptibilities can arise from centrosymmetric and noncentrosymmetric molecular structures possessing extended pi-electron delocalization, such as in the porphyrin ring systems. When these molecules are immobilized in an applied electrical or optical field, third-order nonlinear processes can occur, such as third harmonic generation (THG) and self-focusing. Also, a refractive index property change can result from third-order nonlinear processes. If the refractive index change is of sufficient magnitude and has an ultrafast response time, an optical limiting effect can be effected as required for eye protection from incident laser irradiation.

Measurements of Chi(3) on a material either in solution or dispersed in a solid polymeric phase contain both real and imaginary components. The real component is responsible for the nonlinear refractive index change. This brings about an intensity-dependent change of the phase of the laser beam. The imaginary component is responsible for nonlinear absorption either by a two photon process or by excited state absorption.

This work focuses on the synthesis and measurement of third-order susceptibility, Chi(3), of a class of compounds called metallo-tetrabenzporphyrins (TBP) that have recently been shown to exhibit high Chi(3) values.[1,2] Both real and imaginary components of Chi(3) have been determined. In addition, preliminary experiments with these compounds incorporated into polymeric lenses has shown, for the first time, that optical limiting can be demonstrated in a device.

EXPERIMENTAL

The synthetic procedures for the preparation of the zinc tetrabenz-porphyrins and the zinc tetra-1,2-naphthoporphyrin followed standard methodologies reported previously[3,4].

The third-order NLO susceptibility, Chi(3), was measured by the degenerate four-wave mixing (DFWM) technique using tetrahydrofuran (THF) solutions of the porphyrins. The laser source is a Nd:YAG mode locked laser of 1064 nm. It is frequency-doubled to give 17 ps full width at half maximum (FWHM) single pulses, with repetition rates of 10 to 100 Hz. The energy of each pulse was 25 mJ and the intensity was approximately 4 MW/cm^2. Figure 1 shows the layout of the DFWM experiment.

The laser source used in the measurement is split into three separate beams, B1 through B3, which interact to produce a fourth beam, B4. Approximately 75% of the laser source beam is split into two pumping beams, B1 and B2, of equal intensities; the remaining portion of the laser source becomes a probing beam, B3. The delay units are adjusted so that

B1 and B2 are colinear and all three beams overlay in time and space in the sample cell (1-2 mm path length). The pumping beams, B1 and B2, generate an electrical field which induces changes in the index of refraction of molecules in their paths. This refractive index change is maximum when B1 and B2 are in phase and zero when they are out of phase. The interaction creates a grating along the sample cell path. The probing beam, B3, interacts with this grating to generate a conjugate beam, B4, opposite in direction to B3. The intensity value (I) of the conjugate beam, B4, is used in equation (1) to determine Chi(3) of the solution relative to that of a standard.

$$\frac{\chi^{(3)}_{Sample}}{\chi^{(3)}_{CS_2}} = \left(\frac{n_{Sample}}{n_{CS_2}}\right)^2 \frac{d\alpha}{1-e^{-d\alpha}} e^{\frac{d\alpha}{2}} \left(\frac{I_{Sample}}{I_{CS_2}}\right)^{\frac{1}{2}} \frac{d_{CS_2}}{d_{Sample}} \qquad (1)$$

In equation (1) n is the index of refraction, d is the path length of the beam in the sample solution, I is the intensity of B4, α is the absorption coefficient of the solution at the laser wavelength.

The Chi(3) of the solution was calculated on the basis of the value for carbon disulfide as standard, which we take as 6.8×10^{-13} esu. The Chi(3) value for the solute is obtained from the plot of Chi(3) vs solute concentration.

$$\chi^{(3)}_{Solution} = \chi^{(3)}_{Solvent} + \left(\frac{n^2+2}{3}\right)^4 \frac{\gamma_{Solute} A}{M} C \qquad (2)$$

In equation (2), n is the index of refraction of solute, γ is the second hyperpolarizability, A is the Avogadro's Number, M is molecular weight, and C is concentration of a solute in g/mL. The index of refraction of the solute is determined by refractometric measurements of its solution.

RESULTS AND DISCUSSION

A series of zinc tetrabenzporphyrins (Table 1, Ia.-Ij.) were prepared as outlined in Figure 2. By varying the nature of the aldehyde, several different structural variations were accommodated in the "meso" or R position on the porphyrin ring system. The macrocyclic compound series

consisted of aliphatic and aryl substituents at the meso position. One example with substitution on the benzo- portion of the molecule is included (Ib). Macrocycle, Ib, has a fully fluorinated tetrabenzo periphery. In all cases the metal coordinated at the center of the ring is zinc. Zinc tetra-1,2-naphthoporphyrin (II, R=H) was prepared as outlined in Figure 3.

The results of measurement of Chi(3) for the series of zinc tetrabenz- and a zinc tetranaphthoporphyrin are presented in Table 1. The effect on Chi(3) of electron delocalization over a larger ring system is shown by comparison of zinc tetrabenzporphyrin (Ia) with zinc tetra-1,2-naphthoporphyrin (II). The annelation of an additional aromatic ring to the benzporphyrin base structure has increased Chi(3) by a factor of three.

The introduction of electron donating groups into the meso-phenyl groups of the zinc tetrabenzporphyrin series led to a progressive rise in the measured Chi(3) value as the relative electron donating of the substituent increased. Interestingly, as the electropositive character of the substituent increased, the value of Chi(3) also increased. These findings are consistent with the fact that greater polarization, whether emanating from electron withdrawal or electron donation, enhances Chi(3).

Hammett linear free energy relationships have been traditionally developed to explain combined resonance and inductive effects on the reactivity of substituted aromatic systems. It is interesting to note, as shown in Figure 4, that a linear relationship is observed when Chi(3) values are plotted against the Hammett sigma constants. The plot shows a V-shape composed of two lines of opposite slope depending on whether the substituent is electron withdrawing or donating. As expected, the greater the absolute value of the sigma constant, the greater is the polarization and Chi(3) is increased proportionately.

Conjugated pi-electron systems such as in the porphyrin macrocycles are important to the manifest of the Chi(3) property. Apparently, the meso substituents contribute a resonance and inductive effect on the ring conjugated pi-electrons. A linear relationship is obtained when the meso-phenyl substituents' Hammett sigma constant values are plotted versus Chi(3). These Hammett sigma constant values are based on combined resonance and inductive contributions of the respective substituents. Steric constraints in the porphyrin ring dictate that the meso-phenyl substituents will be almost orthogonal to the plane of the pi-electron conjugated porphyrinic ring. Thus, the meso substituents probably contribute a predominant inductive effect on the ring with the resonance contribution being a lesser contributor.

Similar linear relationships have been proposed based on theoretical calculations of second-order susceptibility done by Goldfarb and Medrano for substituted benzthiazoles[5]. In addition, Meredith and Stevenson[6] calculated empirical substituent group parameters based on THG experiments and found values that were higher at both ends of the Hammett sigma scale.

Structure correlations that are predictive of nonlinear responses are at present based on an intuitive understanding of the phenomena and have yet to reach a rigorous state. The relationships observed contribute to methods that may be predictive of structures with enhanced values of Chi(3). In related work, we have started to investigate the effects of metal substituents on the TBP system. These results will not only be of interest in the search for compounds with high Chi(3) but will offer comparisons with published theoretical and experimental studies on metallo-phthalocyanines[7], which are similar in structure to the benzporphyrins.

Nonlinear effect is limited not only to the induced change in the polarization but also in the absorptivity. In the mathematical derivation, Chi(3) has real and imaginary parts. The measure of the nonlinear absorptivity appears as the imaginary part of Chi(3). This nonlinear absorptivity is another useful property of these porphyrins in the optical limiting application. To determine the imaginary part of Chi(3), the transmission of laser as a function of laser intensity and fluence was measured, using picosecond and nanosecond pulses[2]. The results of the absorption experiments are summarized in Table 2. For a comparison, the real part of Chi(3) for the porphyrin solutions is listed also.

It was found that at certain fluence level, the % transmission did decrease from approximately 45% to 15% and 60% to 20% for compounds Ia and If, respectively. The decrease was observed from both picosecond and nanosecond laser pulses. It is inferred that it resulted from the absorption of laser by an excited state, which was populated by a one photon absorption and had a cross section larger than the ground state. Such an excited state has a lifetime of 15 ns, the population became appreciable during these short laser exposures. The magnitude of this type of absorptivity is expressed as sigma in the Table 2. For both Ia and If, the values of sigma are approximately equal. Besides the excited state absorption, an absorption due to two-photon process was indicated. The two photon absorption cross section is given as beta in Table 2. It was of interest to note that If has a larger beta than Ia.

Toward the development of optical limiting application devices, the tetrabenzporphyrin Ia was incorporated into the solid polymer matrix polycarbonate, and its optical limiting property was evaluated. Figure 5 shows that the optical limiting effect did occur. A 70% transmittance

reduction was observed in these preliminary experiments. It is important to realize that this is the first demonstration of optical limiting in a prototype system. Thus using a porphyrinic ring system, it is possible to engineer an optical limiting capability.

SUMMARY

A series of zinc tetrabenzporphyrins and a zinc tetranaphthoporphyrin were synthesized for evaluation by a degenerative four-wave mixing technique. A high degree of third-order optical nonlinearity was found. The measured Chi(3) values were in the range of 0.2-2.8 10^{-8}esu, which are among the highest reported to date. The influence of the meso-substituent on the Chi(3) value was examined in relation to the Hammett sigma constant. Metallo-porphyrins incorporating variations in the ring systems will be prepared and evaluated to explore the impact of molecular design on the magnitude of third-order optical nonlinearity.

ACKNOWLEDGMENT

The authors would like to thank Joel Carlson of Natick RD&E Center for the mass spectrometric characterization of all the porphyrins we have reported in this paper.

REFERENCES

1. D.V.G.L.N. Rao, F.J. Aranda, J.F. Roach, and D.E. Remy, Appl. Phys. Lett., 58(12), 1241 (1991).
2. S. Guha, K. Kang, P. Porter, J.F. Roach, D.E. Remy, F.J. Aranda, and D.V.G.L.N. Rao, Optics Lett., 17(4), 264 (1992).
3. D.E. Remy, Tetrahedron Lett., 24(14), 145 (1983).
4. D.E. Remy, G.C. Richard, and S.J. Weininger, U.S. Army Natick Technical Report NATICK/TR-87/029, AD A183957 (1987).
5. I.J. Goldfarb and J. Medrano "Nonlinear Optical Effects in Organic Polymers" 93-99 J. Messier, F. Kajzar, P. Prasad and D. Ulrich ed. Kluwer Academic Publishers, 1989.
6. G.R. Meredith and S.H. Stevenson, ibid. 105-122.
7. James S. Shirk, J.R. Lindle, F.J. Bartoli, C.A. Hoffman, Zakaya H. Kafafi, and Arthur W. Snow, "Materials for Nonlinear Optics" S. R. Marder, J.E. Sohn, and G. D. Stucky ed., 626, 1991 Am. Chem. Soc.

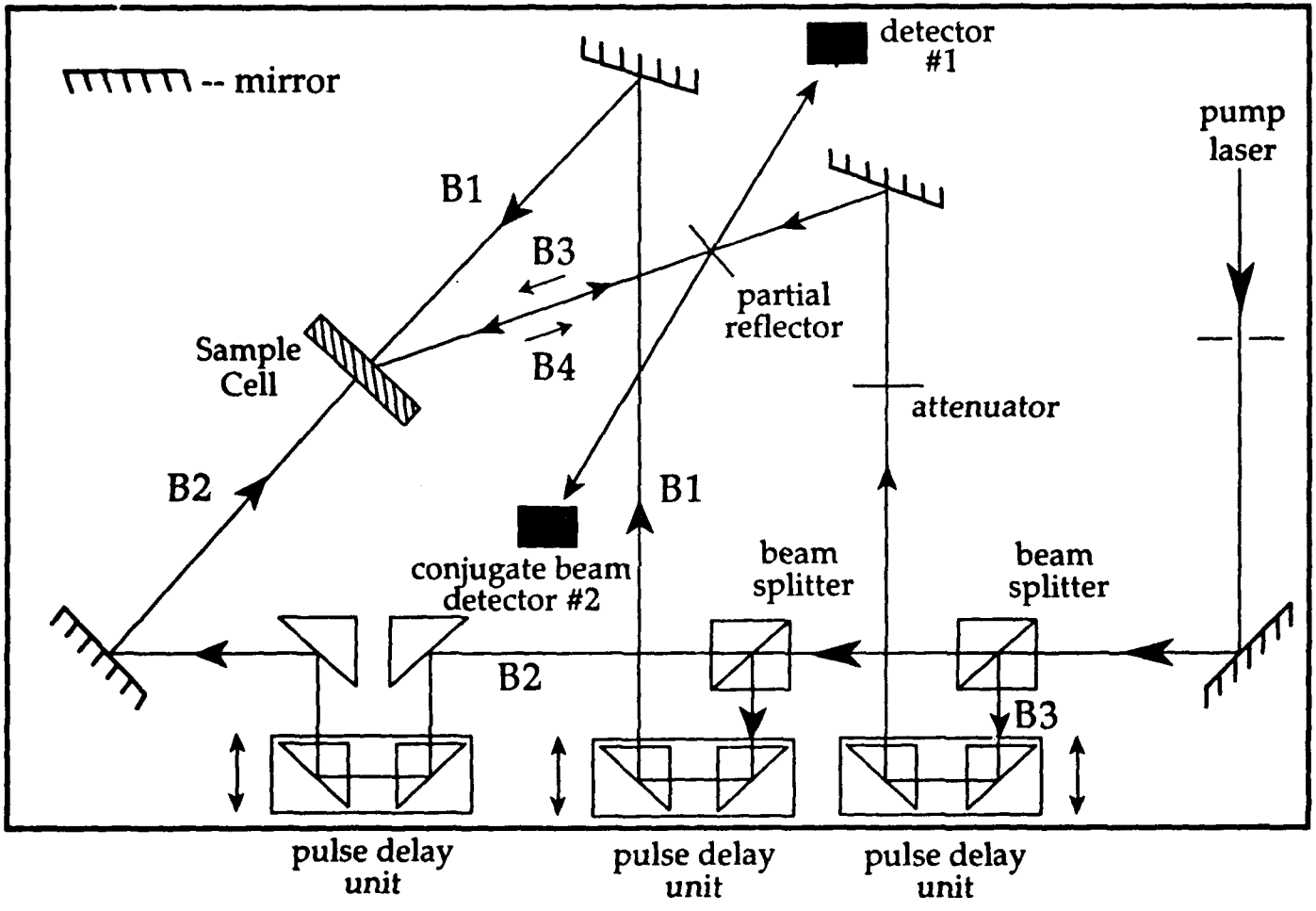
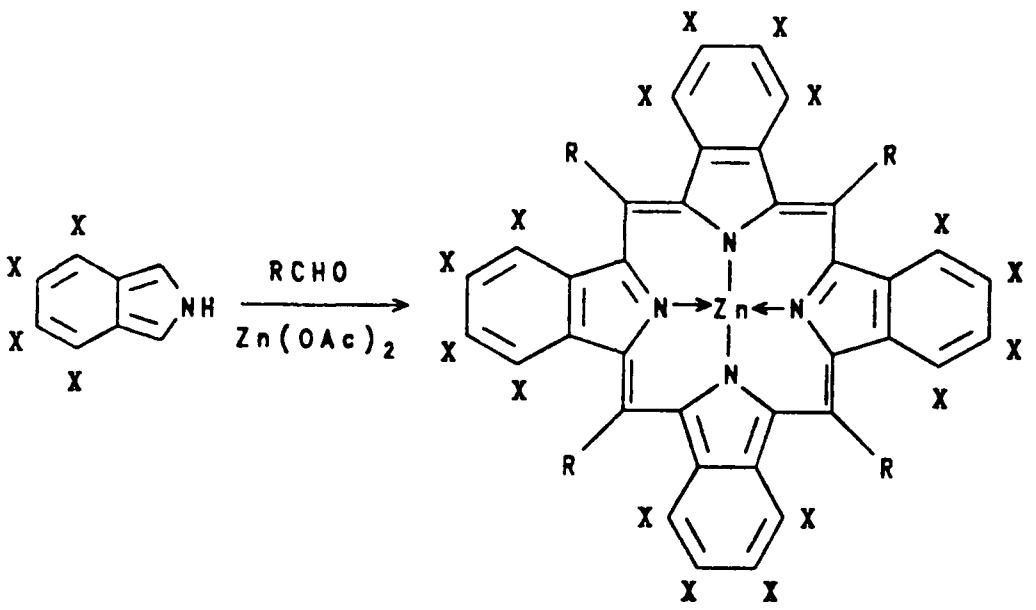


Figure 1
NLO Measurement Layout



I

Figure 2
Preparation of Tetrabenzporphyrins

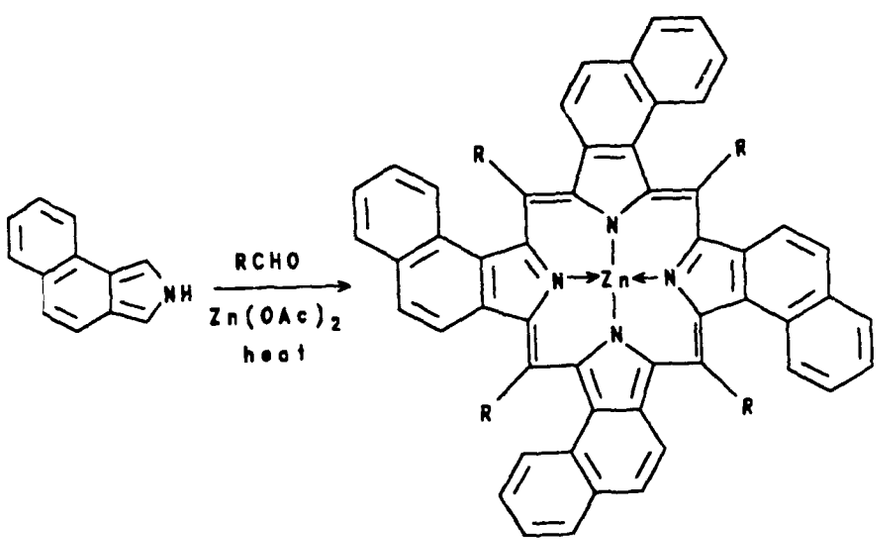


Figure 3
Preparation of Tetranapthoporphyrin

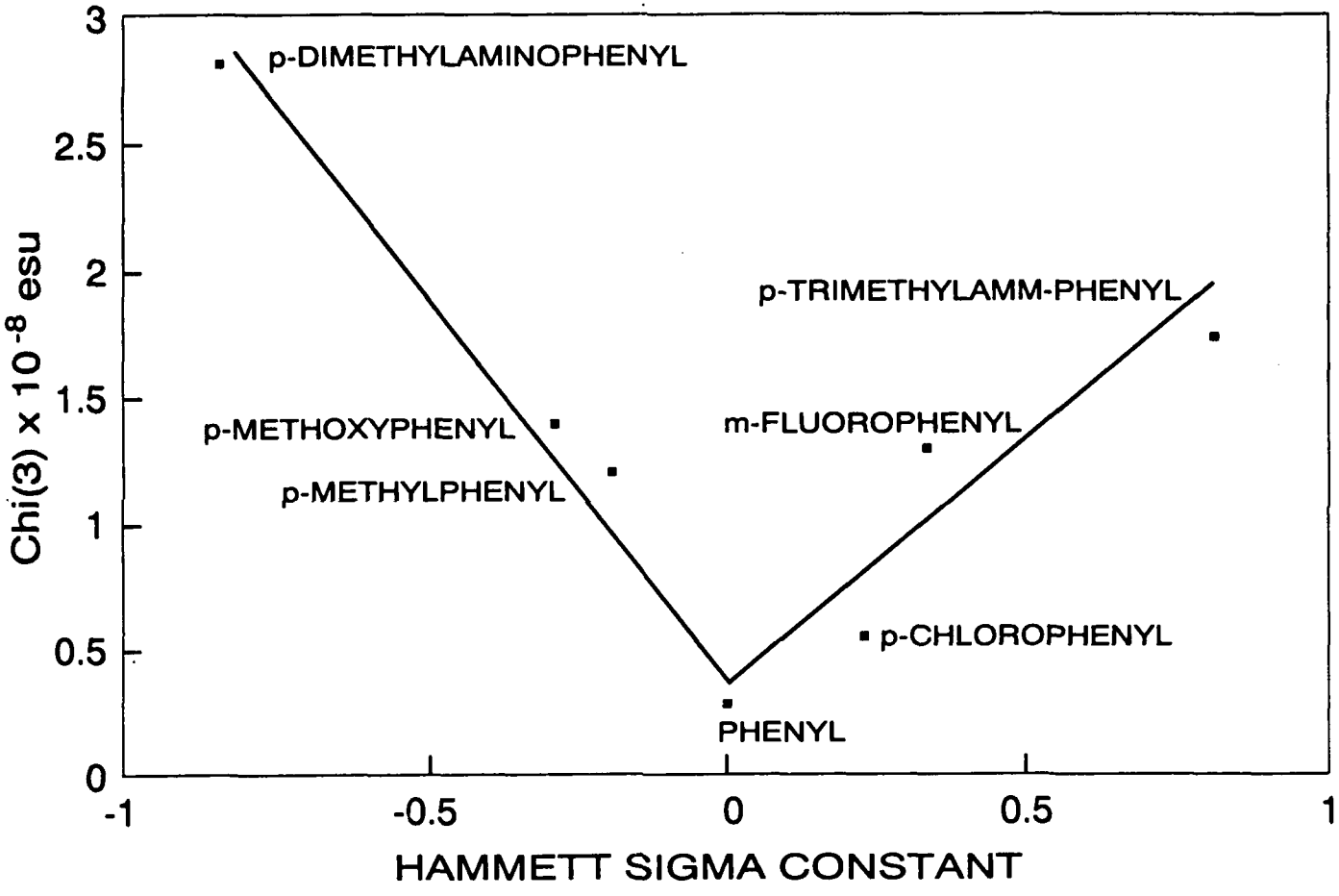


Figure 4
Hammett Plot of Chi(3)

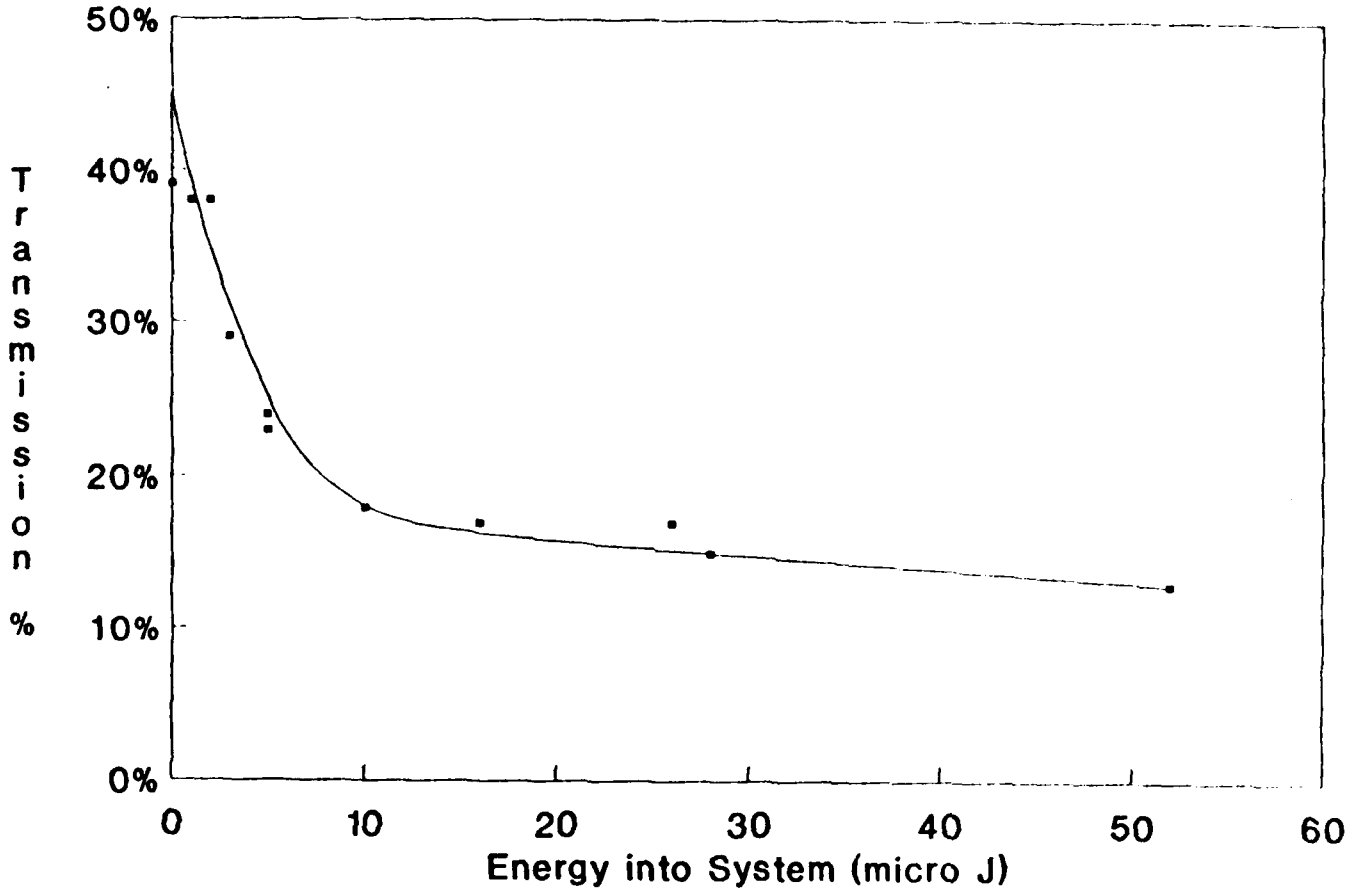


Figure 5
Optical Limiting Property

Table 1. Values of Chi(3) for Porphyrins

ZnTBP (I)	Chi(3) 10^{-8}esu
Ia. R = H, X = H	0.3
Ib. R = H, X = F	0.2
Ic. R = Me, X = H	1.5
Id. R = Ph, X = H	0.3
Ie. R = p-MePh, X = H	1.2
If. R = p-MeOPh, X = H	1.4
Ig. R = p-$N(Me)_2$Ph, X = H	2.8
Ih. R = p-$N(Me)_3$Ph, X = H	1.8
Ii. R = p-ClPh, X = H	0.56
Ij. R = m-FPh, X = H	1.3
ZnTNP (II)	
II. R = H	0.99

Table 2. Chi(3) of TBP Solutions

Compound	Conc. Concentration g/l	Re ($\chi^{(3)}$) 10^{-14} esu	Im ($\chi^{(3)}$) 10^{-13} esu	α cm^{-1}	β cm/GW	σ $10^{-17} cm^{-2}$
Ia	0.46	9.1	≤1.8	3.8	≤0.5	3
If	0.46	7.8	9	2.4	2.5	3

α : Ground state linear absorption coefficient

β : Ground state two photon absorption coefficient

σ : Excited state linear absorption coefficient

High Resolution Image Processing of Geophysical Data with Diffraction Tomography

Dr. Alan J. Witten
Oak Ridge National Laboratory
PO Box 2008
Oak Ridge, TN 37831-6200

LTC Wendell C. King*
Department of Geography and Environmental Engineering
U.S. Military Academy
West Point, NY 10996-1695

Mr. James R. Ursic
U.S. Environmental Protection Agengy - Region 5
77 West Jackson Boulevard
Mail Stop HSRLT-5J
Chicago, IL 60604-3590

Introduction

Interest in geophysical remote sensing spans a wide spectrum of military programs, but all have the common thread of needing ways to develop more and better information concerning the subsurface environment. The environmental cleanup and protection program needs information concerning the location of buried wastes that could contaminate groundwater or through other pathways endanger the environment. A special subset of this problem is the location of unexploded ordnance which is obviously not well suited to intrusive investigative techniques. Detection and location of tunnels has had military significance in several conflicts and continues to be important in our deterrence mission in the Republic of Korea. There is reason to expect that rapid and accurate tunnel detection will be important in the small scale conflicts that are the most likely threats to peace we will face in the immediate future. This list is certainly not a complete set of all possible uses for geophysics, but these examples are sufficient to demonstrate the value of enhanced remote sensing capabilites resulting from the application of advanced signal processing to existing data gathering instrumentation.

This paper will present a method of detection and imaging that has as its input the data collected from the most commonly employed geophysical tools, applies the *Generalized Projection Slice Theorem* (GPST) in the analysis of these data to yield results ranging from greatly improved detection to high resolution images of subsurface features and buried objects. Detection capability can be greatly enhanced by incorporating this more rigorous signal processing procedure into the statistically based maximum likelihood estimator. The paper includes a theoretical development of the new signal processing procedures, results from the implementation of the theory with a number of commonly used geophysical tools, including a presentation of the improved capability for tunnel detection and location estimation and detection of buried ordnance. High resolution imaging of a variety of buried wastes has already been achieved through the application of geophysical diffraction tomography[1,2,3,4], a method that fully exploits the principles of the GPST.

Geophysical Methods

Hazardous waste site characterization is the problem to which geophysics is most widely applied and the purpose to which most developmental research is devoted. The most commonly applied methods are seimic reflection and refraction, ground penetrating radar, electromagnetic induction, resistivity, and magnetometry. Each is discussed below only to the extent needed to understand their basic concepts and uses, and only within the context of the applications previously discussed.

Seismic Reflection and Refraction - In these methods a seismic pulse is generated at the ground surface and the reflected or refracted waves are collected over an array of geophones also on the ground surface. Subsurface features are inferred from changes in slope of the first arriving signal as a function of geophone position. Refraction studies are typically applied to characterize site stratigraphy.

Ground Penetrating Radar (GPR) - GPR, also a wave based method, employs collocated electromagnetic transmitting and receiving antennas which are moved along a line on the ground surface. Data acquired for each measurement line are displayed as a series of parallel plots of amplitude versus time. This manner of presentation is a crude form of imaging a two-dimensional vertical cross-section.

Electromagnetic Induction (EM) - EM systems use two coils: one is a primary coil to impart a magnetic field into the ground and a second coil used to detect the secondary field produced by the currents induced in subsurface conductors by the primary field. A surface map is constructed of the conductivity contours over the surveyed area and locations of significant change are indicative of subsurface features.

Magnetometry - Ferrous objects produce a localized perturbation in the Earth's magnetic field which can be detected by magnetometers. Magnetometry data is displayed

in the same manner as EM data.

Resistivity - Like magnetometry, resistivity uses a static field, however, here this field is created by electrodes. Arrays, similar in configuration to those used in seismic methods, of both monitoring and injection electrodes are used. Since these electrodes must be driven into the ground, resistivity is relatively cumbersome in field applications.

Enhanced Signal Processing

The focus of this section is the discussion of enhanced signal processing methods that can be applied to data acquired by the conventional geophysical tools described in the previous section. In particular, the goal of these signal processing algorithms is to render these tools more quantitative yielding accurate size, shape, and location information. Specifically, this section deals with one approach to signal processing based on an analytic relationship between acquired data and an integral transform of the spatial variations in some subsurface quantity. This relationship is known as the *Generalized Projection Slice Theorem*[5] (GPST) since it represents a generalization to arbitrary wavelength of the *Projection Slice Theorem* which served as the early basis for imaging algorithms used in CT scanners of diagnostic medicine and nondestructive evaluation. Imaging within the GPST has become known as diffraction tomography (DT) and this form of imaging as well as a companion signal processing algorithm for target detection and location estimation, also based on the GPST, will be discussed later in this section.

The Generalized Projection Slice Theorem

Presented here is a derivation of the GPST for the case of propagating scalar waves. The purpose of this derivation is to illustrate the underlying physical concepts and simplifying mathematical assumptions associated with techniques based on this theorem. Subsequently, parallel results for other forms of probing energy employed by the geophysical tools described in the previous section are presented.

In general, signal processing algorithms based on the GPST are implemented in the frequency (rather than the time) domain. For this reason, the derivation of this particular GPST begins with the reduced wave equation

$$\nabla^2 u + k^2 u = k^2 u O(\mathbf{x}), \tag{1}$$

where $u = u(\mathbf{r})$ is the wavefield measured over some contour $\mathbf{r}$, $k = \omega/c_0$ is the wavenumber at frequency ω and background wave speed c_0, and $O(\mathbf{x}) = 1 - c_0^2/c^2(\mathbf{x})$ is known as the object profile and is related, in this case, to spatial variations in refractive index. It should be noted that, along with other factors, the distinction among the GPST's for specific geophysical tools lies in the wavenumber k which distinguishes the form of

probing energy and the object profile $O(\mathbf{x})$ which characterizes the property of the subsurface inhomogeneities that a particular tool will repond to. It should additionally be noted that implicit in eq.(1) is the assumption of a harmonic time dependence $\exp(i\omega t)$ or, alternatively, a mapping from the time to frequency domain by means of a temporal Fourier transform.

The goal of signal processing algorithms based on the GPST is the exploitation of an analytic relationship between the the measured data, or some associated reduced or processed data, to an integral transform of the object profile. This requires an inversion of eq.(1) in order to express the unknown quantity $O(\mathbf{x})$ as some function of the measured quantity $u(\mathbf{r})$. Unfortunately, the inversion problem for eq.(1) is ill-posed and extremely sensitive to small errors in the measured values of u. As a consequence, this equation is linearized by invoking a weak scatter approximation. There are several weak scatter approximations that can be used[6], for illustration purposes, the example derivation considered here is the Born which represents the measured wave field by

$$u(\mathbf{r}) = u_0(\mathbf{r}) + \epsilon u_s(\mathbf{r}), \tag{2}$$

where u_0 is the "illuminating" wave field, u_s is referred to as the scattered wave field, and ϵ is a small parameter included to explicitly indicate the assumption that the deviations of the total wave field, u, from u_0 are small. It is further assumed that the spatial variations in wave speed $c(\mathbf{x})$ do not deviate significantly from the background, c_0, making the object profile, $O(\mathbf{x})$ of order ϵ. This weak scatter approximation yields the linearized version of eq.(1)

$$\nabla^2 u_s + k^2 u_s = k^2 u_0 O(\mathbf{x}) \tag{3}$$

that can also be expressed in the integral equation form

$$u_s(\mathbf{r}) = -k^2 \int d\xi \, G(\mathbf{r} - \boldsymbol{\xi}) u_0(\boldsymbol{\xi}) O(\boldsymbol{\xi}), \tag{4}$$

where G is the Green's function for the scalar Helmholtz operator.

It readily seen from eq.(4) that u_s is derived from measurement and the functions u_0 and G can be computed making the only unknown quantity the object profile $O(\mathbf{x})$. In the form given by this equation, the desired inversion can be accomplished be deconvolution of the integral. This deconvolution can be achieved by representing the Green's function by it's plane wave expansion *[7]

$$G(\mathbf{r}) = \frac{i}{4\pi} \int \frac{d\alpha}{\sqrt{k^2 - \alpha^2}} e^{i(\sqrt{k^2-\alpha^2}\hat{\mathbf{x}} + \alpha \hat{\mathbf{y}})}, \tag{5}$$

*For simplicity, the remainder of the derivation is given in two dimensions. A similar plane wave expansion can be used in the three-dimensional case with a derivation of the corresponding GPST that parallels the two-dimensional one given here.

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are unit vectors in the x and y directions, respectively, and measurements of u (u_s) are made along a line parallel to the y-axis. Taking this measurement line to be $\mathbf{r} = x_R\hat{\mathbf{x}} + \ell\hat{\mathbf{y}}$, (Fig. 1) assuming the incident field is a plane wave propagating in the s_0 direction†, $u_0(\mathbf{x}) = \exp(ik\mathbf{s}_0 \cdot \mathbf{x})$, and using the expression for the Green's function given by eq.(5), (Fig. 1) eq.(4) becomes

$$u_s(\mathbf{r}) = -\frac{ik^2}{4\pi}\int \frac{d\alpha}{\sqrt{k^2-\alpha^2}}\, e^{i(\sqrt{k^2-\alpha^2}x_R+\alpha\ell)} \int d\boldsymbol{\xi}\, O(\boldsymbol{\xi})\, e^{-i(\sqrt{k^2-\alpha^2}\hat{\mathbf{x}}+\alpha\hat{\mathbf{y}}-k\mathbf{s}_0)\cdot\boldsymbol{\xi}}. \qquad (6)$$

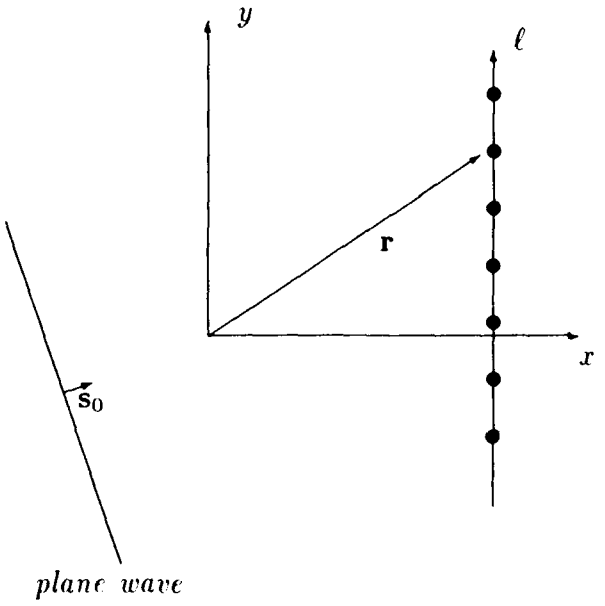


Figure 1: Illustration of the geometry and notation used in deriving the GPST. Measurement locations along the line $\mathbf{r} = x_R\hat{\mathbf{x}} + \ell\hat{\mathbf{y}}$ are indicated by •'s.

The deconvolution of eq.(6) can be accomplished by defining the Fourier transform of the data, u_s, as

$$\hat{u}_s(\kappa) = \int d\ell\, e^{-i\kappa\ell}\, u_s(\ell), \qquad (7)$$

applying this integral transform to eq.(6), and recognizing that

$$\int d\ell\, e^{-i(\kappa-\alpha)\ell} = 2\pi\,\delta(\kappa-\alpha). \qquad (8)$$

† Again for simplicity, plane wave illumination is assumed. It can, however, be shown that, for an arbitrary source beam pattern, plane wave illumination can be synthesized by coherently summing the contributions from a multiplicity of distinct source locations[8,9].

This procedure yields the final form of the GPST

$$\tilde{u}_s(\kappa) = -\frac{ik^2}{2\sqrt{k^2-\kappa^2}}\, e^{i\sqrt{k^2-\kappa^2}x_R}\, \tilde{O}(\mathbf{K}), \tag{9}$$

where

$$\tilde{O}(\mathbf{K}) = \int d\boldsymbol{\xi}\, O(\boldsymbol{\xi}) e^{-i\mathbf{K}\cdot\boldsymbol{\xi}} = \int d\boldsymbol{\xi}\, O(\boldsymbol{\xi}) e^{-i(\sqrt{k^2-\kappa^2}\hat{\mathbf{x}}+\kappa\hat{\mathbf{y}}-k\mathbf{s}_0)\cdot\boldsymbol{\xi}}. \tag{10}$$

The desired result of an analytic expression between measurement and subsurface properties is achieved in eq.(9) which relates the one-dimensional Fourier transform of the scattered field (reduced measured data) to the two-dimensional Fourier transform of the object profile. The key to this result is the plane wave expansion, eq.(5), used in the deconvolution. This step explicitly introduces the concept of coherence by representing the scattered wave as a superposition of plane waves. This aspect of coherence is carried into the GPST where, recalling that $u_0(\mathbf{r}) = \exp(ik\mathbf{s}_0 \cdot \mathbf{r})$, the wave vector $\mathbf{K}$ associated with the integral tranform can be expressed as $\mathbf{K} = k(\mathbf{s} - \mathbf{s}_0)$ which is the difference between a scattered plane wave propagating in the direction $(1/k)(\sqrt{k^2-\kappa^2}\hat{\mathbf{x}} + \kappa\hat{\mathbf{y}})$ and the incident plane wave propagating in the direction $\mathbf{s}_0$. The term multiplying $\tilde{O}$ in eq.(9), $(-ik^2/2\sqrt{k^2-\kappa^2})\exp(\sqrt{k^2-\kappa^2}x_R)$, is the equation for a holographic lens and the imaging procedure discussed in the following subsection is entirely analogous to imaging by optical holography.

Having illustrated the concept and mathematical formulation of the GPST, some general remarks can be made about the GPST's for the suite of geophysical tools discussed in the previous section. First, GPST's are specific to both the form of probing energy and the measurement geometry. While, for most cases, the GPST will involve a Fourier tranform of the data, $\tilde{u}_s(\kappa)$, it will not necessarily relate this quantity to a two-dimensional spatial Fourier transform of the object profile. The actual form of the integral transform $\tilde{O}$ will depend on the definition of wavenumber k. Only for homogeneous propagating waves will k be real and the associated integral transform be Fourier. Table 1 summarizes the nature of the GPST's for the five geophysical tools described in the previous section. It is important to recognize that Table 1 identifies the nature of inhomogeneities that can be resolved by a particular geophysical tool as well as the character of the wave vector $\mathbf{K} = (K_x, K_y)$ which has serious implication regarding the computational ease with which particular types of data can be processed. This topic will be discussed in greater detail in the next two subsections.

method	k	(K_x, K_y)	transform	O
seismics	ω/c_0	(real,real)	Fourier-Fourier	refractive index
GPR	ω/c_0	(real,real)	Fourier-Fourier	refractive index
EM	$\sqrt{-i\omega/\sigma_0}$	(complex,real)	generalized	conductivity
magnetometry	0	(imaginary,real)	Laplace-Fourier	permeability
resistivity	0	(imaginary,real)	Laplace-Fourier	conductivity

Table 1: Summary of GPST relationships for geophysical tools. Here, ω is frequency, c_0 is a background wave speed, and σ_0 is a background conductivity.

Imaging

For wave based tools (seismics and radar), imaging can be accomplished by numerical inversion of the integral transform indicated in eq.(9) or

$$\begin{aligned} O(\mathbf{x}) &= \tfrac{1}{(2\pi)^2} \int d\mathbf{K}\, \tilde{O}(\mathbf{K})\, e^{i\mathbf{K}\cdot\mathbf{x}} \\ &= \tfrac{2i}{(2\pi k)^2} \int \tfrac{d\mathbf{K}}{\sqrt{k^2-\kappa^2}}\, e^{-i\sqrt{k^2-\kappa^2}\, x_R}\, \tilde{u}_s(\kappa)\, e^{i(\sqrt{k^2-\kappa^2}\hat{\mathbf{x}}+\kappa\hat{\mathbf{y}}-k\mathbf{s}_0)\cdot\mathbf{x}}. \end{aligned} \tag{11}$$

The integration over $\mathbf{K}$ can be expressed as a sum over view angles $\mathbf{s}_0$ and an integration over κ to yield the more computational suitable form

$$O(\mathbf{x}) = \frac{2i}{(2\pi)^2 k} \Sigma_{\mathbf{s}_0} e^{-ik\mathbf{s}_0\cdot\mathbf{x}} \int d\kappa\, H(k,\kappa,\mathbf{s}_0)\, \tilde{u}_s(\kappa)\, e^{i(\sqrt{k^2-\kappa^2}\, x+\kappa y)}, \tag{12}$$

where the function H contains the mapping from $\mathbf{K}$ into $\mathbf{s}_0$ and κ along with a filter to remove points of double coverage associated with this mapping. The imaging formula defined by eq.(12) is computationally efficient because the integration over κ can be accomplished by fast Fourier transform numerical procedures.

Imaging by means of GPST and the transform inversion given by eq.(12) has become known as *diffraction tomography* (DT) and it is analogous to holographic imaging using a variety of laser beam illumination directions $\mathbf{s}_0$. The GPST, eq.(9), provides a knowledge of $\tilde{O}$ over a wave vector space $\mathbf{K} = \mathbf{K}(\mathbf{s}_0, \kappa)$. The extent of this coverage and the resulting image quality will be a function of the number and range of illuminating view direction $\mathbf{s}_0$. For any particular view direction, $\mathbf{K}$-space coverage will be a portion of a circular arc. As the view direction changes a region of $\mathbf{K}$-space is "swept" out. For any practical geophysical measurement geometry, there will necessarily be gaps in $\mathbf{K}$-space which will introduce artifacts into the image. The most significant manifestation is an elongation of imaged features along the direction of the gaps. This has the simple physical interpretation that resolution is always better in the direction normal to that of wave propagation than along the direction of propagation. The above mentioned gaps

occur along the predominent direction of propagation and, consequently, optimal image quality can be achieved by utilizing the broadest possible range of view directions.

The previous discussion is appropriate for wave-based methods such as seismics and cross-borehole radar which typically employ independently positioned sources and receivers. One important geophysical tool, GPR, uses a collocated source/receiver pair which is moved in unison. For this measurement geometry, individual incident and scattered wave directions can not be independently controlled and, as a consequence, only a single arc in **K**-space can be realized for a particluar frequency ω. Fortunately, conventional GPR systems are pulsed offering a reasonable source bandwidth. This bandwidth can be exploited by representing the integration over **K** in eq.(11) as a summation over frequencies ω and an integration over the spatial Fourier transform variable κ. By this means, **K**-space coverage is a series of concentric arcs sufficient to yield good tomographic images.

While the wave based imaging procedure described above is easily implemented, imaging with measurements from other geophysical tools is more difficult, less computationally efficient, and, perhaps, impossible. In the case of resistivity, the integral transform to be inverted (Table 1) is a Laplace-Fourier which, again, can be implemented with standard numerical techniques, however, these techniques do not offer nearly the computational efficiency of the the multidimensional Fourier transform methods associated with wave-based measurements. Since resistivity is a "dc" method it operates at zero frequency and, for this reason, there is no opportunity to exploit bandwidth. Resisivity studies, however, commonly employ a variety of injection electrodes which serve the same purpose as variable view angles in wave-based methods and allows sufficient **K**-space coverage to yield good image quality. Magnetometry is also a "dc" measurement, however, this measurement is passive, relying on the Earth's magnetic field as a source. As a result view directions can not be varied, as is the case in resistivity, and **K**-space coverage is insufficient for imaging.

Most EM tools have a fixed coil separation making it impossible to independently vary incident and scattered wave directions. Some EM tools, however, can operate at different frequencies allowing sufficient **K**-space coverage for imaging in a manner similar to that of GPR. The difficulty associated with EM imaging is that the integral transform that must be inverted (Table 1) contains a wave vector component that is complex and, as a result, will require the development of new numerical procedures.

Detection

As a complementary procedure to imaging, the GPST can serve as the basis for target detection and location estimation which, as will be shown later, avoids some of the difficulties and limitations of imaging. The general problem addressed by detection is the quantification of the presence or absence of a known target and, if present, an estimate

of its location. The basic concept of the method described here is known as maximum likelihood estimation (MLE) which effects detection and location by correlating measurements with the known target response for a variety of possible locations $\mathbf{x}_0$. Taking the same measurement line used in the derivation of the GPST, $\mathbf{r} = x_R\hat{\mathbf{x}} + \ell\hat{\mathbf{y}}$, defining the measured data to be $u_s(\ell)$ and the data that would be received if the known target were located at $\mathbf{x}_0$ as $u'_s(\ell;\mathbf{x}_0)$, a measure of the target's presence and location would be the value $\mathbf{x}_0$ that minimized the square of the difference between u_s and u'_s integrated over all measurement points ℓ. This can be represented as the $\mathbf{x}_0$ that minimizes the function F where

$$F(\mathbf{x}_0) = \int d\ell\, |u_s(\ell) - u'_s(\ell;\mathbf{x}_0)|^2. \tag{13}$$

This procedure for target detection can be immediately implemented, however, it requires the numerical simulation and storage of a massive number of target data "templates" $u'_s(\ell;\mathbf{x}_0)$. This approach can be prohibitive, particularly if the target is strongly asymmetric and can assume many different orientations.

One means to make this problem more efficient is to use the GPST. The first step in accomplishing this is using Parseval's theorem to express the integral over ℓ as integral over the transform variable κ. This yields

$$\begin{aligned} F(\mathbf{x}_0) &= \int d\kappa\, |\tilde{u}_s(\kappa) - \tilde{u}'_s(\kappa;\mathbf{x}_0)|^2 \\ &= \int d\kappa\, (\tilde{u}_s\tilde{u}_s^* - 2\tilde{u}_s\tilde{u}_s'^* + \tilde{u}'_s\tilde{u}_s'^*), \end{aligned} \tag{14}$$

where $\tilde{u}_s$ ($\tilde{u}'_s$) is defined in eq.(7) and the $*$ denotes complex conjugation. The $\tilde{u}_s\tilde{u}'^*_s$ term is independent of possible target location $\mathbf{x}_0$ and produces only a constant bias in F. Therefore, this term can be dropped and a new function F_L, known as the log likelihood function defined as

$$F_L(\mathbf{x}_0) = \int d\kappa\, (2\tilde{u}_s\tilde{u}_s'^* - \tilde{u}'_s\tilde{u}_s'^*) \tag{15}$$

and the most probable target location $\mathbf{x}_0$ is the one that maximizes F_L.

The GPST is introduced into the detection signal processing algorithm by using it to represent $\tilde{u}'_s(\kappa)$ in terms of the spatial Fourier transform of the object profile $\tilde{O}(\mathbf{K};\mathbf{x}_0)$ for a target centered at a position $\mathbf{x}_0$. For the specific GPST derived earlier, it is recognized that, for propagating waves, $\tilde{u}'_s\tilde{u}'^*_s$ is a constant and can be neglected. With this and the shift rule for Fourier transforms which can be used to relate the $\tilde{O}(\mathbf{K};\mathbf{x}_0)$ to the Fourier transform of the object profile at some reference location, here taken to be the origin and denoted by $\tilde{O}_0$,

$$\begin{aligned} \tilde{O}(\mathbf{K};\mathbf{x}_0) &= \int d\mathbf{K}\, O(\mathbf{x};\mathbf{x}_0)\, e^{-i\mathbf{K}\cdot\mathbf{x}} \\ &= e^{-i\mathbf{K}\cdot\mathbf{x}_0}\, \tilde{O}_0(\mathbf{K}), \end{aligned} \tag{16}$$

the final form of the log likelihood function becomes

$$F_L(\mathbf{x}_0) = ik^2 e^{-ik\hat{\mathbf{s}}_0\cdot\mathbf{x}_0} \int \frac{d\kappa}{\sqrt{k^2-\kappa^2}}\, e^{-i\sqrt{k^2-\kappa^2}x_R}\tilde{u}_s(\kappa)\tilde{O}_0^*(\mathbf{K})e^{i(\sqrt{k^2-\kappa^2}x_0+\kappa y_0)}. \tag{17}$$

There is a striking resemblence between this result and the imaging formula given by eq.(12). Equation (17) is, in fact, an imaging formula, however, rather than providing an image of of a physical quantity $O(\mathbf{x})$, it yields an image of a statistical measure of target location, namely the log likelihood function. As a result, imaging of F_L has been referred to as the *diffraction tomography maximum likelihood estimator* (DTMLE). The holographic lens aluded to in the discussion of imaging is replaced here by a matched filter holographic lens where features in the data not correlated to the target characteristics as specified by $\tilde{O}_0$ are preferentially rejected. The major difference between the DTMLE and the more simple MLE approach, eq.(13), is that the GPST has been used to eliminate the need to generate a large number of data templates $u'_s(\ell; \mathbf{x}_0)$ replacing these instead with a single target response $\tilde{O}_0^*(\mathbf{K})$ (or, at worst, one response for each possible target orientation), making DTMLE far more computationally efficient.

The most important difference between target detection with DTMLE and DT imaging[‡] is that the DTMLE algorithm uses the GPST directly rather than it's definition of an integral transform that must be inverted to reconstruct an image. The implication of this fact on geophysical remote sensing is that target detection and location estimation can be implemented using precomputed filters and computationally efficient fast Fourier transforms for any geophysical tool. Furthermore, the detection problem uses a priori information, the target characteristics, and, therefore, does not require as much data as imaging. As a result, DTMLE can be used with only a single view or frequency, however, summing eq.(17) over all available views and/or frequencies will improve the results. This point is particularly important for magnetometry since it allows it's use for high resolution target detection where imaging is impossible.

Results

The Generalized Projection Slice Theorem has served as the basis for a number of geophysical imaging and detection algorithms. These are method-specific and the current level of technological maturity depends on both the geophysical tool and measurement geometry.

The most advanced signal processing algorithms are those associated with wave-based measurements. Among these, acoustic geophysical diffraction tomography (GDT) applied in an offset vertical profiling (VSP) (source(s) deployed on the ground surface and receivers emplaced in a vertical borehole) geometry is ready for commercialization. This method has been extensively field tested and an integrated hardware/software system is in use by the U.S. Environmental Protection Agency at Superfund sites. A second

[‡]Another important difference between DTMLE and DT is that a form of the GPST can be derived for DTMLE without the need for a weak scatter approximation. This makes the formulation of DTMLE exact[10].

such system has been built for use by the U.S. Army Toxic and Hazardous Materials Agency. GDT imaging and DTMLE detection algorithms have been developed and field tested for seismic reflection and refraction, cross-borehole (sources in one borehole and receivers in a parallel borehole) radar and seismics, and GPR. More recently, imaging and detection algorithms for resistivity and detection algorithms for magnetometry have been developed and tested on numerically synthesized data with field testing anticipated in the near future. Currently, efforts are being directed towards imaging and detection algorithms for EM measurements with the specific application being the detection of buried unexploded ordnance.

This section presents results, based on both simulated data and actual field measurements, of the applications of enhanced signal processing algorthms. These do not represent the entire extent of experience with GPST signal processing but are selected to be representatitive of potential applications within the scope of this paper.

Figure 2 is a gray scale GDT image of a vertical cross-section within the Korean Demilitarized Zone (DMZ) constructed by the analysis of cross-borehole radar measurements. In this image, variations in electromagnetic wave speed are represented as gray

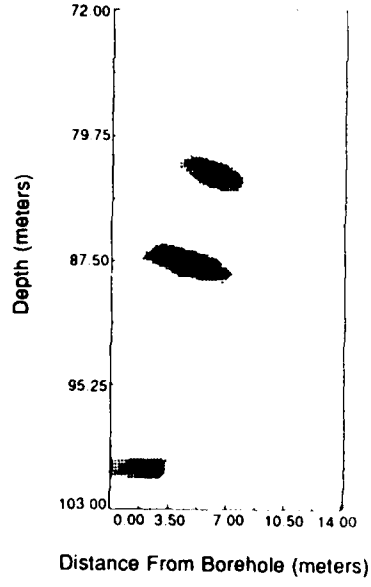


Figure 2: Image of tunnel in the Korean DMZ[11].

levels with darker gray levels associated with greater wave speed. The black area appearing at a depth of 87.5 m and 3.5 m from the source borehole is the image of a known tunnel. The dark gray area approximately 7 m above this tunnel exhibits nearly the size

and contrast of the known tunnel. This feature introduces the concern that a second unknown tunnel exists within this cross-section. It was hypothesized that this feature was not a tunnel, but rather, a fracture zone of the type known to occur in the granite formations within the DMZ. Confirmation of this hypothesis came with application of the DTMLE algorithm to the same data set. Figure 3 is the corresponding image of the log likelihood function, also displayed as gray levels, with darker gray scales associated with increased probability of target location. Notice in this figure that the unknown

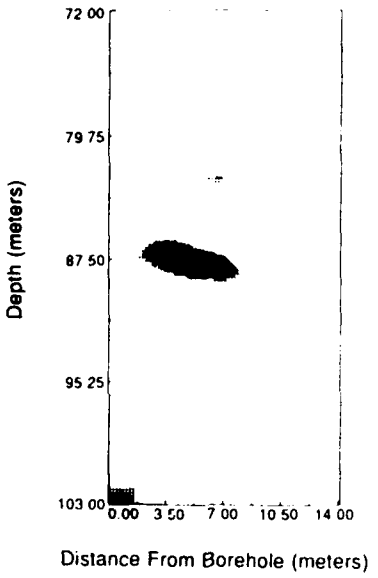


Figure 3: Image of the location probability of a tunnel in the DMZ[11].

feature of concern above the known tunnel is almost gone since it does not match the characteristics of the target, a circular void 2 m in diameter, input to the DTMLE algorithm. The example illustrates the capacity of this algorithm to distinguish target of the same apparent size but different characteristics.

The next result presented serves to demonstrate the capability of the DTMLE approach, in this case used with magnetometer measurements to differentiate targets of differing size and depth. Such differentiation is difficult with magnetometry since the diffusion process associated with the detection of features that perturbed the Earth's magnetic field lines yields considerable ambiguity between target size and depth. This is particularly problematic for the use of magnetometry at buried unexploded ordnance sites since such sites are also expected to be rich in fragments of exploded ordnance that will produce a magnetometer response. These fragments will, in general, be smaller and

shallower than the targets of interest and the problem of concern is the differentiation of unexploded ordnance from fragments. Figure 4 is an illustration of the target con-

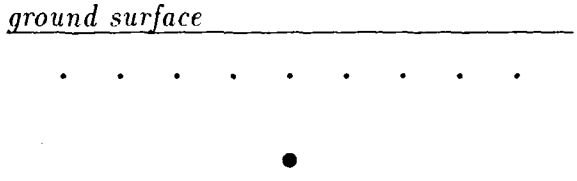


Figure 4: Illustration of the target configuration used for the simulation of magnetometry data.

figuration used for simulation of magnetometry data. Numerous small, shallow circular targets are present along with a single larger circular target at a greater depth. The DTMLE magnetometer algorithm was applied to this data set with the target characteristic being the diameter of the larger target. Figure 5 is the image of the target location probability displayed as gray scales. The darkest region of the figure corresponds to the large target location with no evidence of the presence of the smaller targets. Here the DTMLE matched filter has effected complete rejection of these irrelevant features.

Conclusions

The GPST serves as the basis for signal processing algorithms either to image (GDT) subsurface features or for the detection and location estimation (DTMLE) of known underground targets. A derivation of a particular GPST is given and the relevance of this relationship to various geophysical tools is explored. In addition, several results of the application of these algorithms to both simulated and actual field data are provided to illustrate the resolution that can be achieved and the potential applications for enhanced signal processing using GPST's.

To conclude, it is worthwhile to return to the intended scope of work presented here and endeavor to remark on the most practical geophysical tools and measurement geometries employing enhanced signal processing on the basis of the three applications considered. These findings are summarized in Table 2 and elaborated on below.

A first point that can be made about this table is that various permutations of geophysical tools and measurement geometries employed in a given application with enhanced signal processing is essentially the same as those currently used in geophysical field studies. The use of enhanced signal processing does not affect the suitability of a

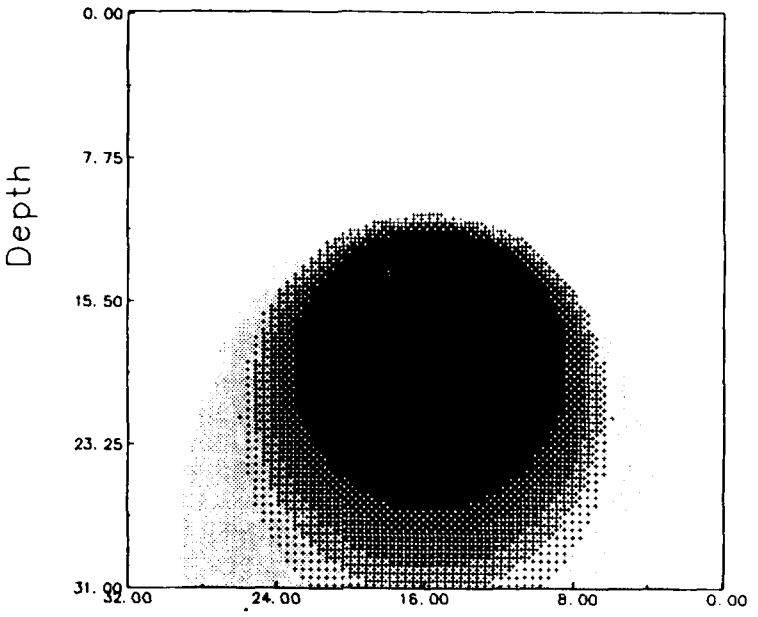


Figure 5: Image of the location probability for a large circular disk in the presence many smaller disks based a simulated magnetometer data.

particular tool to a given application but increases the resolution that can be achieved, provides more quantitative results and can significantly improve the prognosis for success in any field study. One major exception to this is the use of seismic imaging at waste sites with an offset VSP geometry. While seismic reflection and refraction are sometimes used at buried waste sites, they are not employed to locate isolated inclusions such as buried waste. Instead they are used for stratigraphic applications in characterizing the host site. The application of offset VSP GDT offers a seismic method that has proven quite valuable for imaging buried waste where no seismic-based method has been used before.

The resolution indicated in Table 2 differentiates wave based methods from other methods. For the wave based methods, the resolution is wavelength dependent and subsurface features with length scales on the order of one-quarter of a wavelength can faithfully be detected and imaged. For the "dc" tools, the wavelength is infinite, however, here the resolution is proportional to the spacing between adjacent measurement points. Since these methods employ exponentially damped fields, signal-to-noise ratio considerations limit the actual resolution that can be achieved. EM tools use extremely low frequencies and deal with evanescent waves. For this reason, there is the same signal-to-noise ratio concern as the "dc" tools, however, resolution improves with increasing

method	buried waste	tunnels	ordnance detection	resolution
offset VSP seismic	√	√[1]		wavelength
seismic refelection		√[1]		wavelength
cross-hole seismic		√[2]		wavelength
cross-hole radar		√[2]		wavelength
GPR	√	√	√	wavelength
magnetometry	√	√[1]	√	measurement spacing
EM	√	√[1]	√	frequency
resistivity		√[1]		measurement spacing

[1]shallow tunnels

[2]deep tunnels

Table 2: Summary of applicability and resolution for enhanced signal processing.

frequency.

While it is possible to envision new geophysical tools and measurement geometries for use with enhanced signal processing algorithms, the focus of this study is a more practical one. Here, signal processing algorithms are developed that are driven by particular applications and compatible with existing geophysical methods. These algorithms serve to more fully exploit the available data without alteration of field procedures or implementation efficiency. Furthermore, most signal processing algorithms considered here can be executed in real time and in the field with the current state of digital signal processing technology.

References

[1]A.J. Devaney, *IEEE Trans. Geosci. and Remote Sensing*, **GE-22**, 3 (1984).

[2]W.C. King, A.J. Witten, and G.D. Reed, *J. Env. Eng.*, **115**, 166 (1990).

[3]A.J. Witten and W.C. King, *J. Env. Eng.*, **116**, 166 (1990).

[4]A.J. Witten and W.C. King, *Civil Eng.*, **60**, 62 (1990).

[5]A.J. Devaney, *Ultrasonic Imaging*, **4**, 336 (1982).

[6]A.J. Witten and J.E. Molyneux, in *Frontiers in Applied Mathematics*, (SIAM, Philadelphia, 1992), to appear.

[7]P.M. Morse and H. Feshbach, *Methods of Theoretical Physics*, (McGraw-Hill, New York, 1953).

[8]A.J. Witten and J.E. Molyneux, *IEEE Trans. Geosci. and Remote Sensing*, **26**, 409 (1988).

[9]A.J. Devaney and G. Beylkin, *Ultrasonic Imaging*, **6**, 181 (1986).

[10]G.A. Tsihrintsis and A.J. Devaney, *IEEE Trans. Signal Processing*, **39**, 1466 (1991).

[11]A.J. Witten, *Inverse Problems*, **7**, L49 (1991).

YALAMANCHILI

Aerodynamic Heating of Unsteady Hypervelocity Projectile

Rao Yalamanchili, PHD
Close Combat Armaments Center, ARDEC
Picatinny Arsenal, NJ 07806-5000

INTRODUCTION

Recent and novel concepts for launching hypervelocity (HV) projectiles, such as electromagnetic propulsion, are currently being investigated by combined efforts of Army, Navy, Air Force, Marine and Defense Nuclear Agency. It is anticipated that these projectiles will be subjected to severe heat because of the aerodynamic behavior of the projectile traveling at hypersonic speed through dense air, and because of the thermal effects generated during launch. Defense critical technologies list include the following five, pertinent to this project, among twenty technologies: Computational Fluid Dynamics (CFD), Hypervelocity Projectiles, Simulation and Modeling, Parallel Computer Architecture, and Software Producibility. United States is ahead in these technologies except Hypervelocity Projectiles. An accurate modeling of thermodynamic phenomena associated with the launch and flight of the unsteady HV projectile is presented by consideration of CFD models, and models based on transient three dimensional heat conduction in the projectile and equation of motion of projectile. ARO recognized it as a very complicated problem with many facets, each of which offers formidable challenges. Also, accurate simulation by the simultaneous use of CFD model, thermal model, and HV projectile motion model is beyond the scope of existing super-computers because of limited speed and memory. Until then, salient features and individual effects can be examined while preparing for an accurate debut in the near future.

COMPUTATIONAL FLUID DYNAMICS MODEL

Based on the remarks in the previous section, one definitely requires a model based on full Navier - Stokes equations. The geometry of HV projectiles is unknown. There are some general purpose computer programs available for the solution of full Navier - Stokes equations. For example, Ballistic Research Laboratory has a program based on explicit finite

difference techniques. It is used primarily for research and to obtain steady state solution only. NASA -Langley and NASA-Ames have limited computer programs for solution of full set of Navier - Stokes equations. However, limited documentation exists because of continuous development and related changes and still considered as state of the art research tools. Moreover, these tools can't be treated like "Black Box" because of difficulties associated with complex set of nonlinear equations, variable grid generation to match gradients of fluid variables, sensitivity to input of initial and boundary conditions, and finally need of supercomputers. However, it is decided to utilize only -existing codes because of limited resources. Remember to include unsteady effects in selection of code and in performing any analysis. Consider unsteady terms completely without any compromise. Its importance is already documented elsewhere by Yalamanchili(1) during launch process in a conventional weapon system.

The following example (1) illustrates the importance of unsteady Phenomena (effect) and also questions the validity of analogy between momentum and energy transfer for unsteady flow. An accurate simulation of heating of hypervelocity projectile involves simulation not only during its aerodynamic flight but also during its launch from zero velocity to its peak hypervelocity in order to provide proper initial conditions in the unsteady aerodynamic flight. Because of the multidisciplinary phenomena contained in the general case, the numerical modeling results will be difficult to interpret. Therefore, the value of such computations would be greatly enhanced by examining separate simplified cases in which the various phenomena are treated in isolation. Results of such special cases will enhance general understanding and the interpretation of more general results. Hopefully, these results may reveal that some of the effects and phenomena included in the general description may be of minor practical consequence and thus don't need to be included and others that are of greater importance than presently realized. Here, the emphasis is placed on examining the unsteady and analogy effects. Since the state of the art in CFD models and on mechanics of turbulence models is quite limited, it is also common to develop solution procedures for only continuity and momentum equations and then to compute the heat transfer data by the use of analogy between momentum and energy transfer. Such procedures could include all parameters varying with respect to not only time but also streamwise and transverse positions. If the analogy between momentum and energy transfer is valid even for unsteady flow, more accurate skin friction and heat transfer data can be computed than the method discussed at the convenience of simple procedures and also substantial reduction in computational times. Heat transfer by turbulence is qualitatively similar to momentum transfer. However, the molecular diffusivity of heat, i.e. thermal conductivity is different form the molecular diffusivity of momentum, i.e., viscosity. Also the pressure gradients appear in the momentum equation but not in the energy equation except in special cases. It appears that Reynolds quantitative analogy between heat transfer parameters and momentum transfer

parameters is approximate only near the wall in the boundary layer and crude near free boundaries even though the analogy factors may be good enough for data correlation purposes. Therefore, let us compare the results of both large caliber and small caliber projectiles with and without analogy (such as Colburn) between momentum and energy transfer.

The governing equations are a system of nonlinear parabolic partial differential equations with three independent variables for an axisymmetric projectile without angle of attack. A combination of MWR, in particular the method of moments, and method of characteristics is used to reduce the unsteady equations to a system of ordinary differential equations. At the end, one may obtain the numerical solution even by personal computers. Two numerical examples are shown in Figure 1: Typical small caliber and large caliber projectiles. However, the trends remain the same.

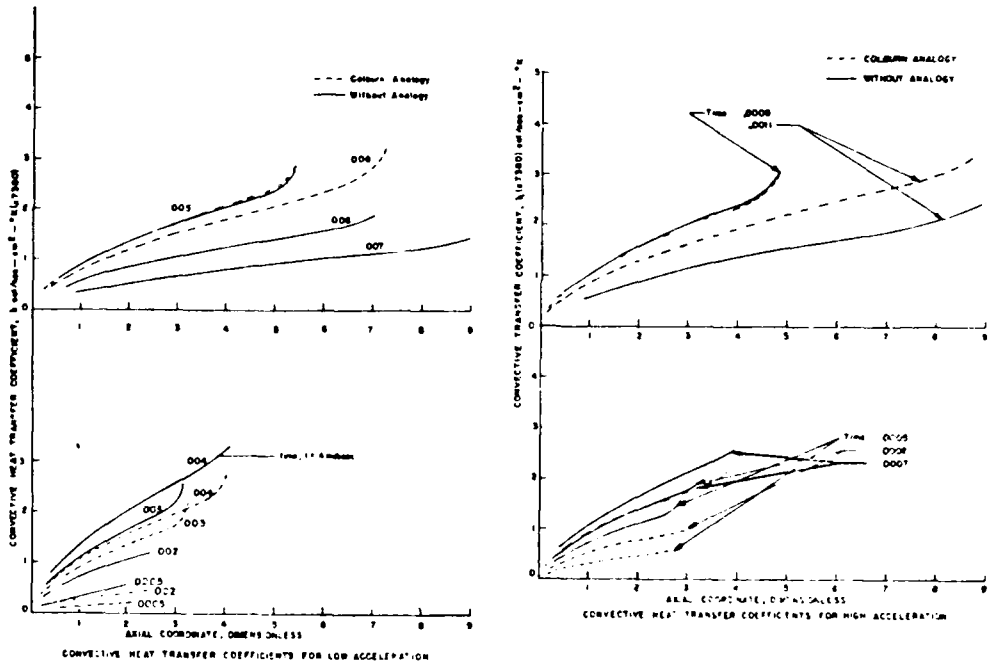


FIG. 1

It is found that the boundary layer parameters and in turn convective heat transfer coefficient vary drastically not only with respect to time but also with axial coordinate. It is not recommended to use an average value. Another observation is that the transient viscous shear stresses are higher than its steady state counterpart. A comparison is made between the results based on continuity, momentum, and energy equations and the results obtained by an improved analogy, such as Colburn, and the solution of continuity and momentum equations. The convective heat transfer results

based on the Colburn analogy underestimate as much as 100 percent for small times and the same analogy overestimates up to 50 percent for large times. Further investigations are necessary to obtain more appropriate analogy, if one exists, for unsteady flow. Until then, it is recommended to avoid utilization of analogy or algebraic treatment of predicting heat fluxes if CFD models are not pursued for unsteady flow.

CFD is applied routinely to complicated steady state flows and it is influencing the design of aerospace structures and will undoubtedly lead the way for the design of HV projectile. CFD is also used to conduct experiments and to interpret the results. Since the CFD is so complex, especially that involve full Navier -Stokes equations, the following are considered as a major research: 1) validation of computer code, 2) investigation of hypersonic flow field influenced greatly by real-gas chemistry, compressibility, boundary layer transition, surface heat transfer, and structure of shock wave effects, 3) scientific visualization of data generated by CFD on workstations, and 4) physics of turbulence by interaction with experimental data.

The unsteady flows are common not only with HV projectiles but also with high performance aircraft flying at large angles of attack. Tremendous potential for research exists in this area in addition to coupling of chemistry models and fluid dynamic models. The more general CFD models are being programed for massively parallel computers and will pave the way for attacking more complex flows.

THERMAL MODEL

The ultimate goal is to prepare a self contained package of simulation from the time a weapon is fired until the time the hypervelocity projectile reaches the target. The thermal model involves convection, conduction, and radiation in addition to other complications such as unsteady, three dimensional and hypervelocity (viscous-inviscid interaction and non-equilibrium flow) effects. For example, convective hear transfer may be modeled from CFD package which is governed by full three dimensional Navier Stokes equations. Typical computer run may take more than 20 hours on a Cray supercomputer in order to simulate one steady state supersonic flow field around a projectile.

The transient three dimensional heat conduction model will provide a means to determine the temperature distribution as a function of location (3 dimensions) and time for any given initial and boundary conditions. The boundary conditions are usually obtained from CFD. There are occasions where there is a strong coupling between convective flow and conduction. No matter what method is utilized, all require greater computer storage and large amounts of computer time. The solution process is not only subjected to these restrictions but also bound to blow-up, in the middle, if

accuracy, stability, and nonoscillation characteristics are not taken into account by proper selection of numerical techniques. If a combined convection and conduction problem is attempted in one-step, the failure in one area can lead to a loosing proposition in both areas. Therefore, a search is initiated to find an accurate, robust and efficient numerical scheme for the solution of transient three dimensional heat conduction problems.

Various numerical methods are in use today. The most popular methods are finite-element (FE), finite-difference (FD), and Weighted residuals (MWR). It is not an easy task to single out the best one. In any case, one has to bring all these methods into the same format in order to make any meaningful comparison. A methodology(2) is established for comparison of various numerical methods, in particular, FE, FD, and MWR.

FINITE ELEMENT ANALYSIS

Wilson and Nickell, following Gurtin's discussion of variation principles for linear initial value problems, confirmed that the function T(x,y,z,t) which leads to an extremum of functional

$$\Omega_*(T) = \tfrac{1}{2}\int_V \left[\rho C_v T^* T + \nabla T^* K^* \nabla T - 2\rho C_p T_o^* T\right] dV - \int_S Q_i n_i^* T \, dS$$

is the solution of the following transient heat-conduction equation:

$$\left(K^* T_{,i}\right)_{,i} - \rho C_p * \frac{\partial T}{\partial t} + \rho^* P = 0$$

with the boundary condition: $K^* T_{,i} - Q_i = 0$

where $T(x,y,z,t)$ = temperature at the spatial point (x,y,z) and at time t

T_o = Initial temperatures

∇T = Gradient of T with respect to spatial coordinates

K = Thermal conductivity

ρ = Material density

Cp = Heat capability of the material per unit mass

$$\hat{Q}_i(x,y,z,t) = \int_0^t Q_i(x,y,z,\tau)\, d\tau$$

V = Volume

* = Convolution symbol defined as:

$$T^* T = \int_0^t T(x,y,z,t-\tau) T(x,y,z,\tau)\, d\tau$$

$$\nabla T^* \nabla T = \frac{\partial T}{\partial x} * \frac{\partial T}{\partial x} + \frac{\partial T}{\partial y} * \frac{\partial T}{\partial y} + \frac{\partial T}{\partial z} * \frac{\partial T}{\partial z}$$

Divide the three dimensional solid body into I axial elements (nodes 0 to I), J transverse elements (nodes o to J) and K normal elements (nodes 0 to K) such that step sizes are same in all three directions. This restriction is introduced to simplify algebraic manipulations involved in

the analysis.

Consider the nodal point (i,j,k), in the range ($0<i<I$, $0<j<J$, $0<k<K$) as shown in Figure 2. The temperature of the nodal point will vary as a function of time, t. There are eight regions around this nodal point. The temperature distribution in any region is a function of spatial coordinates (x,y,z) and surrounding nodal point temperatures. For simplicity, linearity and the same functional distribution are assumed for all elements.

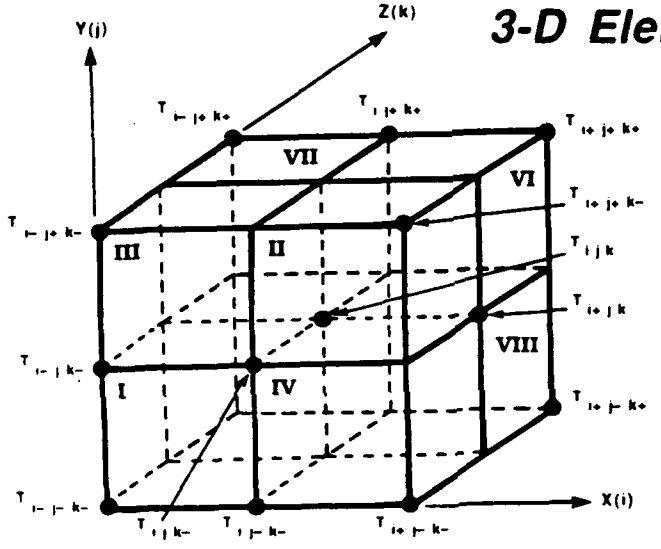


Figure 2a. 3-D Element.

A set of equations involving temperature distributions and corresponding gradients can be derived for all eight regions of a three-dimensional finite element discussed above. One can substitute all equations derived into all three terms of functional, governing equations, and integrate over the volume and take the first variation with respect to $T_{i,j,k}$ at the new time in order to obtain the extremum of the functional. The procedures for the first and third terms of governing equation are same. However, double convolution symbol is involved in the second term. Here, the evaluation of the second term in the governing equation involves integration not only over the volume (and the use of first variation with respect to the nodel temperature, $T_{i,j,k}$ at the new time) but also over the time-step due to the additional convolution symbol. Towards this goal, a linear nodal point temperature variation is assumed within each time-step. Summing up the results of all three terms produces an alternative finite element equation in a form familiar to finite difference community. More details are available (3). The end result is given below.

$$\frac{1}{216}\{(T^{+}_{---}-T^{o}_{---})+4(T^{+}_{--o}-T^{o}_{--o})+(T^{+}_{--+}-T^{o}_{--+})+4(T^{+}_{-o-}-T^{o}_{-o-})$$
$$+16(T^{+}_{-oo}-T^{o}_{-oo})+4(T^{+}_{-o+}-T^{o}_{-o+})+(T^{+}_{-+-}-T^{o}_{-+-})+4(T^{+}_{-+o}-T^{o}_{-+o})$$
$$+(T^{+}_{-++}-T^{o}_{-++})+4(T^{+}_{o--}-T^{o}_{o--})+16(T^{+}_{o-o}-T^{o}_{o-o})+4(T^{+}_{o-+}-T^{o}_{o-+})$$
$$+16(T^{+}_{oo-}-T^{o}_{oo-})+64(T^{+}_{ooo}-T^{o}_{ooo})+16(T^{+}_{oo+}-T^{o}_{oo+})+4(T^{+}_{o+-}-T^{o}_{o+-})$$
$$+16(T^{+}_{o+o}-T^{o}_{o+o})+4(T^{+}_{o++}-T^{o}_{o++})+(T^{+}_{+--}-T^{o}_{+--})+4(T^{+}_{+-o}-T^{o}_{+-o})$$
$$+(T^{+}_{+-+}-T^{o}_{+-+})+4(T^{+}_{+o-}-T^{o}_{+o-})+16(T^{+}_{+oo}-T^{o}_{+oo})+4(T^{+}_{+o+}-T^{o}_{+o+})$$
$$+(T^{+}_{++-}-T^{o}_{++-})+4(T^{+}_{++o}-T^{o}_{++o})+(T^{+}_{+++}-T^{o}_{+++})\}$$

$$=\frac{\theta}{2}\{\frac{1}{9}(T^{+}_{---}+T^{+}_{--o}+T^{+}_{--+}+T^{+}_{-o-}+T^{+}_{-oo}+T^{+}_{-o+}+T^{+}_{-+-}+T^{+}_{-+o}+T^{+}_{-++}$$
$$+T^{+}_{o--}+T^{+}_{o-o}+T^{+}_{o-+}+T^{+}_{oo-}-26T^{+}_{ooo}+T^{+}_{oo+}+T^{+}_{o+-}+T^{+}_{o+o}+T^{+}_{o++}$$
$$+T^{+}_{+--}+T^{+}_{+-o}+T^{+}_{+-+}+T^{+}_{+o-}+T^{+}_{+oo}+T^{+}_{+o+}+T^{+}_{++-}+T^{+}_{++o}+T^{+}_{+++})$$
$$+\frac{1}{9}(T^{o}_{---}+T^{o}_{--o}+T^{o}_{--+}+T^{o}_{-o-}+T^{o}_{-oo}+T^{o}_{-o+}+T^{o}_{-+-}+T^{o}_{-+o}+T^{o}_{-++}$$
$$+T^{o}_{o--}+T^{o}_{o-o}+T^{o}_{o-+}+T^{o}_{oo-}-26T^{o}_{ooo}+T^{o}_{oo+}+T^{o}_{o+-}+T^{o}_{o+o}+T^{o}_{o++}$$
$$+T^{o}_{+--}+T^{o}_{+-o}+T^{o}_{+-+}+T^{o}_{+o-}+T^{o}_{+oo}+T^{o}_{+o+}+T^{o}_{++-}+T^{o}_{++o}+T^{o}_{+++})\}$$

A short nand notation is used for subscripts and superscript. The superscript '+' is used to indicate the temperature at the new time whereas the superscript 'o' indicates the old value. There are three subscripts (i,j,k) to indicate the directions x,y, and z, respectively. The subscript in x-direction can be i-1, i, or i+1. The same subscript in x-direction in short hand notation, becomes -,o, or +. Similar notation is used for j and k along y and z directions. For example,

$$T_{i-1,j,k+1} = T_{-o+}$$

Thus, one can see omission of letters (i,j,k), 1, and commas from all subscripts for clarity and conciseness.

FINITE DIFFERENCE ANALYSIS

Two categories of finite difference (FD) equations have been employed: The explicit FD equations (the temperature at time t is expressed in terms

of temperatures at one time interval, Δt, earlier) and the implicit FD equations. They represent a direct approximation approach to the partial differential equation type of formulation. The finite element (FE) analysis belongs to the class of implicit technique in FD methods. Indeed, Yalamanchili (4,5) proved that FE and FD methods belong to the class of method of weighted residuals, in particular, Galerkin and Collocation methods respectively for transient two dimensional heat conduction problems.

Numerical approximations to solutions of the heat flow equation in three space dimensions may be obtained by the step-wise solution of an associated difference equation. It is the intent of this paper to develop several difference equations that may contain from a minimum of 7 nodes to a maximum of 27 nodal points available in a typical three dimensional element. Of course, the accuracy of these FD equations vary by orders of magnitude. However, it is straightforward to generate a system of algebraic equations and to express it in a matrix form for any chosen FD scheme. Proper numbering of nodes is essential in order to obtain a practical solution even though the matrix is sparse due to an exponential increase in arithmetic operations especially for transient three dimensional problems.

Consider the heat conduction equation in a three dimensional body of length (a), width (b) and height (c) with the following boundary conditions:

$$\frac{\partial T}{\partial t} = \alpha \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right)$$

$$T = T(x,y,z,t) = T^{l}_{i,j,k}$$

$$T(0,y,z,t) = 0 \;,\; T(x,0,z,t) = 0 \;,\; T(x,y,0,t) = 0$$

$$\frac{\partial T}{\partial x}(a,y,z,t) = 0 \;,\; \frac{\partial T}{\partial y}(x,b,z,t) = 0 \;,\; \frac{\partial T}{\partial z}(x,y,c,t) = 0$$

For any chosen finite increment Δx, Δy, Δz and Δt in both finite difference and finite element systems with a fixed value of $\alpha \Delta t(1/\Delta x^2 + 1/\Delta y^2 + 1/\Delta z^2)$, the efforts required to calculate the solution up to a given time is proportional to the number of spatial nodal points raised to the power of three. The number of spatial points changes drastically for multi-dimensional problems. Therefore, a variety of FD schemes, as well as the stability, accuracy, and oscillation characteristics of three dimensional problems are essential for economical and practical reasons. The following analysis is prepared to fulfill such an objective among other considerations. Rewriting the first time derivative in a finite difference form, the governing equation becomes

$$\frac{T^{+}_{000} - T^{0}_{000}}{\alpha \Delta t} = \phi \nabla^2 T^{+}_{000} + (1-\phi) \nabla^2 T^{0}_{000}$$

Where ϕ is a weighted parameter with respect to time and varies between 0 and 1 and the Laplacian term, $\nabla^2 T$, is written as

$$\nabla^2 T = \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}$$

This can be expressed into finite difference form in several ways. Yalamanchili (3) gave details elsewhere. Therefore, only final results are given below. The simplest and the least accurate three dimensional Laplacian can be written as $(O(h^2))$

$$\nabla^2 T = (T_{-00} + T_{0-0} + T_{00-} - 6T_{000} + T_{00+} + T_{0+0} + T_{+00})/h^2$$

where h is the spatial step size. There are only 7 nodes involved out of a possible 27 nodes. Similarly, the most accurate three-dimensional Laplacian can be written as $(O(h^6))$

$$\nabla^2 T = (T_{---} + 3T_{--0} + T_{--+} + 3T_{-0-} + 14T_{-00} + 3T_{-0+} + T_{-+-} + 3T_{-+0} + T_{-++}$$

$$+ 3T_{0--} + 14T_{0-0} + 3T_{0-+} + 14T_{00-} - 128T_{000} + 14T_{00+} + 3T_{0+-} + 14T_{0+0} + 3T_{0++}$$

$$+ T_{+--} + 3T_{+-0} + T_{+-+} + 3T_{+0-} + 14T_{+00} + 3T_{+0+} + T_{++-} + 3T_{++0} + T_{+++})/30h^2$$

All 27 nodal points are involved in this derivation. It is worth to mention another finite difference approximation for a three-dimensional Laplacian:

$$\nabla^2 T = (T_{---} + T_{--0} + T_{--+} + T_{-0-} + T_{-00} + T_{-0+} + T_{-+-} + T_{-+0} + T_{-++}$$

$$+ T_{0--} + T_{0-0} + T_{0-+} + T_{00-} - 26T_{000} + T_{00+} + T_{0+-} + T_{0+0} + T_{0++}$$

$$+ T_{+--} + T_{+-0} + T_{+-+} + T_{+0-} + T_{+00} + T_{+0+} + T_{++-} + T_{++0} + T_{+++})/9h^2$$

Again, all 27 nodel points are involved in this derivation. However, this is $O(h^2)$ accuracy. This may possibly unify various numerical methods for transient three-dimensional heat conduction problems.

Until now, a variety of FD schemes are derived for three dimensional Laplacian term in order to obtain not only more accurate numerical solutions but also to unify and compare FE and FD techniques. In general, it is understood that the higher order scheme yields more accurate solution than a lower order scheme. The accuracy of the numerical results can also be improved by reducing the grid spacing, h. Usually, grid spacing reduction improves the numerical results up to a certain extent. At this point, the numerical results are as accurate as can be with the chosen order of finite difference scheme. Further reduction in grid spacing will lead to increasing round off errors due to enormous increase in number of computations and thus, overall, less accurate results. However, the accuracy of the results can be improved by the use of higher order finite difference scheme.

The accuracy of a numerical solution may also be improved by proper selection of the weighted parameter, ϕ. This parameter also plays a major role in stability and oscillation characteristics of a numerical scheme. The parameter $\phi(0<\phi<1)$ allows a weighted average of sum of three second order spatial derivatives at two discrete times (old and new). An explicit scheme is the result when ϕ is set to zero; otherwise, an implicit scheme will be the result for remaining range of parameter, ϕ.

The general finite difference equation for an unsteady three-dimensional problem can be written as

$$
\begin{aligned}
&AT^{+}_{---}+BT^{+}_{--0}+AT^{+}_{--+}+BT^{+}_{-0-}+CT^{+}_{-00}+BT^{+}_{-0+}+AT^{+}_{-+-}+BT^{+}_{-+0}+AT^{+}_{-++}\\
&+BT^{+}_{0--}+CT^{+}_{0-0}+BT^{+}_{0-+}+CT^{+}_{00-}-DT^{+}_{000}+CT^{+}_{00+}+BT^{+}_{0+-}+CT^{+}_{0+0}+BT^{+}_{0++}\\
&+AT^{+}_{+--}+BT^{+}_{+-0}+AT^{+}_{+-+}+BT^{+}_{+0-}+CT^{+}_{+00}+BT^{+}_{+0+}+AT^{+}_{++-}+BT^{+}_{++0}+AT^{+}_{+++}\\
&=ET^{0}_{---}+FT^{0}_{--0}+ET^{0}_{--+}+FT^{0}_{-0-}+GT^{0}_{-00}+FT^{0}_{-0+}+ET^{0}_{-+-}+FT^{0}_{-+0}+ET^{0}_{-++}\\
&+FT^{0}_{0--}+GT^{0}_{0-0}+FT^{0}_{0-+}+GT^{0}_{00-}-HT^{0}_{000}+GT^{0}_{00+}+FT^{0}_{0+-}+GT^{0}_{0+0}+FT^{0}_{0++}\\
&+ET^{0}_{+--}+FT^{0}_{+-0}+ET^{0}_{+-+}+FT^{0}_{+0-}+GT^{0}_{+00}+FT^{0}_{+0+}+ET^{0}_{++-}+FT^{0}_{++0}+ET^{0}_{+++}
\end{aligned}
$$

All three FD approximations mentioned for three-dimensional Laplacian are substituted into the governing equation in order to obtain the above FD equation. However, the coefficients (A,B,C,D,E,F,G, and H) are different. There are many other FD schemes (3) that can be generated. This is not the intent to exhaust this avenue.

The MWR unifies many approximate methods of solution of differential equations that are in use today including FE and FD methods. Yalamanchili(5) discussed it before with reference to transient heat conduction problems. The beauty of this technique lies in the fact that one can obtain the exact solution, if exists, if the right choice is made in the selection of approximate solution form.

OTHER INVESTIGATIONS

No decision has been made in the selection of methodology for thermal analysis of HV projectile. There are many other reasons why we should take an open - minded approach. Parallel approach always accelerate the process than serial approach provided one has money. No one knows yet what computational fluid dynamic model and code will be adopted for this investigation rather than develop another one from scratch because of limited resources. Still more research is needed regarding mathematical characteristics, such as accuracy, stability and nonoscillation, of various numerical methods. Keep in mind other considerations such as

practicality. One has to repeat thousands of time-steps and also one has to generate and solve thousands of simultaneous equations in each and every time-step. Sometimes one has to resort to splitting methods (3) because of practical considerations. Also note that there is either a step change or a steep change in material properties, as shown in Figure 2, near the interface, such as melting.

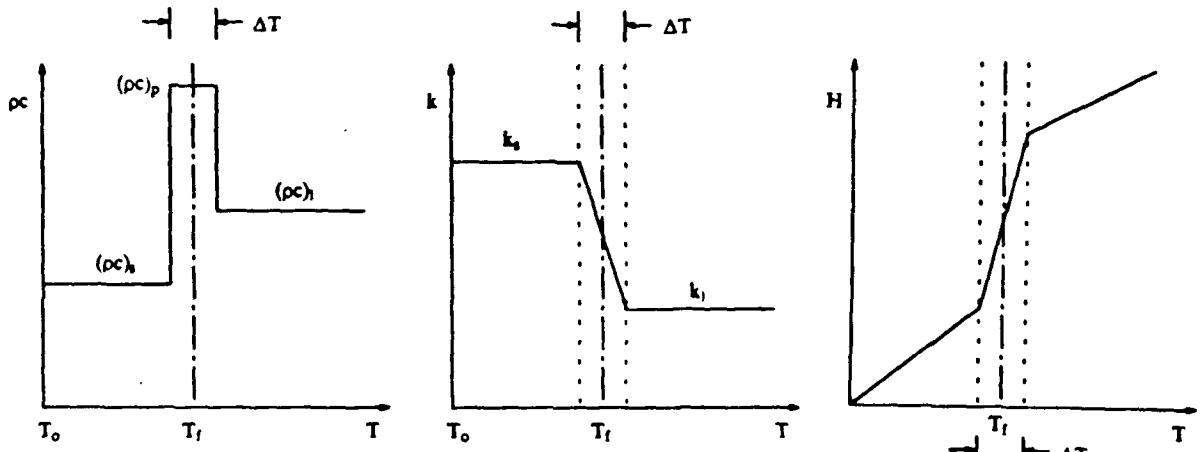


Fig 2(a)ρc, Thermal Conductivity and Enthalpy variations with temperature

A thorough investigation is conducted not only In-House but also outside in the form of two small STAS contracts at Universities. Of course, parallel approach is adopted to speed up the process so that electro-magnetic community can be served sooner without going through common practice of cut and try techniques. It is not the intent to provide everything in detail. Rather only highlights are summarized. The University of Kansas (6) presents a p-version hierarchial axisymmetric solid finite element formulation (similar formulation has also been constructed for planar case) with solid/liquid phase change including radiation effects. In the p-version FE method, the order of approximation for each element of a fixed grid is increased; thereby adding more degrees of freedom to an existing model in order to increase the accuracy of the solution. The p-version FE method yields a much higher convergence rate and requires very simple models which are generated only once. The rate of convergence of the p-version is higher for singular problems than for smooth problems.

Under the University of California (7) STAS contract the capability is developed and established to numerically simulate on the digital computer the melting of projectiles and projectiles parts, such as, fins. The basic equation is well known for a problem involving melting. The difficulties are the geometry, surface gradients near the melting zone, the moving grid or mesh, and the analysis of the surface processes. The first law of thermodynamics for a solid, in finite volume form, can be written as

$$\frac{\partial}{\partial t}\iiint_V \rho C_v T \, dV = \iint_A K \nabla T \cdot d\vec{A}$$

where dV and dA are cell volume and cell area respectively. For the physical problem of melting the equation must be modified to allow for change of control volume in time, and it is necessary to use the geometric continuity equation together with Liebnitz rule in the energy equation. The continuity equation is

$$\frac{\partial}{\partial t}\iiint_V dV = \int_A \vec{V}_b \cdot d\vec{A}$$

where $\vec{V}_b$ is the velocity of the cell boundaries and this velocity is due to the melting of the cell or cells. The new form of the energy equation becomes

$$\frac{\partial}{\partial t}\Big|_b \iiint_V \rho C_v T \, dV - \iint_A \rho C_v T \vec{V}_b \cdot d\vec{A} = \iint_A K \nabla T \cdot d\vec{A}$$

where the derivative is taken with respect to an observer moving with the cell. This observer will see both a time change and a change in cell volume.

The present computer code applies second order approximations in space and time to approximate the derivatives in the above equation, and the mesh can be non-orthogonal. The code also allows for special four and five sided control volumes which are useful at the origin of cylindrical and spherical geometries.

Several simulations are performed on typical projectile configurations including fins. The nonsymmetric case due to angle of attack is shown in Fig 3. For this situation the melting will be greater then on one side of the projectile relative to the other. For early times, there is no significant melting and the temperature gradients are significant in the body and on the surface. The maximum heat transfer occurs in the spherical portion of the projectile, and the maximum temperature occurs in this region. However, the spherical cap of the projectile begins to flatten at later times. Also, the entire windward side of the projectile which is at the melting temperature regresses at a rate faster than the leeward side.

Finally, full three-dimensional simulation is performed for a fin of

Non-axisymmetric solutions

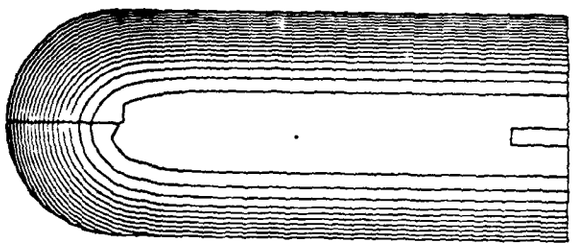


Figure (3B) Early time isotherms

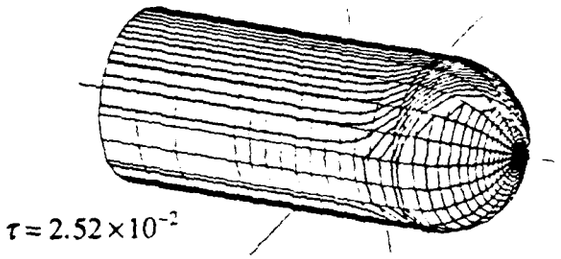


Figure (3c) Early time surface isotherms

Non-axisymmetric solutions

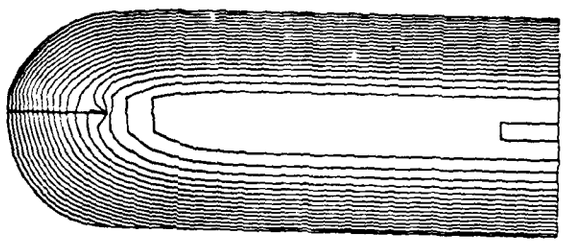


Figure (3b) Late time isotherms

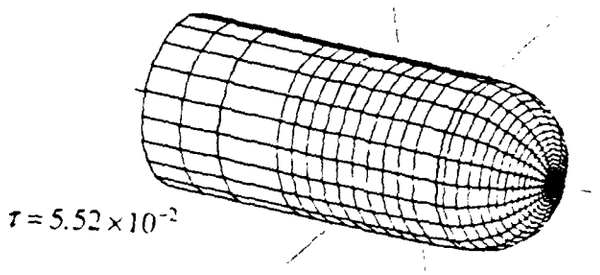


Figure (3C) Late time surface isotherms

KE projectile. The external flow conditions are time-dependent. It is magnified by a factor of twenty for visibility. At a time of 0.66 seconds, there is a small region of material at the melting point on the slanted surface near the last isotherm of Fig 4. The temperature isotherms show that a large portion of the surface is at the melting temperature at 2 seconds and a slight change in fin shape is observed.

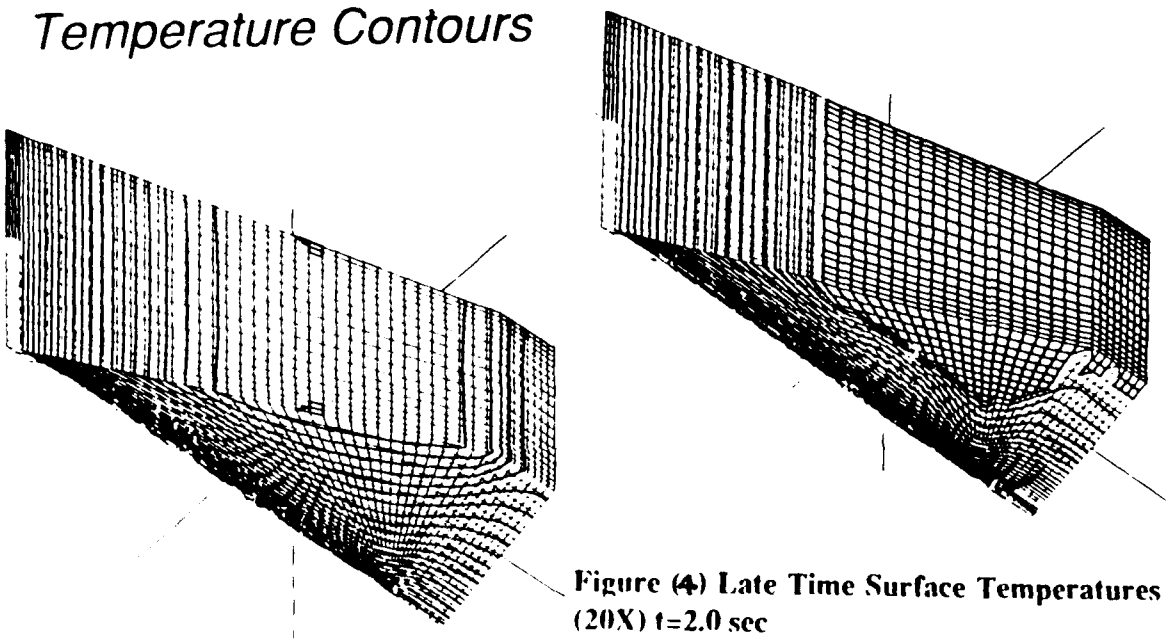


Figure (4) Late Time Surface Temperatures (20X) t=2.0 sec

Figure (4) Early Time Surface Temperatures on the Fin (20X) t=0.66 sec

The results show less heating due to lack of in-bore heating and cold surface initial condition even though the fins were in the combustion chamber at the start of motion. Full transient three-dimensional finite volume computer code with provision for melting is of interest not only at ARDEC but also at BRL. Such complimentary programs are essential for survival because of dwindling resources.

MODEL OF PROJECTILE MOTION

The projectile accelerates during launch and decelerates in flight. The projectile never remains steady. There may be significant uneven pressure distribution around the projectile especially with angle of

attack. One can form an equation of motion of a projectile from Newton's second law. Forces acting on a projectile consists of frictional forces due to viscous boundary layers, pressure forces, and gravitational forces. The second order ordinary differential equation arising form equation of motion can be integrated twice in order to obtain velocity and the location of hypervelocity projectile. The density of air, in space, is a strong function of altitude. However, one can determine the density once the location of projectile is determined. Melting and associated loss of material from the surface complicates the equation of motion further.

GENERALIZED PROJECTILE HEATING MODEL

Until now, fictitious models, such as CFD model, thermal model, and model of projectile motion are discussed. In general, these are not independent. Each and every model require initial and boundary conditions which are obtained from the solution of other model(s). Thus, these models are coupled and to be solved simultaneously in order to make it realistic. However, it is better to develop in modular form. In this way, one can easily upgrade, as needed. More over, present computers may not have enough speed in order to obtain the results in a reasonable time. In this case, one may be able to solve the generalized overall model by the use of more than one computer in parallel processing as in the case of breaking a large number into two prime numbers by the use of idling computers all over the world (8). Thus obtained the results in a few months instead of obtaining same results by operating a medium size computer for more than 200 years.

CONCLUSIONS AND RECOMMENDATIONS

An accurate model of heating of a hypervelocity projectile is discussed. The model is presented in a modular form so that one can update easily as new developments take place in related areas. Various investigators think that they can solve those modular models separately. Also some tried to utilize many assumptions into the governing full Navier -Stokes equations and utilize empirical formulations. What happened to the design of KE projectiles in the past? The fins of KE projectiles were found to be melting and forced to fly in erratic paths due to lack of stability. A quick fix was found by providing ablative coating material so that there wouldn't be any structural damage to the fins. Design of KE projectiles is supposed to be lot simpler than the design of future hypervelocity projectiles because of less severe aerodynamic heating and aerodynamic forces. No matter what type of projectile it is, it is always necessary to design it with mass as small as possible without much factor of safety because of practical considerations such as need to go faster for air defense and/or to carry more explosive power. A design based on sophisticated tool will not only alleviate future mishaps and fix-ups but also will aid in trouble shooting by red teams in addition to achieving

optimization.

Derived a new finite difference equation which didn't appear in the literature in addition to other variants for implicit, Crank-Nicholson, and explicit finite difference schemes for transient three-dimensional heat transfer of hypervelocity projectile based on new as well as other Laplacian finite difference approximations derived. Presented a technical paper based on these findings at 9th Army Conference on Applied Mathematics and Computing, Minnesota Supercomputer Center, University of Minnesota, Minneapolis, MN in June 1991. Presented the importance of unsteady flow analysis during launch in 4th International Computational Fluid Dynamics Symposium at the University of California-Davis in Sep '91. Also presented methodology and progress on Aerodynamic Heating of Hypervelocity Projectile Project to the visitors of Army Research Office, PEO, and ARDEC personnel during ARO Annual Reviews in April '91.

One of the objective behind the derivation of new finite difference approximation for Laplacian is to compare in future, the most popular numerical methods, i.e., finite element and finite difference methods for transient three-dimensional heat transfer problems. It is important to note that there is a need to find a best scheme in case one tackles the combined transient three-dimensional heat conduction and convection problems simultaneously because of limitations on supercomputer's storage and speed in existence today.

Various versions of finite element methodology are discussed. In the p-version of finite element method the order of approximation for each element of a fixed grid is increased; thereby adding more degrees of freedom to an existing model in order to increase the accuracy of solution. The p-version of finite element method yields a much higher convergence rate and requires very simple models which are generated only once. The rate of convergence of p-version is higher for singular problems than for smooth problems. The classical finite element method belong to the class of h-version. The analytical reduction of classical finite element method to a finite difference form is completed and to be presented in a forthcoming Tenth Army Conference on Applied Mathematics and Computing at US Military Academy in June 1992. However, the accuracy, stability and nonoscillation characteristics are not established for both finite element and finite difference schemes. Based on intuition, the h-version finite element method may be inferior to other variations of finite difference techniques. Therefore, p-version may be chosen if finite element methodology is adopted. The basic equation is well known for a problem involving melting and the difficulties are the geometry, surface gradients near the melting zone, the moving grid or mesh, and the analysis of the surface processes. The computer code applies second order approximations in both time and space to approximate the derivatives. The code also allows special four and five sided control volumes which are useful at the

origin of cylindrical and spherical geometries and arbitrary number of blocks of grid.

Various numerical simulations of aerodynamic heating of typical projectiles are performed in a supersonic region. The initial temperature conditions of its flight are assumed same as atmospheric conditions (which is not true). Still, melting temperature is reached at various locations in a typical 2-second supersonic flight.

Developed and established the capability to numerically simulate on a computer the melting of projectiles and projectiles parts, such as fins, due to hypervelocity environment. The resulting transient three-dimensional computer code can be modified easily for both parallel and serial computers because of its modular form. Since this code is independent of stress analysis codes, one can use it as a simple general purpose heat transfer code.

The computer time is about 20 hours for simulation of steady state flow field around the supersonic projectile on supercomputers without the heat transfer into the projectile if full Navier-Stokes model is adopted. Unsteady flight simulation, time-accurate solution, and consideration of heat transfer into the projectile and at the surface takes much longer. The aerodynamic heating of hypervelocity projectile is the most difficult problem with many complexities including several critical technologies listed by DOD. No matter what developments take place including electromagnetic/ electrothermal-chemical efforts, in the near future, the hypervelocity projectiles are stressed up to their critical limit. The complexities of hypervelocity projectile are due to unsteady, three-dimensional, non-equilibrium, and compressible effects, viscous-inviscid interaction, laminar-transition-turbulent flow, mechanics of turbulence modeling, shockwaves/flow separations, and need of supercomputers. Unless one makes genuine efforts with at least 2 man-years In-House, and two consultants from universities no one can make any significant progress in solving one of the most difficult problems in the world.

REFERENCES

1. Yalamanchili, R., "Will Analogy Hold for Unsteady Turbulent Boundary Layers," Proc. of 4th International Symposium on Computational Fluid Dynamics, Univ. of California, Davis, CA (Sep 1991).
2. Yalamanchili, R. and S. Yalamanchili, "Critical Time-Step of Various Numerical Schemes for Transient Heat Conduction," Transactions of the Eighth Army Conference on Applied Mathematics and Computing, ARO Report 91-1, US Army Research Office (1991).
3. Yalamanchili, R. and S. Yalamanchili, "Various Finite Difference Schemes for Transient Three-Dimensional Heat Conduction," Transactions of the Ninth Army Conference on Applied Mathematics and Computing, ARO Report 92-1, U.S.

Army Research Office (1992).
4. Yalamanchili, R.V.S., and S.C. Chu, "Stability and Oscillation Characteristics of Finite Element, Finite Difference, and Weighted Residual Methods for Transient Two Dimensional Heat Conduction in Solids, Journal of Heat Transfer, Trans. of ASME, Vol. 95, Series C, #2, (1973).
5. Yalamanchili, R., "Accuracy, Stability, and Oscillation Characteristics of Transient Two Dimensional Heat Conduction, ASME Paper # 75-WA/HT-85 (1975).
6. Surana, K.S., D.A. Gyorog and A. Bose, "Aerodynamic Heating of Hypervelocity Projectiles: Axisymmetric and Planar Linear and Nonlinear Heat Conduction with Solid/Liquid Phase Change Using P-Version Finite Element Method, The Univ. of Kansas, Lawrence, KS (Sep 1991).
7. Dwyer, H.A.,"Modeling of Melting Around a Hypervelocity Projectile," Final Report under Contract # DAAL03-86-D-0001, Delivery Order 2257, The University of California (June 1991).
8. Lenstra, A., Personal Communication, Bellcore, AT&T, Whippany, NJ (June 1990).

Chemistry of Adhesives at Surfaces

Walter X. Zukas
U.S. Army Materials Technology Laboratory
Polymer Research Branch
Watertown, Massachusetts 02172-0001

For the past several years, the U. S. Army has been engaged in a major effort to improve the quality of adhesively bonded joints in a broad range of materiel[1]. This effort involves both short term problem solving and long term research in adhesion science. As part of this latter effort, in recognition of the critical role of the "interphase" in controlling the performance of adhesive bonds[2,3], we have been studying the effect of adherend surfaces on adhesive cure chemistry. Our objective is a description, at the molecular level, of the structure of this region in support of ongoing efforts to construct predictive models of bonded joint strength and service life.

One of the first indications that an adherend surface might perturb the cure chemistry of an epoxy adhesive is the work of Comyn et al.[4]. Through the use of inelastic electron tunneling spectroscopy, it was concluded that no cure took place in the interphase due to preferential amine adsoption onto the aluminum oxide surface of an aluminum adherend.

Dillingham and Boerio[5] studied the interaction of an amine cured epoxy with a polished aluminum surface by means of X-ray photoelectron spectroscopy and reflection absorption infrared spectroscopy. They concluded that the catalytic action of the acidic hydroxyl groups on the aluminum oxide surface led to a cured resin with a higher crosslink density in this interphase. They inferred that the resultant increased brittleness was responsible for the failure which was observed to occur in this region.

Similarly, Garton et al.[6] have studied the effect of various carbon and carbon fiber surfaces on the course of an epoxy/amine reaction. Using thermal analysis and infrared spectroscopy, they determined that an acidic carbon surface (produced by oxidation) resulted in selective amine adsorption and the catalytic acceleration of its reaction with the epoxy.

Using transmission electron microscopy, Crompton[7] observed that the appearance of a cured epoxy adhesive in contact with an aluminum adherend differed from that in the bulk. A lower density material at the surface was evidenced by a higher electron transparency. It was further observed that the apparent adhesive failure of this joint was actually a cohesive

failure in this very thin "transition region".

Nigro and Ishida[8] used infrared spectroscopy to show that a polished automotive steel catalyzed the homopolymerization of an epoxy resin at the surface.

Although sometimes contradictory, results of these various studies clearly indicate an effect of adherend surfaces on cure chemistry. However, all make use of less than absolute methods to determine the structure of the cured adhesives in contact with these surfaces. We have undertaken an alternative approach to the investigation of this problem. Simplified crosslinking adhesive systems were first investigated, and then a monofunctional amine and epoxy were utilized in order to obtain soluble products amenable to molecular characterization by conventional analytical techniques. Both the crosslinking and model systems were used in conjunction with finely divided aluminum oxides which serve as high surface area models of aluminum adherend surfaces. Some of the initial results of this effort are the subject of this paper.

EXPERIMENTAL

The epoxy resins used in this study consisted of Epon 826, a diglycidyl ether of bisphenol A (DGEBA) resin (Shell Chemical Co.) and the monofunctional glycidyl ether of cumyl phenol (MGECP). The 826 was used as received. MGECP was prepared by the reaction of the sodium salt of cumyl phenol with an excess of epichlorohydrin followed by vacuum distillation of the crude product. MGECP is a clear liquid at room temperature with a purity of greater than 99% by high performance liquid chromatography (HPLC).

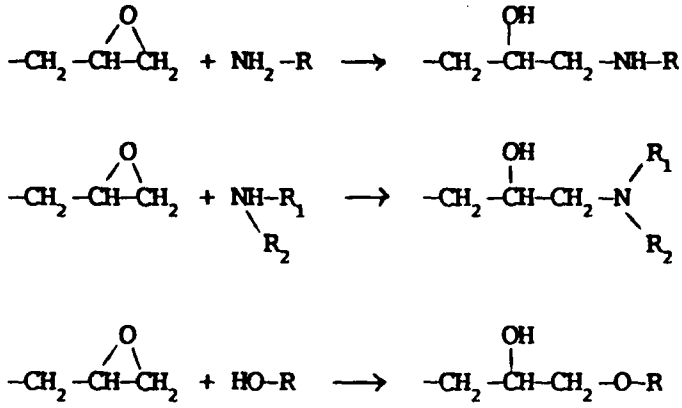
The curing agents used include the primary aromatic diamines 3,3'-diaminodiphenyl sulfone (3,3-DDS) and 4,4'-diaminodiphenyl sulfone (4,4-DDS) (Aldrich Chemical Co.) and the primary aromatic amine 4-aminodiphenyl sulfone (ADS)[9]. The purity of the curing agents was determined to be greater than 99% by HPLC.

Fillers employed in the differential scanning calorimetry (DSC) studies were an activated/neutral alumina, a fused alumina, and powdered aluminum. Fillers for all the additional experiments were a series of aluminum oxides from Aldrich Chemical Co.: activated acidic (A/A), activated neutral (A/N), activated basic (A/B) (all Brockmann I grade), and corundum (crystalline, α-form). The activated aluminas have a surface area of approximately 155 m^2/g with a pore size of 58A at 150 mesh. The corundum was 60 to 200 mesh.

Resin formulations were mixed at room temperature and stored at -18°C if not used immediately. The aluminum oxide of interest was admixed with the resin immediately prior to cure. Approximately 0.5 g of a filled mixture was prepared from which aliquots for further testing were taken. A detailed description of the DSC, HPLC, thermogravimetric (TGA), mass spectral, Fourier transform Raman (FTR), and elemental analyses is available[10].

RESULTS AND DISCUSSSION

The amine cure of epoxy resins generally proceeds through the following reactions:



The reactions represent primary amine addition to epoxy, secondary amine addition to epoxy, and subsequent homopolymerization, respectively. Shechter et al.[11,12] found the first two reactions to dominate the cure of a glycidyl ether epoxy with aliphatic amines. For a similar system Dusek et al.[13] arrived at the same conclusion, but found the third reaction to take place in the presence of excess epoxy. Anderson[14] also found a significant amount of homopolymerization to take place when glycidyl ether epoxies were cured with aliphatic amines. However, when an aromatic amine was used, the amount of homopolymerization became negligible. It was thought that the tertiary aromatic amine which is formed in the second reaction was too sterically hindered to have any catalytic effect on the third reaction. Bell[15] and Byrne et al.[16] also found that no homopolymerization took place in the cure of glycidyl ether epoxies with aromatic amines.

Each of the above three reactions has been reported[17] to release the same molar heat of reaction. DSC measurements can thus be related to the extent of reaction of the epoxy groups and can be used to study the kinetics of cure, providing the total heat of reaction remains constant.

The dynamic DSC results for the various 826/3,3'-DDS mixtures are shown in Table I. Addition of activated aluminum oxide was found to lower the exotherm peak temperature and decrease the T_g of the resulting cured material with higher filler loading levels leading to lower T_g's. These effects were observed for different heating rates and for different epoxy/amine ratios. The large shifts in the exotherm position and shape were not observed for the aluminum or fused alumina filled samples.

The heat of reaction values were calculated from the peak area defined by a straight baseline for the first heating curve added to the area

TABLE I
Dynamic DSC results for 826/3,3'-DDS

Sample	Exotherm peak T(°C)	Heat of Reaction J/g-resin	T_g(°C)
Epoxy/amine 1/1 (5°C/min)			
unfilled	191	396	156
50 phr aluminum	187	514	160
25 phr activated Al_2O_3	187	389	147
50 phr activated Al_2O_3	181	395	141
100 phr activated Al_2O_3	168	393	130
50 phr fused alumina	192	478	160
Epoxy/amine 1/1 (2.5°C/min)			
unfilled	176	485	166
25 phr activated Al_2O_3	169	439	154
50 phr activated Al_2O_3	164	353	139
100 phr activated Al_2O_3	151	321	136
Epoxy/amine 1.5/1 (2.5°C/min)			
unfilled	184	348	134
50 phr activated Al_2O_3	172	350	118

defined by the same limits for the second heating scan. This procedure was used to minimize the effects of baseline curvature. An essentially constant value for heat of reaction is shown in Table I for the unfilled and activated aluminum oxide filled stoichiometric samples heated at 5°C/min. However, the same sample series at 2.5°C/min shows a decreasing heat of reaction with increasing filler level. Also, the aluminum and fused alumina filled samples show a higher heat of reaction than that of the other samples at 5°C/min. These discrepancies were attributed to over 200°C wide baselines used to define the exotherm. Such variation in the total heat of reaction precluded the use of dynamic DSC for the determination of the extent of reaction.

Although not useful for kinetic studies, the dynamic DSC clearly indicated that activated aluminum oxide had a significant effect on the course of the cure process for the 826/3,3'-DDS system as evidenced by changes in both the cure exotherm profile and the T_g of the resulting cured product. To obtain data more amenable to kinetic analysis as well as to avoid possible degradation at high temperatures, isothermal DSC experiments were undertaken.

The isothermal DSC studies were carried out between 120 to 210°C for unfilled stoichiometric mixtures of 826/3,3'-DDS and 826/4,4'-DDS and between 100 to 160°C for the same mixtures filled with 50 phr activated aluminum oxide. A much faster initial rate of reaction was observed for the filled samples at all comparable cure temperatures. Garton et al.[6] similarly observed an acceleration of the epoxy-primary amine reaction in the initial stages of cure when an air oxidized crushed carbon fiber was used as a filler.

The average heat of reaction values from the isothermal DSC experiments are listed in Table II. These values were observed to be

TABLE II
Isothermal heat of reaction and glass transition temperatures

Sample	J/g-resin	T_g(°C)
826/3,3'-DDS	476	155
826/3,3'-DDS/50 phr activated Al_2O_3	478	149
826/4,4'-DDS	478	187
826/4,4'-DDS/50 phr activated Al_2O_3	483	159

essentially the same at approximately 480 J/g for all the mixtures investigated, independent of the presence of filler or type of curing agent. Table II also shows the T_g values obtained for these resin systems. The T_g of 187°C for the 826/4,4'-DDS unfilled resin is in good agreement with the 189°C and 190°C reported by others[18,19]. Inclusion of the filler lowered the T_g of the 3,3'-DDS resin by approximately 5°C, but had a greater effect on the 4,4'-DDS system, lowering the T_g by almost 30°C. Although the final extent of reaction appears to be the same for the filled and unfilled systems, as evidenced by the same total heat of reaction, the lower T_g's indicate a lower crosslink density for the filled systems. This is consistent with the lower T_g observed by Garton et al.[6] for the effect of an acidic carbon black surface and the lower density material observed by Crompton[7] next to an aluminum adherend surface. Garton attributed the lower T_g to a reduced final extent of reaction due to the absorption of amine at the surface.

Since the total heat of a reaction appears to be constant, it is assumed that the extent of reaction is proportional to the partial heat of reaction and the rate of reaction is proportional to the rate of heat evolution. Plots of the rate of reaction vs. extent of reaction for the isothermal cure of a filled and an unfilled 826/3,3'-DDS and 826/4,4'-DDS samples at 130°C are shown in Figure 1. Both resin mixtures exhibit an accelerated initial rate of reaction for the filled samples with the 3,3'-DDS somewhat faster than the 4,4'-DDS. The unfilled samples show the typical autoaccelerated behavior associated with uncatalyzed epoxy/amine mixtures with a maximum rate of reaction at some extent of reaction greater than zero.

Activation energies were derived from plots of the natural logarithm of time to a particular extent of reaction (inversely proportional to the rate) vs. the reciprocal of the isothermal cure temperature. These Arrhenius plots were made for unfilled and filled resin mixtures at 10%, 30%, and 50% extent of reaction. Values for the activation energies derived from these curves are listed in Table III. Accurate activation energy determination for the 10% curves of the filled systems was found to be difficult due to the very high initial rate of reaction. Otherwise, the resins generally show a gradually decreasing activation energy with extent of reaction. The values for the 4,4'-DDS mixtures are also slightly higher when compared to the corresponding 3,3'-DDS mixture. These values correspond well with the 63 kJ/mole reported by Galy et al.[19] from DSC studies of a DGEBA/4,4'-DDS mixture.

Overall, only relatively small differences were observed in the

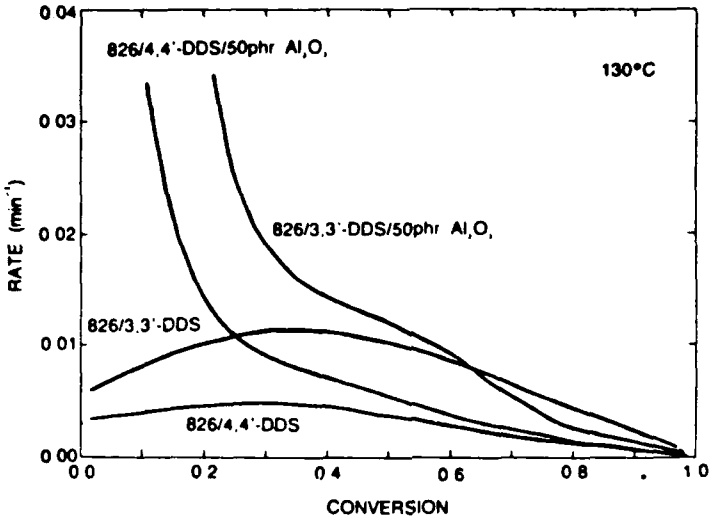


Figure 1. · Isothermal rate vs. extent of reaction traces from DSC at 130°C for unfilled and 50 phr activated aluminum oxide filled stoichiometric resin mixtures.

TABLE III
Activation Energies (kJ/mole) from Isothermal DSC and TBA

Sample	10%	30%	50%	Gelation
826/3,3'-DDS	60	56	53	58
826/3,3'-DDS/50 phr activated Al_2O_3	57	65	60	—
826/4,4'-DDS	64	62	60	61
826/4,4'-DDS/50 phr activated Al_2O_3	54	73	71	—

activation energies when comparing similar extents of reaction for filled and unfilled systems. Although a much faster initial reaction rate was observed for the activated aluminum oxide containing samples, its temperature dependence was similar to the unfilled resin. Riccardi et al.[20] attributed activation energies of 54 to 59 kJ/mole for DGEBA resins cured with ethylene diamine to the catalytic effect of the hydroxyl groups generated in the reaction. The increase in reaction rate observed in this study may thus be similarly due to the presence of catalytic sites, possibly hydroxylic, on the aluminum oxide surface. The change in T_g from unfilled to filled systems implies that the relative rates of the various network forming reactions are affected differently by the aluminum oxide surface.

Since DSC techniques provide no information as to the products of the various reactions, an alternate approach utilizing monofunctional model compounds was employed for this purpose. Stoichiometric mixtures of MGECP

with ADS were investigated using HPLC. The HPLC traces of a typical reaction mixture of MGECP/ADS are shown in Figures 2a, 2b, and 2c after 0, 100, and 400 min at 160°C, respectively. As in previous work on an

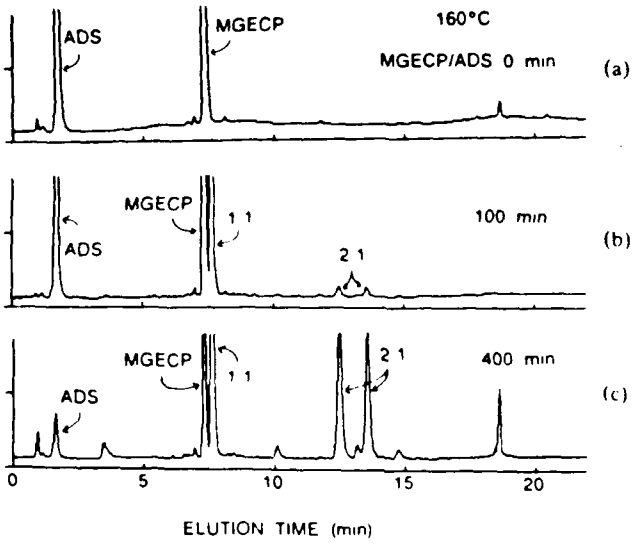


Figure 2. HPLC traces for an unfilled stoichiometric MGECP/ADS mixture heated at 160°C for (a) 0 minutes, (b) 100 minutes, and (c) 400 minutes.

unfilled monofunctional epoxy with ADS[9], these clearly show the formation of the primary amine addition product (1:1) followed by the formation of the diasteriomeric secondary amine addition products (2:1). While other minor peaks are observed in Figure 2c, the dominant reactions are amine addition to epoxy in the unfilled stoichiometric mixture.

Figures 3a, 3b, 3c, 3d, and 3e show the HPLC traces for an unfilled stoichiometric mixture and the same mixture filled with 50 phr corundum, 50 phr A/A, 50 phr A/N, and 50 phr A/B aluminum oxides, respectively, heated for 25 min at 160°C. The addition of corundum to the resin mixture appears to have no effect on the product distribution. A significant difference is observed when the activated aluminum oxides are added. More 1:1 product is produced in the same amount of time and an additional major reaction product is formed with an elution time of approximately 10 min. The HPLC traces of an unfilled MGECP/ADS stoichiometric sample heated for 200 minutes at 180°C and a 50 phr A/N aluminum oxide filled sample heated for 40 min at 180°C are shown in Figure 4a and 4b, respectively. The same major reaction product with an elution time of 10 min is seen in Figure 4b, as observed in Figures 3c, 3d, and 3e, but no MGECP and a significant

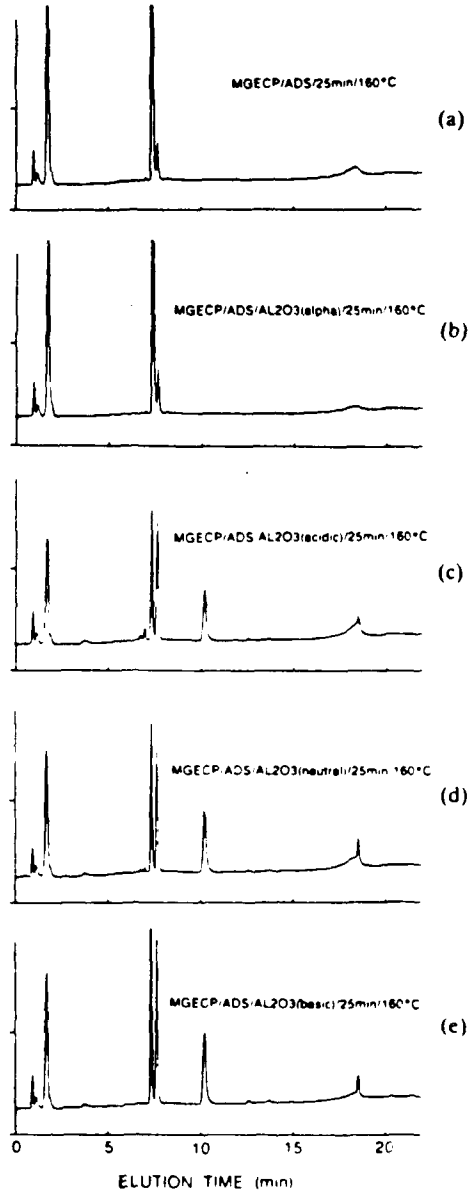


Figure 3. HPLC traces for a stoichiometric MGECP/ADS (a) unfilled, (b) 50 phr corundum filled, (c) 50 phr A/A filled, (d) 50 phr A/N filled, and (e) 50 phr A/B filled mixtures heated for 25 minutes at 160°C.

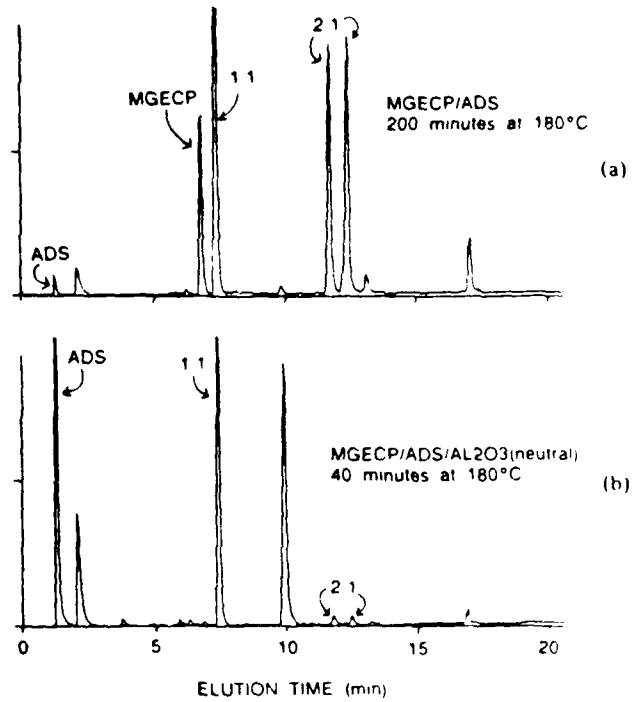


Figure 4. HPLC traces for a stoichiometric MGECP/ADS mixture (a) unfilled and heated for 200 minutes at 180°C and (b) 50 phr A/N filled heated for 40 minutes at 180°C.

amount of ADS remain in the reactant mixture. At this point it was speculated that this new product is the result of a reaction in which the amine does not participate. Additional heating of this sample produced no further change in the product distribution since no epoxy remains. This then represents the fully "cured" sample for these conditions. The activated aluminas thus lead to an acceleration of the epoxy/amine reactions and an entirely different reaction route, perhaps involving only the epoxy compound, leaving an excess of unreacted amine in a filled stoichiometric mixture.

Unfilled MGECP and MGECP filled with 50 phr corundum, 50 phr A/A, 50 phr A/N, and 50 phr A/B aluminum oxides were heated for 50 min at 160°C and the HPLC traces shown in Figures 5a, 5b, 5c, 5d, and 5e, respectively. No reaction takes place in the unfilled and corundum filled MGECP as can be seen by examination of Figures 5a and 5b. Reaction does take place in the activated aluminum oxide filled samples with a reaction product having the same 10 min elution time as that observed for the activated aluminum oxide filled MGECP/ADS mixtures. Amine is clearly not required for the formation of this product.

A surprising result was the apparent lack of effect of the surface pH

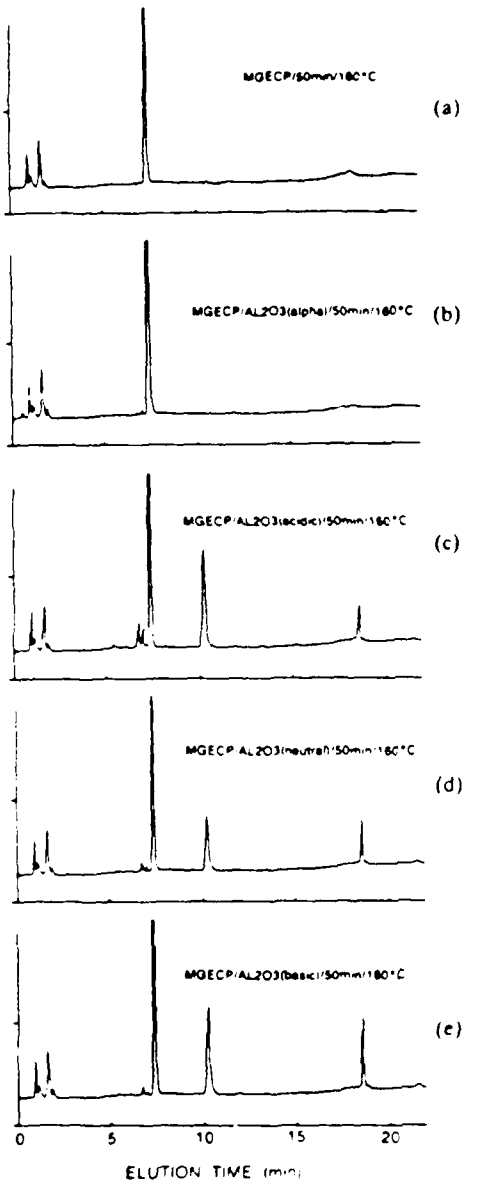


Figure 5. HPLC traces for (a) unfilled, (b) 50 phr corundum filled, (c) 50 phr A/A filled, (d) 50 phr A/N filled, and (e) 50 phr A/B filled MGECP heated for 50 minutes at 160°C.

of the activated aluminum oxides. The HPLC traces for the systems containing the various activated aluminas in Figures 3 and 5 show no dependence of the product distributions on the surface pH. By contrast, Garton et al.[6] observed a strong dependence of the accelerated reactions on the pH of the carbon blacks in their study.

Direct insertion probe mass spectrometry was carried out on the soluble material of the A/N filled MGECP reaction products. The parent ion of the HPLC 10 min elution product was shown to be a dimer of MGECP plus water. The activated aluminas thus appear to accelerate the dimerization and perhaps homopolymerization of the epoxy component.

To test whether this same reaction occurs with a difunctional resin, a sample of 826 was heated both without and in the presence of 50 phr A/N aluminum oxide for 960 min at 160°C. These HPLC traces are shown in Figures 6a and 6b, respectively. Essentially no soluble reaction products

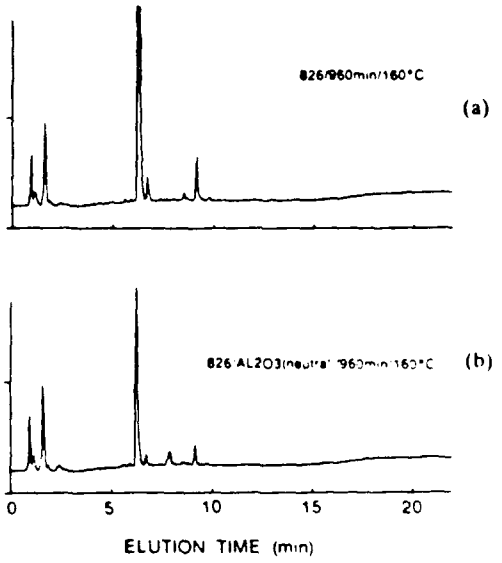


Figure 6. HPLC traces for (a) unfilled and (b) 50 phr A/N filled 826 heated for 960 minutes at 160°C.

were observed, the longer elution time peaks representing mostly the original oligomers in the 826. The homopolymerization process thus did not seem to extend to the difunctional resin. However, a careful mass balance revealed a significant amount of insoluble material remaining on the aluminum oxide.

The solids were recovered from mixtures of MGECP/ADS heated for 25 min at 160°C, MGECP heated for 960 min at 160°C, and 826 heated for 960 min at 160°C all in the presence of 50 phr A/N aluminum oxide. Each was extracted

with methylene chloride and vacuum dried at approximately 60°C, after which it was weighed and TGA performed in air to establish the organic content. These results are shown in Table IV together with a value derived

TABLE IV
Weight uptake by activated A/N alumina recovered from cured resin systems

RESIN SYSTEM	wt% by mass	wt% by TGA	wt% by elemental
MGECP/ADS (after 25 min at 160°C)	9	8	10
MGECP (after 960 min at 160°C)	10	11	—
826 (after 960 min at 160°C)	26	22	—

from the carbon content by elemental analysis. Good agreement between direct mass balance and TGA are shown with more than twice as much insoluble material observed for the difunctional 826 vs. MGECP heated under the same conditions. Accelerated homopolymerization thus seems to be taking place with the 826 since this would rapidly lead to insoluble product formation which would remain on the aluminum oxide surface.

The A/N aluminum oxide from the MGECP/ADS mixture was further studied by FTR. Its spectrum, shown in Figure 7 along with those for MGECP and

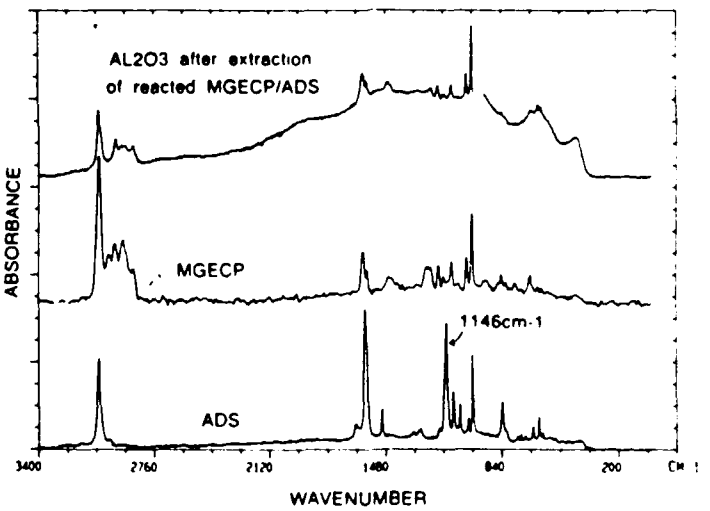


Figure 7. FTR spectra of the solids remaining after extraction of a stoichiometric MGECP/ADS mixture heated with 50 phr A/N for 25 minutes at 160°C, of MGECP, and of ADS.

ADS, is significant for the presence of the absorption bands associated only with the MGECP moiety and the virtually complete absence of the strong absorption band of the sulfone group of the ADS molecule at 1146 cm^{-1}. Elemental analysis of this same material indicated less than 5% of the expected sulfur if a stoichiometric product had remained on the surface. Pyrolysis mass spectrometry similarly showed no sulfur or nitrogen containing species to be present. These experiments demonstrate conclusively that the organic material remaining on the A/N aluminum oxide surface, even in the presence of the amine curing agent, is derived only from the epoxy component.

CONCLUSIONS

DSC experiments on crosslinking epoxy systems show a much faster initial rate of reaction in the presence of an activated aluminum oxide which leads to a lower T_g material. Since total heat of reaction (and thus extent of reaction) remained a constant, a change in the relative rates of the reactions leading to network formation was implied. Preliminary HPLC results using a monofunctional amine and epoxy show faster amine addition to epoxy and a significantly increased rate of epoxy/epoxy reaction(s) in the presence of activated aluminum oxide. This homopolymerization leaves an excess of unreacted amine after complete epoxy conversion in a stoichiometric mixture. Water and/or hydroxyl groups are apparently important to this reaction pathway as evidenced by the inactivity of the crystalline corundum. A surprising result was the lack of effect of the surface pH of the activated aluminum oxides. Organic material was also observed to be bound to the activated aluminum oxide surfaces and consisted of epoxy derived material with little evidence of any incorporated amine.

Whereas previous studies have shown the bulk network structure of epoxy resins cured with aromatic amines to consist primarily of amine/epoxy addition products, this study has revealed that reactions involving only the epoxy become significant in the interphase. Homopolymerization and strong epoxy/surface interactions clearly result in an interphase whose structure differs markedly from that expected from the stoichiometric addition of amine to epoxy. While we are as yet unable to specify the precise structure of this material or the spacial product distribution, it can be readily imagined that perturbations of this sort in an actual bond would effect both strength and durability and must be reflected in the development of any predictive model. Studies are currently underway directed at each of these issues. Particular emphasis is being placed on experiments with the oxides encountered on aluminum prepared for bonding.

ACKNOWLEDGEMENTS

The author wishes to thank Kelly Craven, Stanley Wentworth, David Bulpett, David Dunn, Domenic Macaione, and Michael Sennett for their assistance and helpful discussions.

REFERENCES

1. S. E. Wentworth, in "Adhesion 13", K. W. Allen, ed., pp. 1-7, Elsevier Applied Science, London (1989).
2. L. H. Sharpe, J. Adhesion, 4, 51(1972).
3. National Materials Advisory Board, Reliability of Adhesive Bonds under Sever Environments, NMAB-422, pp. 31-34(1984).
4. J. Comyn, C. C. Horley, D. P. Oxley, R. G. Pritchard, and J. L. Tegg, J. Adhesion, 12, 171(1981).
5. R. G. Dillingham and F. J. Boerio, J. Adhesion, 24, 315(1987).
6. A. Garton, W. T. K. Stevenson, and S. P. Wang, J. Polym. Sci., Chem. Ed., 26, 1377(1988).
7. J. S. Crompton, J. Mater. Sci., 24, 1575(1989).
8. J. Nigro and H. Ishida, J. Appl. Polym. Sci., 38, 2191(1989).
9. W. X. Zukas, D. A. Dunn, M. D. Gilbert, Proc. ACS Div. Polym. Mat. Sci. Eng., 56, 346(1987).
10. W. X. Zukas, K. J. Craven, and S. E. Wentworth, J. Adhesion, 33, 89 (1990).
11. L. Shechter and J. Wynstra, Ind. Eng. Chem., 48, 86 (1956).
12. L. Shechter, J. Wynstra, and R.P. Kurkjy, Ind. Eng. Chem., 48, 94 (1956).
13. K. Dusek, M. Bleha, and S. Lunak, J. Polym. Sci., Polym. Chem. Ed., 15, 2393(1977).
14. H.C. Anderson, SPE J., 16, 1241 (1960).
15. J.P. Bell, J. Polym. Sci. A-2, 6, 417 (1970).
16. C.A. Byrne, G.L. Hagnauer, N.S. Schneider, and R.W. Lenz, Polym. Comp., 1, 71 (1980).
17. G. L. Hagnauer, P. J. Pearce, B. R. LaLiberte, and M. E. Roylance, ACS Symp. Ser. 227, 25(1983).
18. J. M. Barton, Polymer, 21, 603(1980).
19. J. Galy, A. Sabra, and J.-P. Pascault, Polym. Eng. Sci., 26, 1514 (1986).
20. C. C. Riccardi, H. E. Adabbo, and R. J. J. Williams, J. Appl. Polym. Sci., 29, 2481(1984).