DTIC FILE COPY

AD-A203 102

Army Science Conference Proceedings

25-27 October 1988

Volume II

Principal Authors K through R

DTIC ELECTE NOV 15 1988 S CDE

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

88 11 14 018

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

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

SARD-TR

28 October 1988

MEMORANDUM FOR SEE DISTRIBUTION

SUBJECT: Proceedings of the 1988 Army Science Conference

The 16th Army Science Conference was held at Fort Monroe, Virginia, 25-27 October 1988. 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.

cont'd
Pg. VIII

These Proceedings of the 1988 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.

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.

J. R. Sculley
Assistant Secretary of the Army
(Research, Development and Acquisition)

DTIC
ELECTE
NOV 15 1988
D

DISTRIBUTION:
Office of the Under Secretary of Defense for Research and Advanced Technology, Wash, DC 20301-3080
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-ZS, Wash, DC 20310
HQDA, Deputy Chief of Staff for Logistics, ATTN: DALO-ZD, Wash, DC 20310-0400
HQDA, Deputy Chief of Staff for Operations & Plans, ATTN: DAMO-ZD, Wash, DC 20310-0400
HQDA, Director of Information Systems for C4, Wash, DC 20310-0400
HQDA, Deputy Chief of Staff for Personnel, ATTN: DAPE-MR, Wash, DC 20310
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-RDZ, Wash, DC 20310
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
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 63166
USARTA, Ames Research Center, ATTN: SAVRT-D, Moffett Field, CA 94035
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: SAVRT-DD, Ft. Monmouth, NJ 07703
USA Communications-Electronics Cmd, ATTN: AMSEL-TDD, Ft. Monmouth, NJ 07703
Center for Command, Control & Communications Systems, Ft. Monmouth, NJ 07703-5202

Center for Night Vision & Electro-Optics, ATTN: AMSEL-NV-D, Ft. Belvoir, VA 22060-5677
Electronics Warfare RSTA Center, ATTN: AMSEL-EW-D, Ft. Monmouth, NJ 07703-5303
Signals Warfare Lab, ATTN: AMSEL-SW-D, Vint Hill Farms Station, Warrenton, VA 22186-5100
USA Laboratory Command, ATTN: AMSLC-TD, 2800 Powder Mill Road, Adelphi, MD 20783-1145
Atmospheric Sciences Lab, ATTN: SLCAS-DD, WSMR, NM 88002-5501
Ballistic Research Lab, ATTN: SLCBR-OD, APG, MD 21005-5006
Electronics Technology & Devices Lab, ATTN: SLCET-D, Ft. Monmouth, NJ 07703
Electronic Warfare Vulnerability Assessment Lab, ATTN: SLCEW-M-D, WSMR, NM 88002-5513
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-CR, Warren, MI 48397-5000
USA Test & Evaluation Cmd, ATTN: AMSTE-TD, APG, MD 21005-5055
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 Actv, ATTN: STECS-CO, APG, MD 21005-5059
White Sands Missile Range, ATTN: STEWS-CG, WSMR, NM 88002-5000
USA Troop Support Cmd, ATTN: AMSTR-G, 4300 Goodfellow Blvd., St. Louis, MO 63120-1798
Belvoir R&D Center, ATTN: STRBE-ZT, Ft. Belvoir, VA 22060-5606
Natick R&D Center, ATTN: STRNC-T, Natick, MA 01760-5000
US Army Corps of Engineers
Cold Regions Rsch & Eng Lab, ATTN: CRREL-TD, P. O. Box 282, Hanover, NH 03755
Construction Eng Rsch Lab, ATTN: CERL-ZT, P. O. Box 4005, Champaign, IL 61820-1305
Engineer Topographic Labs, ATTN: ETL-TD, Ft. Belvoir, VA 22060
Waterways Experiment Station, ATTN: WESTV, P.O. Box 631, Vicksburg, MS 39180
USA Medical R&D Command, ATTN: SGRD-ZA, Ft. Detrick, Frederick, MD 21701-5012
Aeromedical Rsch Lab, ATTN: SGRD-UAC, Ft. Rucker, AL 36362-5000
Inst of Dental Rsch, WRAMC, ATTN: SGRD-UDZ, Wash, DC 20307-5300
Inst of Surgical Rsch, ATTN: SGRD-USZ, Ft. Sam, 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 21010-5405
Medical Rsch Inst of Environmental Medicine, ATTN: SGRD-UEZ, Natick, MA 01760-5007
Medical Rsch Inst of Infectious Diseases, ATTN: SGRD-UIZ-A, 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 78234-6100
USA Environmental Hygiene Agency, ATTN: HSHB-Z, APG, MD 21010-5422
USA Research Institute for the Behavioral and Social Sciences, ATTN: PERI-BR, 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, Fort Monroe, VA 23651-5000
Combined Arms 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 36205-5020
USA Combined Arms Cen, ATTN: ATZL-SC, Ft. Leavenworth, KS 66027
USA Engineer Cen, ATTN: ATZA-CG, Ft. Belvoir, VA 22060-5331
USA Infantry Cen, ATTN: ATZB-CG, Ft. Benning, GA 31905-5000
USA Intelligence Cen, ATTN: ATSI-SA, Ft. Huachuca, AZ 85613-7000
USA Ordnance Cen, ATTN: ATSL-CMT, APG, MD 21005-5201
USA Signal Center, ATTN: ATZH-CG, Ft. Gordon, GA 30905-5000
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 Agen, Cameron Station, Alexandria, VA 22304-6183
Defense Technical Information Center, Cameron Station, Alexandria, VA 22304-6145
Lawrence Livermore National Lab, ATTN: L-191, P. O. Box 808, Livermore, CA 94550
Los Alamos National Lab, ATTN: Dir for Energy, Research, and Technology, Los Alamos, NM 87545
NASA HQS, Suite 4237, 400 Maryland Avenue, 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

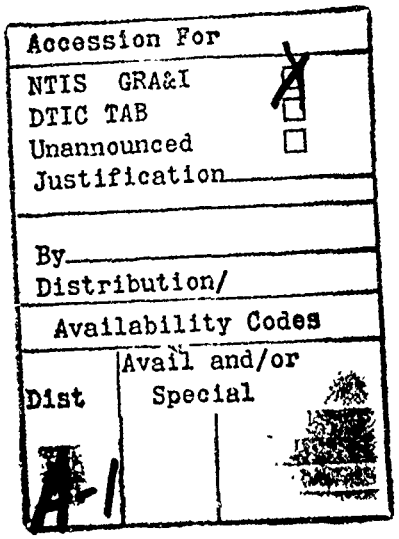
Naval Research Lab, ATTN: Director of Research, Wash, DC 20375
Office of Naval Research, Code 10, 800 North Quincy Street, Arlington, VA 22217-5000
HQ US Marine Corps, Code RD-1, Washington, DC 29380-0001

US Air Force

Air Force Systems Command, ATTN: Technical Director, Andrews AFB, Wash, DC 20334-5000
Air Force Office of Scientific Research, ATTN: Tech Director, Bolling AFB, Wash, DC 20332-6448

Commander

U. S. Army Field Artillery Center, ATTN: ATZR-C, Ft. Sill, OK 73503



PROCEEDINGS

OF THE

1988 ARMY SCIENCE CONFERENCE

FORT MONROE, VIRGINIA

25-27 OCTOBER 1988

VOLUME II

Principal Authors K through R

TABLE OF CONTENTS

PROCEEDINGS OF THE 1988 ARMY SCIENCE CONFERENCE

Author	Title	Vol	Page
Adams, George F.	The A Priori Synthesis of Infrared Spectra for Chemical Agents	I	1
Albert, Donald G.	Experimental and Theoretical Studies of Acoustic-to-Seismic Coupling	I	19
Albizo, Johnnie M.	Hydrolysis of GD and VX by 0.05M/0.10M Copper (II) -N,N,N',N'-Tetramethylethyl-enediamine (TMEN)	I	33
Alexander, C.	See Vezzoli, G.C.	III	393
Alster, Jack	See Iyer, Sury	IV	43
Amos, Roger D.	See Adams, George F.	I	1
Amstutz, Larry I.	See White, Ronald B.	III	471
Anderson, William	Laser Photofragmentation and Fluorescence Studies of CW Agent Simulants	I	39
Andrade, James	See Balkin, Thomas J.	I	69
Andreotti, Raymond E.	See Robbins, Frederick M.	IV	83
Arcidiacono, Steve	See Kaplan, David L.	II	1
Asano, Christopher	Multigrid Time Integration Methods for Parabolic Equations	I	53
Babbitt, Richard W.	Microstrip Devices for Millimeter Wave Electronic Scan Antennas	III	233
Babbitt, Richard W.	Millimeter Wave Microstrip "Drop-In" Circulator	III	241
Bachman, William G.	See Leibrecht, Bruce C.	II	127
Bailey, Charles L.	See Linthicum, Kenneth J.	II	155
Balkin, Thomas J.	Efficacy and Safety of Tri-azolam for Deployment Operations	I	69

Author	Title	Vol	Page
Ballato, Arthur	Acoustic Waves Generated in Insulating and Semiconducting Crystal Wafers by Tangential Electric Fields	I	83
Banderet, Louis	See Rose, Madeleine	II	555
Barnette, B. Diane	See Birkmire, Deborah P.	I	139
Barrett, Ann H.	A Descriptive Model of the Vacuum Infusion Process	I	99
Beardell, Anthony J.	Unicharge Development for Advanced Howitzers	I	113
Beaudry, William T.	See Yang, Yu-Chu	III	519
Belenky, Gregory	See Balkin, Thomas J.	I	69
Bhattacharjee, Aperba	See Sadoff, Jerald C.	III	1
Binder, Michael	Novel High Energy/Power Cells Based on Condensed SO_2 Electrolyte	I	127
Birkmire, Deborah P.	Perceptibility of Military Vehicle Silhouettes	I	139
Bonometti, R.	See Vezzoli, G.C.	III	393
Boyd, Victoria	See Smejkal, Ruthann M.	III	159
Brecht-Clark, Jan	See Headley, Donald B.	I	435
Breuer, Eli	See Smejkal, Ruthann M.	III	159
Brierly, Joseph E.	Optimizing System Availability to Spare Part Budget	I	153
Brooks, Marguerite E.	See Krishnamurthy, Thaiya	II	63
Bruno, John D.	See Tobin, Mary S.	III	367
Bulpett, David A.	Chemical Component Monitoring of High Performance Elastomers	I	165
Bundy, Mark L.	A Thermal Shroud Design for Reducing Firing Induced Barrel Distortion	I	177

Author	Title	Vol	Page
Bunte, Steven W.	Structure, Hydrogen Bonding, and Ion-Pair Formation in Aqueous Liquid Monopropellants as a Function of Concentration, Temperature and Pressure	I	193
Burke, Donald S.	See Hoke, Charles H.	I	487
Burke, Donald S.	Human Immunodeficiency Virus: Progression Correlates Directly with Virus Replication in the Infected Host	I	205
Busey, Brent R.	Detection Applications of a Fiber Optic Waveguide Sensor	IV	1
Canik, James J.	See Kark, John A.	II	15
Cardello, Armand	See Rose, Madeleine	II	555
Carignan, Y.P.	See Vladimiroff, T.	III	407
Cartland, Harry E.	See Johnson, Thomas H.	I	535
Celmins, Ilmars	See Danberg, James E.	I	241
Chiang, Peter K.	See Smejkal, Ruthann M.	III	159
Chionis, Michael	See Vogel, Raymond	III	419
Choi, Chang S.	See Prask, Henry J.	II	481
Christensen, Scott	See Marini, Mario A.	II	247
Christesen, Steven D.	See Long, S. Randolph	II	169
Chubb, Douglas W.J.	Advanced Spatial Problem Solving Methodology Using the Functional Binary Decomposition Spatial Representation	I	211
Clark, William W.	Photorefractive Materials as Optical Limiters	IV	13
Coltharp, David R.	See Jones, Patricia S.	IV	55
Commerford, T.R.	See Fitzgerald, B.E.	I	305
Conrad, Raymond W.	See Russell, Stephen D.	II	583

Author	Title	Vol	Page
Cook, Charles F.	See Kwiatkowski, Joseph H.	II	101
Cora, Elsa M.	See Friedlander, Arthur M.	I	321
Croisant, William J.	A Correlation for Characterizing Nonlinear Electric Field Transients at the Interior Surface of Thin-Walled Ferromagnetic Cylinders	I	225
Cymerman, Allen	See Hamilton, Allan J.	I	377
Danberg, James E.	Prediction and Measurement of the Aerodynamic Characteristics of a Family of Projectile Configurations	I	241
Davies, F.G.	See Linthicum, Kenneth J.	II	155
Davies, Mark A.	See Bunte, Steven W.	I	193
Deaso, Robert E.	See Sawyer, Geoffrey S.	IV	97
Desper, C. Richard	Computer Molecular Modelling as a Polymer Materials Development Tool	I	255
Dickson, Peter W.	See Miller, Robert F.	II	313
Duncan, Fred	See Keniston, Richard C.	II	47
Dutta, Mitra	See Heath, Linda S.	I	463
Dutta, Piyush K.	Performance of Laminated Composites in Cold	I	269
Dwyer, Harry A.	See Sturek, Walter B.	III	291
Egglestone, Graeme T.	See Song, John W.	III	191
Ehrlich, John J.	See Russell, Stephen D.	II	583
Elmer, Frank J.	Development of Voltage Controlled Ceramic Phase Shifter	I	283
Enriquez, John	See Keniston, Richard C.	II	47
Evans, George	See Robbins, Frederick M.	IV	83
Farr, Beatrice J.	See Simutis, Zita M.	III	103

Author	Title	Vol	Page
Farrell, Dennis	See Dutta, Piyush K.	I	269
Feggins, Kristal J.	Intelligent Load Management	I	289
Feickert, Carl A.	See Croisant, William J.	I	225
Fifer, Robert A.	See Bunte, Steven W.	I	193
Filler, Raymond L.	See Rosati, Vincent J.	II	541
Fisher, Jamie M.	See Siron, Robert E.	III	117
Fitzgerald, B.E.	An Objective Computerized Color Measurement System	I	305
Forch, B.E.	See Miziolek, Andrzej W.	II	341
Francesconi, Ralph	See Rose, Madeleine	II	555
Friedlander, Arthur M.	Molecular Interactions of Anthrax Toxin Components with Target Cell Receptors	I	321
Frykman, Peter	See Rose, Madeleine	II	555
Fukai, Konusuke	See Hoke, Charles H.	I	487
Fulton, Laura O.	Estimation of the Probability of Error in Human/ATR Classification Performance	I	331
Gallo, Benedict J.	See Walker, John E.	III	441
Gardner, Henry S.	See van der Schalie, William H.	III	379
Gelber, Nathaniel S.	See Iyer, Sury	IV	43
Genoni, Thomas C.	See Johnson, Thomas H.	I	535
Gilman, Sol	See Binder, Michael	I	127
Givens, J.H.	See Suarez, J.E.	III	303
Goode, Michael T.	See Busey, Brent R.	IV	1
Gordon, Richard K.	See Smejakl, Ruthann M.	III	159
Gowenlock, David	See Walker, John E.	III	441
Grabowsky, Aaron	See Beardell, Anthony J.	I	113

Author	Title	Vol	Page
Graham, Scott E.	Enhancing Reserve Component Maintenance Performance with Computer-Based Training	I	345
Gross, Clark L.	See Meier, Henry L.	II	287
Gupta, Neelam	A Foliage-Smart Optical Sensor	IV	29
Gupta, Raj K.	Effectiveness of U.S. Army Extended-duration Topical Insect/Arthropod Repellent and Permethrin-Treated Battle Dress Uniform in Tropical Environment	I	361
Haberkorn, W.E.	See Feggins, Kristal J.	I	289
Hamilton, Allan J.	Alterations in Central Nervous System Function in Soldiers at High Altitude	I	377
Handy, Nicholas C.	See Adams, George F.	I	1
Hannon, John P.	See Maningas, Peter A.	II	233
Harman, Joan	See Simutis, Zita M.	III	103
Harper, Bruce G.	See Winters, Douglas R.	III	503
Harris, Lee E.	The Low Pressure Combustion of Nitramine Propellant and the Ignition of Low Vulnerability Tank and Artillery Ammunition	I	393
Harrison, Ralph J.	Application of Knowledge Based Techniques to Materials Research	I	407
Hawley, Robert J.	See Heweston, John F.	I	475
Hayes, Pamela G.	Computer-Based Resistance Analysis of Fighting Positions for Mobile Weapon Systems	I	421
Headley, Donald B.	Continuous Operations of Artillery Crews in NBC and non-NBC Environments	I	435
Heaps, Melvin G.	Battlefield Atmospherics	I	449

Author	Title	Vol	Page
Heath, Linda S.	Characterization of Reactive Ion Etch Damage for Advanced GaAs Device Structures	I	463
Heath, Linda S.	See Kwiatkowski, Joseph H.	II	101
Heavey, Karen R.	See Nietubicz, Charles J.	II	369
Herrera, Dan	See White, Ronald B.	III	471
Hetherington, John G.	See Jones, Patricia S.	IV	55
Hewetson, John F.	Detection of T-2 Mycotoxin Metabolites in Urines of Exposed Rats. Comparison of a Potentially Fieldable Kit With a Laboratory Assay	I	475
Hock, V.F.	See Suarez, J.E.	III	303
Hoke, Charles H.	A Successful, Large Scale, Field Efficacy Study of Japanese Encephalitis Vaccine in Thailand	I	487
Hoover, Sharon R.	See Krusinger, Alan E.	II	79
Hoskin, C. Francis	See Walker, John E.	III	441
Hui, Philip	See Beardell, Anthony J.	I	113
Hulthage, Ingemar	See Harrison, Ralph J.	I	407
Iafrate, Gerald J.	See Kwiatkowski, Joseph H.	II	101
Innis, Bruce L.	See Hoke, Charles H.	I	487
Iyer, Sury	Research Toward Development of More Powerful Explosives	IV	43
Jackson, Gerald W.	Motion Base Simulator Control via Analytical Prediction Techniques	I	503
Jang, Sei Joo	See Elmer, Frank J.	I	283
Jasinski, Jerry P.	See Desper, C. Richard	I	255
Jatanasen, Sujarti	See Hoke, Charles H.	I	487

Partial contents of this volume include:

Author	Title	Vol	Page
Jenkinson, Howard A.	GaAs/AlGaAs Superlattice Structures Grown by Molecular Beam Epitaxy and Their Application to Infrared Guided-Wave Devices	I	519
Johnson, Barry W.	See Feggins, Kristal J.	I	289
Johnson, Thomas H.	Xenon Chloride Laser Kinetics	I	535
Jokl, A.L.	See White, Ronald B.	III	471
Jones, Jason F.	See Bulpett, David A.	I	165
Jones, Patricia S.	Masonry Structures - Will They Stand Up Against Terrorist Bombs?	IV	55
Juhasz, Arpad A.	See Oberle, William F.	II	401
Kahn, Narayan K.	See Burke, Donald S.	I	205
Kalafut, John	See Dutta, Piyusk K.	I	269
Kalb, Joel T.	See Price, G. Richard	II	495
Kamimori, Gary	See Balkin, Thomas J.	I	69
Kaplan, David L.	Microbial Polysaccharides for Military Material Applications	II	1
Kark, John A.	Sickle Cell Trait as a Risk Factor for Sudden Death in Basic Training	II	15
Karsh, Robert	See Birkmire, Deborah P.	I	139
Kearney, Frank W.	Intelligent Systems for Construction and Armament Production	II	29
Keller, Harry L.	See Sousk, Stephen F.	III	205
Keniston, Richard C.	The Sergeant Major Study: Health Risk Assessment By Clinical Laboratory Parameters	II	47
Kenyon, R.	See MacDonald, Carolyn	II	217

Contin. p. ix

Author	Title	Vol	Page
Kern, Richard P.	See Simutis, Zita M.	III	103
Kotchansenee, Sa-Ong	See Hoke, Charles H.	I	487
Kotlar, Anthony J.	See Anderson, William R.	I	39
Krishnamurthy, Thaiya	Identification of the Active-Site in a Neurotoxic Snake Venom by Affinity Labelling and State-of-the Art Tandem Mass Spectrometry Technology	II	63
Kroutil, Robert T.	See Adams, George F.	I	1
Krusinger, Alan E.	An Empirical Surface Temperature Model with Automatic Type-Day Classification	II	79
Kuhl, Frank P.	Tracking Maneuvering Helicopters Using Attitude Measurements	II	91
Kwiatkowski, Bonnie	See Heath, Linda S.	I	463
Kwiatkowski, Joseph H.	Nanometer Device Fabrication Techniques for Advanced Military Microelectronics	II	101
Lanzerotti, Yvonne D.	Fractal Properties of Fracture Surfaces of Energetic Materials	II	113
Laorakapongse, Thanom	See Hoke, Charles H.	I	487
Larson, Donald C.	See Jenkinson, Howard A.	I	519
LeClaire, Ross D.	See Templeton, Charles B.	III	329
Leet, D.M.	See Suarez, J.E.	III	303
Leibrecht, Bruce C.	The Eyes Have It: Contact Lens Impact on Performance of Armor Troops	II	127
Lenef, Alan	See Vogel, Raymond	III	419
Leppla, Stephen H.	See Friedlander, Arthur M.	I	321
Leupold, Herbert A.	A Helical Free-Electron Laser Structure for Hyper-power Devices	II	141
Levine, Leslie	See Rose, Madeleine	II	555

Author	Title	Vol	Page
Linthicum, Kenneth J.	Use of Remote Sensing to Predict the Threat of Vector-Borne Diseases to Military Deployment	II	155
Locke, Mark C.	See Sousk, Stephen F.	III	205
Long, S. Randolph	Ultraviolet Laser-Based Detection and Identification of Chemical Agent Simulants	II	169
Long, S. Randolph	See Anderson, William R.	I	39
Lottero, Richard E.	Computational Predictions of Close-In Blast Loading from Bare Spherical Charges	II	185
Lukas, Jeffrey H.	Evoked Potential Technique for Predicting Performance Under High Mental Workloads	II	203
Lukaszek, Theodore	See Ballato, Arthur	I	83
Lupton, H.	See MacDonald, Carolyn	II	217
MacDonald, Carolyn	Initial Evaluation in Humans of a Live-Attenuated Vaccine Against Argentine Hemorrhagic Fever	II	217
Macpherson, A.K.	See Vladimiroff, T.	III	407
Maiztegui, J.	See MacDonald, Carolyn	II	217
Maningas, Peter A.	Hypertonic Saline - Dextran Solutions for the Far Forward Resuscitation of the Combat Casualty	II	233
Margerum, Eugene A.	See Krusinger, Alan E.	II	79
Marini, Mario A.	Novel Derivatives of Human Hemoglobin Ao for Possible Use as Emergency Resuscitation Fluids	II	247
Martin, Samuel K.	See Kark, John A.	II	15
Matta, Joseph E.	An Experimental Investigation of the Aerodynamic Breakup Behavior of Heated Newtonian Liquids by Ambient Air	II	261

Author	Title	Vol	Page
Mayer, Jean	See Kaplan, David L.	II	1
McClennen, William H.	See Snyder, A. Peter	III	169
McGowan, Joeseph	The Exploitation of GPS to Satisfy Precise Platform Requirements	II	275
McKee, K.	See MacDonald, Carolyn	II	217
Medina, Frank	See Marini, Mario A.	II	247
Meegan, J.	See MacDonald, Carolyn	II	217
Meier, Henry L.	Niacinamide Prevents Sulfur Mustard (HD)-induced Pathology and Biochemical Changes in Human Lymphocyte and Mixed Human Leukocyte Preparations	II	287
Messier, Donald R.	Oxynitride Glasses and Glass Fibers	II	301
Meuzelaar, Henk L.C.	See Snyder, A. Peter	III	169
Miller, Mary Jo	See Clark, William W.	IV	13
Miller, Miles C.	Destabilizing Moments Produced by a Viscoelastic Fluid in a Spinning and Coning Cylinder	IV	69
Miller, Robert F.	Expert Systems in the Tactical Environment	II	313
Mirelli, Vincent	3-D Model Based Machine Vision	II	327
Miziolek, Andrzej W.	A Novel Laser Igniter Based on Resonant Multiphoton Excitation	II	341
Miziolek, Andrzej W.	See Anderson, William R.	I	39
Moerkirk, Robert P.	See Ross, Raymond L.	II	571
Monahan, Timothy P.	See Heath, Linda S.	I	463
Morrow, Scott	See Lanzerotti, Yvonne D.	II	113
Mosier, Daniel	See Vogel, Raymond	III	419

Author	Title	Vol	Page
Muller, August J.	The Competitive Edge in Organic Synthesis: Improved Procedures for the Preparation of Important Compounds Containing Fluorine	II	354
Mullins, Linda F.	See Lukas, Jeffrey H.	II	203
Munro, Ilse	See Rose, Madeleine	II	555
Nash, Carolyn J.	See Vollmerhausen, Richard H.	IV	111
Nietubicz, Charles J.	Supercomputer Applications in Projectile Aerodynamics	II	369
Niles, Franklin E.	See Heaps, Melvin G.	I	449
Nisalak, Ananda	See Hoke, Charles H.	I	487
Nixon, Mark W.	Improvements to Tilt Rotor Performance Through Passive Blade Twist Control	II	385
Nord, Roy D.	See Schmitz, Edward J.	III	51
Oaks, Edwin V.	See Stover, Charles K.	III	277
Oberle, William F.	A Realistic Evaluation of Traveling Charge	II	401
O'Donnell, Vincent M.	See Balkin, Thomas J.	I	69
Opalka, Klaus O.	Real-Time Flow Control in Large Blast/Thermal Simulators	II	417
Oro, J. Barrera	See MacDonald, Carolyn	II	217
Ozolins, Aivars	Stochastic High-Resolution Modeling Support for Live-Fire Test Programs	II	433
Paolella, Arthur C.	See Ross, Raymond L.	II	571
Park, Ok-Choon	See Pollack-Nelson, Carol	II	463
Parrish, D.	See MacDonald, Carolyn	II	217
Patterson, Stanley P.	See Russell, Stephen D.	II	583
Pearson, Richard J.	See Opalka, Klaus O.	II	417

Author	Title	Vol	Page
Pennise, Christine A.	See Tobin, Mary S.	III	367
Perez, Ray S.	See Pollack-Nelson, Carol	II	463
Peters, C.	See MacDonald, Carolyn	II	217
Petrali, John P.	See Meier, Henry L.	II	287
Pillalamarri, Krishna	See Birkmire, Deborah P.	I	139
Pinto, James J.	See Lanzerotti, Yvonne D.	II	113
Piscitelle, Louis J.	Mathematical Investigation of The Reaction-Diffusion Equations for Heterogeneous Catalytic Systems	II	449
Pollack-Nelson, Carol	Stages of Skill Acquisition and Training Device/Simulator Fidelity	II	463
Poole, Paula	See Rose, Madeleine	II	555
Posey, David M.	See Kark, John A.	II	15
Potenziani, Ernest	See Leupold, Herbert A.	II	141
Prask, Henry J.	Residual Stress Measurements in Armament-System Components by Means of Neutron Diffraction	II	481
Price, G. Richard	Weapons Design and the Inner Ear: Critical Insights From Mathematical and Physiological Models	II	495
Ramsley, A.O.	See Fitzgerald, B.E.	I	305
Redfield, Robert R.	See Burke, Donald S.	I	205
Redmond, Daniel P.	See Balkin, Thomas J.	I	69
Reinig, Karl D.	Rapid Backscatter Simulation of Complex Targets	II	511
Resnick, I. Gary	See Winters, Douglas R.	III	503
Rigsbee, J.M.	See Suarez, J.E.	III	303
Robbins, Frederick M.	A Catalytically Active Metal Ion-Polymer Complex	IV	83

cont'd pg. xiv

Author	Title	Vol	Page
Robbins, Frederick M.	See Walker, John E.	III	441
Robertson, Struan R.	Numerical Modeling of Contaminant Dispersion in Air by Bouyancy-Driven Flows Within Fabric Structures	II	525
Rodkey, William G.	See Maningas, Peter A.	II	233
Rosati, Vincent J.	Vibration-Immune Quartz Crystal Oscillators	II	541
Rose, Madeleine	Effects of a NBC Nutrient Solution on Physiological and Psychological Status During Sustained Activity in the Heat	II	555
Ross, Edward	See Barrett, Ann	I	99
Ross, Raymond A.	See Ross, Raymond L.	II	571
Ross, Raymond L.	Advanced Gunn Oscillators for Millimeter-wave Smart Munition Systems	II	571
Rossman, David I.	See Muller, August J.	II	354
Ruehle, Charles J.	See Kark, John A.	II	15
Russell, Stephen D.	Stimulated Brillouin Scattering Experiments With Flashlamp-Pumped Dye Lasers	II	583
Rutledge, Louis C.	See Gupta, Raj K.	I	361
Sadoff, Jerald C.	Immunologic Prevention and Treatment of Septic Shock	III	1
Sahu, Jubaraj	Numerical Simulation of Three Dimensional Transionic Flows	III	15
Sahu, Jubaraj	See Nietubicz, Charles J.	II	369
Salamo, Gregory J.	See Clark, William W.	IV	13
Sandus, Oscar	See Iyer, Sury	IV	43
Sangawhipa, Nathirat	See Hoke, Charles H.	I	487
Sarid, Dror	See Jenkinson, Howard A.	I	519

Author	Title	Vol	Page
Sarkis, Randall G.	See Jenkinson, Howard A.	I	519
Sausa, Rosario C.	See Miziolek, Andrzej W.	II	341
Sausa, Rosario C.	See Anderson, William R.	I	39
Sauter, David P.	Three-Dimensional Hazard Zone	III	29
Sawka, Michael	See Rose, Madeleine	II	555
Sawyer, Geoffrey S.	Algorithm Evaluation of the Multi-Function Target Acquisition Processor	IV	97
Schlosser, Kevin	See McGowan, Joeseph	II	275
Schmidt, Bernard	Application of LADAR/FLIR Multi-Sensor Imagery for Improved Aided Target Recognizer Performance	III	37
Schmitz, Edward J.	The Personnel Allocation Simulation System (PASS)	III	51
Schneider, David P.	Designing Tomorrow's Warheads Today	III	65
Schwering, Felix K.	A Simple Model of the Back-Scatter Enhancement for Random Rough Surfaces	III	79
Semasko, Deborah A.	See Ross, Raymond L.	II	571
Shaffer, Dolores	See Sawyer, Geoffrey S.	IV	97
Sharp, Edward J.	See Clark, William W.	IV	13
Shih, Tsung-Ming	Comparison of Cholinergic Effects of HI-6 and 2-PAM in Soman Poisoning	III	89
Shih-Thornton, Jane	See Beardell, Anthony J.	I	113
Shukitt, Barbara	See Rose, Madeleine	II	555
Sigal, Asher	See Danberg, James E.	I	241
Silber, Leo M.	See Wilber, William D.	III	487
Simon, Lisa	See Balkin, Thomas J.	I	69
Simpson, Thomas B.	See Tobin, Mary S.	III	367

Author	Title	Vol	Page
Simutis, Zita M.	Improving Soldiers' Basic Skills	III	103
Siron, Robert E.	Effect of Formulation on Composite Propellant Mechanics and Dilatational Behavior	III	117
Slagg, Norman	See Iyer, Sury	IV	43
Slavin, Michael	Processing, Microstructure and Property Relationships of Armor Ceramics	III	131
Slawson, Thomas R.	Large Deformation Behavior of a Reinforced Concrete Structure Subjected to Blast Overpressures	III	145
Smejkal, Ruthann M.	Design of Potent Antidotes for Chemical Warfare Agents Based on Physical and Mechanical Quantifiers	III	159
Smith, Doran D.	See Kwiatkowski, Joseph H.	II	101
Smith, Doran D.	See Heath, Linda S.	I	463
Snell, Shawn	See Marini, Mario A.	II	247
Snyder, A. Peter	Biomarker Detection by Curie-Point Pyrolysis in Combination with an Ion Trap Mass Spectrometer	III	169
Soicher, Haim	Traveling Ionospheric Disturbances (TID's) at Mid-latitudes: Solar Cycle Phase Dependence	III	177
Solow, Rikki	See Templeton, Charles B.	III	329
Song, John W.	Relationships Between Failure Mechanisms and Ballistic Impact Resistance of Composite Helmets	III	191
Sousk, Stephen F.	Science and Technology for Cargo Handling in the Unstructured Field Environment	III	205
Stanley, Ann E.	A Laser-Induced Synthesis of Cadmium Sulfide	III	219

Author	Title	Vol	Page
Starcher, Barry	See Walker, John E.	III	441
Steger, Joseph L.	See Sahu, Jubaraj	III	15
Stern, Richard A.	See Babbitt, Richard W.	III	241
Stern, Richard A.	See Babbitt, Richard W.	III	233
Stevens, M. Merrill	Application of Remote Sensing to Playa Surface Characteristics for Military Purposes	III	249
Stewart, Nora Kinzer	Model for Analyzing Cohesion and Effectiveness: The South Atlantic Conflict of 1982	III	265
Stover, Charles K.	Development of Improved Recombinant DNA Techniques in Support of the Effort for a Recombinant Scrub Typhus Vaccine	III	277
Sturek, Walter B.	Prediction of In-Bore and Aerodynamic Heating of KE Projectile Fins	III	291
Sturek, Walter B.	See Weinacht, Paul	III	455
Suarez, J.E.	RF Plasma Assisted Depostion of Conductive Oxide Coatings for Cathodic Protection Anodes	III	303
Sutherland, Robert A.	Methods of Radiative Transfer for Electro-Optical Obscuration Modeling	III	315
Szafraniec, Linda L.	See Yang, Yu-Chu	III	519
Tasesky, Jovan	See Marini, Mario A.	II	247
Taub, Irwin	See Barrett, Ann	I	99
Taysing-Lara, Monica	See Heath, Linda S.	I	463
Taysing-Lara, Monica	See Kwiatkowski, Joseph H.	II	101
Templeton, Charles B.	An Antibody to Prevent the Effects of Brevetoxin Poisoning in Conscious Rats	III	329
Teng, Nelson H.	See Sadoff, Jerald C.	III	1

Author	Title	Vol	Page
Tessler, Alexander	Higher-Order Bending Theory for Thick Composite Laminates with Particular Emphasis on Finite Element Analysis	III	337
Tischler, Mark B.	Advancements in Frequency-Domain Methods for Rotorcraft System Identification	III	351
Tobin, Mary S.	Optical Properties and Uses of Gallium Arsenide Doping Superlattices	III	367
Tompkins, Robert E.	See Oberle, William F.	II	401
Tracy, Carl A.	See Slavin, Michael J.	III	131
Trad, Laurie A.	See Hamilton, Allan J.	I	377
Tucker, C.J.	See Linthicum, Kenneth J.	II	155
Turner, Janice B.	See Stanley, Ann E.	III	219
Valdes, James J.	See Shih, Tsung-Ming	III	89
Vanderhoff, John A.	See Bunte, Steven W.	I	193
van der Schalie, W.	New Methods of Toxicity Assessment in Military Relevant Applications	III	379
Vezzoli, Gary C.	Excitonic and Spin Fluctuation Mechanisms for High-T_c $Y_1Ba_2Cu_3O_{7-}$ Superconducting Oxides: Experimental Data on Hall Effect, J_c and Magnetic Properties	III	393
Viechnicki, Dennis J.	See Slavin, Michael J.	III	131
Vladimiroff, T.	Dynamical Properties of A Collapsing Spherical Cavity	III	407
Vogel, Raymond	An Intelligent Spectroscopic System for Real-Time Evaluation of Atmospheric Conditions Affecting Hydrogen in Weldments	III	419
Vollmerhausen, Richard	Design Criteria for Night Pilotage Sensors	IV	111

Author	Title	Vol	Page
Vrabel, Michael J.	An Analytic Model for Aerodynamic Shapes	III	429
Wade, Charles E.	See Maningas, Peter A.	II	233
Walker, John E.	Soman Hydrolyzing and Detoxifying Properties of an Enzyme From a Thermophilic Bacterium	III	441
Walker, John E.	See Robbins, Frederick M.	IV	33
Wannemacher, Robert W.	See Hewetson, John F.	I	475
Ward, J. Richard	See Albizio, Johnnie M.	I	33
Ward, J. Richard	See Yang, Yu-Chu	III	519
Ward, Joseph S.	See Simutis, Zita M.	III	103
Warren, Philip	Rose, Madeleine	II	55
Watkins, Wendell R.	The Target Contrast Characterizer	IV	127
Weber, Daniel J.	See Miller, Miles C.	IV	69
Weber, Robert	See Vogel, Raymond	III	419
Weinacht, Paul	Computation of the Roll Characteristics of Finned Projectiles	III	455
Weir, Michael R.	See Keniston, Richard C.	II	47
Whalley, Christopher	See Shih, Tsung-Ming	III	89
White, Kevin J.	See Oberle, William F.	II	401
White, Ronald B.	Diesel Electric Drives for Tracked Vehicles	III	471
Whittenberg, John A.	See Headley, Donald B.	I	435
Wilber, William D.	High-Sensitivity Millimeter wave Measurements of Magnetic Susceptibility in Cubic and Hexagonal Ferrites	III	487
Wiley, Bonnie	See Kaplan, David L.	II	1
Wilkins, Clifford E.	See Kuhl, Frank P.	II	91
Williams, Robert A.	See Feggins, Kristal J.	I	289

Author	Title	Vol	Page
Williams, Ronald D.	See Feggins, Kristal J.	I	289
Williams, Timothy J.	See Sawyer, Geoffrey S.	IV	97
Winslow, Robert M.	See Marini, Mario A.	II	247
Winters, Douglas R.	An Improved Automated Procedure for the Estimation of Blood Blood Cholinesterase Activities	III	503
Wolfe, Allan	See Lanzerotti, Yvonne D.	II	113
Wood, Gary L.	See Clark, William W.	IV	13
Woodbury, Charles F.	See Robbins, Frederick M.	IV	83
Wortman, John D.	See Lottero, Richard E.	II	185
Wren, Gloria P.	See Oberle, William F.	II	401
Wright, D. Craig	See Burke, Donald S.	I	205
Yang, Yu-Chu	Use of Strong Nucleophiles to Decontaminate Mustard	III	519
Zavada, John M.	See Jenkinson, Howard A.	I	519
Zegna, Angelo	See Marini, Mario A.	II	247
Ziegler, Elizabeth J.	See Sadoff, Jerald C.	III	1
Zywiol, Harry J.	See Jackson, Gerald W.	I	503

Microbial Polysaccharides for Military Material Applications

David L. Kaplan, Dr., Jean Mayer, Mrs., Bonnie Wiley, Miss,
and Steve Arcidiacono, Mr.
U.S. Army Natick Research Development and Engineering Center
Natick, MA 01760-5020

Introduction

Biopolymers produced by fungi and bacteria offer many advantageous properties that make them attractive as new sources of materials. Some of these properties include: (1) the ability to form transparent films and fibers; (2) low oxygen permeable films for improved food storage; (3) biodegradable films; (4) controllable (through fermentation and processing) polymer molecular weight distribution and therefore physical properties of films and fibers; (5) edible polymers with potential for nutritional sustenance; (6) potential for genetic engineering for improved control over polymer molecular weight distribution; and (7) ability to be chemically derivatized to gain desirable properties, such as controlling water solubility or the addition of functional groups.

The polysaccharides pullulan, chitosan, levan and elsinan are currently under study in our laboratory. The first phase of the work is to evaluate the influence of culture conditions and processing steps on the molecular weight distribution and yields of the biopolymers. Batch and continuous cultures, a variety of media and media components, pH controls and processing variables are evaluated. Environmental and processing factors that are critical in controlling the molecular weight distribution of these polymers are identified. The second phase of the work is to evaluate the physical/chemical properties of these defined molecular weight fractions to identify potential applications to our material needs. Films made from these different molecular weight fractions have been evaluated for physical properties including tensile strength, flexibility, oxygen permeability and others. Fibers are being extruded in a solvent bath process and are evaluated for physical properties (tensile strength, flexibility, etc.) and for morphology using electron microscopy.

Pullulan - Pullulan is a biopolymer, which is released into the extracellular medium as a secondary metabolite by the dimorphic fungus Aureobasidium pullulans (de Bary) Arnaud during the yeast-like phase of the growth cycle, because it is not utilized during routine metabolism.[1,2] Pullulan is a linear α-D-glucan, predominantly 1,4-linked maltotriose with some maltotetraose units, connected by 1,6-linkages between the terminal glucosidic residues of the trisaccharide.[3,4]

Nutrient requirements for pullulan elaboration as reported in the literature include the following observations: hydrolyzed starch yielded the largest amount of pullulan;[5] pH, nitrogen source, and nitrogen limitation affected pullulan elaboration,[6,7] and pH and phosphate concentration affected the MW of pullulan.[8] Culture incubation varied from a few hours to seven days, and most results were reported for batch studies. However, most of this information has not been assembled into a cohesive study. Only a few authors have reported the MW distribution of the pullulan elaborated in their studies.[8,9,10] No work has been done to correlate fermentation/processing conditions to molecular weight distribution and subsequently to functional properties of films and fibers made from these different fractions.

Chitosan - Chitosan is a cationic polymer consisting of β-1,4 linked 2-amino-2-deoxy β-D-glucan. Chitosan rarely is found in nature. The primary occurrence is in the cell walls of some fungi, particularly the Zygomycetes which contain chitosan as a cell wall component,[11] in addition to chitin. Chitosan has been identified in the genus Mucor,[12] Aspergillus,[12] and Phycomyces.[13] Commercially, chitosan is derived from chitin which is obtained from shellfish. The strong alkali digestion required to deacetylate chitin to produce chitosan and the variability in source material for the chitin (i.e., shellfish) leads to variable physiochemical characteristics of the commercial chitosan product. The use of a fungus as a source of chitosan avoids the strong alkali digestion. Therefore, less concern exists about affecting molecular weight distribution of the final product. The use of a fungus also has the potential to provide closer control over the physiochemical properties of chitosan because of the ability to control fermentation and processing parameters. In addition, the potential exists for genetic manipulation using the fungal system.

The degree of acetylation of the glucosamine polymer determines whether it is chitin or chitosan and influences its solubility properties. Chitosan produced from shellfish chitin ranges from near 0% to 50% acetylation,[14] and cell wall chitosan isolated from M. rouxii has been reported to contain a 5% to 10% acetyl content.[15] The chitosan content of the M. rouxii cell wall is reported to be 33% for the

mycelia, and 28% for yeast-like forms.[11] Previous work concerning chitosan from fungal sources has not dealt with the aspect of molecular weight distributions of the polymer product. White's[15] choice of extraction parameters was based solely on chitosan yield.

Elsinan - Elsinan is a linear α-D-glucose comprised of α-1,4 and α -1,3 linkages. Elsinan is produced by the fungus Elsinoe fawcettii as an extracellular neutral polysaccharide.

Levan - Levan is an extracellular neutral β-2,6 D-fructose polymer backbone with β 1,2- and β 1,6- D-fructose branches. Levan is produced by the bacterium Erwinia herbicola.

Materials and Methods

Organisms - Eight strains of Aureobasidium pullulans, A. pullulans var. melanigenum, or A. mansonii were evaluated in preliminary studies to compare pullulan elaboration. Strain NRRL-Y 6220 A. pullulans was chosen for further studies, based on product color and yield, when compared with the other cultures. For chitosan production, Mucor rouxii (ATCC 24905) was maintained on yeast extract-peptone-glucose (YPG) agar slants, pH 5.0 and stored at 4°C. For elsinan production eight strains of Elsinoe were evaluated for elsinan yield and molecular weight, and Elsinoe fawcettii was chosen for further study. For levan production, four different organisms were evaluated for synthesis of the biopolymer, and Erwinia herbicola was used in further studies.

Culture Conditions - A variety of different media were evaluated including nutrient rich to nutrient poor or well defined. In addition to different media compositions, other variables evaluated for influence over biopolymer molecular weight distribution, yield, and quality included incubation period, pH controls, different carbon sources, different nitrogen sources and concentrations, and different phosphate sources and concentrations. Both batch and continuous cultures were studied. A BioFlo Model C30 Fermentor (New Brunswick) and a Magnaferm Model MA-114 Fermentor with an Automatic pH Controller and Pump Module (New Brunswick) with a Constant Speed Controller (Cole-Parmer Instrument Co., Chicago, IL) were used for one- and 10-liter continuous studies. Small-scale batches involved shake cultures of 500 to 750 mL in 2800 mL Fernbach vessels, or 50 to 100 mL in 250-mL DeLong flasks. Large-scale batch cultures consisted of 10 liters of culture medium in a 14-liter fermentor vessel on a Magnaferm MA-114.

Processing and Purification - Processing and purification procedures were developed for each of the polymers. A representative extraction scheme is shown for pullulan in Table 1. Extraction of chitosan from the

fungal cell wall was based on modifications of a procedure developed by White et al.[15] These changes included the elimination of lyophilization of NaOH treated (cell wall) material, homogenization of the cell wall material with a Waring blender in 2% acetic acid prior to refluxing, and an increase in cell wall-acid ratio to 1:≥100. These modifications of White's method were examined in relation to their effect on molecular weight distribution of chitosan and yield of product. Other changes were evaluated visually by using Lugol's stain. Modifications included pretreatment of mycelium by soaking or boiling in water; placement of mycelium in -70°C acetone to disrupt the cell wall; sonification of mycelium homogenate prior to NaOH treatment, also for cell wall disruption; and homogenization of cell wall material prior to acetic acid extraction by refluxing.

Analytical Methods - Determination of biopolymer MW distribution and dispersity was performed on a Waters 150-C ALC/GPC Gel Permeation Chromatograph (Waters). Standards and samples were run through three Bio-Gel (Bio-Rad Laboratories, Richmond, CA) columns; a TSK-60, separating in the 40,000 to 8,000,000 molecular weight range, and two TSK-50 columns, effective from 4,000 to 800,000 molecular weight. The TSK-60 column preceded the two TSK-50 columns in line. A third order calibration curve was generated correlating MW distribution with retention time on the columns. Standards and samples were solubilized at a concentration of 0.1% in the carrier solvent, which consisted of an aqueous solvent of sodium acetate, 0.1 M; acetic acid, 2% (v/v), and sodium azide, 0.05% (w/v). The instrument was adjusted to 1.0 mL/min flow-rate. The injection volume was 200 to 300 mL and the run time was 40 minutes.

The degree of acetylation of chitosan was determined by three methods, infrared spectrophotometry,[16] first derivative of UV absorbance spectrum,[17] and titration.

Results

Effect of Culture Conditions on Biopolymer Molecular Weight Distribution and Yield - An example of the influence of culture conditions on the molecular weight (MW) distribution and yield of pullulan is shown in Figure 1. Information such as this was developed on each of the polymers for each of the culture condition variables studied. In the example shown, the highest MW distribution product was attained on the first day of incubation, but the greatest yield was afforded after seven days of incubation.

As another example, the effect of different carbon sources on pullulan yield and molecular weight is shown in Table 1. Fructose used as

a carbon source resulted in the highest MW distribution of pullulan, but the use of sucrose yielded a greater amount of product. The use of soluble starch resulted in the highest yield of pullulan; however, the MW distribution of the product was relatively low. The processing procedure (shown in Figure 1) uses ROCCAL II (suggested by Dr. E. T. Reese of this Center, unpublished data) in processing the polymer was with reference to the work of Scott[18] and that of Bouveng et al.[19,20] who used long-chain quaternary ammonium compounds to precipitate acid-soluble glycans and β-linked glucans from α-linked glucans. The neutral α-linked glucans remain in solution, and after the precipitated products are removed by centrifugation, can then be precipitated by solvents such as ethanol or acetone. Fuller's earth or activated carbon can also be used to adsorb the quaternary ammonium compound from the solubilized polymer; also, these substances aid in removing the discolored products excreted into the medium by older cultures.[5,8,18]

In summary, defined MW fractions of pullulan or the other biopolymers studied were produced by varying fermentation conditions such as constituents of the culture medium, pH, length of incubation, etc. Pullulan biopolymer products with weight average MWs from 100K to 4 million have been produced with a dispersity of around two. The appropriate conditions were then established to produce sufficient quantities of low (below 500 thousand weight average MW), medium (1 to 2 million weight average MW), and high (above 2 million weight average MW) molecular weight pullulan products for subsequent evalution of fibers and films.

Chitosan was produced with weight average molecular weights ranging from 200K up to a million and dispersities varied from five to seven. The yields of the chitosan product were approximately 30% to 35% of the cell wall and 5% to 10% of the total biomass. Efforts are currently underway to develop cell free and recombinant systems for chitosan by cloning the key biosynthetic enzymes involved in chitosan biosynthesis in the fungus.

A number of growth and processing variables influenced the yield of biomass and the weight average molecular weight of the chitosan. These included growth conditions such as the length of incubation, culture vessel volume, the form of inorganic nitrogen salt in defined medium, and medium component concentration. The processing variables affecting the yield of chitosan extracted were the type and strength of acid and homogenization of cell wall material in acid before refluxing.

In complex media, weight average MW distribution increased rapidly over 72 hours of incubation, then declined as length of incubation increased. It may be that the chitosan is degraded or turned over in the

cell wall, or modified in the cell wall making extraction more difficult. There is documentation of the production of chitosanases in a strain of M. rouxii[21] which undergoes autolysis. In the strain of Mucor used in the study, autolysis did not begin until seven days. We have found that the decline in weight average MW begins prior to seven days. It is not believed that chitosanases are responsible. The chitosan must become accessible for the enzyme to be functional. Chitosan does not appear to become increasingly accessible, as the amount of chitosan extracted from the cell wall does not increase with time of incubation.

Elsinan was produced with weight average molecular weights ranging from 500K to over 4 million with a dispersity of around two, and work is continuing on the characterization of levan samples produced under the variety of culture conditions.

Derivatives - Chemical derivatives of the different polymers are being synthesized and are under evaluation for physical/chemical properties. Derivatization reactions involve cross-linking reactions for reduced water solubility for food packaging and other applications, the addition of nitrogen functional groups onto the polymer chain to change functional/reactive properties for chemical decontamination applications, and the addition of phosphorus groups for fire retardancy applications.

Physical/Chemical Characterization of Biopolymers - Aside from the characterization of molecular weight distributions and dispersity of the polymer products by gel permeation chromatography, a number of other parameters are being evaluated. Solubility properties and melting points are being determined; oxygen permeability of films cast from the biopolymer solutions are being evaluated and compared with synthetic films currently used in food packaging systems; flexibility, extensibility, percent elongation, and tensile strength are being determined for films and fiber material applications; and adhesive binding properties are being determined for the polymer solutions for material bonding applications and slow release finishes. Biodegradability of the polymers in plate tests is also determined.

Tables 3,4, and 5 illustrate some of these data for pullulan and chitosan. In Table 3 it can be seen that pullulan and chitsan films have very low oxygen permeability when compared with synthetic films currently used in food packaging applications. Lower oxygen permeability means slower rates of oxidation of food components and, therefore longer storage stability of stored rations. The data in Table 4 illustrate some of the physical properties of films cast from different molecular weight fractions of pullulan and illustrates the decrease in tensile strength and flexibility at higher molecular weights. Plasticizers are under

evalution to modify these properties. Table 5 presents comparative data for pullulan and current food packaging films used by the military.

Discussion

Biopolymers were considered as new sources for materials because of the following characteristics: their unusual physical/chemical properties when compared to synthetic polymers; the fact that they are readily biodegradable; biopolymer molecular weight distribution and resulting physical properties should be controllable through fermentation and processing step adjustments; there are no toxic side products or off-gases produced during their application; ability to form clear resistant films and fibers; and potential for genetic manipulation to further improve polymer properties.

The work we have presented in summary here has shown that by controlling fermentation and processing conditions, we can control or predict the molecular weight distribution of the biopolymer. This capability can then be translated into functional characteristics associated with each of these molecular weight fractions. The functional characteristics associated with these fractions can then be matched to end-item material applications needs.

With the continuing need for new materials that will provide improved functional characteristics over currently used synthetic polymers, with the increasing need for disposable plastics, and with the continuing need for materials with nontoxic characteristics and natural products, biopolymers offer an excellent new source of materials to meet some of our material needs.

Acknowledgement: We thank Drs. Elwyn Reese and Marv Mandels of this Center for their expertise and support.

References

1. HOOG, G. S. de, and E. J. HERMANIDES-NIJHOF. The Black Yeasts and Allied Hyphomycetes. Studies in Mycology, Centraalbureau voor Schimmelcultures, Baarn, Netherlands, 15: 222 pp. (1977).

2. ROMANO, A. H. Dimorphism, Chapter 7, In The Fungi, Volume II, The Fungal Organism, Edited by G. C. Ainsworth and A. S. Sussman, Academic Press, New York, 181-209 (1966).

3. GORIN, P. A. J., and E. J. HERMANIDES-NIJHOF. The Chemistry of Polysaccharides of Fungi and Lichens, Chapter 6, In The Polysaccharides, Volume 2, Edited by G. O. Aspinall, Academic Press, New York, 365-409 (1983).

4. SANDFORD, P. A., and J. BAIRD. Industrial Utilization of Polysaccharides, Chapter 7, In The Polysaccharides, Volume 2, Edited by G. O. Aspinall, Academic Press, New York, 412-490 (1983).

5. KATO, K., and M. SHIOSAKA. United States Patent 3,827,937. Process for the Production of Pullulan (1975).

6. CATLEY, B. J. The Rate of pH and Nitrogen Limitation in the Elaboration of the Extracellular Polysaccharide Pullulan by Pullularia pullulans. Appl. Microbiol. 22: 650-654 (1971).

7. RAMOS, S., and I. G. ACHA. A Vegetative Cycle of Pullularia pullulans. Trans. Br. Mycol. Soc. 64: 129-135 (1975).

8. KATO, K., and M. SHIOSAKA. United States Patent 3,912,591. Process for the Production of Pullulan (1975).

9. CATLEY, B. J. Pullulan, a Relationship Between Molecular Weight and Fine Structure. FEBS Letters, 10: 190-193 (1970).

10. UEDA, S., K. FUJITA, K. KOMATSU, and Z. NAKASHIMA. Polysaccharide Produced by the Genus Pullularia, I. Production of Polysaccharide by Growing Cells. Appl. Microbiol., 11: 211-215 (1963).

11. BARINICKI-GARCIA, S., and W. J. NICKERSON. Isolation, Composition, and Structure of Cell Walls of Filamentous and Yeast-Like Forms of Mucor rouxii. Biochim. Biophys. Acta, 58: 102-119 (1962).

12. ROUGET, C. Des Substances Amylacees Dans les Tissues des Animaux, Specialement des Articules (Chitine). Compt. Rend. Acad. Sci. Paris, 48: 792-795 (1859).

13. KREGER, D. R. Observations on Cell Walls of Yeasts and Some Other Fungi by X-Ray Diffraction and Solubility Tests. Biochim. Biophys. Acta, 13: 1-9 (1954).

14. FOSTER, A. B. and J. M. WEBBER. Chitin. In Advances in Carbohydrate Chemistry, Volume 15, Edited by M. L. Wolfrom and R. S. Tipson. Academic Press, New York, 371-393 (1960).

15. WHITE, S. A., P. R. FARINA and I. FULTON. Production and Isolation of Chitosan from Mucor rouxii. Appl. Environ. Microbiol., 38: 323-328 (1979).

16. SANNAN, T., K. KURITA, K. OGURA, and Y. IWAKURA. Studies on Chitin:

7. I.R. Spectroscopic Determination of Degree of Acetylation. Polymer, 19: 458-459 (1978).

17. MUZZARELLI, R. A. A. and R. ROCCHETTI. Determination of the Degree of Acetylation of Chitosans by the First Derivative Ultraviolet Spectrophotometry. Carbohydrate Polymers, 6:461-472 (1985).

18. SCOTT, J. E. Aliphatic Ammonium Salts in the Assay of Acidic Polysaccharides from Tissues. In Methods of Biochemical Analysis, Vol VIII, Edited by D. Glick, Interscience, New York, 145-196 (1960).

19. BOUVENG, H. O., H. KIESSLING, B. LINDBERG, and J. MCKAY. Polysaccharides Elaborated by Pullularia pullulans. Part I. The Neutral Glucan Synthesized from Sucrose Solutions. Acta Chem. Scand., 16: 615-622 (1963).

20. BOUVENG, H. O., H. KIESSLING, B. LINDBERG, and J. MCKAY. Polysaccharides Elaborated by Pullularia pullulans. Part III. Polysaccharides Synthesized from Xylose Solutions. Acta Chem. Scand., 17: 1351-1356 (1963).

21. REYES, F., R. LAHOZ, M. J. MARTINEZ, and C. ALFONSO. Chitosanases in the Autolysis of Mucor rouxii. Mycopathologia, 89: 181-187 (1985).

Table 1. Pullulan Processing Procedure

1. Culture Medium Neutralized with 1N NaOH
2. Dilution with ROCCAL (1%)
3. Centrifuge 12,000 rpm, 20 min.
4. Dilute Supernatant 1 to 2 with acetone (stirring)
5. Let Stand at Room Temperature or Refrigerate Overnight at 5 C
6. Decant Water/Acetone Solution
7. Wash Precipitate Several Times with Acetone
8. Filter Over Vacuum
9. Dry over Calcium sulfate, desiccate

Table 2. Effect of Carbon Source on Pullulan Yield and Molecular Weight Using NRRL-Y 6220 Aureobasidium pullulans

Carbon Source (10%)	Initial pH	Final pH	Yield (%)	Mol. Weight (X 1000)	Dispersity
Fructose	5.44	3.76	27.0	1122	2.5
Sucrose	5.44	3.69	34.4	895	2.4
Maltose	5.44	4.08	25.4	881	2.1
Corn Syrup	5.44	3.85	29.3	840	3.3
Dextrose	5.44	3.91	23.3	563	1.8
Lactose	5.44	3.79	5.8	518	1.6
Sol. Starch	5.44	4.24	70.3	137	1.4
Dextrin	5.44	2.99	21.6	9	1.1

Table 3. Oxygen Permeability of Biopolymer Films

Sample	cc oxygen/100 square inches (645 square cm) in 24 hours
Pullulan	
low molecular weight	
1.5 mil	0.21
6.5 mil	0.03
medium moleculr weight	
1.5 mil	0.24
6.5 mil	0.03
commercial sample	
1.5 mil	0.00
Others	
mylar (0.5 mil)	0.54
chitosan (1.5 mil)	0.12
polypropylene (3 mil)	13.90

Table 4. Physical Properties of Pullulan and Chitosan/Pullulan Films.[a]

Sample	Weight[b] (g)	Thickness (mm)	Flexibility (cm)	Extension (mm)	Elongation (%)	Tensile Strength (MPa)
Pullulan						
low MW	1.6	0.035±0.007	8.7±1.3	2.1±0.2	1.6±0.2	40.21±7.72
	6.0	0.163±0.043	>15	4.7±0.9	3.7±0.7	57.70±12.03
	1.6[c]	0.025±0.004	9.6±0.7	1.4±0.4	1.1±0.3	41.81±10.38
med. MW	1.6	0.038±0.009	8.5±1.3	2.6±0.7	2.1±0.6	58.71±14.58
	6.0	0.168±0.049	>15	2.6±0.2	2.1±0.2	38.16±5.31
high MW	1.6	0.033±0.007	10.4±1.0	0.6[d]	0.5[d]	9.96[d]
	6.0	---[e]	---	---	---	---
Chitosan/Pullulan (50/50 wt/wt)						
Neutralized	1.6	0.048[d]	---	2.8[d]	2.2[d]	57.34[d]
Untreated	1.6	0.043[c]	---	2.4[c]	1.9[c]	48.26[c]

[a] N=4 to 6

[b] weight per 6.25 square inches (16 square cm)

[c] Ueda media

[d] N=1

[e] no data

Table 5. Physcial Characteristics of Films.

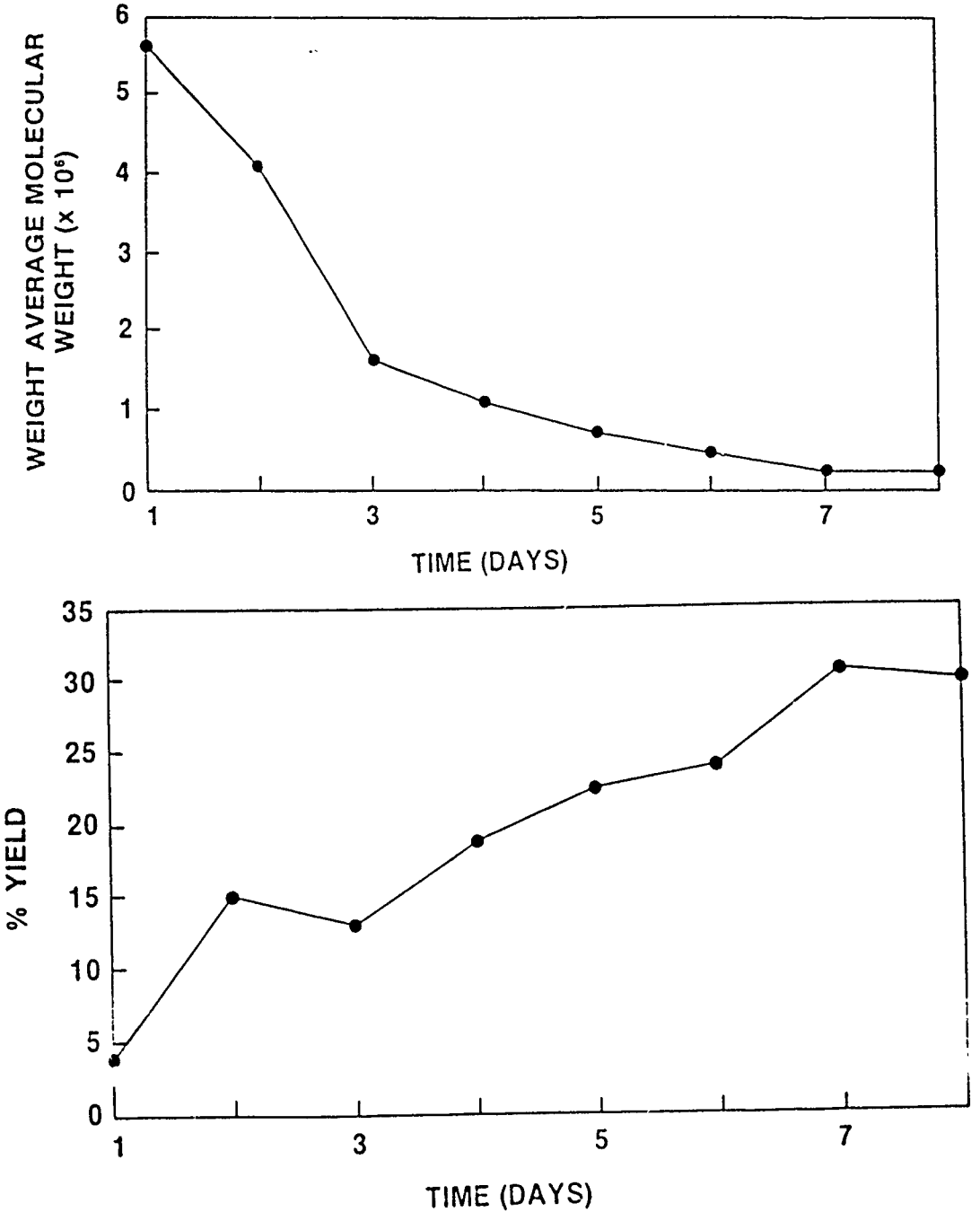
Property	Pullulan	Mylar[a]	Polypropylene[a]	Foil[a]
Tensile Strength (lb/in^2)	4 - 8 K	25 - 33 K	25 - 30 K	--[b]
Elongation (%)	1 - 4	120 - 140	60 - 100	--
Heat Seal Range (^oF)	N.D.[c]	not heat sealable	requires coating	--
MVTR (gm/24 hr/ 100 square inch at 100 ^oF, 90% RH)	N.D.	1.3	0.3	0.05
Oxygen Transmission mil/100 square inches 24 hr)	0 - 0.2	5	110	350
Resistance to Grease and Oils	Good	Good	Good	--
Max. Temp. (^oF)	250	400	250	--
Min. Temp. (^oF)	-90	-80	-60	--
Clarity	transparent	transparent	opaque	opaque
Biodegradability	Yes	No	No	No
Water Solubility	soluble	insoluble	insoluble	insoluble

[a] values from Packaging Encyclopedia and Yearbook, Vol. 30, Nov. 4, 1985

[b] information not provided

[c] not determined

Figure 1. Weight Average Molecular Weight of Pullulan vs. Time (top) and Percent Yield of Pullulan vs. Time (bottom).



Sickle Cell Trait as a Risk Factor for Sudden Death in Basic Training.

JOHN A.KARK, COL, DAVID M. POSEY, LTC, SAMUEL K. MARTIN, MAJ,
JAMES J. CANIK, MAJ, CHARLES J. RUEHLE, LTC
WALTER REED ARMY INSTITUTE OF RESEARCH,
ARMED FORCES INSTITUTE OF PATHOLOGY, WASHINGTON, DC, 20307-6000

Unexpected sudden deaths following collapse during military basic training pose a serious problem to training commands. Extensive regulations are required to prevent exertion-related deaths, because the sensible medical management of basic training requires firm constraints on the means used to produce optimal conditioning in a short time. This natural conflict between realistic understanding of human biologic limitations and military goals for performance has made it hard for the Army to sustain optimal management of training. When deaths occur considerable effort is required to investigate the cause of death and to determine whether correctable problems contributed to the trainee's demise. Occasionally deaths from unexplained cardiac arrest, heat stress, heat stroke, or rhabdomyolysis (muscle necrosis) occur despite correct conduct of training, because of unusual individual susceptibility.

In May of 1982 we were consulted concerning a cluster of training deaths among black recruits with sickle-cell trait at an Army training center in the Southeast. These cases appeared consistent with the then unsubstantiated hypothesis that sickle-cell trait contributes to exertional deaths. Individuals with sickle-cell trait have one normal gene for hemoglobin (Hb), Hb A, and one abnormal Hb S gene. Previous medical surveys have suggested that individuals with Hb AS genotype are essentially healthy except for a slightly increased incidence of hematuria and minor urinary tract infections among females, in sharp contrast to those with Hb SS genotype, who have sickle cell disease and are not elligible for the military (1-3). However sustained anaerobic exercise produces hypoxia, acidosis, dehydration, and fever, which cause sickling of erythrocytes in persons with sickle-cell trait (1-4). This might result in tissue infarction because sickled erythrocytes are abnormally rigid and occasionally are unable to pass through the narrowest blood vessels. Case reports describe young black men with sickle-cell trait, none of whom had a preexisting cause for illness, who died (12 cases) or suffered life-threatening renal failure following collapse during exercise (8 cases) (1-6). Seventeen of these cases were recruits in military basic training.

Most of these patients suffered from exertional rhabdomyolysis, heat stroke, or heat stress with acute renal failure, although a few died suddenly with cardiac arrest before diagnostic evaluation was possible. However, more unexplained deaths due to exertion from heat stress, rhabdomyolysis, or sudden cardiac arrest have been reported in members of the military or athletes who did not have Hb S (1-3). Since there is no means of distinguishing histologically between inconsequential postmortem sickling and serious antemortem vascular obstruction by sickled erythrocytes, the only way to determine whether hemoglobin AS is truly associated with death in such cases would be to establish that sickle-cell trait posed an increased risk of death during comparable exposures (1-3). Case reports have not established an association between sickle-cell trait and exercise-related mortality.

We examined all deaths in 1977-1981 during basic training of enlisted recruits in the U.S. Armed Forces in order to calculate the relative risk of sudden death in comparable populations with and without sickle-cell trait who experienced similar physical stress (7). The number of recruits according to race, age, and sex was known, and the frequency of hemoglobin AS was estimated with the use of well-established prevalence rates among American recruits. An accurate list of deaths had to be compiled from hospitals at the training centers. All natural deaths during basic training were investigated by means of a critical review of clinical data, eyewitness accounts, and autopsy, which included performance of hemoglobin electrophoresis if the recruit was black. Cases were classified on the basis of sudden or non-sudden onset and the presence or absence of a likely preexisting cause for death. This report presents estimates of natural death rates among subjects with and without sickle-cell trait, notes the association of hemoglobin AS with deaths of sudden onset during exertion that were unexplained by a predisposing cause, estimates the attributable risk of recruit exertional deaths for sickle-cell trait, and contrasts the effect of age on the mortality rates for those with and without Hb AS (8).

METHODS

The study population

The study population consisted of all enlisted recruits aged 17 to 34 years who entered basic training from January 1, 1977 through December 31, 1981. Descriptions of the recruits were obtained from the Defense Manpower Data Center and Accessions Operations, Department of Defense. Annual rates of entry of black and nonblack recruits were subdivided according to sex and age. The prevalence of hemoglobin AS was 8 percent among 20,000 black American recruits and 0.08 percent among 19,000 nonblack American recruits, similar to prevalence rates for American civilians (7). These rates were used to calculate the numbers of black and nonblack recruits in the study population who had hemoglobin AS.

Identification of Deaths and Records

We attempted to obtain an accurate list of recruit deaths for the study population from data bases maintained by the Armed Forces Institute of Pathology (AFIP), the Defense Manpower Data Center, the National Personnel Records Center, the Naval Medical Data Services, the Patient Administration Systems and Biostatistics Activity, and Casualty Affairs Offices of the Armed Services. However, basic trainee cases were included with the much larger set of all deaths between the end of basic training and qualification for an MOS. We therefore tabulated the recruit deaths among the study population from the Pathology and Patient Administration records of the 17 hospitals which served the 15 basic-training centers and confirmed these lists by critical comparison with the data bases and the service records. Military hospitals routinely conduct a thorough study of natural deaths of recruits. Unusually complete records were available, including the following for more than 90% of our cases: (1) the death certificate, (2) the complete autopsy protocol with toxicology reports, (3) the clinical records from entry to death, (4) eyewitness accounts from the police investigation, and (5) the AFIP autopsy consultation (only for 51/63 cases).

Classification of manner and cause of death

Sixty-two of 85 recruit deaths were considered to be natural after critical review to exclude deaths by accident, suicide, or homicide. In contrast to sudden deaths among civilians (7), immediate cardiopulmonary resuscitation, rapid transport by ambulance, and institution of life support on an intensive care unit were routinely provided to recruits collapsing at military training sites. Sudden natural death was therefore defined as death from an illness producing an irreversible critical condition within one hour of onset. This definition included cases in which survival was extended by the continuous use of life-support systems. Sudden deaths were further divided into those thought to be explained or unexplained by a known preexisting cause. Important attributed causes for such unexplained deaths in young adults are silent structural heart disease, epilepsy, intra-cranial bleeding, asthma, medications, and drug abuse (7). Our case collection was similar except for the absence of deaths attributed to epilepsy, medications, or drug abuse (confirmed by toxicology). Sudden unexplained deaths included those due to exertional heat stress, heat stroke, or rhabdomyolysis as well as those in which the mechanism was completely unknown, which are generally classified as cardiac deaths. The remaining deaths with slower onset were classified as non-sudden natural deaths. Causes of these deaths in the study population included pneumonia, meningitis, viral infections, structural heart disease, and systemic diseases missed during the physical examination at entry, which are similar to causes reported previously for young adults (7).

Diagnosis of sickle cell trait

Testing for hemoglobin AS was performed and the results were reported during the acute illness or postmortem in all 26 black recruits with natural deaths, but in only 3 of 36 nonblack recruits with natural deaths. The histologic findings were reviewed for sickled erythrocytes in all cases submitted with tissues or slides to the AFIP. Screening of black recruits for Hb S at entry into the Navy and Air Force was recorded in the medical records when positive. All subjects with sickle-cell trait were identified both by sickling morphology postmortem and, more important, by a pattern of Hb electrophoresis consistent with Hb AS genotype, i.e. less than 50 percent Hb S hemoglobin. Sickle cell disease was excluded from the cases by this pattern and by the absence of the characteristic lesions in the spleen (3).

Statistical methods

Death rates (per 100,000 recruit accessions) were computed for the categories of death (non-sudden, sudden explained, and sudden unexplained) among subgroups defined by race (black vs. nonblack) and by the presence or absence of sickle-cell trait (hemoglobin AS vs. non-hemoglobin S). In order to test the possible association between hemoglobin AS and risk of mortality, the relative risk for each category of death (sudden death [explained and unexplained], non-sudden death, and natural death) was calculated as the death rate for a hemoglobin AS group defined by race (black, nonblack, and all races) divided by the death rate for the corresponding non-hemoglobin S racial group (9). Assuming that the number of deaths followed Poisson distributions (the small-rate assumption), significance tests for the difference in death rates and 95 percent confidence intervals (95 % C.I.) for relative risks were obtained with the exact conditional method (10). For observed relative risks of zero an upper 97.5 % C.I. was calculated. All stated P values are two-sided. The attributable risk (or risk difference) of death as related to Hb AS was calculated as the death rate for black recruits with Hb AS minus the death rate for black recruits without Hb S (9). The effect of age on relative risk of death for hemoglobin AS was examined by stratifying death rates according to age, choosing five groups with approximately equal numbers of cases with Hb AS. The age distribution of black recruits showed little difference from that of nonblacks. Age-specfic death rates were calculated under the assumption that recruits with Hb AS had the same age distribution as the those without Hb S (7). Trends in relative risk with age were examined with the maximum-likelihood method for testing a common relative risk against a trend (11). Because of the small number of cases of sudden unexplained death during exertion (12 with Hb AS and 16 controls), this analysis was extended to similar cases with an adequate autopsy report from U.S. military recruit deaths during 1968-1987, yielding 12 additional cases with Hb AS and 15 additional cases without Hb S. We do not believe that all cases of unexplained death during exertion were found for 1968-1987. Therefore we could not calculate absolute death rates or even compare death rates for the two Hb phenotypes. However, the age distribution of recruits

has remained nearly constant. Making adjustments for small changes, one can calculate the predicted fraction of deaths for each age group based on its proportion in the recruits. The ratio of observed to predicted death rates should change between age groups in proportion to the unknown absolute age-specific death rates. These ratios were examined graphically for a trend with age and the significance of apparent trends was tested by the Chi-square distribution (9).

RESULTS

Table 1. Distribution of Deaths during Basic Training (1977-1981), According to Hemoglobin Phenotype, Race, and Category of Death

Category of Death	With Hemoglobin AS		Without Hemoglobin S		Total
	Black	Nonblack	Black	Nonblack	
Sudden Death					
Unexplained	12	0	5	11	28
Explained	1	0	5	8	14
Non-sudden Death	0	0	3	17	20
Total sudden deaths	13	0	10	19	42
Total of all deaths	13	0	13	36	62
No. of Recruits ($\times 10^{-3}$)	37.3	1.3	429	1617	2084

Table 1 presents the 62 study deaths in their classification by Hb phenotype, race, and category of death, with the corresponding numbers of recruits in the study population shown on the bottom row. Thirteen recruits with Hb AS died during basic training. All were black and all had sudden deaths related to exercise. Twelve deaths were classified as unexplained, and one as explained. Forty of the 42 sudden deaths among the entire recruit population presented as collapse during or closely following exercise. The other two deaths occurred in nonblack recruits without Hb S. Thirty-five sudden deaths related to exertion occurred during scheduled military training, three during strenuous activity in the first few days at the training camp before the start of formal physical training, and the remaining two during recreational athletics. Annual rates for sudden death did not exhibit a consistent trend during the five years studied (Fig. 1).

FIGURE 1. Annual Sudden Death Rates

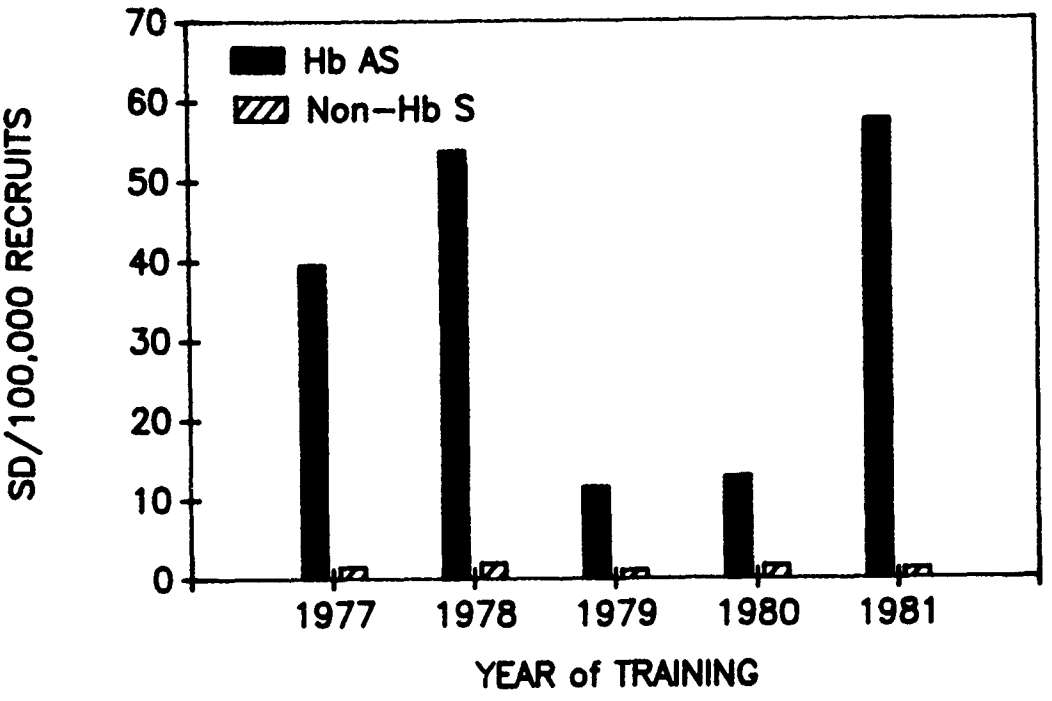


Table 2 summarizes the death rates and the relative risk for subjects with or without Hb AS. Among black recruits (Table 2 A) the relative risk of sudden unexplained death was 28, with no significant increased risk of sudden explained death or non-sudden death. The increased relative risk of sudden death or natural death was due to the large contribution of sudden unexplained deaths. There were no deaths in the small group of nonblack recruits with Hb AS. The corresponding relative risks among the nonblack groups were all zero and 95 % C.I.'s were broad, extending from 0 to at least 300 (data not shown).

TABLE 2. The Relative Risk Ratio and its Statistical Significance

A. Risk of Death among Black Recruits with Hb AS vs Controls without Hb S

Category of Death	Death Rate		Relative	95% CI	P Value
	Hb AS	Non-Hb S	Risk		
Exertional death					
Unexplained	32.2	1.2	27.6	9-100	<0.001
Explained	2.7	1.2	2.3	0.05-21	NS.
Non-sudden death	0	0.7	0	0-28	NS
All sudden deaths	34.9	2.3	15.0	6-38	<0.001
All deaths	34.9	3.0	11.5	5-27	<0.001

B. Risk of Death among all Recruits with AS versus Those without Hb S

Category of Death	Death Rate		Relative	95% CI	P Value
	Hb AS	Non-Hb S	Risk		
Exertional Death					
Unexplained	31.1	0.8	39.8	17-90	<0.001
Explained	2.6	0.6	4.1	0.2-27	NS
Non-sudden death	0	1.0	0	0-9	NS
All sudden deaths	33.7	1.4	23.8	11-47	<0.001
All deaths	33.7	2.4	14.1	7-26	<0.001

Among recruits of all races (Table 2 B) the relative risk of sudden unexplained death was 40. Again, the risk of sudden death or natural death was signficant, but the risk of sudden explained death or non-sudden death was not significant. The higher relative risks found among all recruits were due to higher sudden death rates for black recruits without Hb S than for nonblacks without Hb S. However, this relative risk was not statistically significant (7).

To examine the sensitivity of the relative risks shown in Table 2 to the estimated prevalence of Hb AS, similar calculations were made assuming an improbably high prevalence of Hb AS among both the black and nonblack recruits. If 12 percent of black recruits had Hb AS, the relative risks of sudden unexplained death, sudden death, or natural death would have been 18, 9.5, and 7.3, respectively. These risk ratios would still be highly

significant (P<0.001) with corresponding 95 % C.I.'s of 6 to 64, 4 to 24, and 3 to 17. If 0.16 percent of nonblack recruits had Hb AS, the relative risk of death for Hb AS among the entire recruit population would only decrease by 3 percent in each of the categories of death shown in Table 2. Furthermore, the absence of identified deaths among such a nonblack population with Hb AS, consisting of 2600 recruits, would not have suggested a statistically significant difference in risk when compared with the deaths in the black group with Hb AS, which averaged one sudden unexplained death per 3200 recruits.

None of the subjects with Hb AS had evidence of sickle cell disease by hemoglobin electrophoresis or histology of the spleen. The median fraction of Hb S among the 12 recruits with sudden unexplained death was 39 percent and the extremes were low values of 30, 35, 35, and 36 percent, and high values of 41.5, 42, 42, and 44 percent. Thus very high values of % Hb S (>43%) were not over-represented among deaths and very low values (<36%) were not under-represented (7).

Table 3 summarizes age-specific death rates and relative risks for sudden unexplained death among those with and without Hb AS. Table 3 demonstrates an increasing trend in the rate of unexplained sudden death for Hb AS recruits, but no trend in death rate with age for recruits without Hb S (Fig. 2). The trend in relative risk with age was significant (P<0.04)(11). Age-specific rates for sudden explained deaths and non-sudden deaths could be examined for trends only among the recruits without Hb S; these death rates showed no significant trend with age.

TABLE 3. Age-Specific Death Rates and Relative Risk

Age Group	Percent of Recruits[+]		Sudden Unexplained Deaths		Death Rate[*]		Relative Risk
	Hb AS	Non-Hb S	Hb AS	Non-Hb S	Hb AS	Non-Hb S	
17-18	43.4	47.4	2	9	12.0	0.9	12.9
19-20	31.9	30.4	2	4	16.3	0.6	25.2
21-22	12.2	11.0	2	0	42.6	0.0	--
23-25	7.9	7.0	4	2	131	1.3	94.2
26-30	3.8	3.4	2	1	136	1.4	95.1
31-34	0.9	0.8	0	0	0	0	--
Total	(100)	(100)	12	16	31.1	0.8	39.8

P<0.04 for test for trend with age.

[+]The total number of recruits with hemoglobin AS was 38,600; the total number without hemoglobin S was 2,046,000. [*]Per 100,000 recruits.

FIGURE 2. Age-Specific Death Rates: Sudden Unexplained Deaths per 100,000 Recruits, for 1977-1981.

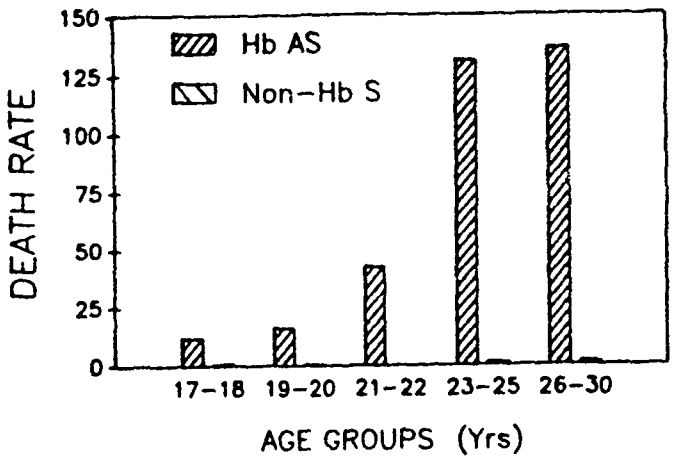


Table 4 summarizes the ratios of observed to predicted death rates, based on the fraction of the population at risk in each age group for the larger time period, 1968-1987. These ratio's were normalized and graphed (Figure 3). This figure illustrates the apparent trend with age in the risk of unexplained exertion-related death for sickle-cell trait (significant at $P<0.001$ by the Chi-square distribution) and the absence of a trend for controls without Hb S ($P<0.3$ by the Chi-square distribution). Similar calculations were made from data for the better studied period, 1977-1981, and were graphed in the same manner (Fig. 4). Comparison of Figures 3 and 4 shows that the two data sets are consistent with each other.

TABLE 4. Sudden Unexplained Deaths: Relative Death Rates Across Age

Age Group	Percent of Recruits		Sudden Unexplained Deaths Hb AS			Hb AA		
	Hb AS	Non-Hb S	Obs.	Pred.	Ratio	Obs.	Pred.	Ratio
17-18	43	47	4	10	0.4	12	15	0.8
19-20	32	30	5	8	0.6	13	10	1.3
21-22	12	11	6	3	2	2	3	0.7
23-25	8	7	6	2	3	3	2	1.5
26-30	3.8	3.4	3	1	3	1	1	1
31-34	0.9	0.8	0	0	-	0	0	-
Total	(100)	(100)	24	(24)	-	31	(31)	-

FIGURE 3. Relative age-specific death rates for exertional unexplained deaths, setting rate of the youngest age group arbitrarily at 1.0 for both hemoglobin phenotypes, using data from 1968-1987 (24 versus 31 cases).

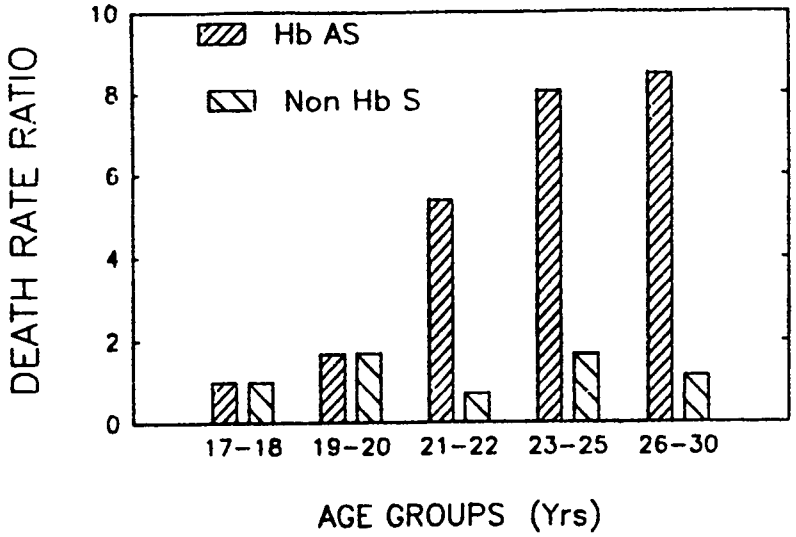
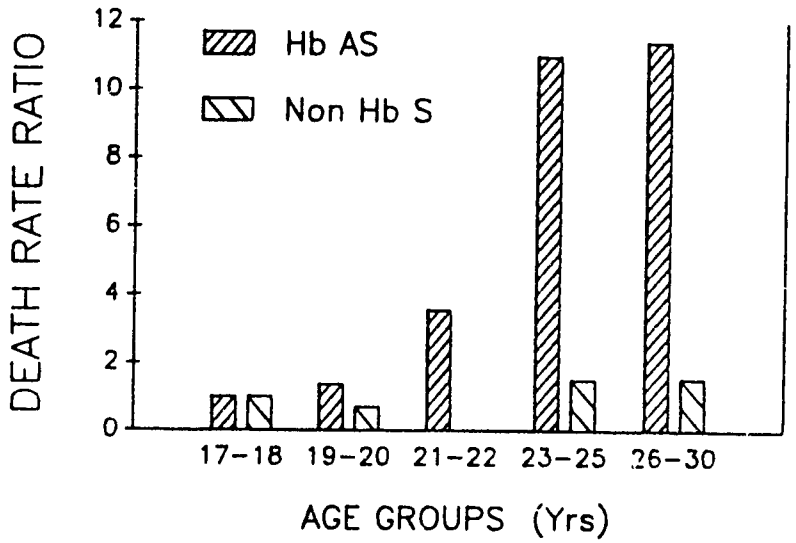


FIGURE 4. Relative age-specific death rates for exertional unexplained deaths, setting rate of the youngest age group arbitrarily at 1.0 for both hemoglobin phenotypes, using data from 1977-1981 (12 versus 16 cases).



DISCUSSION

We have identified sickle-cell trait as an important risk factor for sudden death in exercise, by estimating the rates of natural death in a cohort of 2.1 million recruits who entered U.S. Armed Forces basic training during 1977-1981. The risk of exertion-induced sudden unexplained death was 28 and 40 times higher among those with Hb AS, as estimated among black recruits and among recruits of all races, respectively. About 43% of unexplained exertional deaths in military basic training occured among recruits with sickle-cell trait. The risk of unexplained exertional death attributable to Hb AS among black recruits was 31 deaths per 100,000, a high death rate considering the age distribution of recruits. Charache, for example, claimed that this risk was lower than the corresponding age-specific annual death rate from homicide for urban blacks in Baltimore (100/100,000 per year)(4). However, he failed to note that we did not calculate death rates per unit time, because of our perception of this risk as due to cumulative stress. The risk attributable to sickle cell trait among recruits would correspond to a death rate of 160/100,000 per year. sickle-cell trait per year. However, as expected, Hb AS was not significantly associated with sudden explained or non-sudden deaths. The risk of sudden death (both explained and unexplained) among the subjects without Hb S was approximately twice as high in black as in nonblack recruits, consistent with larger national mortality surveys for the same age range (7). The number of nonblack recruits with Hb AS was too small to provide a reliable estimate of risk associated with Hb AS in this group.

Attempts were made in this study to eliminate potential sources of bias by ensuring that identification of cases of natural death was complete, that such cases involved only recruits in basic training, that classifications of the type of death and hemoglobin phenotype were accurate, that demographic data on the recruits (numbers according to race, sex, and age) were accurate, and that conclusions were not dependent on unreasonably low estimates of the prevalence of Hb AS. In contrast to many retrospective studies in which reliance on the death certificate alone has resulted in erroneous diagnosis, the present study determined the manner and cause of death from full autopsy protocols, including biochemical studies, clinical data, and contemporaneous critical review by the AFIP staff. Accurate recognition and description of illness of sudden onset, without selection bias, would be expected because recruits are constantly under observation and eyewitness accounts of such events are consistently reported. Moreover, subclassification of natural deaths was not essential to demonstrate the risk associated with sickle-cell trait, since the relative risk of natural death had a high level of significance. Precise figures for the populations at risk were obtained from contemporaneous Department of Defense records of recruit entering the services. The prevalence of Hb AS was not measured in this study, but consistent values have been observed in similar populations, including military recruits.

The relative risk of death in persons with Hb AS remained statistically significant even when one assumed a much higher prevalence than has been reported among American blacks.

All of the 12 cases of sudden unexplained death in subjects with Hb AS occurred during exertion. These cases could be divided into deaths presenting with acute cardiac arrest of undefined mechanism and deaths related to exertional heat stroke, heat stress, or rhabdomyolysis. The data were insufficient to test whether Hb AS was more closely associated with any specific subset of sudden unexplained deaths. It has been suggested that higher fractions of Hb S would be an important risk factor for complications related to sickling (1-3). However, the reported cases with complications related to sickle-cell trait have shown little or no difference in the % Hb S distribution from that observed among healthy individuals with sickle-cell trait (1-3), as was the case in this study (Results). The metabolic consequences of heat stroke, rhabdomyolysis, or cardiac hypotension would be expected to provoke polymerization of Hb S, which in theory could increase mortality by causing tissue infarction (6). However there is no evidence which distinquishes between the possibilities that sickling directly causes increased mortality or that sickle-cell trait is only indirectly related to death. For example, some unrecognized genetic variant linked to Hb S might be the actual cause of increased risk.

There was a significant trend with age in the relative risk of sudden unexplained death in subjects with Hb AS, because the risk-ratio numerator (the rate of sudden unexplained death among recruits with Hb AS) increased with age while the risk-ratio demoninator (the rate of sudden unexplained death rate among recruits without Hb S) did not change substantially with age. Since this trend could be eliminated by the movement of a few cases from the older age groups, which were very small, we tested it further by collection of additional cases from 1968-1987. Analysis of this larger data set confirmed the unique age trend for risk with sickle-cell trait at a high level of statistical significance. In addition, the new data confirmed the rather steep slope for this trend, an 8 to 12-fold increase in going from age 17 to age 27. These observations imply that the pathogenesis of exertional death without prior disease also differs between individuals with sickle-cell trait and those without Hb S. This supports the conclusion that death rates differed by hemoglobin phenotype and provides a clue to mechanism.

It is tempting to explain the age dependence of the risk in persons with Hb AS as a consequence of partial hyposthenuria (attributed to silent renal papillary necrosis), because this defect in urine concentrating ability progresses with age (3, 7). However, it remains to be shown that the extra fluid loss due to this limitation of renal function would be clinically important. Other potentially important factors which could increase the risk of complications of sickling, such as viral infection of

the upper respiratory tract with transient hypoxia (6), elevated hematocrit, and transient causes of dehydration, would not appear to explain the increasing risk with age.

Our study appears to challenge the conventional view of sickle-cell trait as a benign genetic variant (except for rare complications during substantial hypoxia). However, whether the risk associated with sickle-cell trait is important depends upon the range of activities sufficient to provoke risk, the possible restriction of risk to a subset of those with trait, and the extent to which this risk is preventable. The risk of sickle-cell trait might be very much less for civilian life, since sports, recreation, and labor seldom are as stressful as basic military training with respect to rapid physical conditioning, the severity of exposure to heat stress and dehydration during exertion, and the duration of physical training. In editorial comment on our study, Sullivan stated that athletes were not at risk since no complications were found in previous studies of athletes with sickle-cell trait (12). Moreover, subjects with sickle-cell trait or with Hb AA had similar exercise performance and metabolic responses to exercise in controlled studies of exercise to exhaustion, including observations upon recruits in basic training (3, 4). No complications were noted following exercise at sea level despite in vivo sickling in all subjects with Hb AS. However, less than 200 people with sickle-cell trait have been observed in the descriptive surveys quoted by Sullivan and less than 50 in the physiologic measurements of exercise, which were accomplished rapidly in a pleasant envirnoment (4,7). Prior studies of the medical risks of sickle-cell trait have examined populations which were much to small to detect an excess death rate of about 1 per 3000 (1-4, 7), especially if one assumes that mortality might be lower for activities which are less severe than military basic training. The largest studies examined approximately 600 enlisted men in the Navy with sickle-cell trait (14) and 5000 hospitalized veterans with sickle-cell trait (15). There are little data bearing on the potential risk for advanced military training or the possibility that risk might be restricted to a susceptible subset, who could be eliminated from the military by morbidity or motality during entry training. The level of unexplained exertional death we observed for Army recruits without Hb S during 1977-1981 appeared excessive, especially since deaths from heat stroke and heat stress are considered largely preventable. Our data suggest that inadequate hydration, failure to reduce activity appropriately for hot, humid environments, use of inappropriate clothing, excessively intense physical effort, and failure to acclimatize adequately prior to training contributed to the majority of these deaths. Perhaps the mortality associated with sickle-cell trait might be largely preventable by correct application of measures required for the safe conduct of training regardless of Hb phenotype. In 1982 Army recruit training commands attempted substantial improvements in the conduct of basic training, at the time of a memorandum communicating our concern that excessive deaths with sickle-cell trait

might indicate a need to improve the medical management of training. It will be important to determine whether death rates for recruits have fallen subsequently. Further clinical and epidemiologic study of exertion-related deaths in military training may allow us to prevent the excess mortality associated with sickle-cell trait and also reduce training deaths for the great majority of recruits who do not have hemoglobin S.

REFERENCES

1. Sears DA. The morbidity of sickle cell trait: a review of the literature. Am J Med 1978;64:1021-36.
2. Diggs LW. The sickle cell trait in relation to the training and assignment of duties in the Armed Forces: III. Hyposthenuria, hematuria, sudden death, rhabdomyolysis, and acute tubular necrosis. Aviat Space Environ Med 1984;55:358-64.
3. Serjeant GR. The sickle cell trait. In: Sickle cell disease. New York: Oxford University Press, 1985:329-37.
4. Charache S. Sudden death and sickle cell trait. Editorial. Am J Med 1988;84:159.
5. Jones SR, Binder RA, Donowho EM Jr. Sudden death in sickle cell-trait. N Engl J Med 1970;282:323-5.
6. Koppes GM, Daly JJ, Coltman CA Jr, Butkus DE. Exertion-induced rhabdomyolysis with acute renal failure and disseminated intravascular coagulation in sickle cell trait. Am J Med 1977;63:313-7.
7. Kark JA, Posey DM, Schumacher HR, and Ruehle CJ. Sickle-cell trait as a risk factor for sudden death in physical training. New Engl J Med 1987;317:781-7.
8. Kark JA., Canik JJ, Martin SK, Hicks CU. Age dependence of the risk of sudden death associated with sickle cell trait. Abstract. Blood 1987;70: 64a.
9. Schlesselman JJ. Case-control studies. New York: Oxford University Press, 1982:32-33, 40-1.
10. Ederer F, Mantel N. Confidence limits on the ratio of two Poisson variables. Am J Epid 1974;100:165-7.
11. Breslow NE. Elementary methods of cohort analysis. Internat J Epid 1984;13:112-5.
12. Sullivan LW. The risks of sickle-cell trait. Editorial. 1987. New Engl J Med 1987;317:830-1.
13. Hoiberg A, Ernst J, Uddin DE. Sickle cell trait and glucose-6-phosphate dehydrogenase deficiency: effects on health and military performance in black navy enlistees. Arch Int Med 1981;141:1485-8.
14. Heller P, Best WR, Nelson RB, Becktel J. Clinical implications of sickle-cell trait and glucose-6-phosphate dehydrogenase deficiency in hospitalized black male patients. N Eng J Med 1979;300:1001-5.

Intelligent Systems for Construction and Armament Production (U)

*Frank W. Kearney, Mr.
U.S. Army Construction Engineering Research Laboratory
Champaign, IL 61820-1305

1 Background

The genesis for this work in intelligent systems came from the CERL R&D in real time weld quality control which was reported on at this conference in 1980. In that work, sensors and associated micro-computer based monitoring and control systems were developed to provide quality control information as the weld was in process rather than in the conventional after the fact nondestructive inspection systems. These "post mortem" quality assurance procedures result in a disproportionately high cost of rework whereas in real time monitoring, flaw inducing conditions are detected and the welding process can be modified or terminated to preclude defective welding. The unique sensor systems developed in this work, which have been patented and licensed to a private firm, provided a rudimentary form of computer cognition for welding control systems that are suitable in the development of an intelligent control system.

About the same time that the welding R&D was in progress, CERL became concerned with the imminence of automated construction and its subsequent impact on Corps of Engineer construction practices. A study performed by CERL indicated several automated construction systems were under development in foreign countries that would probably make their appearance on the American construction scene by 1990. To rigorously investigate this problem CERL began active participation with academia, industry and other government agencies that shared an interest in the impact of automation in construction. One of the salient resolves of this group was to learn and profit from the mistakes made in the chaotic robotization of manufacturing processes in the beginning of the mid 70s continuing into the 80s; robots were haphazardly employed as a panacea to reduce manufacturing costs through direct replacement of a human in a task without considering the necessary process modifications to accommodate to automata. Several amusing stories emerged from this period such as robots for spray painting of autos painting each other instead of the cars. As a starting point in obviating this occurrence in the construction industry

it was determined that emulation of flexible manufacturing concepts was advisable. In order to implement these concepts multitask "robots" would be involved which necessitated a control hierarchy involving artificial intelligence.

The confluence of these two major technology thrusts coupled with the need for intelligent systems in AMC applications provided an impetus that has resulted in an extensive and fruitful effort. Four technical specialties are molded in the accomplishment of the objectives of this work, i.e. artificial intelligence, automatic control systems, computer cognition and system integration. It is the purpose of this paper to present a montage of this trans-discipline effort. Hence, the philosophical essence of achieving intelligent control will be emphasized while the references will provide a source for more rigorous treatment of particular specialties.

2 Objectives

There are dual objectives to this work although separate they are synergestic and the work packages involve distinct timeframes but are concurrent in execution. The first objective can be referred to as the inventive objective and is to determine a global abstract causal model that is generic to the design of intelligent systems. This would be a type of meta system similar to the approach in knowledge engineering. The second objective, the innovative objective, is to utilize the technologies developed in the laboratory in real world applications such as described in Section 7. It is the experiential factors of the field implementation that develop and enhance the global, innovative objective. Although most of the laboratory work is related to an intelligent welding control system, this is not the ultimate objective; rather, the welding system is serving as a testbed for concepts and concept validation.

3 Root Problems

Three onorous difficulties are present in this endeavor, they are partly semantic and quasi-technical. The first has to do with the definition of an intelligent system itself. Many purveyors of robots attach the adjective "intelligent" to their device when actually they are referring to a higher degree of end effector tactility or a simpler trajectory programming scheme. There have been endless discussions regarding the definition of an intelligent system or the description of an intelligent action. For this presentation the "podium preogative" will be invoked and an intelligent system will be defined as one which has the knowledge of the process resident in the control system and a meta knowledge control function. Stated another way, intrinsic artificial intelligence must function as a supervening agent in the control of the automated system; the level of this AI can be quantified as shown later

and an analogous machine intelligence quotient can be ascribed for comparative purposes.

The second difficulty arises from what is artificial intelligence and what is an expert system. For example, there are two quality parameters in welding defined as:

$$\text{Heat Input} = \frac{V \times I}{S}$$

Where: V - arc voltage; I - arc current; S - weld travel speed

$$\text{Nugget Area} = \frac{I^M}{S^N}$$ M & N empirical constants.

Depending on the nature of the weld, one of these can be the controlling quality parameter. However, under certain conditions a welding engineer may decide to switch control from heat input to nugget area which is an intelligent decision that is based on his innate knowledge and experience. If the control system implemented this decision autonomously is the system intelligent; an extension of the Turing criterion would indicate the affirmative. Even the action of a self-tuning adaptive control system which modifies the control algorithm to accommodate changes in the process environment implies primitive intelligence. The resolution of the existence of machine intelligence and/or the threshold of intelligence must depend on the particular situation.

Another ponderable problem area is the reconciliation of control engineering protocols and procedures with the methods of artificial intelligence. In the former very austere and precise methods are implied whereas in the later a less mechanistic method is employed. For example the control engineer would design a system to control a weld preciously to 18 volts, 370 amps, 9.5 inch travel speed to obtain a quality weld whereas the AI objective, although being the same, would be described as "good," "acceptable." Although in the example given above the models associated with heat input and nugget area would be the primary guidance for control system design, while the decision to chose heat input or nugget area would be a function of the artificial intelligence module based on the more abstract concept of "good," "acceptable" weld. The situation can be compared to the methods of rigorous deterministic mathematical methods and the method of fuzzy sets.[2] This dichotomy was described by Dr. Phil Hart of Massachusetts Institute of Technology in 1983.

Resolution of these root problems is a tenuous task for the principal investigator for intelligent system design and requires the wisdom of Solomon and the patience of Job.

4 Approach

The basic model for this work is an hierarchical control system as schematically shown in Figure 1. The hierarchical ascension is from primitive, proportional integral derivative (PID) control to self tuning adaptive control and culminating in the artificial intelligence/expert system level. The complexities of hierarchical control are myriad but will be condensed to a degree sufficient to be relevant to the discussion in this paper. For a cogent exposition of the principles of hierarchical systems, the treatment by Albus[1] is superb. Concepts associated with his Cerebellar Model Algorithmic Computer (CMAC) are seminal and any work in the domain of intelligent systems bears some degree of propinquity.

The hierarchy shown in Figure 1 can be reduced to two major functional modules. One incorporating the mechanistic control system, the other encompassing the AI expert system portion. This is shown in Figure 1 with annotations showing differences in operational time and labels providing an analog to human behavior. The intelligence module operates in an exception mode. It is transparent to the control system except when the process variations exceed the capabilities of the highest level in the control system; it is then that the intelligent module supervenes and modifies control actuations. In the normal mode the AI module is monitoring the system both through the cognition system and the processed or derived functions within the control loop.

It became obvious that a special control system design suitable for AI supervention was needed for achievement of the objectives. Initially, brute force, gain scheduling was tried. Although this was somewhat successful, instability resulted in all but the slowest processes. In 1983 methods devised by Kokotovic[2] have been utilized and through continuing work with his graduate students, the present level of success was made possible.

As in the case of hierarchical control structure there is a functional complexity gradation in a computer cognition system as shown in Figure 2. At first it may seem trivial to represent a sensor system in this fashion but for intelligent systems it is necessary to distinguish between computer cognition and a multiple sensor array. The sensors essential to the operation of the intelligence module are included in the cognition category whereas ancillary sensors such as limit snitches, shaft and coders, etc. not relevant to intelligent control are excluded from the cognition category. Although Figure 2 shows a cognition system in a discrete block form, it is not sometimes apparent where the processing function occurs as will be seen in the discussion of the control system. Part of this occurs in the implementation of the algorithm. In each of the applications discussed later the cognition aspects of the application will be delineated.

5 Artificial Intelligence/Expert Systems for Intelligent Control

Two subsets of artificial intelligence are currently being used at CERL in intelligent system design. They are pattern recognition and knowledge base expert systems; for the application examples discussed in this paper the later will be considered.

Nearly every technical journal and several popular periodicals have published articles describing expert systems and their applications. An expert system is conventionally shown as in Figure 3; it is composed of a knowledge base containing both deterministic and heuristic knowledge coupled with an inference system that performs the logical operations that lead to the conclusion. Two caveats regarding expert systems: first they are not a sophisticated data base management system. Data as contained in a data base in not knowledge although it can be processed using induction methods to become knowledge. Second, the architecture of the inference engine is not similar to that of a data base management system. Algorithmically the inference engine is based upon predicate calculus whereas the data base management system is based on relational algebra. As mentioned the knowledge base consists of deterministic knowledge such as contained in textbooks, operational manuals, troubleshooting guides, etc. and heuristic knowledge or "gut-feeling procedural knowledge." To input knowledge to the computer memory it must be in correct format; it is relatively easy to format textual deterministic knowledge but considerably more difficult to do the same with heuristic knowledge. Because of this, most of CERL's knowledge engineering research has been on methods to explicate the heuristics of humans.

The most common form of expert system is of the deductive type, that is, the system reasons from given rules and knowledge input to a conclusion. In the case of the intelligent control system, the input is from both the cognition system and information regarding the instantaneous state of the control loop.

In the course of this work a severe limitation appeared that threatened to thwart the achievement sought in this work. The problem is not unique to intelligent system design but is endemic in nearly all applications of high speed microprocessors and that is the inability of the human to devise methods that keep pace with the capabilities of computer based systems. In the case of the intelligent welding work the computer cognition provides more processed information that can be utilized because of inadequate process knowledge. For example, it is not clear how to incorporate acoustic emission data into the control strategy; this is true for other sensor systems as well.

One approach used to circumvent this impediment is to utilize inductive systems whereby processed data from the cognition system would be used to generate rules and probabilistic knowledge engineering methods to augment the current state of knowledge.[2]

6 Control System

In order to form an overview of the intelligent welding system a brief description of the weld process involved and the principle sensor systems will be described.

The welding process is capital intensive, prone to quality control difficulties and often unpleasant for the human operator. During the welding process changes in parameters, consumables and the weld arc atmosphere can occur without the operator's knowledge. These changes may result in thermal damage to the base materials and defects (e.g., hydrogen induced cracking, porosity, embrittlement, lack of fusion and penetration) which seriously reduce the strength and service life of the welded joint. The cost of locating and repairing these defects constitutes a significant portion of the total weld fabrication cost. Not surprisingly, a large effort is underway to automate welding and improve weld process control.

In very simple welding systems open loop control is employed. Figure 4 shows the basic components of such a system. The operator sets the output by appropriately adjusting the demand input. Due to external disturbances and variations in parameters inherent to the weld process, the output of the system will drift from the desired value. And so to maintain a desired output, the operator must manually readjust the demand input. In most circumstances frequent readjustment of the demand input is impractical. And so one fixes the input and is forced to accept variations in the output.

The next step in controller sophistication is to employ a closed loop system as shown in Figure 5. An array of sensors is employed to measure the salient output variables of the weld process (e.g., weld pool width, bead temperature, arc current). These measurements are compared against the demand values to form an error signal. The controller is designed to minimize the error signal in the presence of external disturbances and variations in the weld process parameters. In addition the controller is chosen to provide a desired dynamic response of the output variables due to changes in the demand input. In general, for a given set of weld process parameters, one can select a controller design which optimizes both the dynamic response and the minimization of the error signal. Unfortunately, this design is typically suboptimal for other sets of process parameters. And so a fixed controller design in a closed-loop system is by no means the final solution to weld process control.

The performance of a closed loop system in the presence of time varying process parameters is enhanced through the use of an adaptive control scheme. Figure 6 shows the basic components of this system. The parameter estimator monitors the weld process inputs and outputs and uses these observations to calculate the process parameters. The controller is then tuned for optimal performance given the state of the process parameters.

The merge of the adaptive control system with an intelligent supervisory system is shown in Figure 7. The AI system determines the demand input for a given process as well as fine tune the controller. It is important to realize the separation of time scales in this figure. The parameter estimates and controller posses time constraints on the order of milliseconds. On the otherhand, the intelligent system oversees the process on a much slower time scale, as indicated in Figure 1.

A detailed block diagram of the intelligent welding system currently under development at CERL is shown in Figure 8. This figure illustrates the multi-input multi-output nature of the weld process. Currently we employ the gas metal arc welding process (GMAW) in which the electrode material is consumed by the arc and deposited on the workpiece. The primary inputs to the process are heat and metal deposition. These inputs in turn are specified by the wirefeed rate and travel speed. The primary outputs of the system are weld width and bead temperature. In addition, arc current, wire feed rate and travel speed are also measured. The demand inputs to the system consist of bead width and cooling rate. These variables were chosen since they primarily determine the quality of the resulting weld.

At CERL considerable attention has been given to developing sensors capable of measuring various welding parameters and resultant weld characteristics in real time. This development has progressed on three fronts.

First, a process data system (PDS) has been developed to measure arc current, voltage and travel speed. In the PDS, a Hall effect device is used to obtain arc current values, voltage is measured at the welding head, and one of a variety of methods (tachometer, shaft encoder, etc.) is used to measure travel speed. The analog output from these devices is digitized and the data used by a microprocessor to determine whether the process is being maintained within preset limits. Secondary information about the weld quality, such as heat input and nugget area may also be computed. The PDS generates this data in real time, and can be used to stop welding, or alert an operator when an out-of-limits event occurs; it can also be used to drive a feedback control system.

The second sensor system, which was developed in conjunction with the Radio Research Laboratory of the University of Illinois, is an optical data system (ODS) capable of detecting variations in weld arc chemistry. During welding, loss of shielding gas, contaminated electrodes, contaminated shielding gas, or any of a number of other difficulties which may be associated with welding consumables can cause defects in weldments ranging from porosity to cracking. The detection of variations in weld arc chemistry which are the cause of such problems is not a simple matter. Because of the extremely high temperatures near the welding arc, a remote sensing method is required.

Remote arc sensing has been accomplished by the development of an opto-electronic method for observing the arc and noting variations in its composition. The high excitation energies in the arc plasma make spectroscopic analysis an attractive method of obtaining qualitative and quantitative information about events during welding. To this end, equipment capable of real-time evaluation of the spectral features of the welding arc was developed by the Construction Engineering Research Laboratory at the University of Illinois.

The third sensor system was developed over the past year. It consists of a Reticon line scan camera and a Vanzetti optical pyrometer. The former device is used to measure the weld pool width while the latter measures bead temperature. It is important to point out that both these devices are available "off the shelf." Only minor modifications and additions are necessary to incorporate them into the welding system.

7 Intelligent Systems Applications

Three AMC operations are excellent candidates for intelligent systems; work is in progress to implement them.

a. Explosive Ordnance Disposal (EOD)

Presently the "render safe" procedures for ordnance disposal is conducted by a three-man team and is extremely hazardous. Because "blow-in-place" procedures are usually not acceptable because of a cascading effect through detonation of adjacent explosives or location of the ordnance near critical facilities, meticulous hand operations are employed. PM-AMMOLOG discussed with CERL the feasibility of utilizing intelligent automation systems to perform these tasks and remove the soldier from this lethal operation. It was determined that an automated abrasive waterjet metallurgical procedure would be suitable. The initial work involves teleoperation with subsequent phases being a completely autonomous operation.

A unique robot design concept was developed for this project. It was necessary to minimize the weight of the device to be placed next to the ordnance because it had to be transported in a small pickup truck and handled by two soldiers without the aid of mechanized handling systems. Conventional robots have a payload to structural weight ratio of 15 or 20 to 1 in order to insure the rigidity required for preprogrammed trajectories. Because the operational loads for field EOD operations using the waterjet would be in the order of 50 pounds the resultant weight of a conventional robot would be in excess of 1000 pounds, which certainly is not manageable by two humans. With the unique "chopstick" design and the innovative trajectory control system, the automated device weighs slightly more than 100 pounds.

Because there are certain critical limits to the cutting or drilling procedures to preclude unwanted explosion, special sensors to detect these limits are required.

As mentioned the first phase will be teleoperated with the second phase autonomously operated. This will involve image acquisition and analysis coupled with an expert system to initiate, direct and terminate the disposal operation. The expert system to be used will be that developed by the Electronics, Technology and Devices Laboratory (EDTL) and is described in a paper presented at this conference - Expert Systems in the Tactical Environment.

This application demonstrates the requirements for successful design and implementation of an intelligent automated system. Rather than retrofit an off-the-shelf robot to perform a task presently done by humans, the entire procedure is modified for automation. The co-mingling of computer cognition (image analysis and acquisition, process sensors), mechatronics (chopstick robot, SP control system) and intelligence (ETDL expert system) are integrated into a system to perform a specific function. However, the concepts of an open system are maintained such that this assemblage is easily adaptable to a variety of construction processes and AMC armaments applications.

b. Project REARM Intelligent Robotics

The Rock Island Arsenal in Rock Island, Illinois, has established two robotic welding stations utilizing two CYBOTECK 1200 robots. Presently, the welding stations utilize preprogrammed trajectory weld operations with conventional power source control for voltage, current, wirefeed speed, etc. The Arsenal, TACOM, CERL and the Idaho National Laboratory have embarked on a program to develop an intelligent, open system for multitask welding capability to accommodate various base metals, welding procedures and geometries. This task differs from the EOD work in that retrofitting an intelligent system to an existing operation is involved. Since the

control elements for the CYBOTECK robots are closely matched to the mechanical characteristics of the robot, the existing control system becomes the primitive level in the intelligent hierarchy. Coupling the intelligent system to this extant control system is simplified somewhat by the inherent feature of the system mentioned in Section 6.

This application also illustrates the differentiation between computer cognition and an assemblage of multiple sensors. Because this is primarily armor steel welding, hydrogen contamination is the most serious malefactor; this causes embrittlement and cracking in this type of steel. Therefore, an opto-electronic sensor developed by CERL to detect hydrogen in the arc will be utilized, however, it will not be a part of the cognition system; the INL sensors and the CERL arc imaging sensors will be the primary cognition input.

c. Automated Forging and Swaging Operations

Metallurgical processing of alloys for special weapons are complicated by drastic material property changes due to temperature changes during the processing. Visual acuity developed by experienced operators is the principle quality control method used in the manual control of these operations. The Materials Technology Laboratory in Watertown, Massachusetts, obtained a PRAB robot and a General Electric control system to develop a prototype system to automate the swaging and forging operations for these special alloys. CERL was called upon to assist in this project with the task to incorporate appropriate cognition and intelligence to successfully accomplish the objective. Although CERL developed an isotherm imaging and analysis system for this project, the metallurgical knowledge base to utilize these measurands was lacking. As in the automated welding work this posed a serious impediment to the successful accomplishment of this objective. As in the project REARM application, the primitive control system was in place and the intelligent system would be superimposed. Although the mechanization of the PRAB robots could be accomplished with the hybrid configuration, another factor expanded the complexity of the system. Operation of the forging and swaging presses depended entirely on the heuristics of the operator so that complete automatic operation would not be possible without duplicating these heuristics in the overall system. It is proposed to instrument the presses and utilize an inductive learning system to replicate the human function.

8. Summary

In the April 1954 issue of Control Engineering the lead article was entitled: "Is it Adaptive." These were the initial years for a superior control scheme, i.e. adaptive control. It would appear that this is an appropriate time for the query, "Is it Intelligent," to become prominent

in planning automation. A philosophy which leads to the proliferation of facile but unintelligent automata will be deleterious not for just segments of industries as occurred in the early 70s and early 80s, but for the two principal industries in this country - defense and construction. The need for intelligent automation is obvious - stamina (24 hour operation, no illness), flexibility (multi-tasking), robust (strength equivalent to more than one human) - all attributes that will contribute to considerable productivity improvement. It is hoped that this paper will indicate some points of departure for planning intelligent systems.

FOOTNOTES

[1]The concepts of inventive research and innovative research were initially suggested by the Technical Director of CERL, Dr. L. R. Shaffer to more aptly connote the scope and mission of CERL R&D rather than traditional "basic" or "applied" research terminology which implies a high degree of specialization which is incongruent with cross-discipline R&D.

[2]There is a paper entitled "Expert Systems In the Tactical Environment" being presented at this conference by Dr. R. F. Miller of the Electronics and Devices Laboratory (ETDL) that uses Bayesian statistical methods in an expert system.

CITED REFERENCES

(1)Kokotovic, P. V. "The Sensitivity Point Method in the Investigation and Optimization of Linear Control Systems," in J. B. Cruz, Jr., ed., System Sensitivity Analysis. Stroudsberg, PA: Dowden, Hutchinson and Ross, Inc., 1973.

(2)Albus, James S. Brains, Behavior, and Robotics. Peterborough, NH: McGraw-Hill, 1981.

GENERAL REFERENCES

Barr, A. and E. Feigenbaum. The Handbook of Artificial Intelligence. Los Altos, CA: William Kaufmann, 1981-82. Three-vol. compendium of techniques and systems.

Buchanan, B. G. and E .H. Shortliffe. Rule Based Expert Systems. Reading, Mass.: Addison-Wesley, 1984.

Evans, G. M. and F. Weyland. "Diffusible Hydrogen as a Criterion of the Quality of Welding Additives," Deutsch. Ver. fur Schweisstehnik, No. 50, p. 21-33, 1978.

Gardner, C. S. and F. Kearney. "Electro-Optic System for Non-Destructive Testing of Field Welds," Proc. Conference on Developments in Welding Processes and Consumables, to be published.

Hayes-Roth, Frederick. Building Expert Systems. Reading, MA: Addison-Wesley, 1983.

Kearney, F. "Nondestructive Testing for Field Welds: Real Time Weld Quality Monitor - Field Tests," CERL Technical Report M-295, June 1981.

Kokotovic, P. V., J. V. Medanic, M. I. Vuskovic and S. P. Bingulac. "Sensitivity Method in the Experimental Design of Adaptive Control Systems," Proceedings of the Third Congress of the International Federation of Automatic Control, London (London: Institution of Mechanical Engineers, 1966), p. 45B1-45B.12.

Lundin, C. P. "The Significance of Weld Discontinuities - A Review of the Current Literature," WRC Bulletin 222, December 1976.

Norris, M. E. "Microprocessor Controlled Weld Arc Spectrum Analyzer for Quality Control and Analysis," CERL Technical Manuscript M-317, June 1982.

Raphael, B. The Thinking Computer: Mind Inside Matter. San Francisco: W. H. Freeman, 1976.

Riedle, B. D. and P. V. Kokotovic. "Integral Manifolds of Slow Adaption," IEEE Trans. on Automatic Control, Vol. AC-31, No. 4, p. 316-324, April 1986.

Winston, P. H. Artificial Intelligence. Reading, MA: Addison-Wesley, 1977.

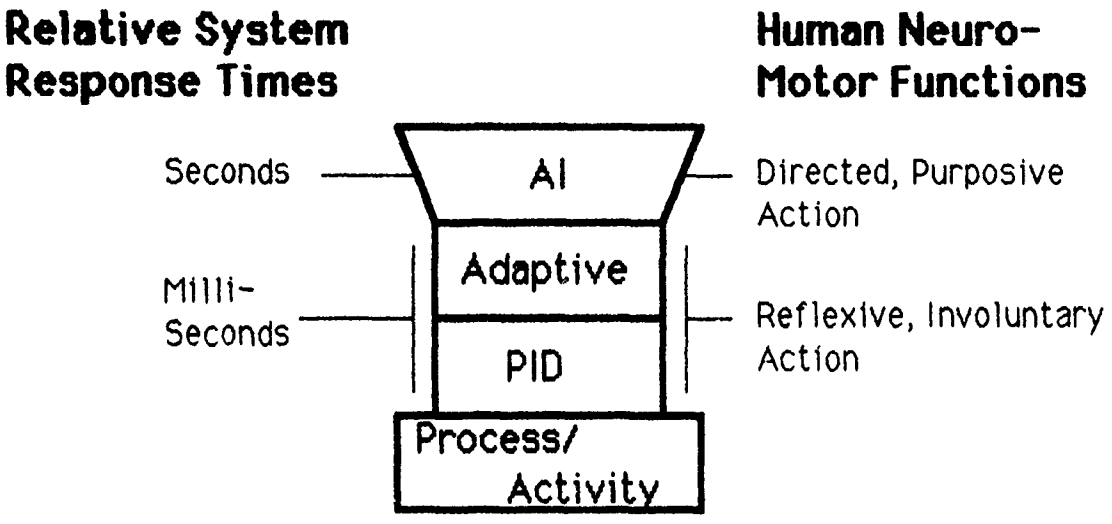


Figure 1 Human Analogues for AI/Control Interactions

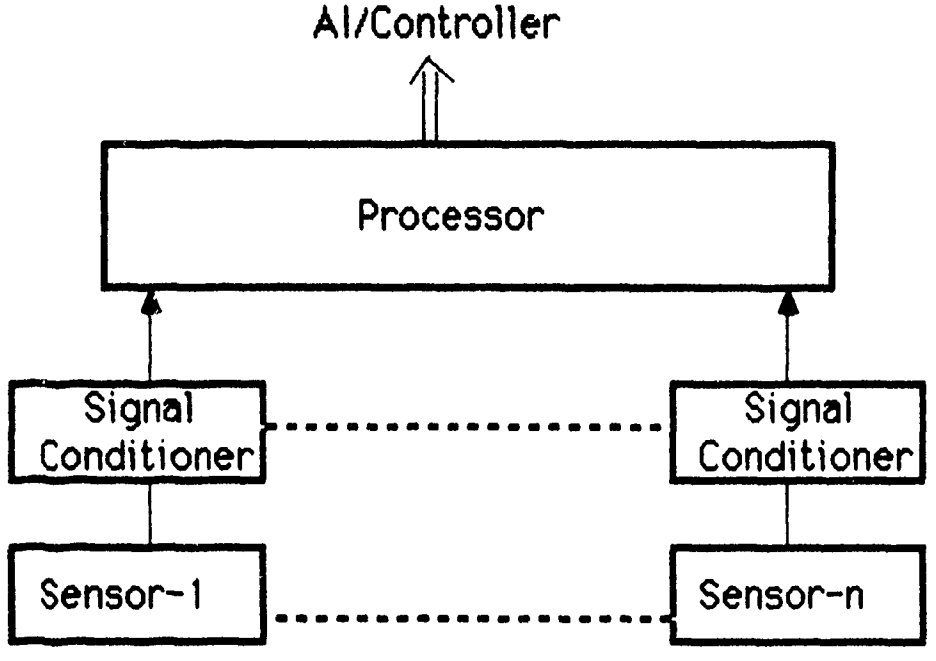
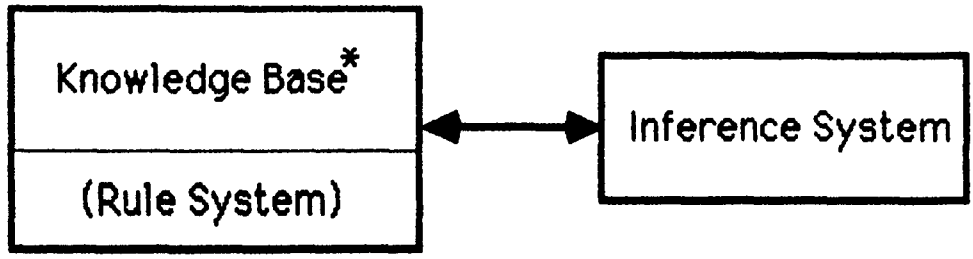


Figure 2 Multiple Sensor Configuration for Computer Cognition System; Output Function is the Distinguising Parameter.



*Knowledge = Facts + Experience

Figure 3. Expert System Topology

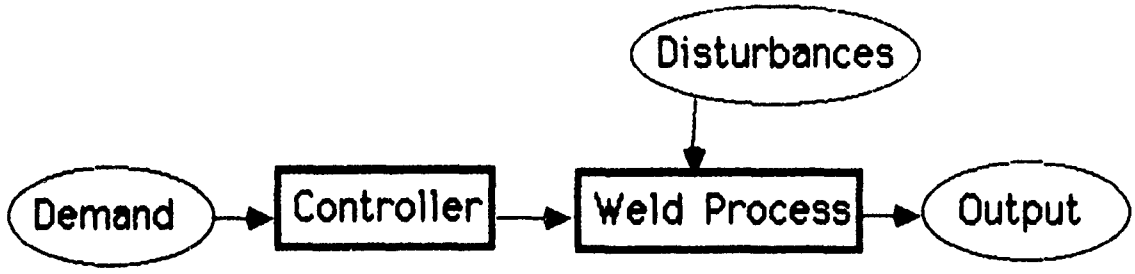


Figure 4. Open Loop Control System.

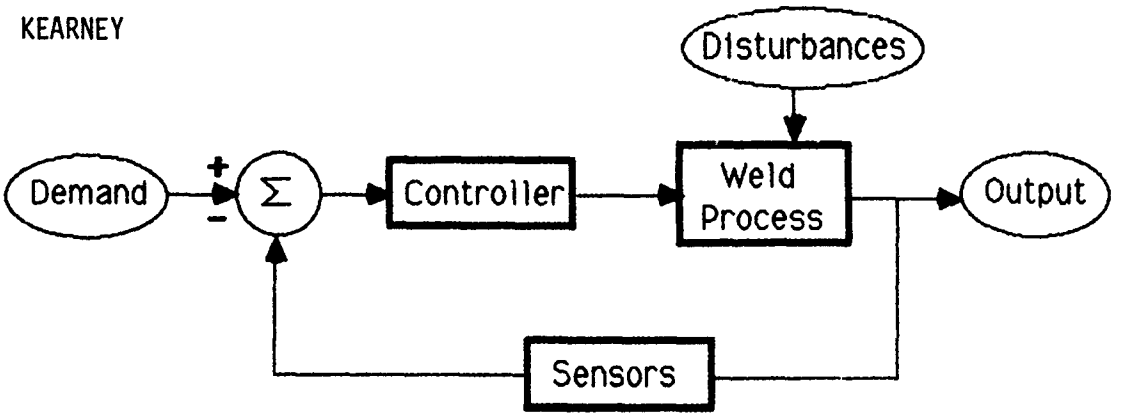


Figure 5. Closed Loop Control System.

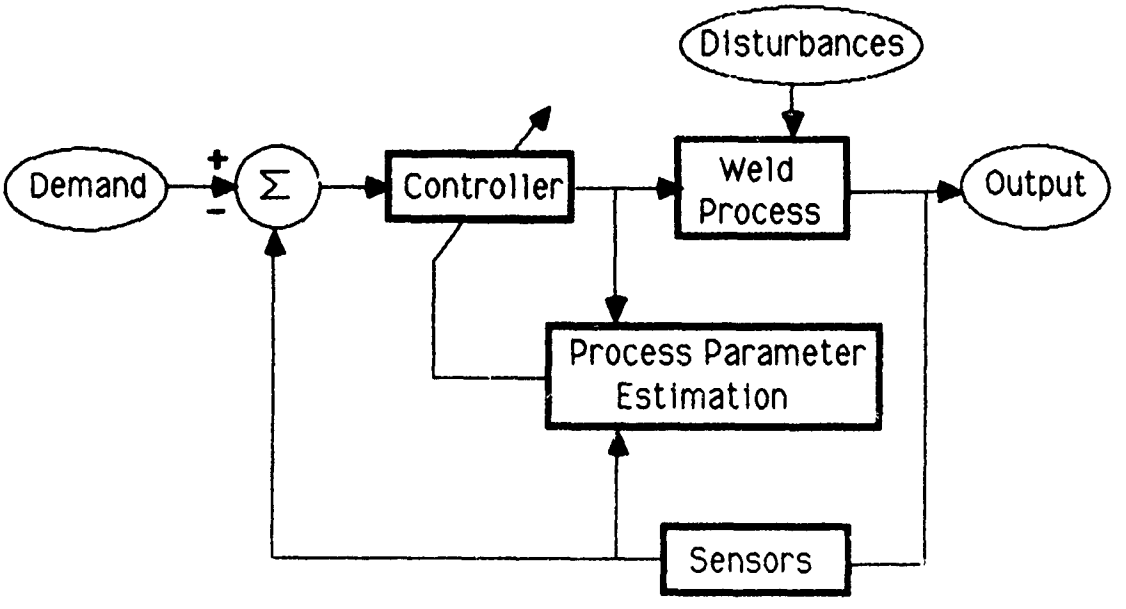


Figure 6. Adaptive Control System.

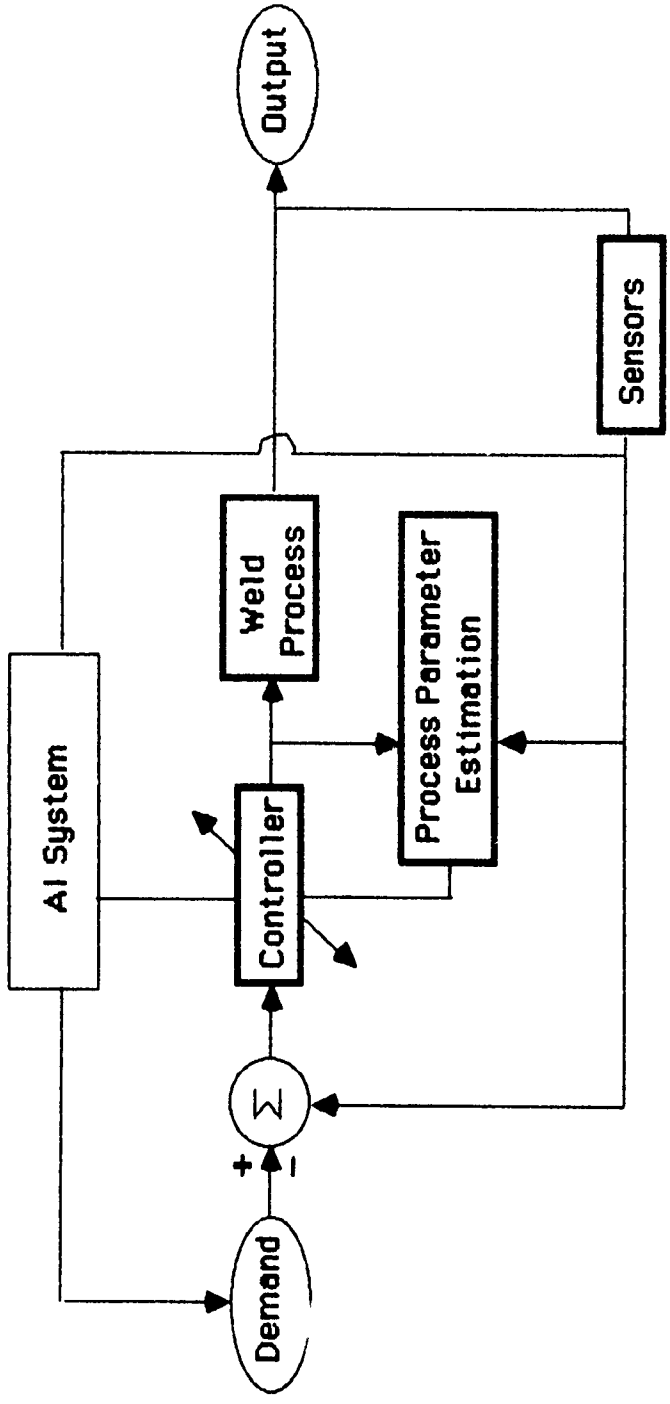


Figure 7. The Merger of an Adaptive Control System and an Intelligent Supervisory System

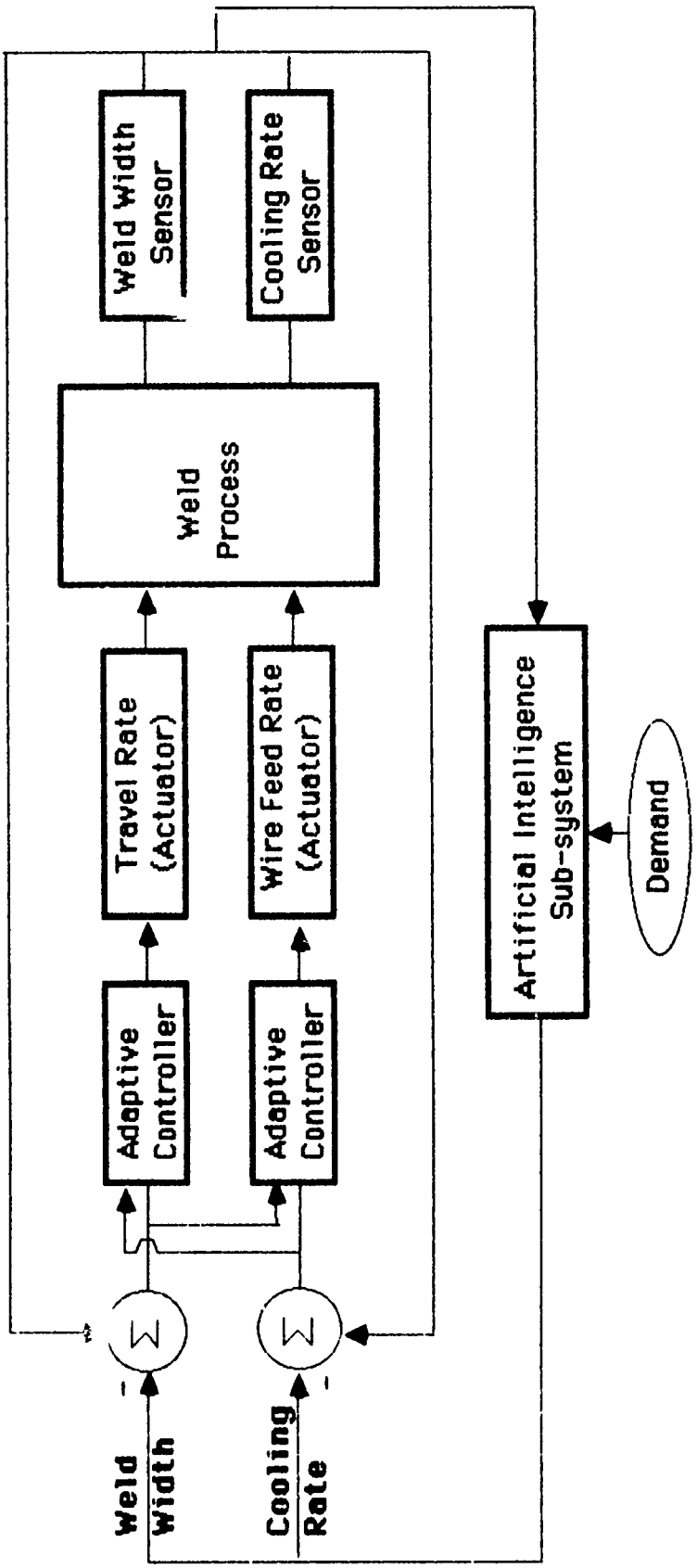


Figure 8. CERL Intelligent Welding System.

The Sergeant Major Study: Health Risk Assessment By Clinical Laboratory Parameters

Richard C. Keniston, MAJ, MC*
Michael R. Weir, COL, MC**
John Enriquez+
Fred Duncan#
Departments of Pathology* and
Clinical Investigation+, and
Health Fitness Center#,
William Beaumont Army Medical Center
El Paso, Texas 79920-5001
and Department of Pediatrics**,
Madigan Army Medical Center
Tacoma, Washington 98431-5406

The Army is concerned about the health of its military and civilian employees, and also about the health of dependents and retirees [1,2]. Physical and mental fitness is stressed, with an emphasis on educational programs and activities that promote good health habits. Part of this program includes an attempt to identify individuals at risk for health problems, before those health problems become manifest. This is the aim of the periodic physical examination, and particularly of the over 40 physical.

As detailed in AR 600-63, the Health Risk Appraisal (HRA) is another instrument designed to identify individuals at increased risk. The HRA includes the drawing of blood for glucose (GLU) and total cholesterol (CHO). However, from a statistician's point of view, neither of these laboratory tests is very sensitive or specific for identifying individuals with diabetes or cardiovascular disease. For example, using a critical value for CHO of 224 mg/dL (Lipid Research Clinics, LRC, method), CHO has a sensitivity of only 63% and a specificity of only 67% in diagnosing cardiovascular disease [3]. We have observed that most individuals who die of whatever cause, including cardiovascular disease, have normal or reduced CHO levels. Also, particularly in the younger activity duty individuals and in teenage dependents, gamma glutamyl transferase (GGT) levels, which rise in those who chronically abuse ethanol, are more likely to identify individuals at risk than either CHO or GLU.

We use the Technicon SMAC, a multichannel biochemical analyzer, to measure CHO and GLU for the HRA. At the same time, therefore, we also obtain a value for GGT, and for 19 other parameters, in addition to the CHO

and GLU values. We wondered if we could improve the sensitivity and specifity of the cardiovascular disease risk assessment using some of these other parameters (such as fasting total triglycerides (TRI), in conjunction with an HDL cholesterol (HDL)). Also, we reasoned that we could identify individuals at risk for other conditions, such as malnutrition, liver disease, or cancer using different combinations of the already measured parameters.

Finally, we have read and personally observed that smokers, heavy drinkers, and patients with a wide variety of diseases, including cardiovascular disease and cancer, have decreased levels of vitamin B6 [4-8]. So we decided to include plasma pyridoxal 5'-phosphate (PLP) in our study (PLP is the active cofactor form of vitamin B6).

Smokers and heavy drinkers have higher levels of cardiovascular disease, cancer, and all cause mortality than nonsmokers or light drinkers [9]. Therefore, we reasoned that those parameters which best correlated with smoking status and heavy alcohol consumption would also best correlate with disease risk and mortality risk. With these principles in mind, we developed optimal ranges for each parameter measured, and then used these optimal ranges and the known pathophysiology of each type of disease to develop a profiling system, which we hoped ultimately would be completely objective, that is, based only on blood parameters. In a subset of the subjects, we also evaluated the complete blood count (CBC) parameters, in addition to the 22 SMAC serum chemistries, serum HDL, and plasma PLP.

METHODS

Study Population: The subjects were sergeant major candidates from three classes (28-86, 29-87, and 30-87), of the USA Sergeants Major Academy, Biggs Army Airfield, El Paso, Texas. This included 757 males and 17 females. Mean and median age was 39 yrs, with a standard deviatiion of 3.0 yrs. These individuals were already mandated to have the HRA.

Subjective Data: In addition to the HRA questionnaire, a second questionnaire was handed out at the beginning and near the end of each class (approximately 5 months apart). The following information was obtained: age, sex, race, smoking habits (CIGS/day), drinking habits (drinks/mo.,ALC), vitamin supplementation (mg/day of vitamin B6, B6S), medications, medical diagnoses, profiles, exercise history, dieting, and fasting status.

Objective Data: Blood was drawn from each individual at the beginning and the end of each class. For all 773 participants, the following tests (with abbreviations and units) were run: sodium (SOD, mEq/L), potassium (POT, mEq/L), chloride (CHL, mEq/L), total carbon dioxide (CO2, mEq/L), glucose (GLU, mg/dL), blood urea nitrogen (BUN, mg/dL), creatinine (CRE, mg/dL), uric acid (UA, mg/dL), total protein (PRO, gm/L), albumin (ALB, gm/L), alkaline phosphatase (ALP, IU/L), GGT (IU/L), lactic dehydrogenase (LDH, IU/L), aspartate aminotransferase (AST, IU/L), calcium (CAL, mg/dL), inorganic phosphorus (PO4, mg/dL), CHO (mg/dL), TRI (mg/dL), high density

lipoprotein cholesterol (HDL, mg/dL), and vitamin B6 (PLP, nmoles/L). In addition, 108 members of Class 28.1 and 257 members of Class 30.1 had complete blood counts (CBC) done, which included white blood cell count (WBC), red blood cell count (RBC), hemoglobin (HGB, gm/dL), hematocrit (HCT,%), and platelet count (PLT). All members of Classes 29 and 30 also had serum iron (IRO, mcg/dL) and alanine aminotransferase (ALT, IU/L) levels done.

Calculated Parameters: The following calculated parameters were also evaluated: CHO/HDL Ratio (HDLR), very low density lipoprotein cholesterol (VLDL = 0.16xTRI, VLDL always less than 80), low density lipoprotein cholesterol (LDL = CHO - VLDL - HDL) , LDL/HDL Ratio (LDLR), LDH/ALB Ratio (LDHR), vitamin B6 factor (B6F = PLP - (34 + (4xB6S))), and logPLP (LPLP).

Determining Optimal Ranges for Each Parameter: Literature [3,10-15] and the authors' clinical experience were used to assign optimal ranges. For many parameters, the optimal range centered on the reference range. For example, for POT, the optimal range was 4.0-4.5 mEq/L, and the reference range is 3.5-5.0 mE/L. For other parameters, such as CHO, ALB, and PLP, the optimal range was considerably shifted from the center of the reference range, because significant proportions of individuals with values within the reference range nevertheless manifested significant health problems.

Preparation of Profiling Instrument: First, the data for each parameter from Class 28 (and later, Classes 29 and 30 as well) were broken down by smoking status (CIGS) and by drinking status (ALC). The means, standard deviations, and distributions of each parameter were compared, and discriminant analyses were done, using CIGS and ALC as the dependent parameters. This process identified those parameters which best discriminated between smokers and nonsmokers and drinkers and nondrinkers. Since smokers have considerably higher illness rates and mortality rates than nonsmokers, and the same is true for heavy drinking (9), we selected these parameters as the best ones to use in setting up the profiling system.

For cardiovascular risk, we chose CHO, TRI, HDL, HDLR, PLP, and added factors (which we expect ultimately to be able to eliminate) for CIGS and ALC. For alcohol-liver disease, we chose ALC, GGT, AST, ALP, and PLP. For diabetes risk, we chose GLU, with a caveat about fasting and post-collection-to-processing interval. For cancer risk, we emphasized CIGS, but also included PLP, ALB, CHO (low CHO), and LDH; hopefully, we will ultimately be able to substitute other objective blood parameters for CIGS. For nutrition, we chose ALB, TRI, BUN, CHO, and PLP.

We were well aware that some of the CBC parameters, such as WBC, might considerably improve our risk assessment [16] .

Estimating Ten-Year Mortality Risk: The male candidates were grouped into 20 subgroups, based on reported smoking and drinking status. For smoking status, there were four categories: never smoked (NS), exsmoker (ES), light smoker (LS, 20 or fewer CIGS/day), and heavy smoker (HS, more than 20 CIGS/

day). For drinking status, there were five categories: nondrinker (ND,=1), occasional drinker (OD,=2, one or two drinks/mo.), light drinker (LD,=3, 3 to 10 drinks/mo.), moderate drinker (MD,=4, 11 to 30 drinkers/mo.), and heavy drinker (HD,=5, more than 30 drinks/mo.).

Using the 480 males from the first two classes, each member of each of the 20 subgroups was assigned a 10-year mortality risk (MORT RISK) based on literature values for similar aged men [9]. A discriminant analysis was then performed, using only blood parameters, plus B6S. This gave an idea as to which parameters were most important in overall mortality risk.

Use of the Profiling Program: All subjects who had complete data sets were given profile reports, which included recommendations. If the value for a given parameter fell within the optimal range, that individual was scored a "5" for that parameter. Any value outside the optimal range was assigned a score of from "1" (for extremely low values) to "9" (for extremely high values). The individual disease risk scores were in turn derived from the composite of the scores of the individual component parameters. A nutrition score less than 4.0 indicated a serious problem with nutrition, while any other disease risk score greater than 6.0 indicated an increased risk for the disease in question. A risk score greater than 7.5 indicated a serious health problem, requiring prompt medical attention.

Testing the Profiling System: Individuals who manifested a serious health problem during the course of the study, such as a heart attack (AMI), or who were picked up for driving while intoxicated (DWI), were noted, and the ability of the individual's profile to predict the problem was evaluated.

Effect of Vitamin B6 Supplementation (B6S) on Serum Lipid Profiles: Volunteers who had low initial plasma PLP (less than 15 nmoles/L) were given 50 mg/day of pyridoxine HCl. Changes in CHO, HDL, and HDLR were noted.

RESULTS

Effect of Smoking, Drinking, and Vitamin B6 Supplementation on Clinical Laboratory Parameters: The first table illustrates the mean values for certain key parameters for six groups of subjects, separated first on the the basis of smoking status, second on the basis of alcohol status, and third on the basis of initial vitamin B6 supplementation. Ideally, PLP, HDL, and ALB should be as high as possible, within certain limits, and CHO, TRI, LDL, HDLR, ALP, and WBC (not illustrated) should be as low as possible, again within certain limits. In all categories, the Smokers have the worst biochemical profile, while the Nonsmokers and the Vitamin B6-Supplemented Group (which includes many smokers) have the best biochemical profiles.

The results of the discriminant analyses for Smoking status and for Drinking status are shown in Tables 2 and 3, respectively. When CBC data was not included, and only unsupplemented males were used in the analysis,

the key discriminant parameters, in order of correlation with the discriminant function, were PLP, HDLR, ALC, ALB, AST, LDH, and HDL. The derived function classified approximately 75% of the unsupplemented males correctly by Smoking status, but misclassified most of the vitamin B6-supplemented smokers (who had on the average much better lipid profiles than the non-supplemented smokers) as nonsmokers.

When the CBC parameters were included in the analysis, and all males, supplemented and unsupplemented, from Class 30 were used for the analysis, the key discriminant parameters for Smoking status, in order of correlation with the discriminant function, were WBC, CO2, B6F, RBC, PO4, BUN, LDH, CHO, UA, ALC, B6S, LDHR, HCT, CHL, and GLU. The derived function correctly classified 90% of the Class 30 males by Smoking status, and also correctly classified 81% of the subjects from Class 28, including 88% of the females.

As far as Drinking status was concerned, discriminant analysis did not do as well as just guessing that everyone was a drinker. However, all of the heavy drinkers were correctly classified as drinkers, and most of the drinkers misclassified as nondrinkers were occasional or light drinkers. Using unsupplemented males only, and no CBC parameters, the key discriminant parameters for Drinking status, in order of correlation with the discriminant function, were HDL, CIGS, GGT, PLP, LDH, and ALP. Only 62% of the unsupplemented males were correctly classified by Drinking status, and 75% of the vitamin B6-supplemented nondrinkers were misclassified as drinkers (again in part because of their much better lipid status).

Using all males from Class 30, and including the CBC parameters, the following were the key discriminant parameters for Drinking status, in order of correlation with the discriminant function: PLP, GGT, CRE, GLU, LPLP, PO4, IRO, ALB, and SOD. Of the Class 30 males, 69% were correctly classified by this function, including over 90% of the heavy drinkers. Only 63% of the individuals in Class 28 with a CBC were correctly classified, including 88% of the females.

Calculated Ten-Year Mortality Risk (MORT RISK): The following equation for MORT RISK was derived, using all 480 males in Classes 28 and 29:

(1) MORT RISK (%) = 11.31 - 0.015xPLP + 0.258xHDLR + 0.042xB6S - 0.131xALB + 0.012xGGT

Table 4 shows these 480 males broken down by deciles of MORT RISK, with the means for each of the discriminant parameters by decile, as well as the percentage of individuals within each decile who were smokers. Table 5 breaks the subjects in Table 4 down by smoking and drinking status, and and gives the number and percent of each group that are at Low, Moderate, and High Overall MORT RISK. The Nonsmokers who were not heavy drinkers were mostly at low risk, while the Smokers and Heavy Drinkers were mostly at moderate to high risk.

Another Ten-Year MORT RISK was also derived, for use by those who do not have an assay for PLP available; it includes the subjective parameters CIGS and ALC:

(2) MORT RISK (%) = 3.559 + 0.192xCIGS + 0.011xGGT - 0.174xALC + 0.079xLDLR

The Health Risk Assessment Profile (HRAP): A copy of the HRAP is shown in Figure 1. This is a copy of an actual report on one of the candidates who had a heart attack, and who would have been missed by the HRA CHO screen; his CHO was 215 mg/dL. This report includes the name, a Nutrition Score, four disease risk scores, and an Overall Health Assessment Score (not shown in Figure), followed by an explanation of the numerical profiling system (1 to 9, with "5" being optimal). Then there is a Percent 10-Year MORT RISK (which averages about 6.0%), followed by the profile score for each of the five key discriminant parameters (CHO and HDL are the components of the HDLR). Next, recommendations are given for improving one's health, based on the individual's profile. Finally, the F1 value classifies the individual by most likely Smoking status, and the F2 value classifies the individual by most likely Drinking status, based on the discriminant functions previously discussed. Each individual is also given a group HRAP, similar to Tables 4 and 5, so that he can see how his own profile compares with that of his peers.

Individuals With Serious Health Problems: Correlation with HRAP and Conventional Screen (HRA): There were nine individuals who manifested serious health problems during (or in one case shortly before) the study. All five individuals with cardiovascular problems would have been picked up by the HRAP, whereas two would have been missed on the basis of routine CHO measurements (CHO = 183 and 215; the former had type IV hyperlipidemia, with a very low HDL and angina, and the latter has already been discussed in the preceding paragraph). Two of three subjects picked up for DWI were identified as being at increased risk for alcohol-liver disease on the HRAP. These individuals would have had to report that they were heavy drinkers on the HRA questionnaire to have been picked up by that screen; in this study, five heavy drinkers stated on the questionnaires that they were nondrinkers. The one individual with confirmed diabetes mellitus was identified by both the HRA and the HRAP, since both rely on the serum GLU, and his GLU was markedly elevated.

Problems With Glucose and Cholesterol Measurements: Figure 2 shows the trends in the medians for four lipid parameters, GLU, and GGT over time. From the first blood draw on Class 28 (July 1986) to the last blood draw on Class 30 (Dec 1987), there were significant trends for three of the lipid parameters. CHO and LDL rose, and HDL fell. We believe that these trends probably represent systematic errors over time in the analysis technique, rather than indicating that Class 30 averaged less healthy than Class 28. In each of the three classes, the second median TRI was significantly higher than the first, presumably due to a higher percentage of non-fasting prior to the second draw. There were no trends for GGT. The trend with GLU was especially interesting: classes drawn in the summer (28.1, 29.9, 30.1) had lower median GLU values than the corresponding classes (28.9, 29.1, 30.9) drawn in winter. This led to the failure to

detect four individuals with borderline diabetes in two of the classes drawn in summer.

Figure 3 compares the CHO values by percentile for the Lipid Research Clinics (LRC) study [17] and for this study. The Technicon SMAC method consistently gives higher values for CHO at all percentiles [18]. Therefore, we could not use the suggested LRC cutoff values, but rather had to develop our own cutoff values, or else convert our values to the LRC method values:

(3) LRC CHO = 0.842xSMAC CHO + 15.672

We decided to do the former (develop our own cutoff value, at 240 mg/dL). Using an HDLR of greater than 5.0 as being indicative of latent cardiovascular disease in those not actually manifesting it, we came up with a sensitivity of 70% and a specificity of 94% for an isolated CHO of 240 mg/dL or greater. This misclassified only 27 people as being at increased risk for cardiovascular disease who probably are not at increased risk (have HDLR of less than 4.0), including one of the authors, who had one of the highest HDL values in the study, and no other risk factors for heart disease. On the other hand, 89 individuals with CHO less than 240 mg/dL either had an HDLR greater than 5.0, or, in one case, even had a heart attack, and would have been missed using this cutoff value. The best combination of sensitivity and specificity (85% and 87%, respectively) occurred at a CHO of 220 mg/dL.

Effect of Vitamin B6 Supplementation on Lipid Parameters: Seventy-eight subjects started taking pyridoxine HCl, at 2 to 50 mg/day. While there was no significant change in mean CHO, HDL, or HDLR values with B6S, there was a significant decrease in the variance of the HDLR (p less than 0.05), and the mean values all trended toward more optimal values. Those with an HDLR value greater than 4.0 tended to have a falling HDLR with B6S, and those with an HDLR less than 3.0 tended to have an increasing HDLR with B6S. We called this phenomenon "collapse in the variance", and feel that this may represent an equilibrium-seeking phenomenon. CHO closed in on a value of 217 mg/dL, which was very close to the value giving lowest all-cause mortality in our five-year study here at WBAMC. Also, as shown in Figure 4, there was a highly significant inverse relationship between change in PLP and change in HDLR (p less than 0.005).

DISCUSSION

A new profiling instrument, the Health Risk Assessment Profile (HRAP), is introduced, which hopefully will be more reliable than the current Health Risk Appraisal (HRA). Specifically, the HRAP depends primarily upon clinical laboratory parameters, rather than a questionnaire, to arrive at a series of health risk scores. The optimal score for any parameter is a "5", and anything greater than "6" or less than "4" indicates a potential health problem. Similarly, the optimal risk score is a "5", and any risk score greater than "6.0" or less than "4.0" (for Nutrition Score) indicates

a potentially serious health problem. Risk scores are numerical, and thus also correlate with severity of risk, those greater than "7.5" being considered indicative of very high risk in that disease category. Recommendations are provided to assist in managing any health problem identified.

Advantages of the new HRAP system include the following: (1) More clinical, objective information is considered (multiple parameters instead of one relatively insensitive and nonspecific parameter, CHO, for cardiovascular disease risk); (2) Risk for other potential problems (poor nutrition, alcohol-liver disease, cancer, and overall health risk) are included; (3) A ten-year MORT RISK is provided, which should improve if therapy is initiated and that therapy is effective, and has no adverse side effects; (4) Specific, individualized recommendations are given; (5) The individual who has been profiled receives an individual report, which alerts him to his own individual health problems, and reassures him if he has been doing well; and (6) A group report is also provided, which shows the individual where he stands in relationship to his peers in overall health risk, and which emphasizes just how seriously habits such as smoking and heavy drinking can influence his health. Another advantage is that the numerical profiling system is easy for an untrained screener to use: without any knowledge whatsoever of medicine, a person can recognize that a "9" or a "1" requires prompt medical followup.

Disadvantages of the HRAP include: (1) More tests must be run, including an HDL and, ideally, a PLP and a CBC, which requires more time, money, and personnel; (2) Certain test results, such as GLU and TRI, still depend on proper fasting (at least 12 hrs); (3) Certain test results, such as GLU and CO2, may be inaccurate if the day is hot or there is a delay in processing the blood sample; (4) Currently, some subjective input, such as CIGS and ALC, is still required, although eventually, when a CBC is incorporated into the HRAP, this may not be necessary; (5) The precise sensitivity and specificity of each component of the HRAP, like that of the HRA, is still not known, and will require a long-term study; and (6) It takes considerable time and effort to prepare each HRAP report.

The HRAP approach attempts to look at the basic biochemical abnormalities that underlie each individual's health problems, rather than "risk factors" or "group risk". For example, a particular heavy smoker may be biochemically just as healthy as the average nonsmoker, while another individual with none of the conventional risk factors may be at very high risk (e.g., some of the nonsmokers with type IV hyperlipidemia, who have completely normal CHO levels). Neglected parameters, as PLP, GGT, or WBC, which seem to have considerable prognostic significance [5,6,16], are also to be included.

In summary, we would like to say that the Health Risk Assessment Profile system is still evolving. We are still looking at a variety of problems, and are attempting to optimize and simplify our approach, as much as possible. We are particularly concerned that the current "conventional wisdom" approach may do as much harm as good. The current emphasis

on CHO in cardiovascular disease risk seems to have diverted attention from the role of poor nutrition in all diseases [4,5,19-22]. We have noted, for example, that none of the widely-quoted CHO-lowering studies [23-25] have resulted in a decrease in overall mortality rates, mortality due most probably to treatment-induced complications rising to meet any decrease in mortality due to ischemic heart disease [26]. Significantly, all of the conventional therapies for reducing high cardiovascular risk (diet, anti-hypertensive medications, lipid-lowering drugs, aspirin) have resulted in a significant decrease in plasma PLP (vitamin B6) levels, and no overall decrease in Ten-Year MORT RISK. We would like to see an overall improvement in the health of the Army personnel, not just a shift in the causes of morbidity and mortality. We feel that well thought-out studies and profiling instruments such as the HRAP may be important in achieving this objective.

Due to space limitations, we have not been able to include a copy of the HRAP computer program in this report, but a copy is available from the senior author.

REFERENCES

1. Family Fitness Handbook, DA Pamphlet 350-21, November 1984.

2. Army Health Promotion, AR 600-63, November 1987.

3. Naito HK: The clinical significance of apolipoprotein measurements, J Clin Immunol 9:11-20, 1986

4. Gruberg ER, Raymond SA: Beyond Cholesterol: Vitamin B6, Arteriosclerosis and Your Heart, N.Y.; St. Martin's Press, 1981.

5. Keniston RC, Becker W, Reyna T, Enriquez J: Prognostic significance of undeproteinized plasma pyridoxal 5'-phosphate, In: Clinical and Physiological Applications of Vitamin B6, J Leklem, R Reynolds, eds. N.Y.: AR Liss, 1988 (In press).

6. Campbell RA, Keniston RC: Hyperpolyaminemia and vitamin B6, In: Vitamin B6: Its Role in Health and Disease, R Reynolds, J Leklem, eds. N.Y.: AR Liss, 1985, pp 347-386.

7. Serfontein WJ, Ubbink JB, DeVillers LS, Becker PJ: Depressed plasma pyridoxal 5'-phosphate levels in tobacco-smoking men, Atherosclerosis 59:341-346, 1986.

8. Lumeng L, Li TK: Vitamin B6 metabolism in chronic alcohol abuse. Pyridoxal 5'-phosphate levels in plasma and effect of acetaldehyde on pyridoxal phosphate synthesis and degradation in human erythrocytes, J Clin Invest 53:693-704, 1974.

9. Klatsky AL, Friedman GD, Siegelaub AB: Alcohol and mortality, a ten-year Kaiser-Permanente experience, Ann Int Med 95:139-145, 1981.

10. Solberg HE: Establishment and use of reference values, In: Textbook of Clinical Chemistry, NW Tietz, ed. Philadelphia: WB Saunders Co, 1986, pp 356-386.

11. Dales LG, Friedman GD, Siegelaub AB, Seltzer CC: Cigarette smoking and serum chemistry tests, J Chron Dis 27:293-307, 1974.

12. Ryback RS, Eckardt MJ, Flesche B, et al: Biochemical and hematological correlates of alcoholism and liver disease, JAMA 248:2261-2165, 1982.

13. Friedman RB, Anderson RE, Entine SM, Hirshberg SB: Effects of diseases on clinical laboratory tests, Clin Chem 26:1D-476D, 1980.

14. Product Labelling for the Technicon SMAC High Speed Computer-Controlled Biochemical Analyzer, Vol. II. Methods. Tech. Pub. UA3-030B32, September, 1976.

15. Hamfelt A: Age variation of vitamin B6 metabolism in man, Clin Chim Acta 10:48-54, 1964.

16. Grimm RH, Neaton JD, Ludwig W: Prognostic importance of the white blood cell count for coronary, cancer, and all-cause mortality, JAMA 254:1932 -1937, 1985.

17. Coronary Risk Handbook, American Heart Association. Estimating Risk of Coronary Heart Disease in Daily Practice. 1973.

18. Blank DW, Hoeg JM, et al: The method of determination must be considered in interpeting blood cholesterol levels, JAMA 256:2867-2870, 1986.

19. Eaton SB, Konner M: Paleolithic nutrition: a consideration of its nature and current implications, N Engl J Med 312:283-289, 1985.

20. Engstrom AM, Tobelmann RC: Nutritional consequences of reducing sodium intake, Ann Int Med 98:870-872, 1983.

21. Cohen LA: Diet and cancer, Sci Amer 257:42-48, 1987.

22. Roe DA: Drug-Induced Nutritional Deficiencies, Westport, Conn: AVI Publishing Co, Inc, 1978.

23. Frick MH, Elo E, et al: Helsinki heart study: Primary prevention trial with gemfibrozil in middle-aged men with dyslipidemia, N Engl J Med 317:1237-1245, 1987.

24. The Lipid Research Clinics coronary primary prevention trial results. I. Reduction in incidence of coronary heart disease, JAMA 251:351-364, 1984.

25. Oliver MF: Serum cholesterol - the knave of hearts and the joker, Lancet 14 Nov 1981, pp 1090-1095.

26. Pinckney ER: Statistical analysis of Lipid Research Clinics program, Lancet 14 Feb 1981, p 249.

Table 1. Mean Values for Main Discriminant Parameters by Smoking, Alcohol, and Vitamin B6 Supplementation Status

	Parameter	Nonsmokers	Smokers	Nondrinkers	Drinkers	No B6S	B6S +
1	Number	396	334	143	580	564	162
2	Age	39	39	39	39	39	39
3	Cigarettes	0	25	8	12	12	8
4	Alcohol	2	3	1	3	3	3
5	B6 Supple...	3	2	4	2	0	12
6	PLP	63	37	58	50	35	109
7	Cholesterol	216	228	220	223	223	219
8	Triglycerides	110	128	113	120	120	113
9	HDL	53	50	49	53	51	54
10	LDL	141	152	148	146	148	142
11	HDLRH100	408	456	449	421	437	406
12	GGT	26	28	20	29	27	28
13	AST	23	22	22	23	22	23
14	LDH	157	154	159	156	157	153
15	ALP	80	87	83	84	84	80
16	Albumin	49	48	49	49	49	49

Table 2. Discriminant Analysis on Smoking and Nonsmoking Status, 375 Unsupplemented Males

Parameter	Wilk's Lambda	Significance
1. Plasma PLP	0.88059	0.0000
2. HDL Ratio	0.82355	0.0000
3. Alcohol Status	0.78134	0.0000
4. Albumin	0.77102	0.0000
5. AST	0.75960	0.0000
6. LDH	0.75694	0.0000
7. HDL Cholesterol	0.75345	0.0000

Table 3. Discriminant Analysis on Drinking and Nondrinking Status, 375 Unsupplemented Males

Parameter	Wilk's Lambda	Significance
1. HDL Cholesterol	0.97527	0.0023
2. No. Cigarettes	0.95272	0.0001
3. GGT	0.93705	0.0000
4. Plasma PLP	0.92814	0.0000
5. LDH	0.92062	0.0000
6. ALP	0.91711	0.0000

Table 4. Discriminant Parameters as a Function of Calculated 10-Year MORT RISK, by deciles (n = 480 males). Means are indicated.

Decile	Calculated Mortality	Plasma PLP	Daily B6S	HDL Ratio	Serum Albumin	Serum GGT	Percent Smokers
1s	4.35	131.5	10.65	3.46	51.3	18.1	12.50
2nd	5.11	73.1	3.21	3.85	51.0	22.0	29.17
3rd	5.50	92.0	1.83	3.85	50.0	20.8	33.33
4th	5.75	38.9	1.46	4.02	49.2	16.3	39.58
5th	5.95	37.4	1.25	4.40	49.6	27.2	50.00
6th	6.11	34.8	1.96	4.28	48.2	22.5	39.58
7th	6.28	30.9	1.27	4.63	48.2	26.7	39.58
8th	6.46	33.5	3.25	4.99	47.6	23.5	62.50
9th	6.80	25.8	1.17	5.61	47.4	34.3	79.17
10th	7.54	23.9	2.33	6.73	46.9	61.2	85.42

Table 5. Distribution of 10-Year MORT RISK by Smoking and Drinking Status

Group	Low Risk	Moderate Risk	High Risk
1. Nonsmokers, not Heavy Drinkers	143 (59.3%)	90 (37.3%)	8 (3.3%)
2. Smokers, not Heavy Drinkers	62 (30.0%)	83 (40.1%)	62 (30.0%)
3. All Heavy Drinkers	8 (12.5%)	12 (37.5%)	16 (50.0%)

Number of subjects (percent). Low Risk - MORT RISK 6.0% or less; Moderate Risk - MORT RISK 6.01 to 7.00%; High Risk - MORT RISK greater than 7.0%.

Figure 1. HRAP Individual Profile Report

LAST NAME FIRST NAME

NUTRITION SCORE = 4.8

CARDIOVASCULAR RISK SCORE = 6.875

ALCOHOL - LIVER DISEASE RISK SCORE = 5.6

DIABETES RISK SCORE = 5

CANCER RISK SCORE = 5.8

EXPLANATION OF RISK SCORES
5 = OPTIMAL
4,6 = NORMAL, BUT NOT OPTIMAL
3,7 = ABNORMAL, BUT NOT SERIOUS
2,8 = ABNORMAL, MODERATELY INCREASED RISK, SEE PHYSICIAN
1,9 = ABNORMAL, VERY HIGH RISK, SEE PHYSICIAN IMMEDIATELY

PERCENT TEN-YEAR MORT RISK = 6.02

PLP = 3 HDL = 5 CHO = 6
ALB = 5 GGT = 5

RECOMMENDATIONS

STOP SMOKING
PAY MORE ATTENTION TO YOUR DIET
YOU ARE AT INCREASED RISK FOR CARDIOVASCULAR DISEASE. WATCH YOUR INTAKE OF CHOLESTEROL AND SATURATED FAT

F1 = - 2.22 WHICH SUGGESTS SMOKER
F2 = + 4.56 WHICH SUGGESTS DRINKER

Figure 2. Trends in Median Values for Lipid Parameters, Glucose, and GGT

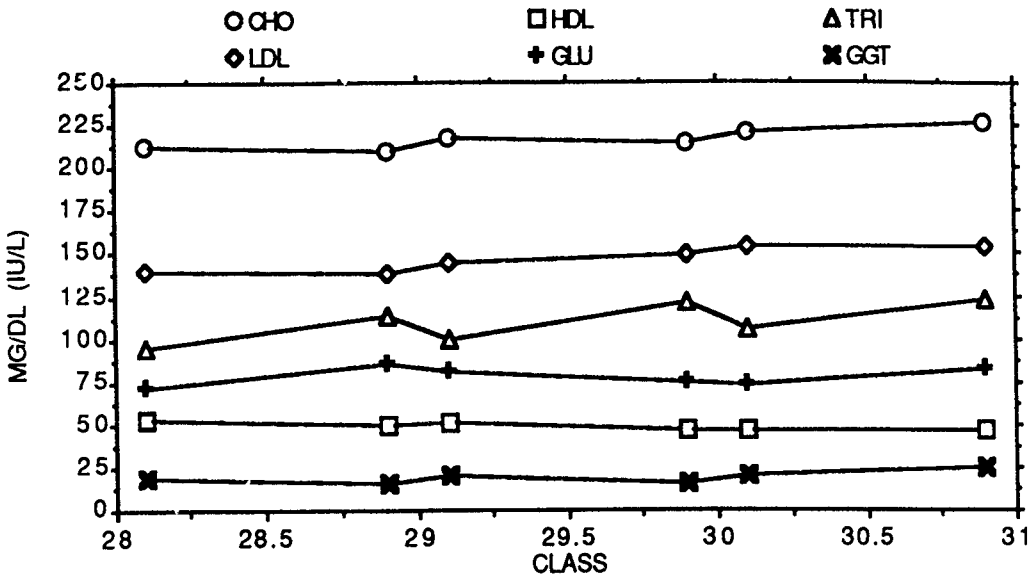


Figure 3. Comparison of Lipid Research Clinics (LRC) and WBAMC (SGT MAJ STUDY, Technicon SMAC) Methods for Total Cholesterol. 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles are plotted.

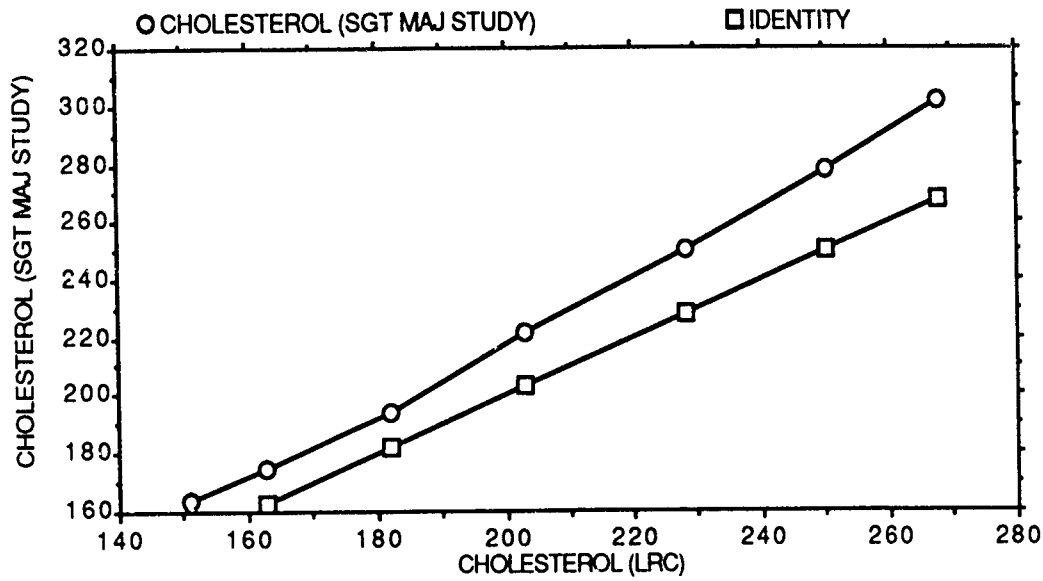
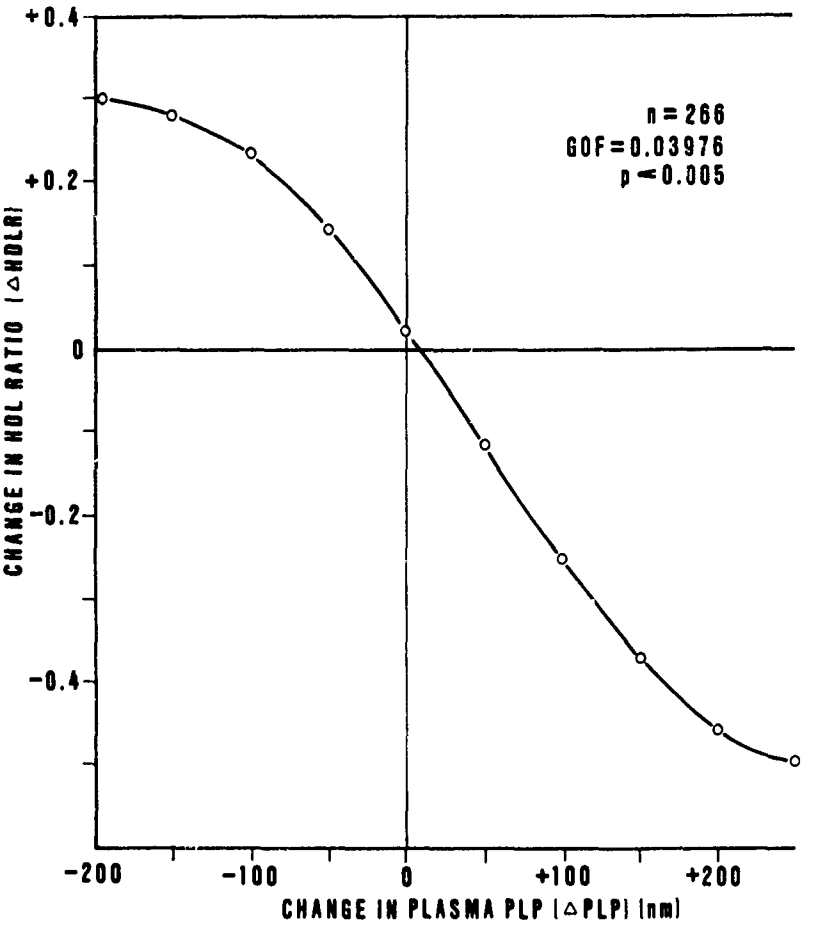


Figure 4. Effect of Change in Plasma PLP Level on HDL Ratio



IDENTIFICATION OF ACTIVE-SITE IN A NEUROTOXIC SNAKE VENOM BY AFFINITY LABELLING AND STATE-OF-THE-ART TANDEM MASS SPECTROMETRY TECHNOLOGY

Thaiya Krishnamurthy & Marguerite E. Brooks
Research Directorate
US Army Chemical Research, Development and Engineering Center
Aberdeen Proving Ground, MD 21010-5423 USA

ABSTRACT

α-Bungarotoxin, a neurotoxic snake venom, is a protein of molecular weight 7,994 Daltons. The sequences of all the amino acids in the molecule were deduced using nanomole quantities of the toxin and utilizing the state-of-the-art fast atom bombardment tandem mass spectrometric techniques.Active-site of the venom was identified by affinity labelling followed by the sequencing of the labelled molecule.

INTRODUCTION

Snake venoms are complex mixtures of substances with various types of toxic properties. Some are neurotoxic and others are cardiotoxic, cytotoxic, or ion channel blockers. Active components of several snake venoms have been deduced to be proteins. Even though, a large number of venoms have been studied by conventional sequencing methods, accurate accounts of their structures and specific active-sites, etc., are unknown. These are vital for the development of specific methods of detection and antibody production, etc.

Recently, we devised a strategy based on the tandem mass spectrometric techniques in combination with the HPLC and enzymatic methods in order to determine the sequences of the amino acids in several snake venoms. We have applied it to verify the amino acid sequences in α-bungarotoxin, reported in the literature.[1] The intact toxin was subjected to reduction and carboxymethylation followed by specific enzymatic cleavages. The products were separated over a microbore HPLC column and lyophilized. The fractions and their corresponding methyl esters were ionized under fast atom bombardment conditions and their corresponding masses were determined. Daughter mass spectra of selected fractions containing oligopeptides and their methyl esters provided their corresponding sequences of amino acids present in the molecule. Logical compilation of all the observed sequences provided the total sequence data of the toxin. Active-site of

α-bungarotoxin was then detected by labelling the purified venom with azo compounds, followed by the identification of the active-site from the sequences of the labelled molecule. Identification of the oligopeptide with the label, resulting from the enzymatic cleavage of the labelled α-bungarotoxin, could be applied for the specific detection of the venom in samples, and development of specific antibody, etc.

This type of approach to determine the sequences of amino acids in a peptide or protein is helpful even when blocked and unusual amino acids are present, when conventional sequencing methods fail. In addition, the amino acid sequences of the entire molecule even with affinity labels can be determined rapidly and more importantly with accuracy. The determination of the amino acid sequences in proteinaceous toxins and venoms, accurately and rapidly, is the initial and vital step to be undertaken for the detection of such larger biomolecules in samples and treatment of inflicted personnel.

EXPERIMENTAL METHODS

The chemicals mentioned in this paper, including α-bungarotoxin, were purchased either from Sigma Chemicals (St. Louis, MO) or Pierce Chemicals (Rockford, IL). All solvents were of HPLC grade and bought from Burdick and Jackson (Muskegon, MI). Trypsin and chymotrypsin were purchased from Sigma Chemicals and Staphylococcus Aureus Protease V8 from Miles Laboratories (Elkart, IN).

The HPLC analysis was performed on an Applied Biosystems Model 130A Separation System (Foster City, CA) using the microbore RP-300 Aquapore Octyl column (2.1 mm x 3 cm, 10 μM particles) from Brownlee Labs, Rainin Instruments Co. (Woburn, MA). Mass spectra[2] and daughter spectra were recorded either on a triple quadrupole mass spectrometer[2] or on a tandem quadrupole Fourier Transform (QFT) mass spectrometer[3]

REDUCTION AND CARBOXYMETHYLATION OF α-BUNGAROTOXIN: The venom (50 nMoles) in a polypropylene vial (2 ml) was dissolved in 50nM ammonium bicarbonate (500 μl) and treated with 100nMole/ul dithiothreiotol solution (50 μl) and left at 37°C for 30 minutes with occasional agitation. The reaction mixture was then treated with iodoacetic acid solution in freshly degassed 50 nMole ammonium bicarbonate (50 nMole/μl; 50 μl). The pH of the mixture was then adjusted to 8.6 using solid ammonium bicarbonate. The vial was sealed in a nitrogen atmosphere and heated in dark at 37°C for 30 minutes. The mixture was lyophilized overnight and the residue was treated with water (200 μl) and lyophilized in order to remove the excess buffer. The residue was dissolved in 5% acetic acid (300 μl) and divided into three equal portions in 3 separate polypropylene vials (2 ml) and lyophilized again. The residues containing ca. 17 nMoles were used for digesting with different enzymes.

PREPARATION AND PURIFICATION OF LABELLED α-BUNGAROTOXIN: A solution containing α-bungarotoxin (6 nMole) in borate buffer (16 μl) was treated with 4-azidophenylglyoxal (15 nMole) or 4-fluoro-3-nitrophenylazide (15 nMole) and incubated at 37°C in the dark for 2 hours. The reaction products were then cooled

and loaded on a microbore Vydac C-18 column in order to purify the labelled molecule. The column was then eluted with 0-90% acetonitrile (0.1% TFA)/0.1% aqueous TFA in 30 minutes with a flow rate of 0.25ml/min. Column effluent was monitored at 214nM.

PROTEOLYTIC OR ENZYMATIC DIGESTION: All three portions of residues obtained as above were dissolved in 50nM ammonium bicarbonate solution (100 μl) treated with 5 ug of enzyme (1-2% w/w enzyme/protein ratio) and incubated at 37°C for 24 hrs. The three enzymes used in this investigation were trypsin, chymotrypsin and S.aureus. The digests were then lyophilized in order to arrest the reaction and the residues were dissolved in 35-45 μl of 5% aqueous acetic acid and separated by HPLC. Similarly the carboxymethylated, labelled α-bungarotoxin was also subjected to cleavage with trypsin.

HPLC SEPARATION OF OLIGOPEPTIDES FROM THE ENZYMATIC DIGESTS: The product in 5% acetic acid (35-50 ul) was loaded on an Aquapore column. A 40 minute linear gradient of 0% to 60% acetonitrile containing 0.085% trifluoroacetic acid in 0.1% aqueous trifluoroacetic acid was employed to separate the mixtures. Fractions containing each peptide were collected and divided into three equal portions and all portions were lyophilized. Fractions containing the mixtures of oligopeptides were further chromatographed over a C-18 column using the above elution conditions.

PREPARATION OF METHYL ESTERS: A solution of 2N methanolic HCl was prepared by adding 16ml of acetyl chloride dropwise to 100 ml of methanol. After standing at ambient temperature for 5 minutes, 200 ul of the methanolic HCl solution was added to each of the HPLC fractions in order to convert them into their corresponding methyl esters. The reaction was kept at room temperature for 2 hrs with occasional shaking and excess reagent was removed by lyophilization.

MASS SPECTROMETRY: Mass spectra were recorded on a triple quadrupole mass spectrometer assembled with components from Finnigan-MAT (San Jose, CA)[2] by using 1 μl of a 5% acetic acid solution containing the peptide at 0.5-1.0 nMole level to 1 μl of thioglycerol on a gold-plated, stainless steel 2 mm probe tip. The peptides were sputtered out of the liquid matrix by bombardment with 6-8 kev Cs^+ ions generated with a cesium ion gun[4], mounted directly on the source. The masses of the peptides in each sample were measured by operating Q3 in the rf and dc mode and the other quadrupoles in the rf mode only in order to focus all the ions up to m/z 1,800 Daltons into Q3. The masses of the ions were determined by scanning Q3 between the mass ranges, 400-1,800 Daltons.

When the mass of the oligopeptides in the enzymatic digests exceeded 1,800 Daltons, the products were analyzed in the tandem quadrupole Fourier Transform mass spectrometer.[3] Samples were ionized using a liquid matrix and a cesium gun (Antek, Palo Alto, CA).[4] Experimental details are described elsewhere in detail.[3]

TANDEM MASS SPECTROMETRY: An aliquot (1 μl) of oligopeptide solutions were used to determine their corresponding sequences. The sample was ionized as before.

The molecular ion of interest was chosen by Q1 and sent to collision chamber(Q2), which was filled with argon to a pressure of 5 mTorr. The fragment ions formed as a result of collision in Q2, which was operated in the rf mode only, were focused into Q3. The Q3 were scanned between mass ranges m/z50-1,800 at a scan rate of 200 Daltons/second. The collisionally activated dissociation (CAD) mass spectra were obtained by summing up the acquired six scans. Thus, the CAD (daughter) spectra of the fragments present in all of the HPLC fractions were obtained.

Some of the MS/MS spectra of larger oligopeptides (mw >1,800 Daltons) were recorded in a VG Zab-SE/Zab-SE 4 sector double focussing tandem mass spectrometer with the courtesy of VG Instruments, Manchester, UK.

RESULTS AND DISCUSSION

The amino acid sequences obtained by the conventional sequencing method (Edman Degradation) indicated the molecular weight of the intact protein to be 7,994 Daltons.[1] The amino acid composition[1] of α-bungarotoxin is listed in Table I. There are no potential glycosylation sites in the reported sequence. Ten cysteine moieties and 5 disulfide bonds were also reported.[1] The theoretical isoelectric point is 7.97 and the estimated half life is approximately 30 minutes.[1]

The entire approach pursued by us for the determination of sequences of amino acids in α-bungarotoxin is illustrated in the scheme. The disulfide bridges present at the cysteine moieties of the intact α-bungarotoxin (50 nMoles) were reduced by treating the protein with dithiothreitol, followed by the introduction of carboxymethyl groups at each of the SH groups by treating the mixture with iodoacetic acid. The molecular weight of the carboxymethylated toxin was calculated to be 8,584 Daltons.

The products were divided into three equal parts each containing 17 nMoles) and each of them was subjected to enzymatic cleavage. One portion was digested with trypsin in order to cleave the bonds specifically at lysine and arginine. The bonds at glutamic and aspartic acids were cleaved by digesting the second portion with S.aureus. Similarly, the bonds at leucine, tyrosine, tryptophan, phenylalanine were cleaved by digesting the third portion with chymotrypsin. The digests were incubated overnight at 37°C and lyophilized to stop the cleavage and remove the enzyme from the products. The residues were dissolved (separately) with 40-45 μl of 5% aqueous acetic acid, loaded on a microbore RP- 300 column and the components were separated, as specified in the experimental section. The LC chromatogram of the trypsin digest is shown in Figure 1. Eluent corresponding to each peak was manually collected in polypropylene tubes, and lyophilized. Each residue was treated with 5% acetic acid (30 μl), agitated with Vortex and centrifuged. Two 10 μl portions from each vial were collected in separate vials and lyophilized. One of the two portions were used for the preparation of corresponding methyl esters of polypeptides and the other for recording the higher mass daughter spectral data etc. Methyl esters of all the fractions were prepared by treating the residues with methanolic HCl. Main beam spectra of all of the fractions and their corresponding methyl esters were recorded. The molecular weights of most of the fractions were found to exceed the mass range of the triple quadrupole mass spectrometer. The masses of the molecular ions of those fragments were determined from their Fourier Transform mass spectra (Figure 2).

The fractions containing the oligopeptides exceeding the mass range m/z 1,800

Table I. AMINO ACID COMPOSITION

AMINO ACID	RATIO
Alanine	5
Arginine	3
Asparagine	2
Aspartic acid	2
Cysteine	10
Glutamine	1
Glutamic acid	4
Glycine	4
Histidine	2
Isoleucine	2
Leucine	2
Lysine	6
Methionine	1
Phenylalanine	1
Proline	8
Serine	6
Threonine	7
Tryptophan	1
Tyrosine	2
Valine	5

SCHEME

INTACT TOXIN(SNAKE VENOM)

|

Reduction and Carboxymethylation

|

Enzymatic Cleavage

|

HPLC Separation of Products

|

Collection and Lyophilization of Fractions.

|

Conversion of Aliquots of Fractions to Methyl Esters

|

Fast Atom Bomardment Mass Spectra of All Fractions and the Corresponding Methyl Esters.

|

Collisionally Activated Dissociation (Daughter or MS/MS) Mass Spectra of Selected Fractions and Derivatives

|

Determination of Sequences in Individual Peptides.

|

Compilation of Sequences of All Fragments.

|

Sequences of Amino Acids in

α-BUNGAROTOXIN

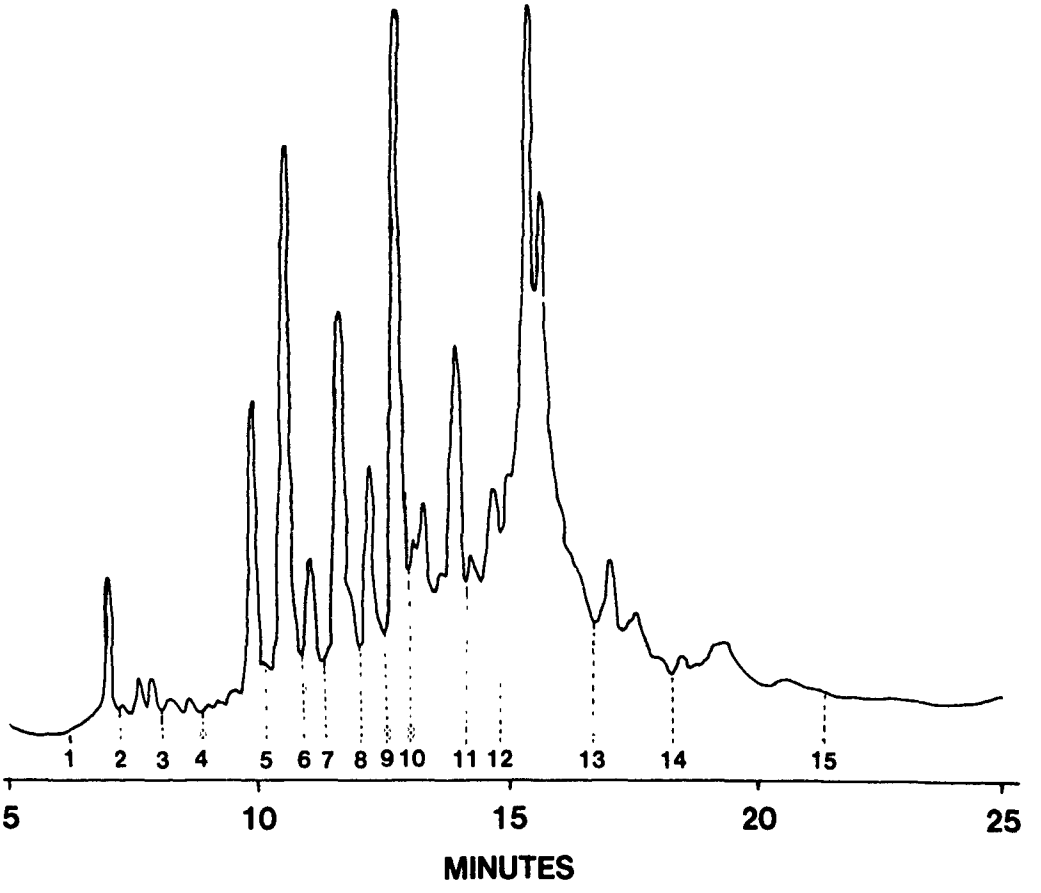


Figure 1. HPLC SEPARATION OF TRYPTIC DIGEST OF CARBOXYMETHYLATED α-BUNGAROTOXIN

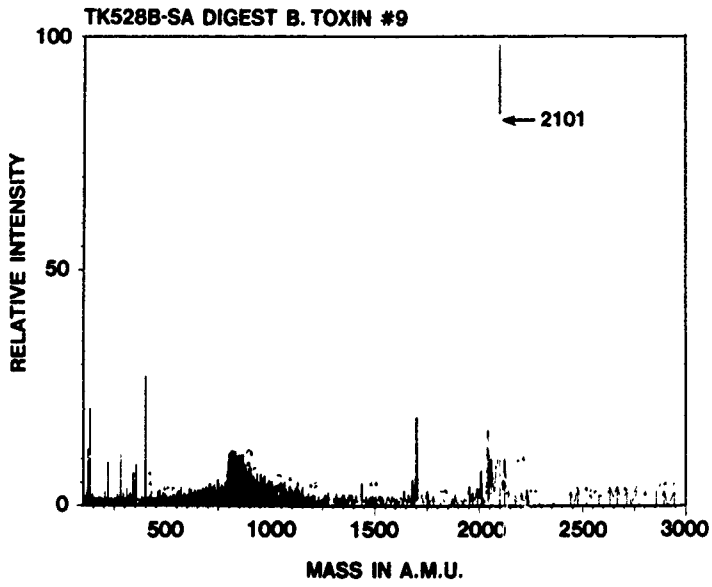


Figure 2. FAST ATOM BOMBARDMENT FOURIER TRANSFORM MASS SPECTRUM OF S. AUREUS FRAGMENT

VVELGCAATCPSK

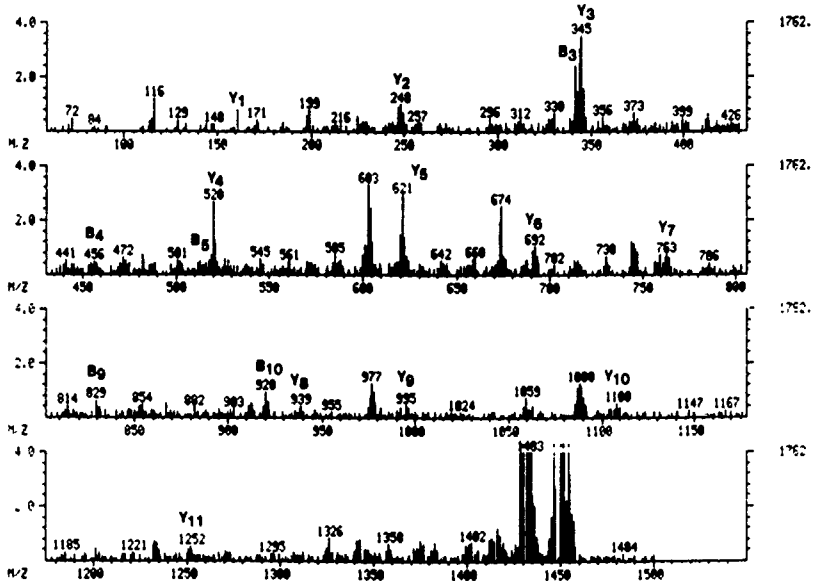


Figure 3. COLLISIONALLY ACTIVATED DISSOCIATION MASS (DAUGHTER) SPECTRUM OF METHYL ESTER OF TRYPTIC DIGEST FRAGMENT (m/z 1450).

Daltons were subjected to a second enzymatic cleavage in order to cleave them into smaller fragments, suitable to analyze by triple quadrupole mass spectrometer. Enzymatic digestions were carried as stated above and the products were lyophilized. Main beam mass spectra of all of the products were recorded. Similarly, their corresponding methyl esters were also digested with adequate enzymes, lyophilized, and treated with methanolic HCl in order to esterify the new oligopeptides formed during the second enzymatic digestion. The masses of the methyl esters were also determined from their main beam mass spectra.

The CAD (daughter) mass spectra of all of the oligopeptides with mass ranges less than 1,800 Daltons, and their corresponding methyl esters, were recorded using the triple quadrupole mass spectrometer. Daughter spectrum of a methyl ester of a tryptic fragment (m/z 1450) is shown in Figure 3. The CAD spectra of some of the larger fragements prior to a second enzymatic cleavage were recorded on a 4 sector magnetic high resolution tandem mass spectrometer (Zab SE/Zab SE) with the courtesy of VG Instruments, Manchester, UK. The CAD spectrum of a S.aureus fragment thus obtained is illustrated in Figure 4.

The sequences of the oligopeptides were determined from the B-type and Y"-type[5] ions observed in their corresponding daughter ion spectra. The sequences of the peptides thus deduced are listed in Table II. Logical compilation of all of the observed sequences led to the total sequences of the amino acids in α-bungarotoxin (Figure 5). It should be noted that distinction between amino acids with the same molecular weights and fragmentation patterns under CAD conditions is not possible by the tandem mass spectrometric technique. Thus, the presence of leucine and isoleucine in the bungarotoxin could not be differentiated from our findings, so it was made from the reported sequences[1], which are also illustrated in Figure 5. The sequences of the amino acids, 9-11 (from the n-terminus), deduced by us differed from the reported sequences.[1] The sequences of 12 amino acids from the n-terminus of the intact toxin were also determined using Associated Biosystems (ABI) gas-phase amino acid sequencer. The observed sequences corraborated our results.

4-Azidophenylglyoxal(APG)

m.w. 175; label 147

4-Fluoro-3-nitrophenylazide(FNPA)

m.w. 182; label 154

One nanomole of α-bungarotoxin in borate buffer was digested with three nanomoles of p-azidophenylglyoxal, specific for modifying arginine. The product was purified by chromatographic separation over a microbore C-18 HPLC column and lyophilized. The residue in ammonium bicarbonate (50nMole) buffer was cleaved with trypsin (1 μg). The digest containing the mixture of oligopeptides was separated over a RP-300 microbore column. All fractions were lyophilized, dissolved in 5%

114	213	375	512	613	714	785	886	973	1070
Ile	Val	Cys	His	Thr	Thr	Ala	Thr	Ser	Pro
2100	1986	1887	1725	1588	1487	1386	1315	1214	1127
1183	1270	1341	1440	1541	1703	1800	1897	1954	2100
Ile	Ser	Ala	Val	Thr	Cys	Pro	Pro	Gly	Glu
1030	917	830	759	660	559	397	300	203	146

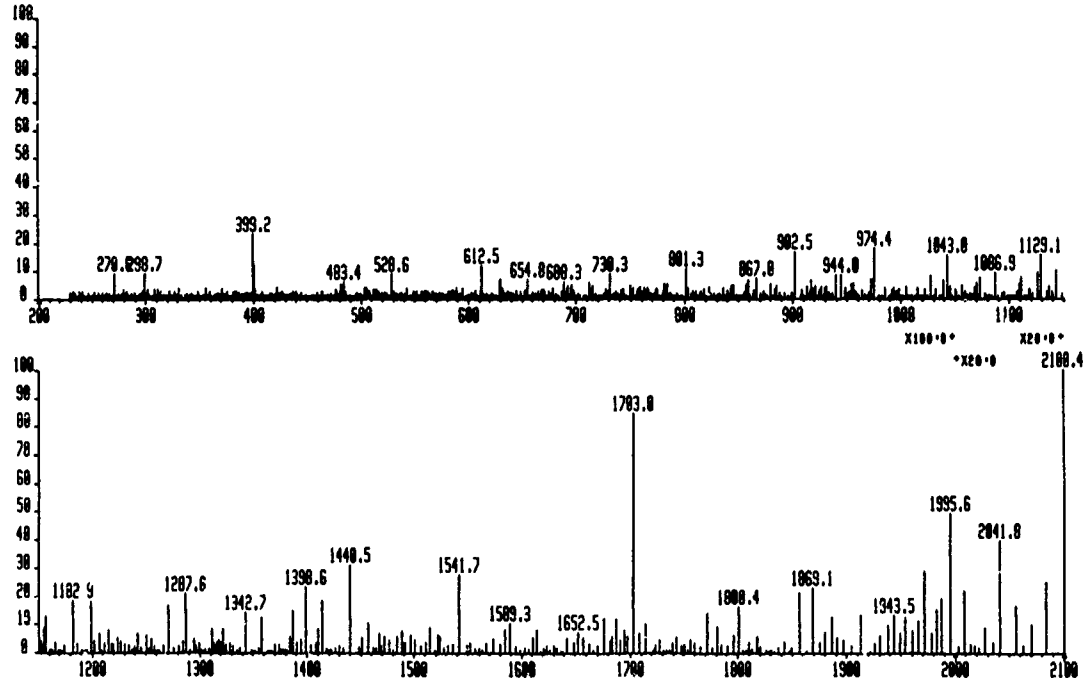


Figure 4. DAUGHTER SPECTRUM OF S. AUREUS DIGEST (m/z 2100)

Table II. PROTEOLYTIC DIGESTION PRODUCTS

ENZYME	m/z* CALCULATED	m/z* OBSERVED	SEQUENCES OF FRAGMENTS
Chymotrypsin	2328	2328	IVCHTTATSPISAVTCPPGENL
	2653	2653	IVCHTTATSPISAVTCPPGENLCY
	513	513	CDAF
	1136	1136	CSSRGKVVEL
	2308	2308	MWCDAFCSSRGKVVELGCAA
	2461	2461	CSSRGKVVELGCAATCPKKPY
	2404	2404	EEVTCCSTDKCNHPPKRQPG
Trypsin	2809	2809	IVCHTTATSPISAVTCPPGENLCYR
	1322	1322	MWCDAFCSSR
	1393	1393	VVELGCAATCPSK
	1578	1578	VVELGCAATCPSKGK
	1619	1619	KPYEEVTCCSTDK
	2922	2922	DAFCSSRGKVVELGCAATCPSKKPYE
S.aureus	2101	2101	IVCHTTATSPISAVTCPPGE
	2671	2671	NLCYRKMWCDAFCSSRGKVVE
	1712	1712	LGCAATCPSKKPYE
	2668	2668	NLCTRKMTCDFCSSRGKVVE

aqueous acetic acid (5 μl) and analyzed by tandem quadrupole Fourier Transform Mass Spectrometry. Similarly, one nanomole of the unlabelled α-bungarotoxin as the control, was digested with trypsin, separated and all fractions were analyzed by QFT mass spectrometer. The oligopeptides obtained from both digests were identical indicating that APG-α-bungarotoxin was not formed.

Similarly, labelling of the toxin with 4-fluoro-3-nitrophenylazide was carried out. One of the LC fractions, which is highlighted in Figure 6A, was found to contain an oligopeptide with molecular weight of 2,193 (Retention time, 14.5 minutes). There were also other ions in this fraction, all of which were observed in the tryptic digest of the unlabelled molecule. This fraction was further purified by chromatographing over a C-18 column (0-40% 0.1% TFA in acetonitrile/0.085% aqueous TFA in 40 minutes). All fractions were analyzed as usual and one of the fractions contained only the labelled oligopeptide (Retention time, 18.9 minutes; m.w. 2,193). The molecular weight of this fragment without the label was calculated to be 2,038.

Computer analysis using PEPED[6] software was performed in order to acquire the projected sequence of the residue (m.w. 2,038) in the bungarotoxin. The projected sequences are shown in Table III. The solution containing the labelled fragment was lyophilized and the residue was treated with methanolic HCl (50 μl) and left at ambient temperature for two hours with occasional agitation. The product was lyophilized and dissolved in 5% aqueous acetic acid (5 μl) and analyzed using QFT mass spectrometry. An ion with m/z 2,248 was observed indicating the addition of four ester methyl groups to the labelled fragment. The last two entries in Table III with masses 2,041 were ruled out based on the fact both would have added on 6 methyl groups on methylation. The second and third fragments would only have resulted from the S.areous and chymotrypsin cleavages of α-bungarotoxin, respectively. In addition, all except the first entry have a higher molecular weight than the observed one. Hence, it was concluded that the first projected sequence of the labelled fragment shown in Table III would be a more adequate one. This approach could also be applied for the detection of α-bungarotoxin in samples. Investigations towards the elucidation of labelled fragment's structure and development of rapid, sensitive, and selective method of analysis are underway.

CONCLUSION

Tandem mass spectrometric techniques in combination with affinity labelling and enzymatic cleavage methodology could successfully be applied for the rapid and accurate determination of the total sequences of amino acids in larger toxic biomolecules such as snake venoms and for the identification of their active-site(s). Presence of amino acids with blocked n-terminii, other uncommon amino acids, and amino acids with affinity labels also could easily be detected by this technique. Identification of the specific fragment containing the affinity labels also could be used in the specific detection of the proteinaceous toxins in samples. Even though, the present observed sensitivity in this approach does not meet our requirements, it can easily be rectified by the application of rapidly developing state-of-the-art tandem mass spectrometric technology.

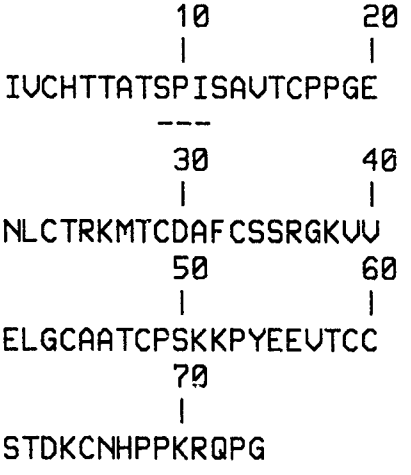


FIGURE 5. SEQUENCES OF AMINO ACIDS IN alpha-BUNGAROTOXIN

TABLE III. PROJECTED SEQUENCES OF AMINO ACIDS IN LABELLED FRAGMENT

RESIDUE	MASS	AMINO ACID SEQUENCE	# of FREE COOH GROUPS
34-52	2038.3	SSRGKVVELGCAATCPSKK	4
38-55	2040.4	KVVELGCAATCPSKKPYE	4
6-24	2041.3	TATSPISAVTCPPGENLCY	4
39-53	2041.3	VVELGCAATCPSKKPYEE	6
40-57	2041.3	VELGCAATCPSKKPYEEV	6

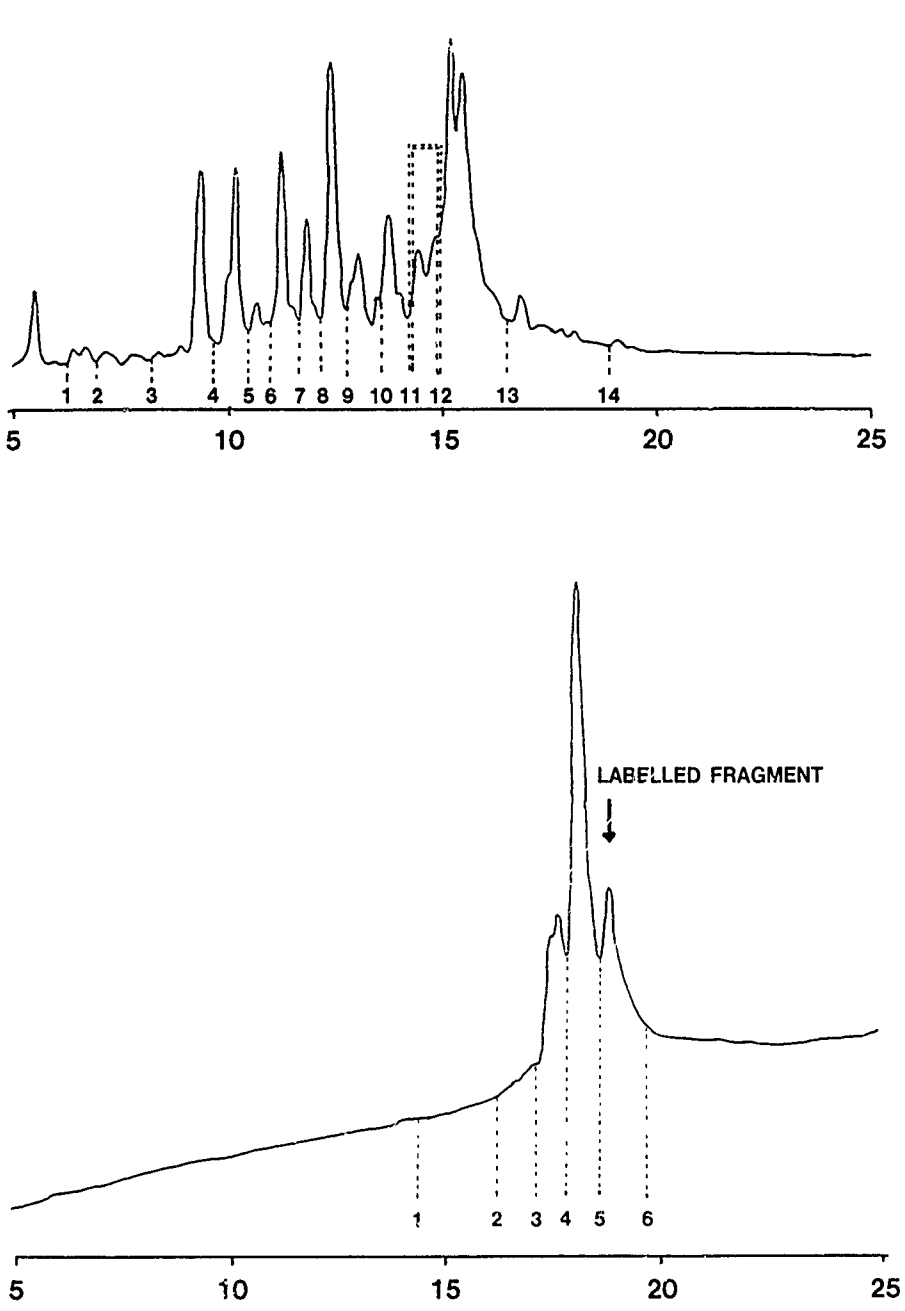


FIGURE 6. HPLC SEPARATION OF TRYPTIC DIGEST OF CARBOXYMETHYLATED LABELLED α-BUNGAROTOXIN

ACKNOWLEDGMENT

Other contributors to this paper are: Prof. Donald F. Hunt and Dr. Jeffrey Shabanowitz of University of Virginia, Charlottsville, VA 21903; and Dr. Shiuan Chen of Beckman Research Institute, City of Hope, Duarte, CA 91010.

REFERENCES.

1. D. Mebs, K. Narita, S. Iwanaga, Y. Samejima, and C. Y. Lee, Z. Physiol. Chem., **353**, 243-262 (1972).

2. D. F. Hunt, A. B. Giordani, J. Rhodes, D. A. Herold, Clin. Chem.(Winston-Salem, N. C.), **28**, 2387-2392 (1982).

3. D. F. Hunt, J. Shabanowitz, J. R. Yates III, N. Z. Zhu, D. H. Russel, and M. E. Castro, Proc. Natl. Acad. Sci. USA, **84**, 620-623(1987)

4. W. A. Berth, K. M. Straub, A. L. Burlingame, Anal. Chem., **54**, 2029- 2034 (1982).

5. P. Roepstorff and J. Fohlman, Biomed. Mass Spectrum., **11**, 601 (1984).

6. John Yates III, University of Virginia, Charlottsville, VA, Personal Communications, 1988.

An Empirical Surface Temperature Model with Automatic Type-Day Classification

*Alan E. Krusinger, Mr., Sharon R. Hoover, Mrs.,
and Eugene A. Margerum, Dr.
U.S. Army Engineer Topographic Laboratories
Fort Belvoir, Virginia 22060-5546

The Army, and other services, are committed to the use of thermal infrared (IR) sensing, either alone or in conjunction with other sensors, to locate targets on the battlefield because thermal infrared sensing is remarkably well adapted to locating warm vehicles against cooler backgrounds.

Thermal infrared surface temperature prediction models are necessary to deal with the challenge of automatic, or Assisted Target Recognition (ATR), false target discrimination, and forward looking infrared (FLIR) image interpretation. In addition, the development of hyperspectral imagery (each scene in 220 spectral bands) requires these models to exploit that capability in the longer wavelength portions of the electromagnetic spectrum (3-5 um, 8-12 um).

Thermal infrared sensing is not without problems, and background clutter (large temperature variation with small spatial variation) causes a serious problem with the use of thermal imagery in target location and identification. At certain times and places, the background may be the same temperature or warmer than all, or parts, of a military target. Vehicles can successfully hide thermally in the right background, at the right time of day.

The use of energy budget "first-principles" models has so far dominated the effort to predict surface temperatures. Unfortunately, they tend to be very complex and iterative, and they require large quantities of input data that are unavailable to the field Army, as in the models by Balick, et al.[1]

Other current models, or tactical decision aids (TDA's), like those done by Higgins[2] and Higgins, et al.[3] are of a very restricted type, using "snapshot data," and still having an extensive list of inputs (5 system, 5 target/background, 26 site/meteorological). We have concluded

that these models are not usable outside of a carefully measured, calibrated test site, and that they cannot be simplified for tactical use.

APPROACH

Our previous work on the effects of meteorological variables on the surface temperature indicated the great complexity of the modeling problem, a complexity which seemed to require "first-principles" models. We decided, however, to try to find an interim solution with simplified model inputs to fill the void of useful models.

Our approach was to let the model evolve from our data base, i.e., the model is empirical. The model is based upon "type-days" of relatively unique sky cover and soil moisture conditions. We postulated that in a given climate, in a given season, on a given type-day, meteorological variables would be repeatable, and backgrounds would have repeatable diurnal temperature curves.

The model is being developed from meteorological, radiometric and temperature data taken around the clock, for several years, at a temperate climate site in Northern Virginia, and in cooperation with the U.S. Department of the Interior, Geological Survey, Geologic Division, Astrogeology Branch, Flagstaff, Arizona, and the U.S. Department of Agriculture at a semiarid site in New Mexico. A second remote site is being outfitted at Yuma, Arizona, in 1988, and data collection in other climates is being planned.

DESCRIPTION OF THE WORK

The project entailed assembling composite or type-days of similar weather and sky cover conditions and doing a regression computation on the associated values for background temperature through the diurnal cycle. The seven "type-days" selected for the summer season, temperate climate are listed in Table 1, and together with the climate and season, these are the model inputs.

Table 1
The Seven Type-Days

1. Clear	Dry Surface Soil
2. Clear	Wet Surface Soil
3. Partly Cloudy	Dry Surface Soil
4. Partly Cloudy	Wet Surface Soil
5. Overcast	Dry Surface Soil
6. Overcast	Wet Surface Soil
7. Overcast, Rain	Wet Surface Soil

Surface soil moisture affects surface temperature greatly and had to be addressed. We didn't expect the field Army to be measuring soil moisture, so we made the soil moisture input "dry" or "wet," a simple observation of bare surface soil.

The backgrounds, target and temperature differences that were available at the temperate site are listed in Table 2.

Table 2
Backgrounds, Target and Temperature Differences

1. Cut Grass	Temperature
2. Bare Soil	Temperature (silty sand)
3. Uncut Grass/Weeds	Temperature
4. Gravel	Temperature
5. M114 Armored Reconn. Vehicle	Temperature
6. M114-Cut Grass	Temperature Difference
7. M114-Bare Soil	Temperature Difference
8. M114-Uncut Grass	Temperature Difference
9. M114-Gravel	Temperature Difference

The temperature differences (items 6-9) were put in so that we would have an accurate difference value and not have to subtract one regression curve from another to determine thermal contrasts. All temperatures used in the study are degrees Celsius. Because emissivities of backgrounds cannot be measured practically, we used effective blackbody temperatures throughout, just as a tactical system would do. Blackbodies theoretically do not exist, but they would not reflect and would emit all energy they received.

In the first attempt, we subjectively sorted days and parts of days from our 33,750 half-hourly observations from 1984 and 1985 into the seven type-days. This formed composite 24-hour periods of data taken from the same season but from several years, for each type-day. The variables used for subjectively sorting type-days are shown in Table 3 and in Figure 1 they are shown plotted together, as we use them. This is a five day plot from 26 September-1 October 1986. No rainfall occurred in this period. Day 270 is partly cloudy and dry. Day 271 is overcast and dry. Day 273 is clear and dry.

Table 3
Meteorological Variables Used to Sort Type-Days

1. Short-Wave Incoming Radiation (Swi) (0.2-2.8 um)
2. Long-Wave Incoming Radiation (Lwi) (4-50 um)
3. Air Temperature
4. Absolute Humidity
5. Bare Soil Surface Radiometric Temperature
6. Precipitation
7. Wind Speed
8. Soil Moisture Near The Surface

In the next iteration, classification of days into type-days was done automatically. Quarter-day periods were selected as the minimum time period for consecutive observations with sufficient continuity. Daylight periods (0600-1200, 1200-1800) were sorted mainly on the basis of the area under the short-wave incoming radiation, or insolation curve. Large area indicated clear conditions (Figure 1, Day 273). Medium area indicated partly cloudy conditions (Figure 1, Day 270). Small area indicated overcast conditions (Figure 1, Day 271). The variation of the Swi curve, a measure of the smoothness of the curve, was computed to assist in classification. High variation indicated partly cloudy conditions and low variation indicated clear conditions. The soil moisture maximum of the quarter-day period was compared to a threshold value to determine "wet" or "dry" surface soil for both day and night periods.

Nighttime periods (1800-2400, 2400-0600) were sorted mainly on the area under the long-wave incoming radiation curve. Long-wave incoming radiation is affected by clouds, air temperature, and moisture in the air. High Lwi area indicated overcast conditions; medium area indicated partly cloudy conditions; and low area indicated clear conditions. The variation of the Lwi curve was also computed. High variation indicated partly cloudy conditions and low variation indicated overcast or clear conditions. Quarter-day maximum absolute humidity indicated overcast conditions when it was high, and clear conditions when it was low. A combined Lwi/Absolute Humidity variable was created for classification because they are so interrelated. Thresholds for the above mentioned variables were determined by trial and error and comparison with the subjectively classified data. This process is still ongoing and incomplete.

The partly cloudy type-days had considerable surface temperature variation, because of the variability of conditions contained in this category. In the subjective method, we digitized the envelopes from a plot of the data. Now we will determine a top and bottom envelope for each partly cloudy background by using maximum and minimum values for each time period. These envelopes should represent the maximum (sunlit)

and minimum (shaded) conditions inherent in the partly cloudy type-day. Future refinements should subdivide this type-day category into a partly cloudy and a mostly cloudy condition.

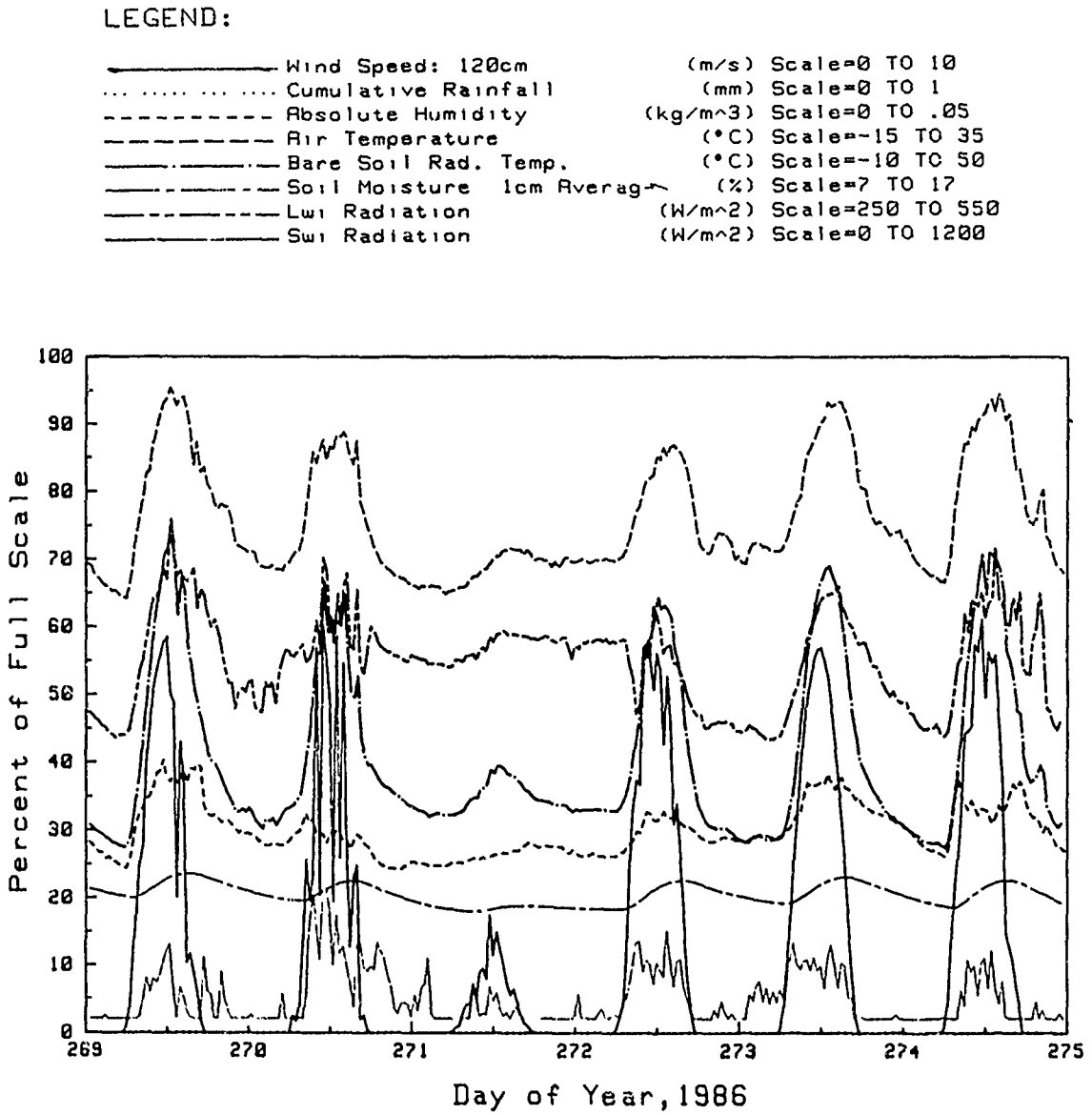


Figure 1. Eight parameter weather description with bare soil surface temperature for a five day period, 26 September-1 October 1986.

RESULTS

A computer program was written to calculate a smooth curve from the data in the form of a truncated Fourier Series. The coefficients of the trigonometric series shown in Equation 1 were adjusted by applying regression conditions. Nine terms were used for the equation to produce diurnal curves most similar in shape to real data.

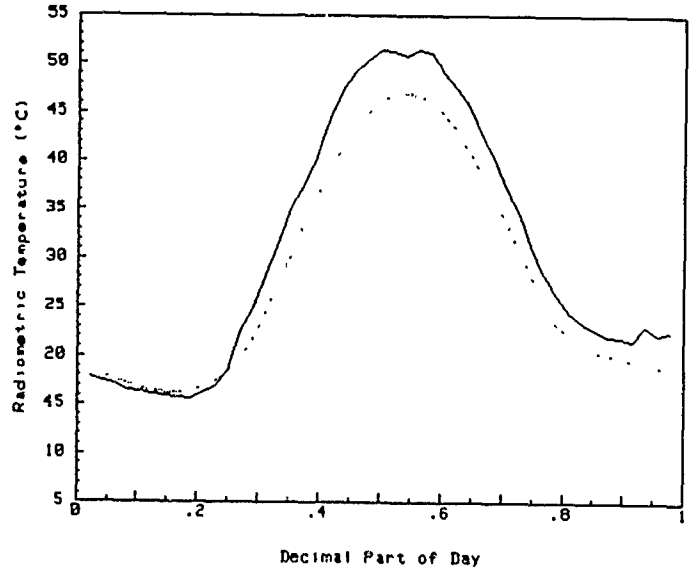
$$y(x) = b_0 + b_1 \sin x + b_2 \cos x + b_3 \sin 2x + b_4 \cos 2x + b_5 \sin 3x + b_6 \cos 3x + b_7 \sin 4x + b_8 \cos 4x \quad (1)$$

A program that compares the predicted model curve with raw data from the data base was used to test the model by plotting both measured and predicted values, using 1987 measured data (Figures 2-5). None of the days picked had constant weather conditions, so the model curves sometimes deviated from measured values. Results using the automatic classification technique are not satisfactory yet, due to threshold selection, but they are presented with results from the subjective technique as parts "a," subjective and "b," automatic, for each figure.

Figure 2 is a bare soil background under clear dry conditions in mid June that had persisted for several days. Figure 3 is a bare soil background for clear wet conditions, 25 mm of rain having occurred the previous day. The wet ground is indicative of the previous rainfall that cleaned the air, leaving a true, clear sky. Some clouds after midnight raised plot temperature and clouds in mid morning reduced plot temperature. Figure 4 is a bare soil background in an overcast dry condition with some clouds in the early morning that lowered plot temperature by shading. Figure 5 represents a bare soil background under overcast conditions, with light rain.

This empirical model provides a reasonable value for various background temperatures, a target temperature, and various meteorological variables for typical days. The model comes directly from the data and contains no estimates of critical factors needed to compute the model. The inputs to the model are simple, requiring no measurements. For all the simplicity of the inputs, it does a good job of temperature prediction. This model illustrates the feasibility of this empirical approach.

a



b

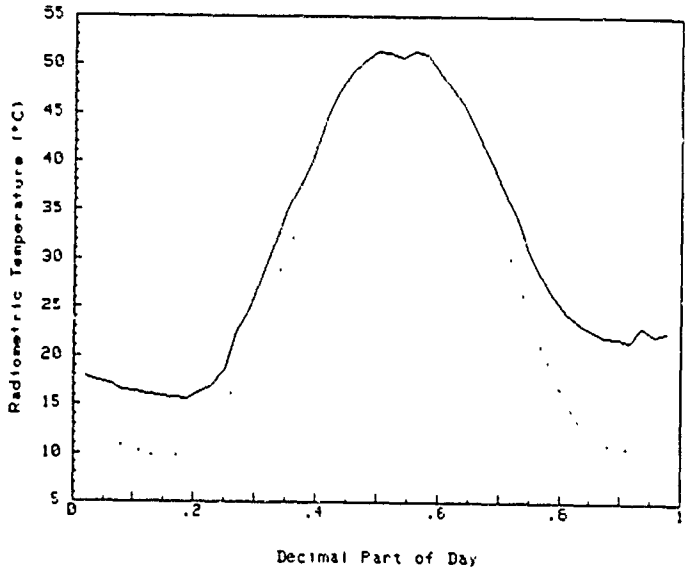
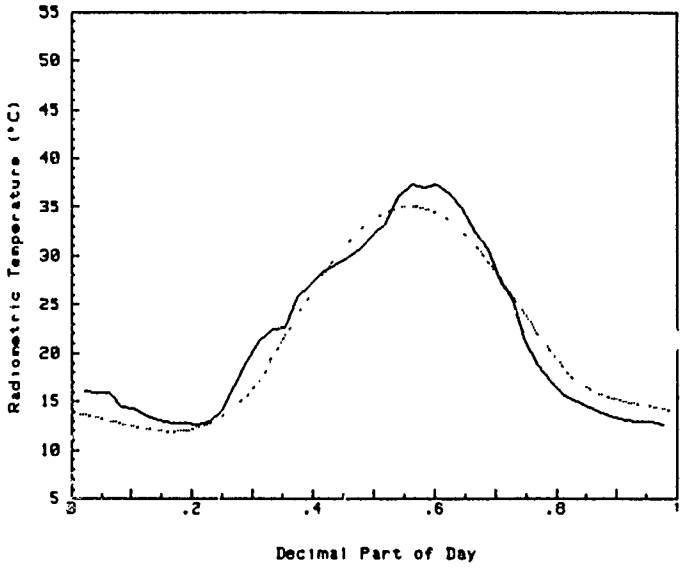


Figure 2. Measured (solid line) and predicted (dotted line) effective blackbody temperatures for bare soil on a clear dry summer day on 19 June 1987. Part a subjective; part b automatic classification.

a



b

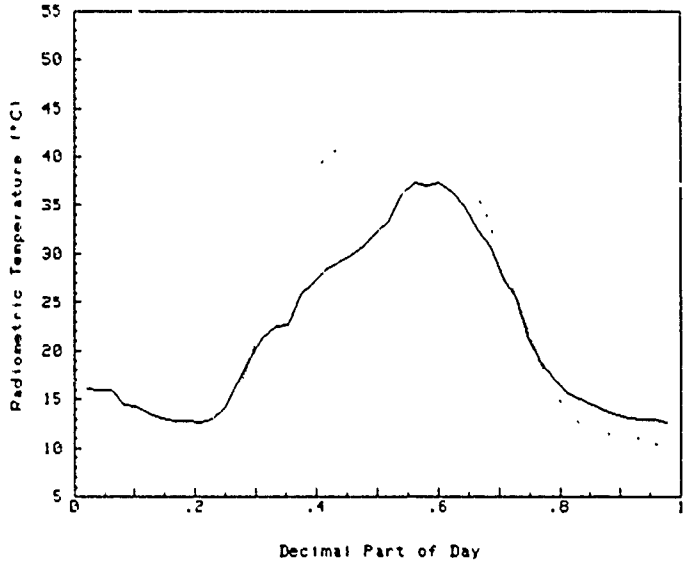
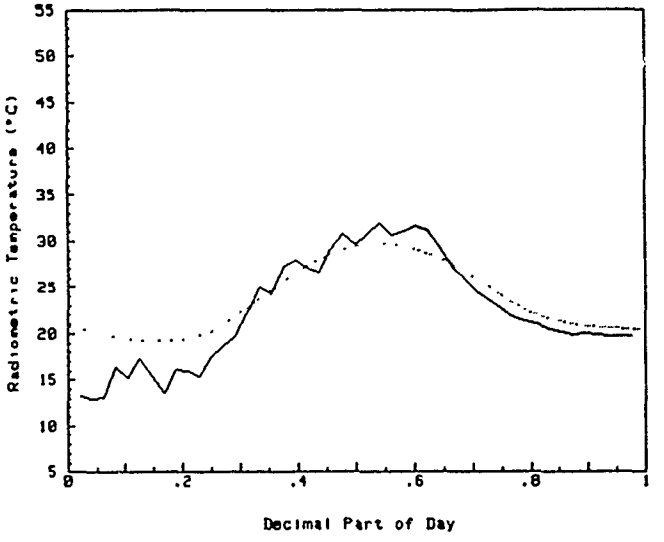


Figure 3. Measured (solid line) and predicted (dotted line) effective blackbody temperatures for bare soil on a clear wet summer day on 5 June 1987. Part a subjective; part b automatic classification.

a



b

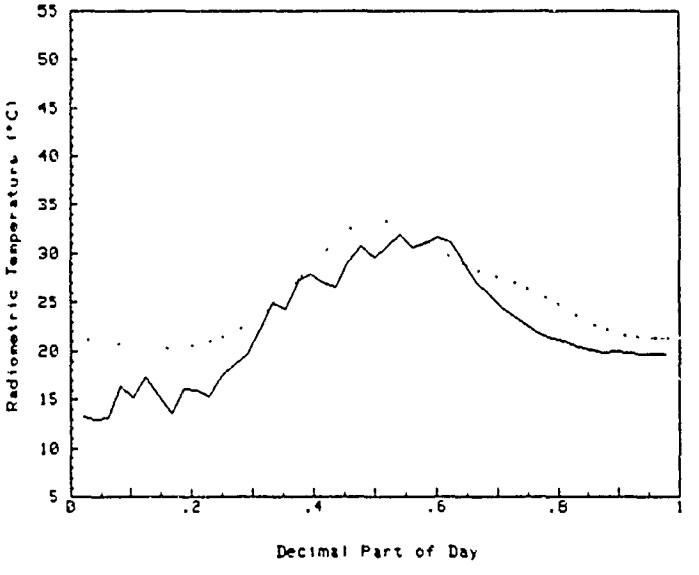
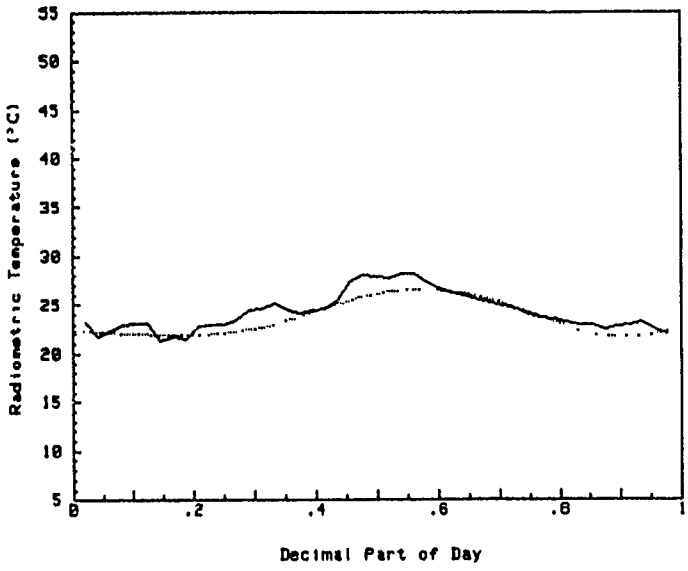


Figure 4. Measured (solid line) and predicted (dotted line) effective blackbody temperatures for bare soil on an overcast dry summer day on 25 August 1987. Part a subjective; part b automatic classification.

a



b

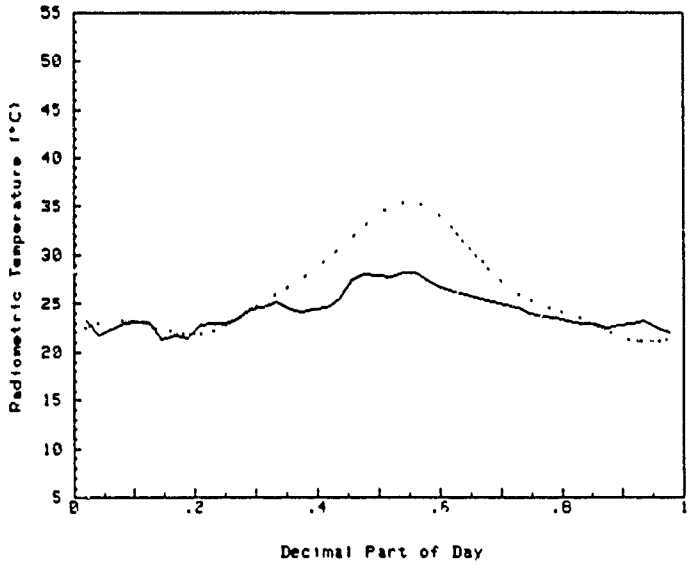


Figure 5. Measured (solid line) and predicted (dotted line) effective blackbody temperatures for bare soil on an overcast rainy summer day on 22 August 1987. Part a subjective; part b automatic classification.

CONCLUSIONS

We have observed a great need for this empirical approach to fill the gap of simple, usable predictive models and to provide data to test other models. We have concluded from this effort that this is a feasible and necessary approach.

These types of data are needed in other climates, and we started data collection near Las Cruces, New Mexico, a semiarid desert, in September 1986. In 1988, another site will be automatically collecting and transmitting data to us from Yuma, Arizona. Another desert site is envisioned in Death Valley, California, and a moist, tropical site is planned for Puerto Rico. We solicit cooperative work in this endeavor.

REFERENCES

1. Balick, L.K., L.E. Link, R.K. Scoggins, and J.L. Solomon, 1981, Thermal Modeling of Terrain Surface Elements. Technical Report EL-81-2, prepared by the Environmental Laboratory, Waterways Experiment Station, in collaboration with Mississippi State Univ., for the U.S. Army Corps of Engineers Waterways Experiment Station, Vicksburg, MS, pp. 13-14

2. Higgins, G.J., 1984, Users Guide for the Operational Tactical Decision Aid (OTDA) for Infrared (8-12 um) Systems - Mark 111 Computer Version, Report No. AFGL-TR-84-0164, Scientific Report No. 18, by Systems and Applied Sciences Corp. (SASC) for the Air Force Geophysics Laboratory, U.S. Air Force Systems Command.

3. Higgins, G.J., J.M. Freni, P.F. Hilton, T.J. Keegan, B.A. Mareiro, M.A. Mickelson, C.N. Touart, and R.F. Wachtmann, 1987, Mark 1 Tactical Decision Aids for Microcomputer Systems, Report No. AFGL-TR-87-0057, Scientific Report No. 32, by ST Systems Corp. for the Air Force Geophysics Laboratory, U.S. Air Force Systems Command.

Tracking Maneuvering Helicopters Using Attitude Measurements (U)

*Frank P. Kuhl
Fire Support Araments Center
Clifford E. Wilkins
Close Combac Armament Center
Armament Research, Development and Engineering Center
Picatinny Arsenal, NJ 07806-5000

1. Introduction:

The object of the tracking and lead-angle prediction work described in this paper is to determine whether optical (e.g., television, infra-red or laser radar) measurements of helicopter fuselage attitude angles and rotor tip-path plane and coning angles can substantially improve the accuracy of standard fire control radars, which can make only target position and velocity measurements. The approach taken here for helicopters is a modification to that taken for fixed-wing aircraft (References 1 and 2) in which the aircraft lift vector is modeled and used in the estimation of acceleration: the magnitude is a function of the attack angle and the direction is perpendicular to the wings. For helicopters the thrust vector of the rotor system is modeled to improve the estimate of acceleration: the magnitude is a function of the rotor coning angle and the direction is perpendicular to the rotor tip-path plane.

A flow chart of the proposed lead-angle calculation system is shown in Figure 1 for an air-defense gun application. The gun is at the origin of the inertial coordinate system (X, Y, Z) and the aircraft coordinate system ($\bar{b}x$, $\bar{b}y$, $\bar{b}z$) is related to the inertial system by the Euler angles. The radar supplies position and position rate (i.e., velocity) to the trajectory estimator to calculate the lead angle as in a standard fire control system. The optical sensor captures a 2-dimensional image (e.g., television or infra-red) or a 3- dimensional range map (e.g., laser radar) that is digitized and given to the recognition system (see, for example, References 3, 4, 5). The class of the aircraft is identified along with its Euler angles, and additionally the tilt and coning angles of a helicopter rotor system.

* Approved for public release. Distribution unlimited.

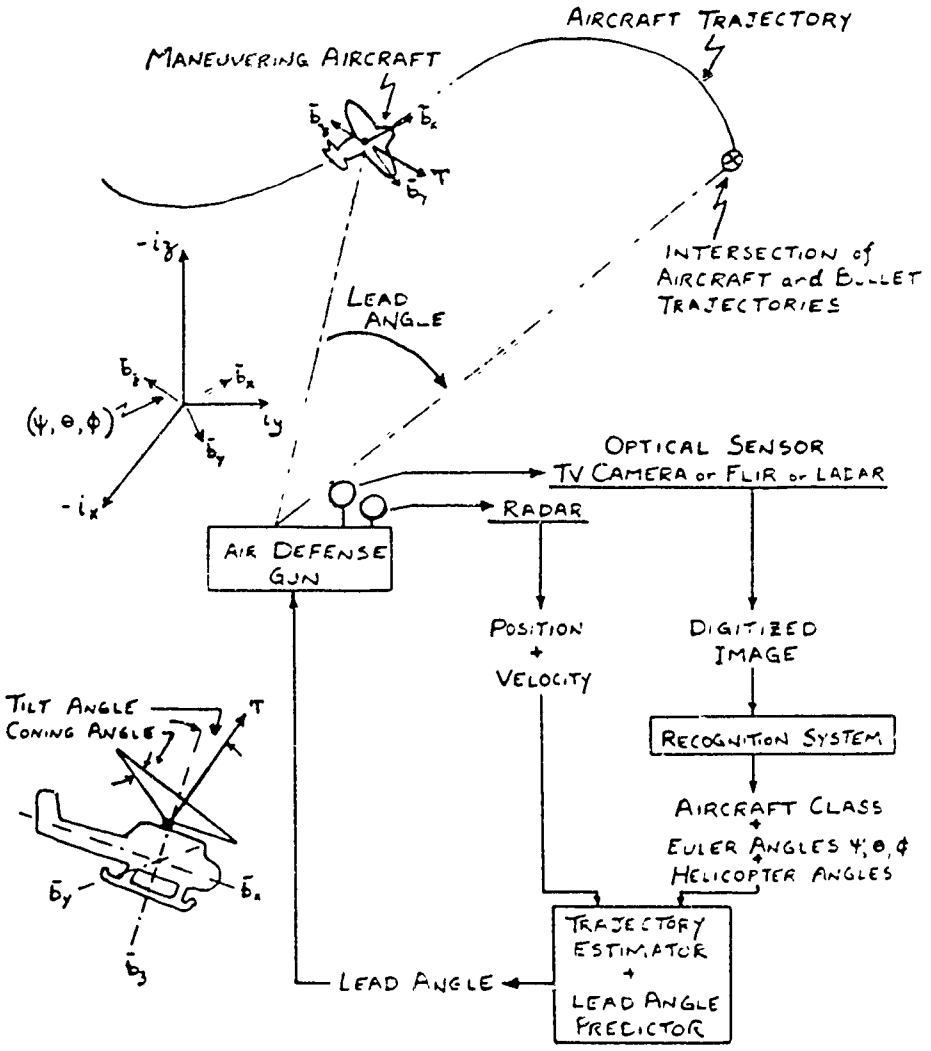


Figure 1. Proposed lead-angle calculation system that supplements radar with an optical sensor.

This information is used to calculate the aerodynamic forces on the airframe by the trajectory estimator in the attempt to overcome the limitations inherent in trying to determine acceleration from position measurements only: noisy and slowly derived estimates.

2. Tracker Models:

The basic approach to aircraft tracking used in this research centers around the extended Kalman filter. The Kalman filter is used because of its recursive nature and the extended Kalman filter is used because of the nonlinearities inherent in transforming the thrust vector, T shown in Figure 1 into the inertial fixed coordinates ix, iy, iz. Equations for the extended Kalman filter can be found in Reference 6. For the helicopter tracking problem, a state vector consisting of 22 states and a measurement vector containing 12 measurements was developed. The parameter definitions are those in References 1 and 7. The vectors are listed in Table 1. The state equations are listed in Table 2. and the measurement equations are given in Table 3. The first 15 states in Table 1. are the same as used and discussed in Reference 1. The functions of the 22 states are as follows: states 1-3 are the body axis system components of the angular velocity of the aircraft with respect to the inertial frame, states 4-6 are the three Euler angles that orient the aircraft with respect to the inertial frame, states 7-9 are the three components of inertial velocity, states 10-12 are the three inertial position components, states 13-15 are acceleration bias states used to improve the modeling of vehicle acceleration, states 16-18 are the rate of change of tip-path plane angles, states 19-21 are the tip-path plane angles, and state 22 is a bias in tip path phase angle.

For the helicopter, thrust and weight are the only forces considered. Thrust is modeled as shown in Figures 2 and 3. The direction of the thrust is always perpendicular to the tip-path plane and the orientation of the tip-path plane with respect to the body is given by the angles a1 and b1. The magnitude of the thrust vector is proportional to the coning angle a0. Equations for modeling of thrust are given in Table 4.

The Maneuver Estimator is a name given to a tracker that uses the fuselage Euler angles and rotor system angles to improve trajectory estimation. It is a 19 state estimator that uses all of the states listed in Table 1. except states 13, 14 and 15. The Maneuver Estimator also uses, but does not require, all 12 measurements. States 13, 14 and 15 are not used because acceleration biases are adequately modeled by the thrust terms that appear in the acceleration equations 7, 8 and 9 of Table 2. The Center-of-Gravity or C. G. Tracker is the name given

State Vector	Measurement Vector
x(1) = p(rad/sec)	z(1) = ϕ(rad) Not used
x(2) = q(rad/sec) Not used	z(2) = θ(rad) in CG
x(3) = r(rad/sec) in CG	z(3) = ψ(rad) Tracker
x(4) = ϕ(rad) Tracker	z(4) = R(ft) Range
x(5) = θ(rad)	z(5) = η(rad) Azimuth
x(6) = ψ(rad)	z(6) = ζ(rad) Elevation
x(7) = $\dot{x}$(ft/sec)	z(7) = $\dot{R}$(ft/sec)
x(8) = $\dot{y}$(ft/sec)	z(8) = $\dot{\eta}$(rad/sec)
x(9) = $\dot{z}$(ft/sec)	z(9) = $\dot{\zeta}$(rad/sec)
x(10) = x(ft)	z(10) = a_1(rad) Not used
x(11) = y(ft)	z(11) = b_1(rad) in CG
x(12) = z(ft)	z(12) = a_o(rad) Tracker
x(13) = b_x(rad) Not used	
x(14) = b_y(rad) in Maneuver	
x(15) = b_z(rad) Estimator	
x(16) = $\dot{a}_1$(rad/sec)	
x(17) = $\dot{b}_1$(rad/sec) Not Used	
x(18) = $\dot{a}_o$(rad/sec) in CG	
x(19) = a_1(rad) Tracker	
x(20) = b_1(rad)	
x(21) = a_o(rad)	
x(22) = c(rad)	

Table 1. State and Measurement Vectors

1	$\dot{p}$	=	$C_1p + C_2q + C_3(b_1+w_1)$
2	$\dot{q}$	=	$C_4p + C_5q + C_6(a_1+w_2)$
3	$\dot{r}$	=	$C_7r + W_3$
4	$\dot{\phi}$	=	$p + q\sin\phi\tan\theta + r\cos\phi\tan\theta$
5	$\dot{\theta}$	=	$q\cos\phi - r\sin\phi$
6	$\dot{\psi}$	=	$(q\sin\phi + r\cos\phi)/\cos\theta$
7	$\dot{x}$	=	$\underline{n_xT/m} + C_8(b_x+w_7)$
8	$\dot{y}$	=	$\underline{n_yT/m} + C_8(b_y+w_7)$
9	$\dot{z}$	=	$\underline{n_zT/m} + C_8(b_z+w_7) + \underline{32.2}$
10	$\dot{x}$	=	$\dot{x}$
11	$\dot{y}$	=	$\dot{y}$
12	$\dot{z}$	=	z
13	$\dot{b}_x$	=	$C_9b_x + w_{13}$
14	$\dot{b}_y$	=	$C_{10}b_y + w_{14}$
15	$\dot{b}_z$	=	$C_{11}b_z + w_{15}$
16	$\ddot{a}_1$	=	$C_{12}\dot{a}_1 + C_{13}a_1 + w_{16}$
17	$\ddot{b}_1$	=	$C_{14}\dot{b}_1 + C_{15}b_1 + w_{17}$
18	$\ddot{a}_0$	=	$C_{16}\dot{a}_0 + C_{17}a_o + w_{18}$
19	a_1	=	a_1
20	b_1	=	b_1
21	$\dot{a}_0$	=	a_0
22	$\dot{c}$	=	$0 + w_{22}$

Note Underlined terms are zero for the C G Tracker

Table 2. State Equations

$$\phi_m = \phi + v_1 \qquad \dot{R} = [x\dot{x} + y\dot{y} + z\dot{z}]/[x^2+y^2+z^2]^{1/2} + v_7$$

$$\theta_m = \theta + v_2 \qquad \dot{\eta} = (x\dot{y} - y\dot{x})/(x^2+y^2) + v_8$$

$$\psi_m = \psi + v_3 \qquad \dot{\xi} = [z(x\dot{x}+y\dot{y}) - \dot{z}(x^2+y^2)]/[(x^2+y^2+z^2)(x^2+y^2)^{1/2}] + v_9$$

$$R = [x^2+y^2+z^2]^{1/2} + v_4 \qquad a_{1_m} = a_1 + v_{10}$$

$$\eta = \tan^{-1}(y/x) + v_5 \qquad b_{1_m} = b_1 + v_{11}$$

$$\xi = \tan^{-1}[-z/(x^2+y^2)^{1/2}] + v_6 \qquad a_{o_m} = a_o + v_{12}$$

Table 3. Measurement Equations

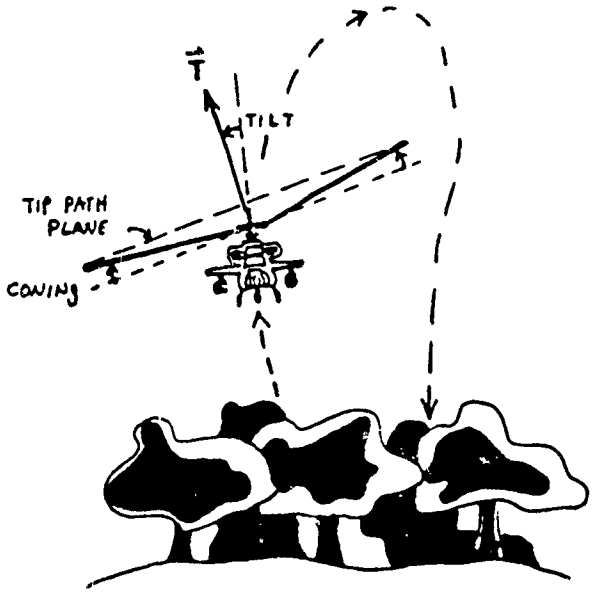


Figure 2. Helicopter Tip-Path Plane

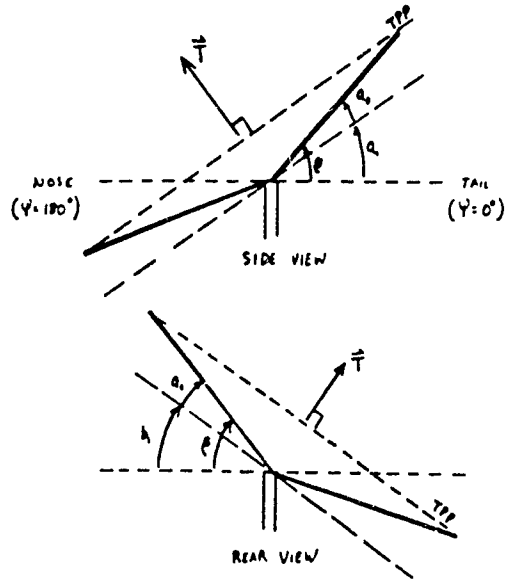


Figure 3. Helicopter Rotor Angles

Thrust vector divided by mass

$$\frac{\vec{T}}{m} = \frac{T\hat{n}}{m} = \frac{T}{m}(n_x\hat{i}_x + n_y\hat{i}_y + n_z\hat{i}_z)$$

where the unit vector $\hat{n}$ is given by

$$n_x = -(\cos\psi\cos\theta)(\sin a_1 \cos b_1)$$
$$+ (\cos\psi\sin\theta\sin\phi - \sin\psi\cos\phi)(\cos a_1 \sin b_1)$$
$$- (\cos\psi\sin\theta\cos\phi + \sin\psi\sin\phi)(\cos a_1 \cos b_1)$$

$$n_y = -(\sin\psi\cos\theta)(\sin a_1 \cos b_1)$$
$$+ (\sin\psi\sin\theta\sin\phi + \cos\psi\cos\phi)(\cos a_1 \sin b_1)$$
$$- (\sin\psi\sin\theta\cos\phi - \cos\psi\sin\phi)(\cos a_1 \cos b_1)$$

$$n_z = +\sin\theta(\sin a_1 \cos b_1)$$
$$+ \cos\theta\sin\phi(\cos a_1 \sin b_1)$$
$$- \cos\theta\cos\phi(\cos a_1 \cos b_1)$$

$$T/m = C_{18}(a_0 + C) + C_{19}$$

Table 4. Helicopter thrust model as a function of fuselage and rotor orientation.

to a tracker that models only translation. It is a 9 state estimator that uses states 7 through 15 of Tables 1. and 2. It uses 6 measured values (4-9 of Table 2.4).

The extended Kalman filter estimates the current aircraft state based upon the measurements. A prediction is made of the future aircraft position by integrating the state equations forward in time. The predicted time into the future is based on the bullet time of flight from the tracker to the target. In this study the bullet time of flight was always exactly two seconds. The integration of the state equations was accomplished in such a way that linear target acceleration over the bullet time of flight was constant. This resulted in a second order prediction algorithm using the exact time of flight.

3. Experimental Results:

The helicopter maneuver studied in this paper is described as follows. The helicopter is traveling in the X direction at 30 m/sec with a 10 knot tail wind. After 8 seconds into the maneuver the pilot pulls the nose up, then pushes the nose down, and pulls it back up again at 12 seconds. Overall, a 10 meter pop-up and dive returning to starting height is accomplished during a 10 second period of the flight maneuver. The maximum incremental vertical acceleration of 0.9g occurs between 10 and 12 seconds into the maneuver. The maneuver lies almost entirely in the X-Z inertial plane starting at X = 2000m, Z = -5m (corresponding to a height of 5m above ground level), and X = 1450m.

Simulated helicopter data was obtained from NASA-Ames Research Center via a private communication (Reference 8). The helicopter simulated is a generic one; i.e., it represented no known helicopter. To provide the 12 measurements required for the tracker program , radar data was simulated from the inertial positions and velocities and all measurements were corrupted with gaussian uncorrelated noise. The root mean square values of this noise were as follows: for all Euler and rotor angles, 35 mrad; for range, 5m; for velocity, 5m/sec; for elevation and azimuth angle, 1.4 mrad; for elevation and azimuth angle rates, 0.5 mrad/sec. The signal to noise ratios for the r.m.s. motion of the Euler and rotor angles in the maneuver were approximately 1, which was not at all high.

The metric of performance used here computes the mean and standard deviation of the azimuth error and elevation error over a 2 second shell burst interval, which extends from 13 to 15 seconds into the maneuver and covers the region of maximum vertical acceleration. With proper tuning of the fuselage and rotor-blade states, it was determined

that the acceleration bias states (13, 14, 15) were of little help in the Maneuver Estimator so they were deleted from further study. This result was significant since it suggests that three state variables can be deleted from the Maneuver Estimator leading to computational savings.

The results for the complete Maneuver Estimator (19 states) with all measurements (12) are summarized below in Table 5.

	Mean Error (mrad)	Std.Dev.Error (mrad)
Elevation Angle	0.48	1.07
Azimuth Angle	0.09	0.15

Table 5. Maneuver Estimator Elevation and Azimuth Angle Prediction Accuracies

The results of deleting certain measurements from the Kalman filter such as tip-path plane measurements, Euler body angles, and velocity are shown below in Table 6.

Measurements Used	Mean Azimuth Error (mrad)	Std. Dev. Azimuth Error (mrad)
All	0.09	0.15
Euler, Pos. and Veloc.	0.02	0.17
Position and Veloc.	-0.27	0.30
Position	-0.17	0.78
Euler and Position	0.76	0.50

Table 6. Maneuver Estimator Azimuth Angle Prediction Accuracies Without Certain Measurements

The results show that there is little penalty in deleting the tip-path plane measurements. This is because the signal level is small compared to the noise. It is believed that the tip-path plane measurements will not be helpful unless they can be measured more accurately than assumed in this experiment or unless the motion of the tip-path plane is larger. Larger tip-path plane motion is possible in larger amplitude maneuvers and in certain helicopters with tetering rotor systems. The effect of body angles can be seen by comparing the second and third entries in Table 6. It is apparent that body angles

do substantially reduce azimuth angle errors. However, body angles were found to have no effect on elevation error. The benficial effects of measuring body angles, while not gigantic in this experiment, can easily become larger when the body angles become larger. Any helicopter can achieve larger body angles by simply maneuvering nonvigorously. It is felt, therefore, that measuring body angles is beneficial to tracking of helicopters.

The results for a tuned C. G. Tracker, typical of present-day radar-based fire control systems which model only linear translation, are summarized below in Table 7 for position and velocity, and position-only measurements.

Measurements Used	Mean Elev. Error(mrad)	Std.Dev. Elev. Error(mrad)	Mean Azimuth Error(mrad	Std.Dev. Azimuth
Pos. + Vel.	-0.87	2.07	-0.39	0.31
Position	-4.24	0.68	-0.23	0.71

Table 7. C. G. Estimator Azimuth and Elevation Angle Prediction Accuracies

These results show that the azimuth and elevation overall errors (mean error $\pm$ standard deviation) are reduced by including the velocity measurement.

A comparison of the best maneuver estimator with the best C. G. tracker shows clearly that the maneuver estimator produces far superior performance. This is shown below in Table 8.

Tracker Elevation	Mean Elevation Error(mrad)	Std. Dev. Error(mrad)
Maneuver Estimator	0.48	1.07
C. G. Tracker	-0.87	2.07
Maneuver Estimator uses same measurements as C. G. Tracker	0.47	1.18

Table 8. Comparison of Elevation Angle Prediction Accuracies of C. G. Tracker with Maneuver Estimator

It is clear that both the mean and standard deviation of the elevation error for the C.G. Tracker are approximately twice those for the maneuver estimator. Now, recall that the Manueuver Estimator use 12 measurements while the C. G. Tracker uses only 6 measurements (i.e., position and velocity). If the measurements used in the maneuver estimator are made the same as for the C. G. Tracker the Maneuver Estimator still outperforms the C.G. Tracker by the same significant 2 to 1 margin. This would strongly suggest that in a low signal to noise environment measuring body angles and blade angles was not as important as including them in the tracker mathematical model. This is a very significant observation that could obviate the use of the supplementary optical sensor in Figure 1 and allow a reduced computational burden in a high-noise tactical environment without a loss of effectiveness of the fire control system.

4. Conclusions:

A new trajectory estimator and predictor for maneuvering helicopters has been investigated for a pop-up and dive maneuver and found to out-perform the present day, radar-based, center-of-gravity tracker by a 2 to 1 margin. The new tracker uses optically derived body attitude and blade system measurements. The rotor measurements were shown to be of limited value in a noisy environment, while the body angles will be considerably more useful in tracking a helicopter performing evasive maneuvers such as loops and rolls. The new tracker still out-performs the C.G. tracker when the body attitude and rotor system angles are not measured indicating that these characteristics should be included in the mathematical model for any new radar-only tracker as well as the optically supplemented tracker proposed here.

References:

1. Andrisani, D., Kuhl, F. P., Gleason, D., "A Nonlinear Tracker Using Attitude Measurements, "IEEE Transactions on Aerospace and Electronic Systems, September 1986, Vol. AES-22, No. 5.

2. Kuhl, F. P. Hutchings, T., Andrisani, D., "Aircraft Lead Angle Prediction," presented at the Eighth Meeting of the Coordinating Group on Modern Control, Oakland University, Rochester, MI, February 4-5, 1987.

3. Kuhl, F. P., Reeves, A. P., Taylor, R. W., "Shape Identification with Moments and Fourier Descriptors", Proc. of 1986 ACSM-ASP Convention, Washington, D.C., March 16-21, 1986.

4. Gorman, J. W., Kuhl, F. P., "Partial Shape Recognition and Orientation Estimation", Proc. of 1988 ACSM-ASP Conv., St. Louis, MO, March 13-18, 1988.

5. Grogan, T. A., Kuhl, F. P., "Evaluation of Walsh Shape Descriptors" ,Proc. of 1988 ACSM-ASP Conv., St. Louis, MO, March 13-18, 1988.

6. Gelb, A., Applied Optimal Estimation, The M.I.T. Press, Cambridge, MA, 1974.

7. Andrisani, D., Kuhl, F., Schierman, J., "Helicopter Tracking Using Attitute Measurements", Proc. of 8th Meeting of Coordinating Group on Modern Control Theory, Oakland University, Rochester, MI, February 4-5, 1987.

8. Personal Communication with Dominick Andrisani (317-494-5135), Purdue University School of Aeronautics and Astronautics, Lafayette, IN.

Nanometer Device Fabrication Techniques for Advanced Military Microelectronics (U)

*Joseph H. Kwiatkowski, Mr., Charles F. Cook, Mr., Monica Taysing-Lara, Ms., Doran D. Smith, Dr., Linda S. Heath, Ms., and Gerald J. Iafrate, Dr.
US Army Electronics Technology and Devices Laboratory (LABCOM)
Fort Monmouth, New Jersey 07703-5000

INTRODUCTION

The twenty-first century military requirements for real-time information acquisition and processing in tactical EW, C3I, and smart munitions establish a need for superlarge-scale and superhigh-speed integrated circuit microelectronics. In the 1990's current technology will be developed to the point of diminishing returns due mainly to anticipated device time delays arising from interconnect resistance, long signal paths, electromigration, and the like. Thus, even at maturation, the existing technologies will fail to provide the required ultrahigh-speed IC technology for military needs.

Previous papers[1-5] have discussed the theoretical basis and physics of a number of devices which promise the possibility of superspeed (picosecond) switching, superhigh-frequency (100GHz to greater than 1000GHz), mm-wave components and oscillators. These new devices have very small feature sizes (100nm to 10nm) and cannot be made by conventional fabrication techniques. They require the use of highly controlled molecular beam epitaxy (MBE) for the growth of different layers, precise electron beam lithography, and highly controlled anisotropic reactive ion etching (RIE).

This paper will give a brief introduction to MBE and RIE as well as a more detailed discussion of our Nanometer Lithography facility. The experimental section will focus on the fabrication of a mini-FET although the procedures can be adapted to a wide variety of devices.

MOLECULAR BEAM EPITAXY

The use of Molecular Beam Epitaxy (MBE) for the fabrication of active device layers has been extensively reviewed.[6,7] Also, a review of conventional microwave and millimeter-wave devices fabricated from epitaxial layers of Gallium Arsenide and its alloys has been published[8].

Essentially, molecular beam epitaxy is an ultrahigh vacuum evaporation process wherein molecular or atomic beams of film constituents are formed by evaporating pure materials from heated effusion cells. The beams, upon passing through shuttered apertures, converge onto a heated substrate where they react to form the semiconductor film. Continuous changes in chemical composition can be achieved by programmed variation of effusion-cell temperature, while abrupt changes are obtained by using shutters which are interposed between effusion cell and substrate. Typical growth rates vary between 10 and 20 nm/min depending upon material arrival rates and substrate temperature. Dopant growth is a two-step process wherein the impurity atom is first absorbed into the substrate surface where it is subsequently incorporated into the growing film. Accordingly, impurity atoms such as Be, Si, Ge, which have low vapor pressures and unity sticking coefficients are most generally used.

At ETDL, we have a Varian model 360 MBE which has been modified to upgrade its capability as well as a Varian Gen II. The Gen II has eight effusion cells enabling the growth of both p and n type GaAs, AlGaAs, and others. It also has the capability of handling up to two inch wafers. The mini-FET's, described in the experimental section were fabricated with a variety of epitaxial layers on both p and semi-insulating substrates. Other layers are grown routinely for such projects as High Electron Mobility Transistors (HEMT's), superlattice research, and a variety of quantum devices.

LITHOGRAPHY

One of the more difficult and extremely essential steps in the fabrication of these devices is the lithographic patterning. Because of the very small feature sizes necessary (<100nm), even conventional electron +beam lithography cannot be used. Until the development of a high voltage lithography system[9], one had the option of modifying a Scanning Electron Microscope (SEM)[9], or a Scanning Transmission Electron Microscope (STEM)[1,10-12]. The latter two options, while being much less expensive, require the integration of independently constructed beam blanker, pattern generator (PG), stage drive, and computer system. We have modified a Philips EM-420T STEM to do the nanometer lithography[1, 11, 12] for the following reasons; (1.) It's high accelerating voltage (to 120 keV), greatly reduces proximity effects[13], allowing nanometer sized features to be written. (2.) The excellent regulation of the lens current and high-voltage power supplies. (3.) The minimum usable spot size (< 2.5 nm) is more than adequate to expose the resist with reasonable dwell times. (4.) It's capability to examine finished devices at high magnification (>50,000X). (5.) Beam telecentricity. (6.) A very accurate specimen stage that can be driven by stepping motors that provide a 2.0 nanometer-per-step resolution capability. (7.) It's use as an analytical Transmission Electron Microscope (TEM) is not impeded, since only one switch converts between the two modes. A schematic of the STEM/lithography system is shown in Figure 1.

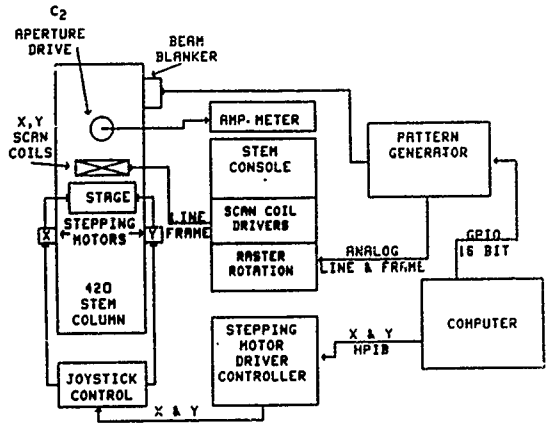


Figure 1. Schematic of the STEM/lithography system.

The pattern generator and the computer aided design (CAD) program were both designed and built in house and interfaced to the beam blanker and the stepping motors which can then be controlled by a joystick or from the computer. The sample size is limited to a 5 x 8 mm rectangle on which the center 2x2 mm is the writing area. This sample size is large enough, however, that up to 110 active areas can be written on one chip, depending on the particular device. For example, the Surface Acoustic Wave (SAW) devices have one device per chip, while the mini-FET has 110.

The CAD program is a fairly simple one consisting of about 260 lines which enables dots, lines, rectangles, grids, circles, and other figures to be written. (Fig. 2-7).

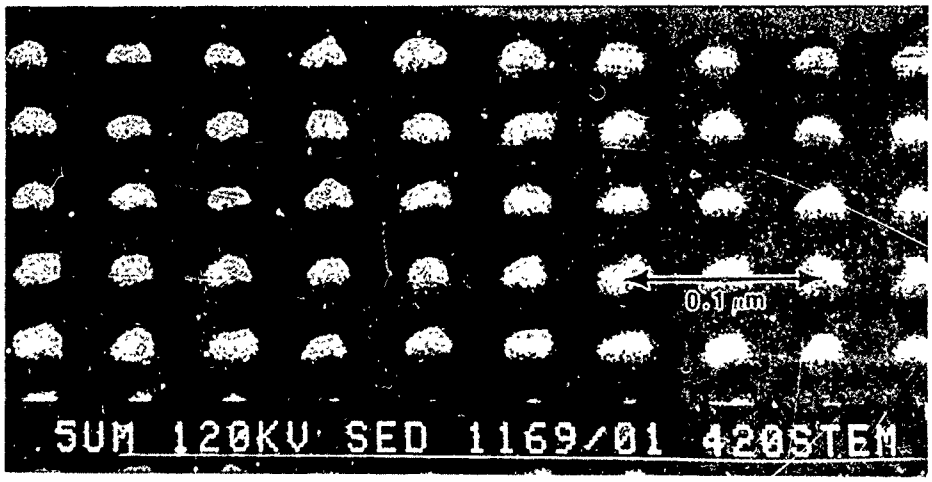


Figure 2. SEM of 20 nm Ag dots spaced 50 nm apart.

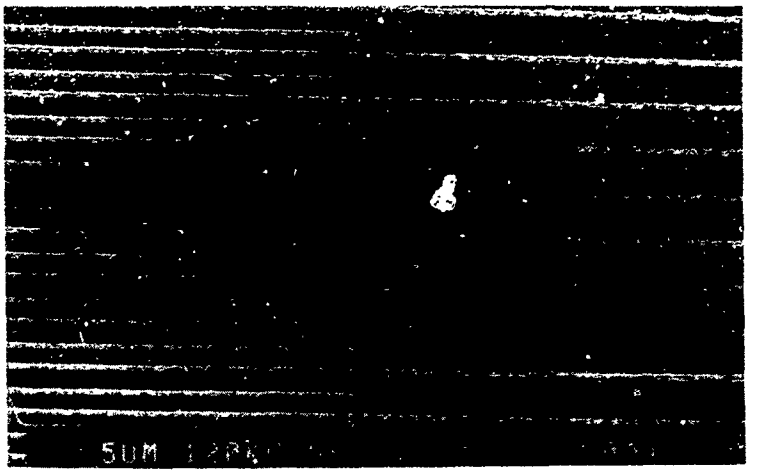


Figure 3. SEM micrograph of a sub-100 nm prototype SAW device.

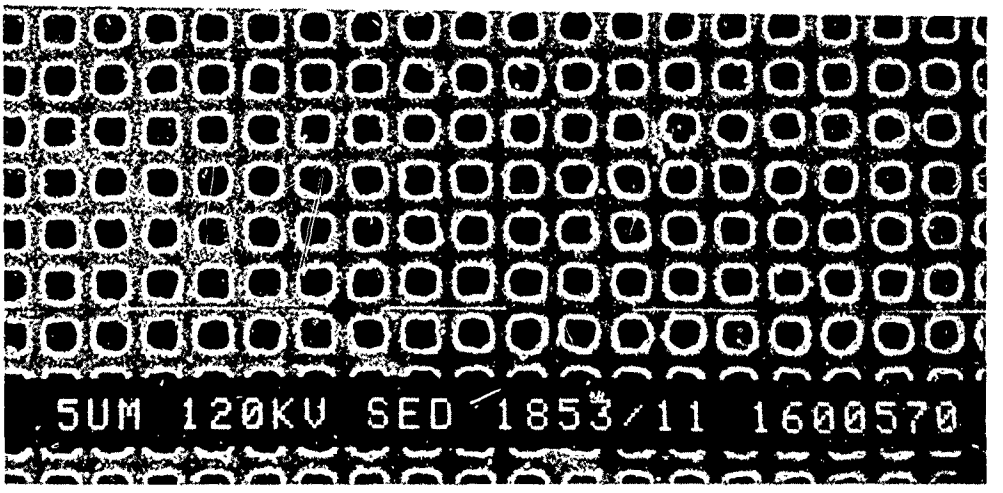


Figure 4. An SEM micrograph of a 127,000 mesh AuPd grid consisting of an 80 nm bar width and a 120 nm space.

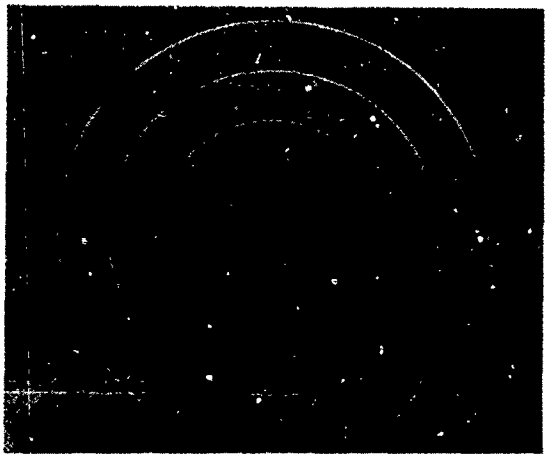


Figure 5. Scanning Electron Micrograph of
100 nm gold/palladium concentric circles

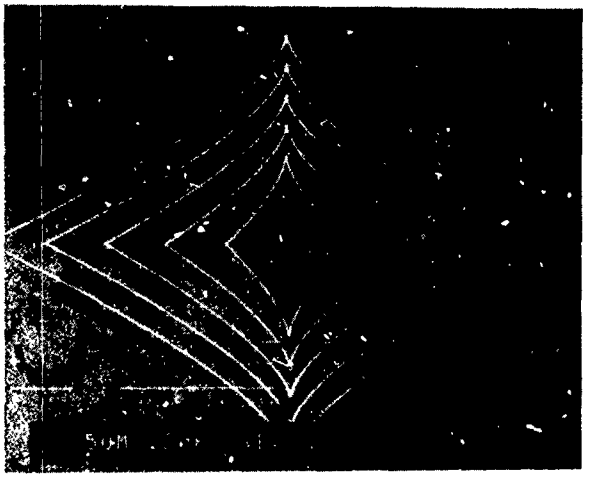


Figure 6. Scanning Electron Micrograph of 100nm non-orthogonal lines
demonstrating unique pattern generation capability

Figure 7. Scanning Electron Micrograph of three leaved rose pattern demonstrating unique pattern generation capability

Since the exposure time depends to a certain extent on the density of the pattern, the program was designed to easily enable changing the dwell time either in a controlled preset fashion, or by simply changing it manually. Once the design is made, it can be centered in the field, or moved to any location within the field. It can then be stored in a file, so that it is easily accessible for subsequent writing. The field sizes are generally from 15 to 100 microns square although they can be made smaller. The program can also be used to automatically move the stage drive to the next location and writing started without intervention by the operator.

Since electrons are used both to see the sample and to expose the resist, unwanted exposure would result if the active area was not protected in some way during the registration procedure. Two different registration techniques have been developed for this system. The first uses a pulse-generator placed in series with the digital beam-blanking data line from the PG to the beam blanker. By carefully adjusting the width and length of the pulse train and triggering on the horizontal scan line, the area of the field to be written can be blanked, while the unblanked portion is used to perform visual alignment of the fiducial marks on the device. Careful rotation and translation adjustments within this mode will give acceptable alignment results. The devices fabricated in this paper have been written using this method.

The second method uses the stepping motor stage controller to accurately measure the distance in x and y between successive devices on a chip[12]. The distance is fed back into the computer along with the number of devices per row and the number of rows. The stage controller moves the stage to the correct position, the beam is unblanked and the device is

written. The beam is then blanked and the stage moves to the next position and so on until the entire chip is written. The advantage of this method is that it is automatic and involves only one manual alignment. The disadvantage at this time is that only one pattern can be input at a time and that the dwell time cannot be set randomly for each device, but only increased or decreased in a preset fashion. This is a problem when attempting to write gates with different lengths on the same chip. Software modifications are currently being investigated to address this problem.

ETCHING

Etching, or the controlled removal of material is a crucial aspect of wafer processing. In the fabrication of some types of MESFET'S, this procedure can be required on no less than four occasions, namely a pre-implant etch, mesa etch, pre-metallization etch, and recess etch. These processes serve to clean the surface, remove unwanted or damaged material, and to define and tailor the active areas in order to fabricate a device with specific properties. Etching can be done either by 'wet' chemical methods or 'dry' processing. Wet chemical methods are still widely used, but dry etching techniques are becoming more widespread. A review of these processes has been done by Woodward.[14] At ETDL, we have both a reactive ion/plasma etcher and a magnetron ion etcher (MIE). A variety of types of structures have been etched including HEMT's, mini-FET's, and grids. The structure shown in Fig. 8 was done with Freon 12 and represents a large structure on the order of 10 microns. This was one of the early attempts.

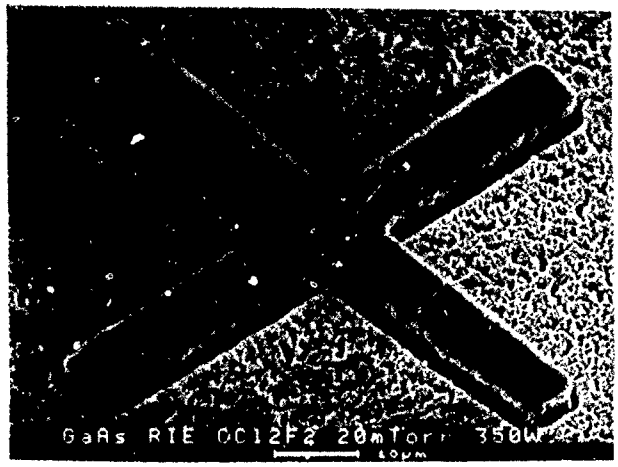


Figure 8. Scanning Electron Micrograph of large (10micron) feature etched with a Freon 12 Reactive Ion Etch

The grid shown in Fig. 9 was done in a HCl reactive etch plasma and represents linewidths on the order of 70nm. This represents one of the smallest

structures etched to date in our laboratory.

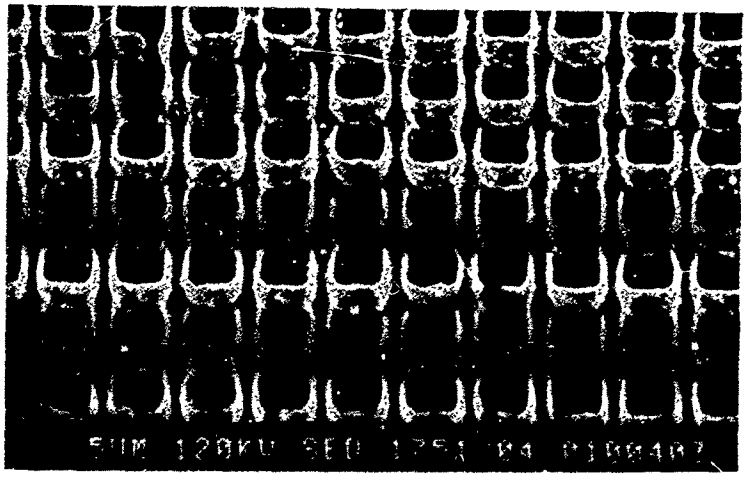


Figure 9. Scanning Electron Micrograph of Grid shown in Fig. 4 after reactive ion etching with a hydrogen chloride plasma

EXPERIMENTAL

This experimental section describes the fabrication of a mini-FET, whose scematic is shown in Fig. 10, but can be adapted to a variety of other devices with the appropriate modifications.

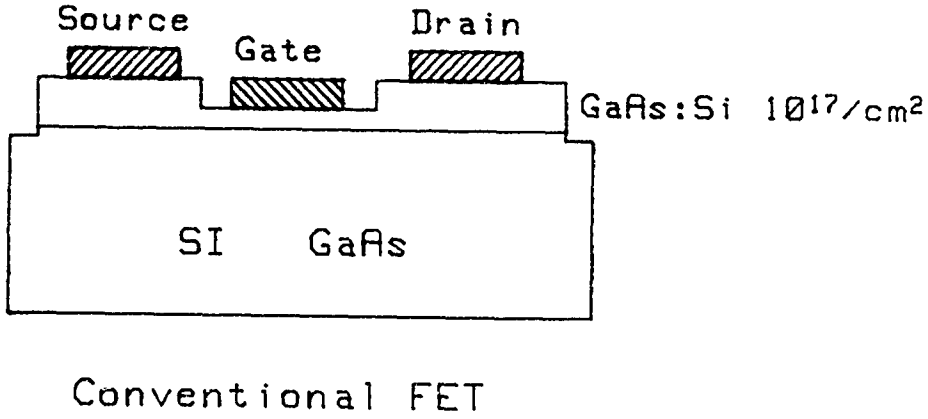


Figure 10. Schematic of conventional field effect transistor showing source, gate, and drain regions and layers.

The wafers used in the fabrication of the mini-FET's were either p-type or semi-insulating. 0.9 micron of undoped GaAs followed by either 100nm $5X10^{17}$ n-type GaAs or 50nm $1X10^{18}$ n-type(both Si doped) was grown by

MBE. The ohmic metal contact pattern was then defined by conventional photolithography using the chlorobenzene liftoff process. The metals used were 5nm Ni/30nm Ge/60nm Au. The contacts were alloyed using a rapid thermal annealing sequence at 450 C for 60 seconds. A second optical lithography step was used to define the active area, followed by an etch for isolation. The etch used is a combination of phosphoric acid and hydrogen peroxide and leaves a sloping edge to the mesa which allows step coverage. The wafer was then scribed into 5 x 8 mm chips which can fit into the STEM. The chip layout has 110 devices with gate widths of 38, 30, 20 and 10 microns. This can be seen in Fig. 11.

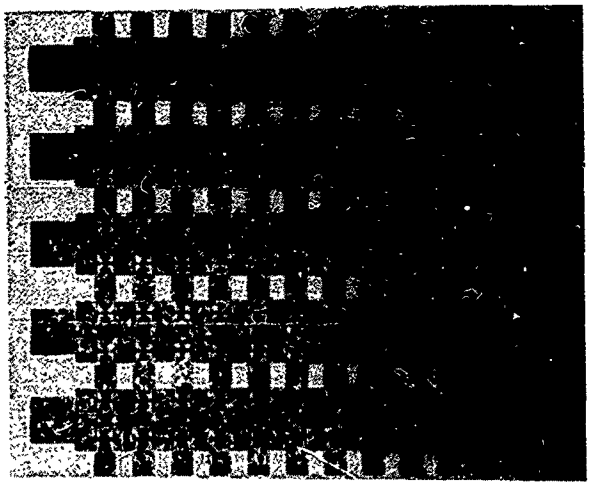


Figure 11. Optical Micrograph of layout of miniFET chip showing 110 devices in a 2mm square

A higher magnification picture showing a 10 micron width with the gate already written is shown in Fig. 12. The resist used was polymethylmethacrylate (PMMA) 950k from KTI diluted to 2% with chlorobenzene. The resist thickness was either 100nm or 150nm depending on the sample. It was found that the thicker resist allowed for easier liftoff without degradation of the linewidth. The thicker resist also allowed for thicker gates. After spinning, the sample was baked at 160 C overnight. The sample was exposed at 120 keV using a beam size of 4nm and dwell times of 300 to 1000 microseconds. Both single and multiple pass lines were made. All lines were made by spacing dots closely together (15-20nm) and overexposing. The resist was developed either in 3:7 2-ethoxyethanol/methanol for 10 sec. or in 1:3 methylisobutylketone(MIBK)/2-propanol for 90-120 seconds. The MIBK process seems to allow more process latitude. The gate metal was then evaporated by e-beam evaporation and the excess lifted off in warm trichloroethylene. Metals used for the gates were either Ti/Pt/Au, Ti/Au, or Al. Thicknesses ranged from 40 to 75nm. Following the gate metal, a gold plate up step was done both to facilitate contact to the gate region and also to facilitate bonding to the contacts. The gate lengths fabricated were from 200nm down to about 60nm. Electrical testing revealed that the

gates were continuous. Further testing of the finished devices is inconclusive at this time.

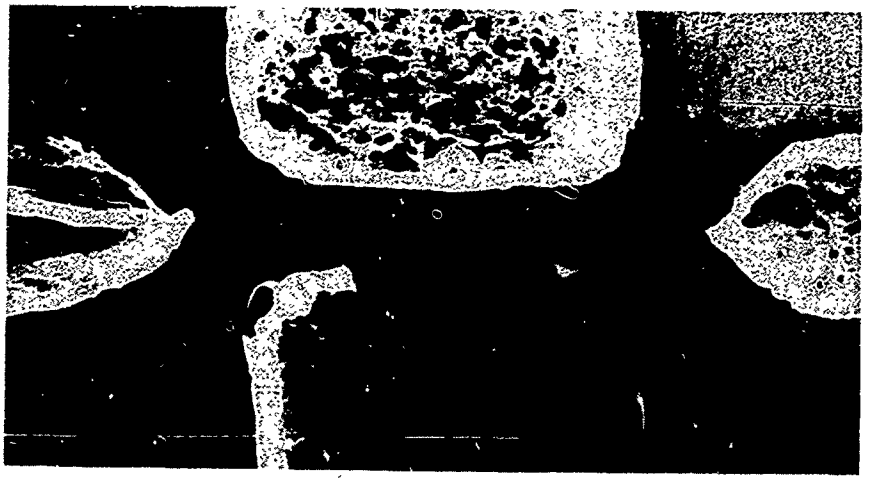


Figure 12. Scanning Electron Micrograph of a gold/palladium gate miniFET with a gate length of 70nm.

SUMMARY

We have developed and demonstrated the capability to fabricate nanometer sized devices by the use of molecular beam epitaxy, novel nanometer lithography, and reactive ion etching. These devices promise the possibility of high-speed switching for high-data rate information processing; they also have the capability of being used in mm-wave microelectronics applications, as well as for studying quantum scale physics.

ACKNOWLEDGMENTS

The authors would like to thank the following for their contributions: P. Newman for growth of the MBE layers, M. Wade for the metal deposition, and D. Eckart and M. L. Saunders for photographic assistance.

REFERENCES

1. G. J. Iafrate, T. R. AuCoin, C. F. Cook, Jr., and J. H. Kwiatkowski, Proceedings 1984 Army Science Conference. Vol 4, P.95.

2. G. J. Iafrate, D. K. Ferry, and R. K. Reich, Surface Science 113, 485 (1982).

3. R. K. Reich, R. O. Grondin, D. K. Ferry, and G. J. Iafrate, Phys. Lett. 91A, 28 (1982).

4. R. K. Reich, R. O. Grondin, D. K. Ferry, and G. J. Iafrate, Electr. Dev. Lett. 3, 381 (1982).

5. G. J. Iafrate in 'The Physics of Submicron Structures' p.301 ed. by H. L. Grubin, K. Hess, G. J. Iafrate, D. K. Ferry Plenum Pub. Corp. 1984.

6. R. L. Ross, D A. Dekanski, R. O. Savage, R. A. Lux, and T. R. AuCoin, Proceedings of the 1985 Producibility of Microwave and Millimeter Integrated Circuits Conference, U. S. Army Missile Command Redstone Arsenal, AL.

7. K. Ploog and K. Graf, 'Molecular Beam Epitaxy of III-V Compounds, A Comprehensive Bibliography 1958-1983', Springer Verlag, Berlin (1984).

8. M. Hassel Shearer, H. Takemura, J. Vac. Sci. Technol. B 4(1), 64 (1986)

9. T. H. P. Chang, M. Hatzakis, A. D. Wilson, A. N. Broers, Electronics, vol. 89, May 12, 1977.

10. H. G. Craighead, R. E. Howard, L. D. Jackel, P. M. Mankiewich, Appl. Phys. Lett., vol. 42, p. 38 (1983).

11. C. F. Cook, Jr., T. R. AuCoin, G.J. Iafrate, Solid State Technol. 28 (10), 124 (1985).

12. C. F. Cook, Jr., J. H. Kwiatkowski, D. D. Smith, H. R. Berkowitz, Mat. Res. Soc. Symp. Proc. Vol. 76 (1987) p. 61.

13. P. M. Mankiewich, L. D. Jackel, and R. E. Howard, J. Vac. Sci. Technol. B 3, 174 (1985).

14. J. Woodward in 'Gallium Arsenide for Devices and Integrated Circuits', p. 143 ed by H. Thomas, D. V. Morgan, B. Thomas, J. E. Aubrey and G. B. Morgan, P. Peregrinus Ltd. London (1986).

Fractal Properties of Fracture Surfaces of Energetic Materials

*M. Yvonne D. Lanzerotti[+], Dr.
James J. Pinto[+], Mr.
Allan Wolfe[++], Prof.
Scott Morrow[+], Dr.
+U. S. Army Armament Research, Development and Engineering Center
Picatinny Arsenal, NJ 07806-5000
++New York City Technical College, Brooklyn, NY 11201-0000

INTRODUCTION

An innovative approach has been devised to study the behavior of energetic materials during high acceleration by using an ultracentrifuge (1-4). Knowledge of the fracture phenomena of energetic materials subjected to high acceleration is important in ordnance and propulsion applications where energetic materials are often subjected to high, fluctuating, and/or sustained accelerations. For example, explosives in projectiles are subjected to high setback acceleration during the gun launch process. Grain fracture of propellants during acceleration can result in an abnormal propellant burning rate. Extensive studies have yielded highly significant results which are of considerable importance in the understanding of major munitions and propulsion problems. In addition, this work is particularly relevant to the future development of insensitive energetic materials subjected to higher acceleration.

The mechanical behavior of explosives and gun propellant has been studied during high acceleration above 10,000 g (1-3). Explosives are found to behave as brittle solids; particles fracture above an acceleration threshold (1). M30 propellant deforms as a function of maximum applied stress (3). Above an acceleration threshold recovery of the propellant deformation is observed under zero stress. Failure of both explosives and gun propellant occurs when the tensile or shear strength is exceeded (1-3).

The fracture behavior of TNT and Composition B has been studied as a function of time, temperature, and loading rate (3). At constant acceleration the fractured mass of TNT and Composition B increases with time and temperature. The fractured mass of TNT increases linearly with loading rate; the fractured mass of Composition B increases with loading

rate but the relationship is not linear.

Chemical reactions have been discovered to occur in TNT, TNT-1% HNS and M30 propellant as a result of fracture and impact during high acceleration (4). Values for acceleration, velocity, time,and load rate at fracture and at impact have been estimated. The reacted samples have been studied with electron paramagnetic resonance (EPR) and x-ray photoelectron spectroscopy (XPS) to verify that chemical reaction has occurred and to investigate properties of the reaction products. This chemical decomposition of TNT appears to be the same as the thermal and uv decomposition of TNT observed with EPR and XPS.

Highly important results germane to mechanical failure of energetic materials during high acceleration have been obtained (1-4). In a projectile, the particles of explosive which break loose are confined to a void. Small particles are more easily ignited and are a significant consideration for possible premature ignition in accelerated explosives. Also significant is the finding that as a result of fracture and impact the energetic materials can undergo chemical decomposition. Understanding these phenomena may be important for solving the problems of abnormal propellant burning and premature ignition or initiation of explosives and propellants.

In this work fractal geometry techniques have been adapted to determine the fractal dimension of the fracture surfaces of energetic materials (5-7). The fractal dimension is a parameter describing the scaling properties of surface topography; that is, how topography or roughness varies with surface size. In contrast, conventional measures of surface roughness, such as the root-mean-square roughness, vary with surface size and are valid only for a given sample size. The mechanical properties of two surfaces in contact depends strongly on the topography of the contacting surfaces. Therefore, knowledge of the fractal dimension has significant implications concerning the scaling of surface topography, mechanical properties of contacting surfaces, and the understanding of mechanical sensitivity tests including friction and impact.

These techniques are also invaluable tools to understand the fundamental physics of the fracture processes of energetic materials and to make quantitative comparisons between different materials. Since the fractal dimension provides concise quantitative descriptions of the fracture surfaces of energetic materials, improvements in cast and composition can be suggested and verified. Such quantitative descriptions are not presently available by any other method. Improved munitions survivability and reduced hazard can be achieved through improved knowledge of energetic material behavior in projectiles.

The fracture surfaces of specially prepared samples are obtained during high acceleration in an ultracentrifuge when the strength of the material is exceeded. Fracture profiles are obtained with optical microscopy or a stylus profilometer. The fractal dimension is determined by the divider method or power spectral analysis.

EXPERIMENT

A Beckman preparative ultracentrifuge model L8-80 with a swinging bucket rotor model SW 60 Ti is used to rotate the sample under study up to 60,000 rpm. The distance of the specimen from the axis of rotation can be chosen as a variable between 6 and 12 cm.

Samples are prepared as follows. Cylindrical polycrystalline plugs of explosive are prepared by pouring about one-half gram of the material into 9-mm i.d. polycarbonate tubes and allowing the material to crystallize. The open-ended sample tube is then joined to a short, closed-end polycarbonate centrifuge tube. The as-cast surface of the explosive faces away from the axis of rotation. The sample experiences a time rate of change of the acceleration up to a maximum acceleration. The sample then decelerates smoothly to zero acceleration. The initial maximum acceleration is less than the fracture acceleration for the material. The maximum acceleration for the sample is then increased systematically in each successive run. Particles break loose from the surface exposed to the acceleration and transfer to the closed-end tube when the strength of the material is exceeded. A hemispherical fracture surface is formed. In successive experiments, most of the prepared sample fractures and transfers to the closed-end tube; a hole is thus produced through the original sample.

Fracture profiles of the holes are obtained by optical microscopy at 30X magnification. The Z,X coordinates of the concave fracture surfaces are measured with a stylus profilometer. Two types of styli were used to measure the Z,X coordinates. One stylus is a diamond tipped, 90° included angle truncated pyramid with a 1.3-um edge in the direction of motion. The other stylus is a sapphire ball with 397-um radius. The Form Talysurf instrument is capable of recording profile lengths from 0.5 mm to 120 cm. The diamond stylus height range is 4.0 mm with 10 nm resolution. The sapphire stylus height range is 8.0 mm with 20 nm resolution.

RESULTS

Closed fracture profiles of melt-cast TNT and Composition B samples have been studied at temperatures of 0 °, 25 ° and 50 °C. At an acceleration of 50,000-200,000 g a hole a few mm in diameter is found to be produced through the original sample. The profile of this hole has been studied with optical microscopy at 30X magnification. An optical

micrograph of a fracture hole produced through a TNT sample at an acceleration of 50,000 g at 50 °C is shown in Fig. 1.

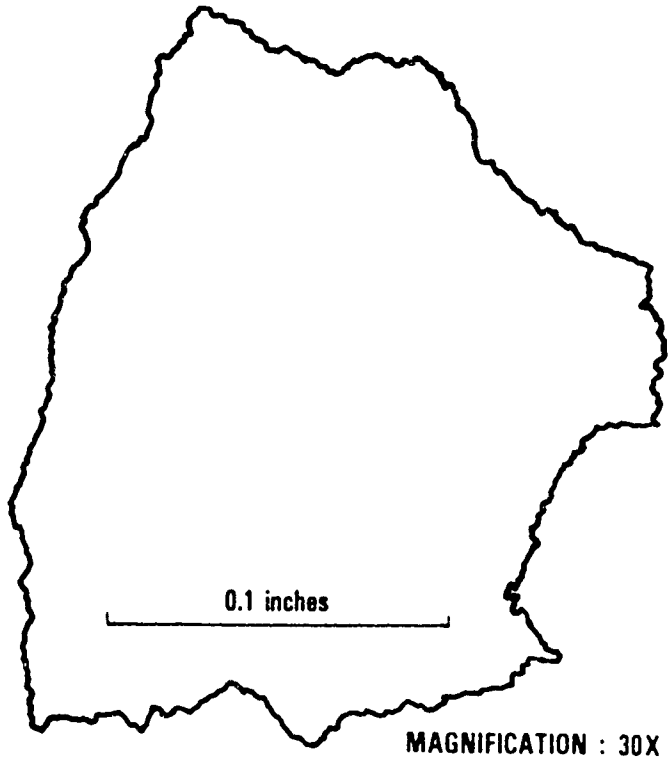


Figure 1. Optical Micrograph of Fracture Hole Through TNT Sample; Acceleration: 50,000 g; Temperature: 50 °C

The closed fracture profiles have been evaluated by the divider technique (8-10). The perimeter of the closed fracture profile is estimated by opening a pair of dividers to some ruler length, r, and walking them around the profile to determine the total number of steps. The perimeter is then equal to the number of steps times the ruler length, r. In general, the perimeter is not exactly equal to an integral number of ruler lengths. Therefore the ratio of the straight-line distance

between the final step and the end point of the perimeter to the ruler length is added to the number of steps (11,12). The perimeter is plotted as a function of ruler length on a log-log plot. The slope of the line is equal to 1-D, where D is the fractal dimension.

The perimeter of the 30X optical micrograph has been estimated with ruler lengths of 0.25, 0.5, 1.0, and 2.0 inches. The actual perimeter of the unmagnified profile is plotted in Fig. 2 as a function of ruler length on a log-log scale for two samples of TNT fractured at 50 °C.

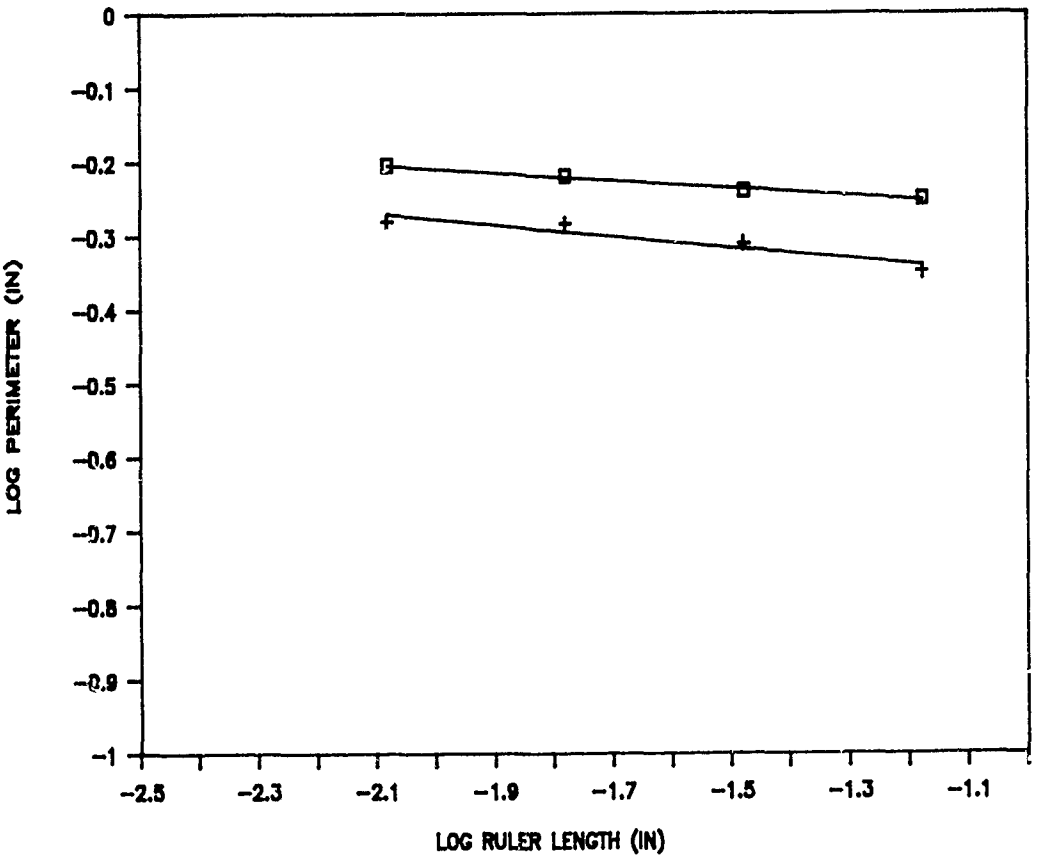


Figure 2. Log of the Perimeter Versus Log of the Ruler Length for Two Samples of Melt-cast TNT Fractured at 50 °C

The upper line (open squares) has been determined from the optical micrograph in Fig. 1. The slope of this line is -0.05. Thus, the fractal dimension of this profile is 1.05 over one order of magnitude in the range of 0.01" to 0.1". The lower line (crosses) has been determined from another fracture profile obtained under the same experimental conditions. The slope of the lower line is -0.08; the fractal dimension of the corresponding fracture profile is 1.08. Values obtained for the fractal dimension of TNT and Composition B as a function of temperature are given in Table 1. With this set of measurements it appears that the fractal dimension of TNT and Composition B increases slightly with temperature in the scaling range of 0.01" to 0.1". These data suggest an inverse relationship with the impact sensitivity of TNT (13).

Table 1. Fractal Dimension of TNT and Composition B as a Function of Temperature

Temperature (°C)	Dimension, D	
	TNT	Composition B
0	1.03	1.02
25	1.05	1.05
50	1.07	1.10

The topography of the fracture surface of TNT has also been studied from wavelengths of 8 cm to 1.0 um using a stylus profilometer. The sample of melt-cast TNT was found to fracture at 41,000 g at 25°C. The crystallites are visually seen to protrude from the surface and range in size from approximately 0.1 to 0.5 mm. The largest wavelength is the length of the profile. The smallest possible wavelength is twice the horizontal measurement spacing corresponding to the Nyquist cutoff frequency (14). In addition, the finite dimension of the stylus imposes a cutoff wavelength equal to the tip radius (15), 397 um for the sapphire stylus. The truncated pyramid diamond stylus cannot resolve a wavelength shorter than twice the length of the edge in the direction of motion, 2.6 um for the diamond stylus (16).

An 8-mm surface height profile of the fracture surface of TNT is shown in Fig. 3. The Z,X coordinates of this profile were measured with the diamond stylus at a horizontal spacing (mean sampling interval) of 0.006 mm. The vertical displacement is plotted as a function of horizontal displacement.

The spectrum analyses are performed by using a prolate spheroidal data window (17,18) in the horizontal space domain prior to the employment

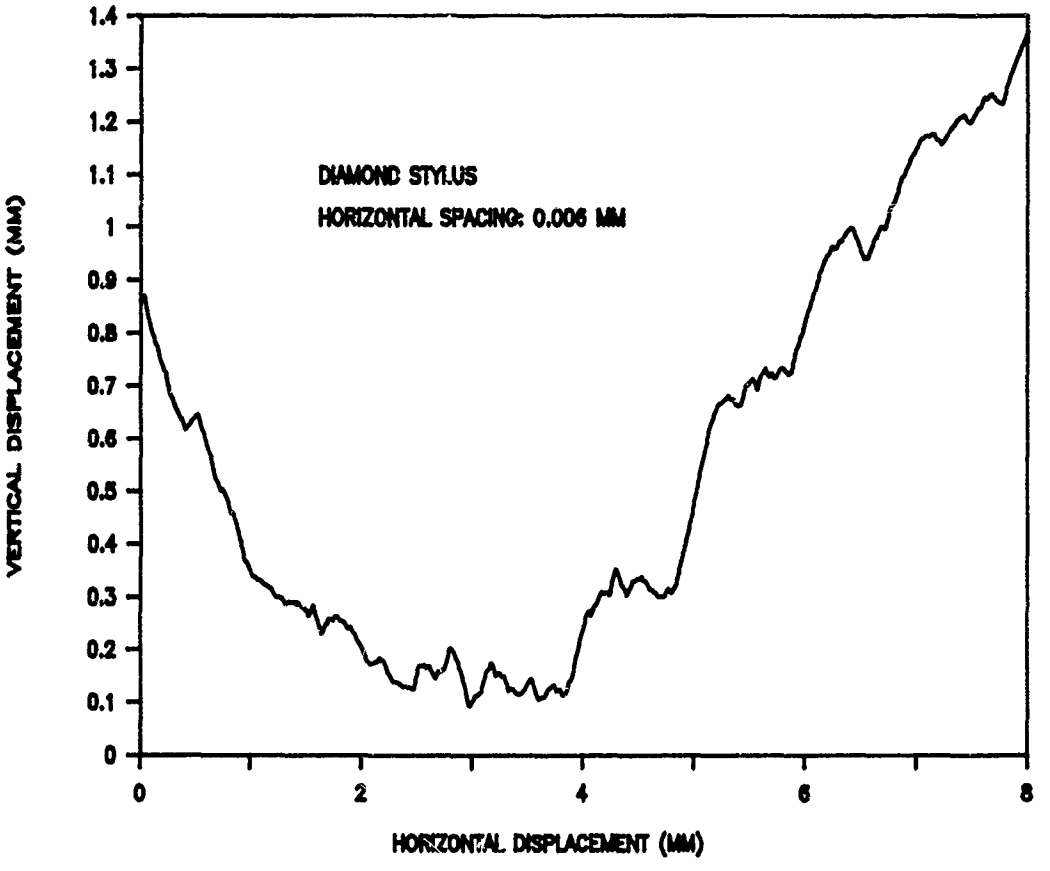


Figure 3. Surface Height Profile of the Fracture Surface of TNT

of a fast Fourier transform algorithm to compute the power spectral density (19). The prolate spheroidal window is used because it is superior to more commonly used windows in analyses of short time series (18). The spatial power spectrum of the fracture surface profile of the TNT sample of Fig. 3 is shown in Fig. 4. The power spectral density is plotted as a function of spatial frequency on a log-log plot. The power spectral density decreases with spatial frequency. At low spatial frequencies the slope (dashed line) of the spectrum is about -3. At high spatial frequencies the slope (solid line) of the spectrum is about -4. The break in slope corresponds approximately to the smallest grain size observed in the fracture surface (0.1 mm).

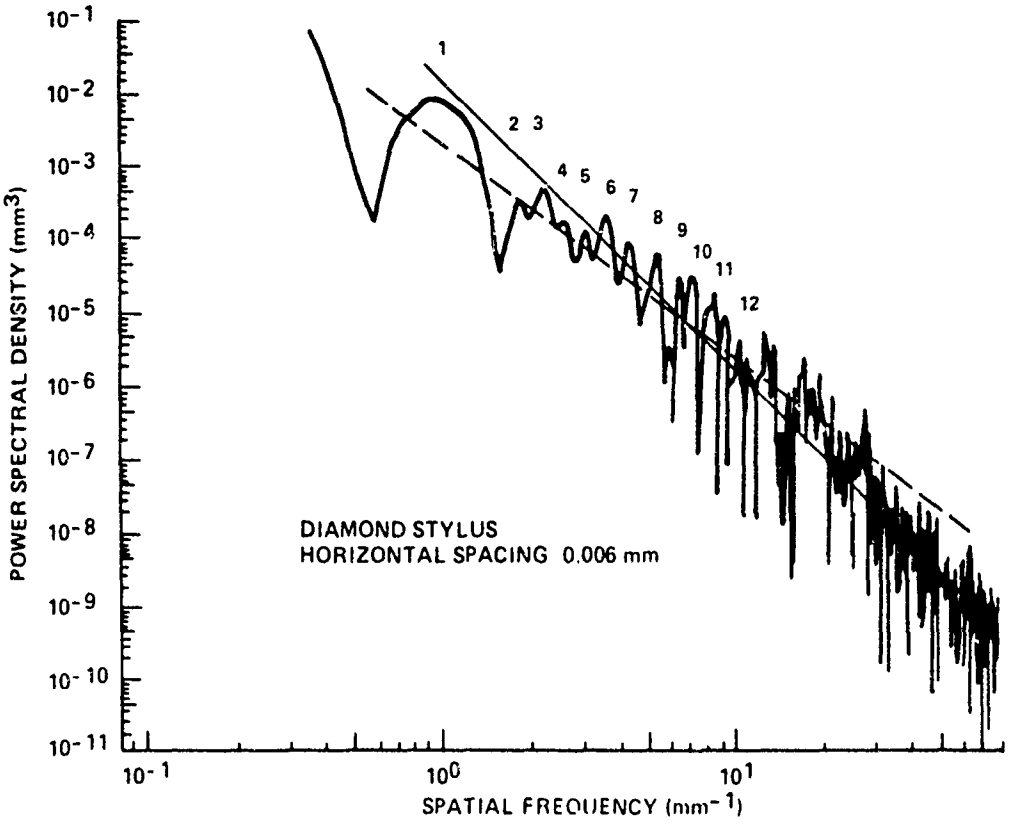


Figure 4. Spatial Power Spectrum of the Fracture Surface Profile of TNT

The fractal functions (8) which serve as models of surface profiles have power spectra whose slopes are given by

$$s = -(5-2D),$$

where D is the fractal dimension of the profile. In addition, the slope must lie in the range $-3 < s \leq -2$ (15) in order to yield reliable results for surface profiles. Therefore at low spatial frequencies the fractal dimension is found to be about 1. At high spatial frequencies, the slope is about -4. For spectral slopes steeper than -3, the relationship between spectral slope and fractal dimension does not hold (15). In this

case the dimension is exactly equal to 1, the topological dimension. Therefore at spatial frequencies greater than 10 mm^{-1} the process is non-fractal. That is, these fracture surfaces are not fractal at wavelengths smaller than the grain size when the profilometer stylus traverses smooth crystal cleavage surfaces.

Peaks are observed in the low frequency region of the power spectrum and correspond roughly to the size of the crystals in the sample. These peaks indicate that much of the fracture is occurring at crystal boundaries. The power spectra peaks are listed as a function of spatial frequency in Table 2. The power spectra peaks appear to be quasi-periodic over the range measured.

The spatial power spectrum has also been calculated from the Z,X coordinates of an independent fracture surface profile of TNT measured with the sapphire stylus. Similar results are obtained. The power spectral density decreases with spatial frequency. At low spatial frequencies the slope of the spectrum is about -3. At high spatial frequencies the slope of the spectrum is about -4. Again the break in slope corresponds approximately to the smallest grain size observed in the fracture surface. Peaks are also observed in the low frequency region of the power spectrum and correspond to the size of the crystals in the sample. The power spectra peaks in this power spectrum are also listed as a function of spatial frequency in Table 2. These spectral peaks also

Table 2. Power Spectra Peaks as a Function of Spatial Frequency

DIAMOND STYLUS HORIZONTAL SPACING = 0.006 MM		SAPPHIRE STYLUS HORIZONTAL SPACING = 0.005 MM	
PEAK NO.	SPATIAL FREQUENCY (MM^{-1})	SPATIAL FREQUENCY (MM^{-1})	PEAK NO.
1	0.9	0.8	1
2	1.7	1.5	2
3	2.1	2.1	3
4	2.5	---	---
5	2.9	2.8	4
6	3.4	3.5	5
7	4.2	4.0	6
8	5.2	4.7	7
9	6.3	---	---
10	7.1	6.7	8
11	8.1	8.3	9
12	8.8	---	---

appear to be quasi-periodic. In addition, many peaks in both spectra from the two individual samples are at approximately similar frequencies.

Peaks have also been observed in the Fourier spectra of fracture

profiles of metals (20,21). The periodicities observed for these peaks have been ascribed to a fundamental fracture unit which has a triangular shape (20).

The root-mean-square (rms) roughness, R, has been calculated for the fracture surface profile of TNT shown in Fig. 3. The roughness has been obtained from data measured by the diamond stylus. The determination has also been made for thirty-one separate subprofiles chosen at random from this surface. A straight line has been fit to the points by the method of least squares. The equation of the line is

$$\log R = (0.83 \pm 0.10) \log L - (0.64 \pm 0.22),$$

where L is the profile length. The best line fit has a linear correlation coefficient equal to 0.83. This is beyond the 99.9% significance level for the thirty-two data points (22). The log of the rms roughness increases linearly with the log of the profile length. Nevertheless each specific value of rms roughness is valid only for the particular subprofile from which it is obtained (14,23). These rms roughness values are plotted as a function of profile length on a log-log plot in Fig. 5.

SUMMARY

The divider method and power spectral analysis techniques have been adapted to determine the fractal dimensions of the fracture surfaces of energetic materials. The fracture surfaces of melt-cast samples of TNT and Composition B are obtained during high acceleration in an ultracentrifuge. When the tensile or shear strength is exceeded at 50,000 to 200,000 g a fracture surface is obtained. Closed fracture profiles of holes produced through TNT and Composition B at 0°, 25° and 50°C have been obtained by optical microscopy at 30X magnification. Evaluation of these closed fracture profiles with the divider technique yields values of fractal dimensions for TNT and Composition B slightly larger than one. These values increase slightly with temperature. These results suggest an inverse relationship with the impact sensitivity of TNT.

The Z,X coordinates of the concave fracture surface of TNT have been measured with both a diamond and sapphire stylus. The power spectra have been calculated from the data using a prolate spheroidal data window in the horizontal space domain prior to the employment of a fast Fourier transform algorithm. Spectra from both samples are found to decrease with spatial frequency. At low spatial frequencies the slope of the spectra is about -3; the fractal dimension is found to be about 1. At high spatial frequency the slope is about -4; the fractal relationship between spectral slope and fractal dimension does not hold and the dimension is exactly equal to 1, the topological dimension. The break in slope corresponds approximately to the smallest grain size observed in the fracture surface

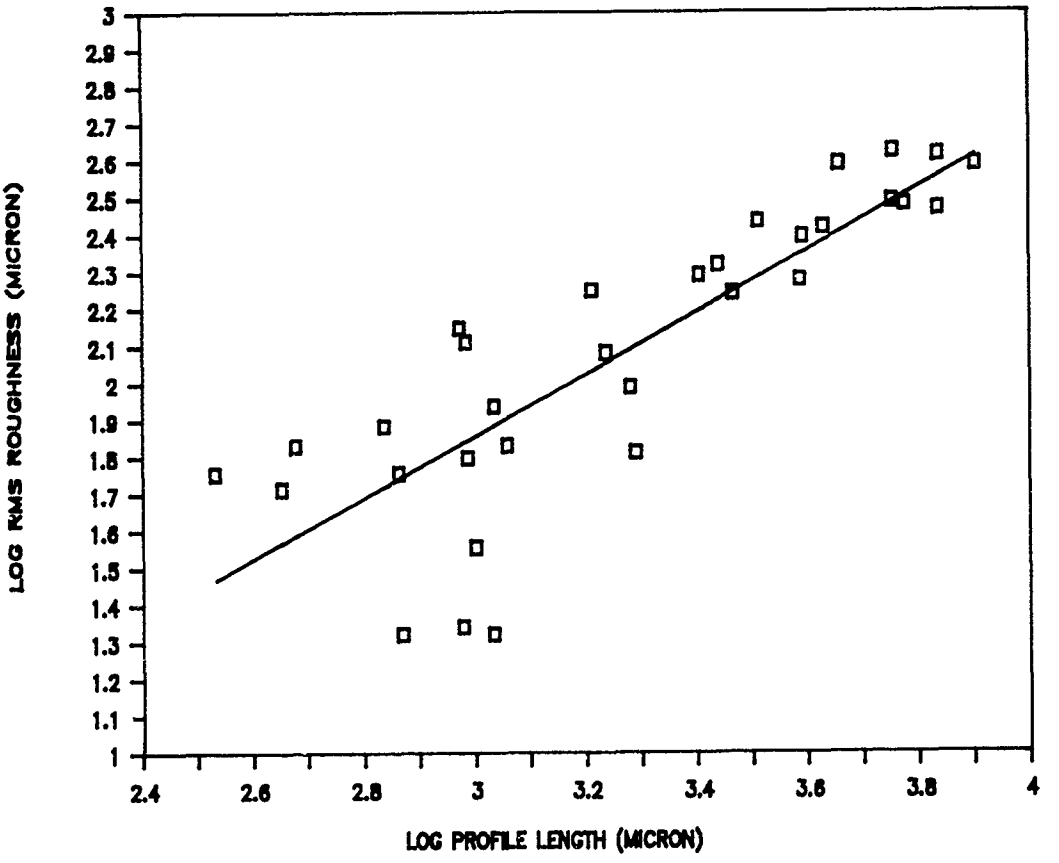


Figure 5. Log of the Root-mean-square Roughness of TNT as a Function of the Log of the Profile Length

(0.1mm). Quasi-periodic spectral peaks at similar spatial frequencies are observed in both spectra and indicate that much of the fracture is occurring at grain boundaries. The log of the root-mean-square roughness of the fracture surface of the TNT is found to increase linearly with the log of the profile length.

CONCLUSIONS

Understanding of the fracture and rupture of energetic materials subjected to high acceleration is a key to better practical designs in

several fields, including ordnance, the extraction industry, and space propulsion. In such applications the materials can often be subjected to high, fluctuating, and/or sustained accelerations. We have presented experimental data on the studies of the fracture of energetic materials under high accelerations and have introduced the use of fractal geometry techniques to characterize the fracture surfaces. We have shown that the fractal and other statistical measures provide new ways to parameterize the fracture surfaces and also provide new insights into the fracture process. We further have demonstrated clearly that at certain spatial wavelengths the fracture statistically occurs at grain boundaries. We believe the experimental and analysis techniques used here have wide applicability in future studies of energetic materials.

REFERENCES

1. M. Y. D. Lanzerotti and J. Sharma, "Brittle Behavior of Explosives During High Acceleration", Appl. Phys. Lett. 39, 455 (1981).

2. M. Y. D. Lanzerotti, "Fracture Phenomena of Energetic Materials During High Acceleration", in Proc. Symposium on Thermomechanical Properties of Energetic Materials and Their Effects on Munitions Survivability, Vol. II, (Naval Weapons Center, China Lake, CA, March 1985), pp. 1-43.

3. M. Y. D. Lanzerotti, "Mechanical Behavior of Gun Propellant During High Acceleration", in Proc. 21st JANNAF Combustion Meeting, Vol. 1, (Johns Hopkins University, Laurel, MD, 1984), pp. 275-280.

4. M. Y. D. Lanzerotti and J. Pinto, "Chemical Reaction of Energetic Materials During High Acceleration", in Shock Waves in Condensed Matter, Y. M. Gupta, Ed., (Plenum Publishing Corp., New York, 1986), pp. 909-916.

5. M. Y. D. Lanzerotti and J. Pinto, "Fractal Dimension of Fracture Surfaces of Energetic Materials", in Fractal Aspects of Materials II, D. W. Schaefer, R. B. Laibowitz, B. B. Mandelbrot, and S. H. Liu, Eds., (Materials Research Society, Pittsburgh, PA, 1986), pp. 133-4.

6. M. Y. D. Lanzerotti and J. Pinto, Bull. Am. Phys. Soc. 32, 937 (1987).

7. M. Y. D. Lanzerotti and J. Pinto, "Fractal Characteristics of Fracture Surfaces of Energetic Materials", in Fractal Aspects of Materials: Disordered Systems, EA-13, A. J. Hurd, D. A. Weitz, and B. B. Mandelbrot, Eds., (Materials Research Society, Pittsburgh, PA, 1987), p. 30.

8. B. B. Mandelbrot, The Fractal Geometry of Nature (Freeman, New York, 1983); also Fractals (Freeman, New York, 1977).

9. B. H. Kaye, Am. Laboratory, 18, 55 (1986).

10. S. R. Brown, Geophys. Res. Lett. 14, 1095 (1987).

11. Y. Termonia and P. Meakin, Nature, 320, 429 (1986).

12. C. A. Aviles and C. H. Scholz, J. Geophys. Res. 92, 331 (1987).

13. S. M. Kaye, Encyclopedia of Explosives and Related Items, PATR 2700, Vol. 9, (U. S. Army Armament Research and Development Command, Dover, NJ, 1986) p. T266.

14. S. R. Brown and C. H. Scholz, J. Geophys. Res. 90, 12,575 (1985).

15. A. G. M. Hunter and E. A. Smith, Wear, 59, 383 (1980).

16. T. R. Thomas, "Stylus Instruments", in Rough Surfaces, T. R. Thomas, Ed., (Longman, London, 1982), p. 16.

17. D. J. Thomson, Spectral Analysis of Short Series, Ph. D. Dissertation, Department of Electrical Engineering, Polytechnic Institute of Brooklyn, Brooklyn, NY, 1971.

18. D. J. Thomson, Proc. IEEE, 70, 1055 (1982).

19. D. J. Thomson, M. F. Robbins, C. F. Maclennan, and L. J. Lanzerotti, Physics of the Earth and Planetary Interiors, 12, 217 (1976).

20. D. E. Passoja and D. J. Amborski, Microstructural Science, 6, 143 (1978).

21. D. E. Passoja and J. A. Psioda, "Fourier Transform Techniques - Fracture and Fatigue", in Fractography and Materials Science, ASTM STP 733, L. N. Gilbertson and R. D. Zipp, Eds., (American Society for Testing and Materials, Philadelphia, PA, 1981), pp. 355-386.

22. P. R. Bevington, Data Reduction and Error Analysis for the Physical Sciences (McGraw-Hill, New York, 1969), p. 310.

23. R. S. Sayles and T. R. Thomas, Nature, 271, 431 (1978).

The Eyes Have It: Contact Lens Impact on Performance of Armor Troops (U)

Bruce C. Leibrecht, LTC*
William G. Bachman, LTC
U. S. Army Aeromedical Research Laboratory
Fort Rucker, AL 36362-5292

Introduction

A large proportion of Army troops are myopic (nearsighted), hyperopic (farsighted), and/or astigmatic. These troops generally require visual correction, since good vision is essential for maximum combat effectiveness. Spectacles provide the standard means for visual correction. Yet spectacles are minimally compatible or outright incompatible with many military systems. Examples of these systems include weapons sights, night vision goggles, head mounted displays, binoculars, and protective masks. In addition, Mother Nature compounds operational problems with rain, sweat, dust, and condensation clinging to spectacle lenses of troops operating visually coupled equipment.

To accommodate the spectacle wearing soldier, visually coupled systems frequently incorporate "dial-in optics." The soldier simply dials adjustable optics, which are part of the system, to obtain corrected vision. Unfortunately, this approach compensates for only nearsightedness and farsightedness, not astigmatism. This shortcoming presents a serious limitation, since a high percentage of personnel requiring spectacles have varying amounts of astigmatism.[1] Of significance to the Army is that unresolved incompatibilities and environmental difficulties for spectacle wearing crewmembers can compromise system effectiveness and limit the manpower pool available for key systems.

Contact lenses, particularly the extended wear varieties, offer an appealing alternative for solving the compatibility and environmental problems faced by spectacle wearing soldiers. However, the published data documenting the impact of contact lenses on soldier performance are very limited.[2,3,4,5] There is a need to assess systematically the operational impact of contact lens wear. This issue is fundamentally important to the establishment of comprehensive Army policy for contact lens use.

In this paper, we present some of the data from the Army's first major field investigation of contact lenses.[6,7] At the time we initiated this study, soft contact lenses (CLs) worn for extended periods of time offered the greatest potential for Army applications. The armor environment provided an excellent setting, given the sighting devices found in M1 tanks and Bradley Fighting Vehicles. A major study objective was to assess the effects of contact lens wear on job performance and environmentally related aspects. Soldiers assigned to an armored division served as subjects while they participated fully in their units' normal activities. Questionnaires captured information on participants' experiences with job related activities, environmental factors, and operational settings.

Materials and methods

After screening candidates for medical suitability, we selected 311 volunteers to participate in the study; 215 wore extended-wear soft CLs for up to 6 months, while 96 served as spectacle-wearing controls. Ranging in age from 18 to 43, all were male soldiers (commissioned officers, noncommissioned officers, enlisted personnel) assigned to the 2d Armored Division located at Fort Hood, Texas. We excluded soldiers from wearing CLs if they had more than -6.00 diopters of myopia, more than +4.00 diopters of hyperopia, or more than 1.25 diopters of astigmatism. Thirty-five of the CL wearers were wearing their own soft CLs at the start of the study or had worn CLs within the preceding 6 months; we will refer to these as "experienced" wearers. Although the remaining CL wearers included 31 participants who had worn contact lenses at some point in the past, we will refer to this larger group as "inexperienced" wearers.

Most of the participants held duty assignments related to armor (M1 tank), mechanized infantry (M2 and M3 fighting vehicles, Improved TOW Vehicle, M106 mortar carrier), and air defense artillery (Redeye, Vulcan, Chapparal). About 20 percent were in support categories, including wheeled vehicle operations, maintenance, medical support, operations, logistics, and administrative support. The median time in the Army was 2.8 years for the CL wearers and 3.5 years for the spectacle wearers.

We used three different types of extended-wear soft CLs: 71 percent water content, 55 percent water content, and 38.5 percent water content. This mix provided high, medium, and low water content lenses in a variety of base curves for reasonably broad fitting capabilities. We instructed the CL participants to wear their lenses continuously for 7 days (plus or minus 1 day), then remove them for cleaning and disinfecting overnight. To guard against protein deposits building up on the lenses, we replaced all CLs after 4 months wear, or sooner if indicated.

During the course of the study, subjects participated without restriction in their units' normal activities, including training in garrison and in the field. We made no attempt to alter any unit's schedule. At the end of the investigation, participants completed a questionnaire addressing operational issues, responding primarily on the basis of their experiences in the study. We used separate questionnaires for CL wearers and spectacle wearers. Questionnaire items dealt generally with visual ability, job or task performance, and difficulties related to environmental factors or operational situations.

Results and discussion

In terms of effectiveness in correcting vision, the contact lenses we used were equal to spectacles. The same proportion of the CL wearers and the spectacle wearers (97 percent of each group) achieved 20/25 visual acuity or better. In practical terms, both groups were able to achieve comparable corrected acuities.

At the conclusion of the study, 135 inexperienced CL wearers, 25 experienced CL wearers, and 84 spectacle wearers had completed questionnaires. Not every individual answered every question. In interpreting the questionnaire results presented below, two tempering considerations are important. First, the corrective lens frame of reference for CL subjects was different than for spectacle wearers, since the latter had no experience with contact lenses. This may have skewed responses of spectacle wearers where relative judgments about acceptability, severity of problems, and so forth were required. Second, the CL wearers generally may have been motivated to present a favorable picture of the contact lenses. This could have influenced them to underestimate the frequency or severity of lens-related problems.

Nearly all of the CL wearers regarded their contact lenses positively, with 94 percent indicating they liked their lenses moderately or very much. This contrasts with 18 percent of the spectacle wearers who stated they liked their spectacles moderately or very much. The reasons reported most often for dislike were that spectacles got in the way and were uncomfortable, and that Army spectacles were ugly.

Both CL-wearing and spectacle-wearing participants were almost unanimously confident (99 percent and 96 percent, respectively) in their ability to see adequately. A large majority of the CL participants (77 percent of the inexperienced wearers, 92 percent of the experienced wearers) felt they could see better with contact lenses than with spectacles. The larger proportion of the experienced CL group in this category is consistent with their greater cumulative CL wearing experience, but also may reflect some self-selection. Only 6 percent of the CL wearers felt they could see better with spectacles than with contact lenses.

We asked the CL participants whether they could see better with contact lenses or spectacles for a variety of visually oriented tasks. These tasks included sighting/aiming and surveillance under different conditions. As Table 1 shows, the proportion of subjects judging they could see better with contact lenses exceeded 75 percent for most of the tasks. This held true for both inexperienced wearers and experienced wearers, the trend being slightly stronger for experienced wearers. The smallest proportions favoring contact lenses (62 percent of the inexperienced wearers, 68 percent of the experienced wearers) occurred for reading and writing. The reason for this most likely lies in the inconvenience of removing contact lenses when unaided vision might be appropriate for close-up work. Not surprisingly, nearly all of the CL

Table 1

Proportion of CL wearers judging task-related visual ability better with contact lenses or spectacles

	Inexperienced CL Wearers			Experienced CL Wearers		
Task	CL better	Spect better	No diff	CL better	Spect better	No diff
Sight/aim rifle	85%	3%	12%	82%	5%	13%
Sight/aim thru optics	91%	3%	6%	95%	0	5%
Surveillance						
- <1000m, naked eye	75%	9%	16%	88%	4%	8%
- <1000m, thru optics	85%	3%	12%	92%	0	8%
- >1000m, naked eye	69%	11%	20%	80%	4%	16%
- >1000m, thru optics	82%	5%	13%	88%	0	12%
Wear prot mask	95%	2%	2%	96%	0	4%
Read and write	62%	8%	29%	68%	8%	24%

subjects favored contact lenses when wearing protective masks. The proportion of respondents favoring spectacles for the various tasks did not exceed 11 percent for either group.

The great majority of the CL subjects (83 percent of the inexperienced wearers, 96 percent of the experienced wearers) agreed wearing contact lenses had improved their overall job performance; 3 percent felt it had not. Of the inexperienced CL wearers, 14 percent neither agreed nor disagreed that job performance had improved.

When we asked CL subjects to compare the relative benefits of contact lenses and spectacles in performing their duties, the response patterns seen in Table 2 emerged. For garrison duties, 82 percent of the inexperienced wearers and 96 percent of the experienced wearers felt contact lenses were at least somewhat better than spectacles. However, for field duties the judgments favoring contact lenses tended to be less strong. Only 2 percent of the CL participants felt spectacles were better than contact lenses for performing duties in garrison; this proportion climbed to 12 percent when field duties were considered.

Table 2

Proportion of CL wearers judging contact lenses vs. spectacles better for overall performance of duties

	Inexperienced CL Wearers		Experienced CL Wearers	
Response	Garrison	Field	Garrison	Field
CL much better	67%	58%	92%	68%
CL somewhat better	15%	20%	4%	28%
No difference	16%	8%	0	0
Spect somewhat better	<1%	7%	4%	0
Spect much better	<1%	7%	0	4%

To clarify the weaker ratings of CLs for field duties, we asked the CL participants if they had encountered difficulties while wearing their contact lenses during field training. Table 3 presents the response patterns and includes data for offduty and garrison settings as baseline conditions. Slightly more than one-third of both CL wearing groups reported lens-related difficulties in the field, compared to 7 percent or less in garrison and offduty environments. These difficulties frequently pertained to environmental factors (e.g. dust) or problems with cleaning the contact lenses. Nearly one in three CL subjects reported substituting their spectacles in place of contact lenses during field training. Significantly, 44 percent of the spectacle wearers reported lens-related difficulties during field training, compared to 15 percent in garrison. Fifty-three percent of the spectacle wearers stated they had avoided wearing their spectacles on occasion, usually during field training or physical fitness training.

Table 3

Proportion of participants reporting difficulties while wearing corrective lenses in selected settings

Setting	Inexperienced CL Wearers	Experienced CL Wearers	Spectacle Wearers
Offduty	7%	0	*
Garrison	5%	0	15%
Field training	34%	36%	44%

* Spectacle wearers were not queried about offduty difficulties.

Table 4 presents the proportions of participants reporting lens-related difficulties when performing various job-related tasks. The proportion of CL wearers noting difficulties did not exceed 8 percent for either the experienced or inexperienced groups. In contrast, spectacle wearing subjects reported substantial incidence of difficulties for several tasks, especially physical training and use of equipment requiring ocular compatibility (e.g. night vision goggles, protective mask). In response to separate items, two of every five spectacle wearers reported difficulty

sighting/aiming a rifle and sighting/aiming with optical devices. In parallel fashion, spectacle wearers frequently reported removing their spectacles to complete tasks for which difficulties commonly occurred (physical training, 35 percent; night vision goggle wear, 69 percent; night sight use, 51 percent). Among CL wearers, removal of contact lenses to complete a task was reported seldom (generally 2 percent or less of both inexperienced and experienced CL participants).

Table 4

Proportion of participants reporting difficulty performing tasks while wearing corrective lenses

Task	Inexperienced CL Wearers	Experienced CL Wearers	Spectacle Wearers
Read map	4%	0	6%
Shoot compass azimuth	2%	0	5%
Assemble indiv wpn	2%	0	4%
Drive vehicle	6%	4%	11%
Fuel vehicle	3%	0	6%
Maintain vehicle	5%	8%	16%
Use night sights	<1%	0	53%
Wear NVG	2%	0	75%
Don/use prot mask	4%	0	67%*
Perform PT	5%	0	39%
Read	7%	8%	5%
Write	4%	0	4%

* Protective mask worn with optical inserts.

Table 5 summarizes the results which emerged when we asked CL participants which corrective lens (spectacles or contact lenses) they preferred for performing various tasks. The proportion of inexperienced and experienced CL wearers preferring contact lenses never fell below 90 percent, except for a simulated combat exercise with minimum sleep, where the proportion was 83 percent for both groups. The latter may relate to CL participants' frequent reports of lens-related difficulties in the field, discussed above. As a whole, these results document a strong preference for contact lenses for all types of activities queried.

Table 5

Proportion of CL wearers preferring contact lenses vs. spectacles for performing various tasks

	Inexperienced CL Wearers			Experienced CL Wearers		
Task	CL	Spect	No pref	CL	Spect	No pref
Routine duties	92%	1%	7%	100%	0	0
Manual labor	93%	1%	6%	100%	0	0
Physical exercise	93%	2%	5%	100%	0	0
Sports activities	92%	2%	6%	100%	0	0
Vehicle ops	92%	3%	4%	100%	0	0
Veh fueling	89%	2%	9%	100%	0	0
Veh maintenance	92%	3%	5%	100%	0	0
Guard/patrol duty	92%	4%	3%	100%	0	0
Night gunnery	95%	3%	3%	94%	0	6%
Simulated combat	83%	11%	6%	83%	11%	6%

We asked all participants if a variety of environmental conditions made wearing their corrective lenses difficult. Figure 1 presents the resulting response patterns. A substantial proportion (greater than 25 percent) of the CL wearers reported lens-related environmental difficulties for only dust, wind, and smoke. Dry air and tear gas were also somewhat problematic for CL wearers. In contrast, among spectacle wearers the occurrence of environmental difficulties was substantial (more than 25% reporting) in 7 of 12 conditions queried. Especially problematic were rain

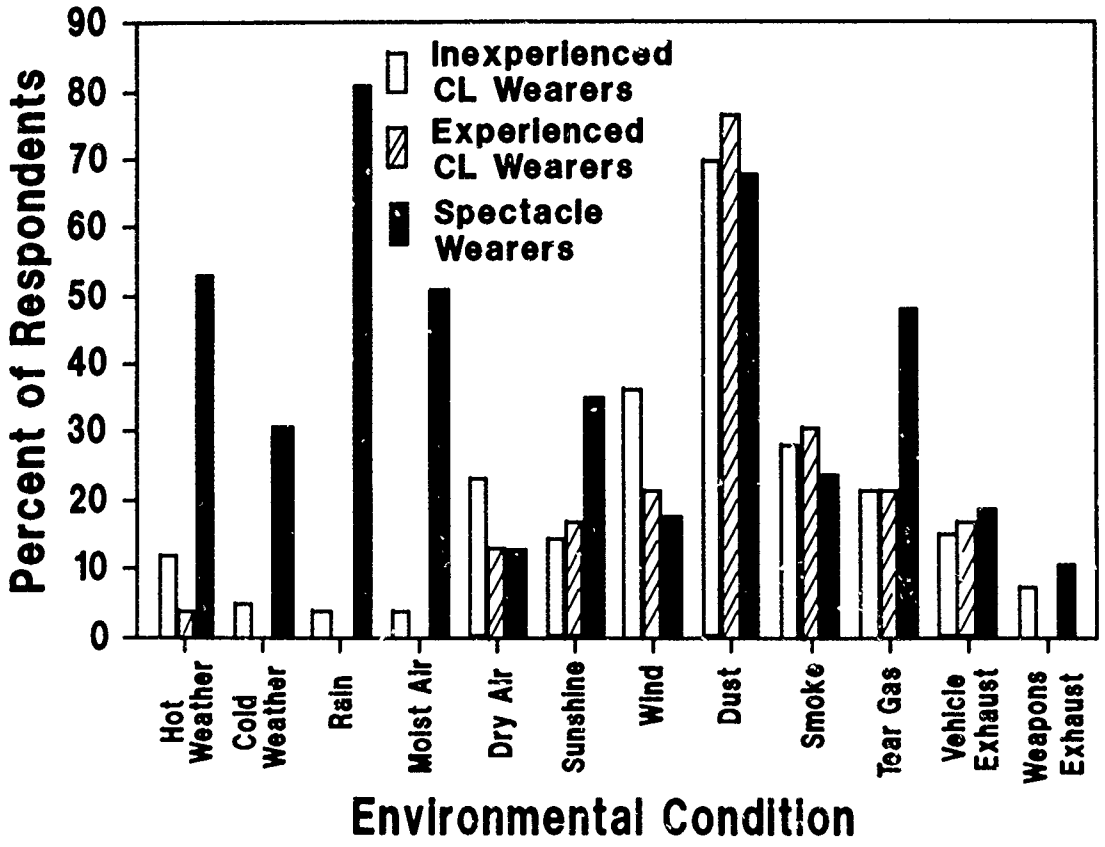


Figure 1. Percent of respondents reporting that various environmental conditions made wearing their corrective lenses difficult.

and dust (81 percent and 68 percent, respectively). The spectacle-related difficulties are understandable in terms of physical problems characteristic of spectacle lenses (rain or sweat streaking, fogging, dust coating, glare, etc.). We can relate the contact lens-related difficulties to ocular physiology (e.g. sensitivity to airborne substances and drying). Difficulties related to tear gas exposure during chemical defense training occurred less frequently among CL wearers than spectacle wearers. Kok-van-Aalphen et al.[8] have reported a similar finding in a study of policemen.

Appearing in Table 6 are the CL wearers' preferences for contact lenses or spectacles in the various environmental conditions queried. The proportion of subjects preferring contact lenses, based on experience during the study, was 70% or greater for every condition except dusty environments. In the latter case, preferences were split evenly between spectacles and contact lenses, with 15 percent of the inexperienced CL wearers expressing no preference. For both inexperienced and experienced CL subjects, the proportion preferring spectacles was less than 10 percent for every condition except dust, smoke, wind, and tear gas. This pattern of results is consistent with the trends for environmental difficulties presented above. Overall these findings reveal a strong preference for contact lenses for all environmental conditions queried, with the exception of dusty environments.

We asked the CL wearers to express their corrective lens preferences for a variety of military situations (e.g. airborne operations, combat operations), regardless of whether they had participated in the respective operations during the study. The resulting preference patterns appear in Table 7. For half the situations, 70 percent or more of the subjects preferred contact lenses. In the remaining cases, a substantial proportion of the respondents checked "don't know." If we remove these subjects from the analysis, the proportion of participants preferring contact lenses was less than 70 percent in only two cases -- airborne operations and air assault operations among inexperienced CL wearers. With "don't know" respondents excluded, 15 percent or less of the subjects preferred spectacles in every case except field training among inexperienced CL wearers (19 percent) and air assault operations among experienced CL wearers (21 percent). Although the judgments for the more rigorous military situations were based largely on estimates rather than actual experience, these findings demonstrate a substantial preference for contact lenses in all operational settings queried.

At the end of the study, nearly all of the CL participants (96 percent of the experienced wearers, 94 percent of the inexperienced wearers) wanted to continue wearing their contact lenses. Only 2 percent of the inexperienced subjects and 4 percent of the experienced subjects did not want to continue wearing their lenses.

Table 6

Proportion of CL wearers preferring contact lenses vs. spectacles in various environmental conditions

	Inexperienced CL Wearers			Experienced CL Wearers		
Condition	CL	Spect	No pref	CL	Spect	No pref
Hot weather	89%	3%	7%	100%	0	0
Cold weather	89%	3%	8%	100%	0	0
Rain	88%	1%	11%	100%	0	0
Moist air	87%	2%	11%	100%	0	0
Dry air	84%	9%	8%	94%	6%	0
Sunshine	88%	4%	7%	100%	0	0
Wind	70%	19%	11%	100%	0	0
Dust	42%	43%	15%	50%	50%	0
Smoke	72%	19%	9%	94%	0	6%
Tear gas	74%	21%	5%	75%	17%	8%
Veh exhaust	75%	7%	18%	89%	6%	6%
Wpns exhaust	79%	6%	15%	94%	0	6%

Table 7

Proportion of CL wearers preferring contact lenses vs. spectacles in various operational settings

	Inexperienced CL Wearers				Experienced CL Wearers			
Setting	CL	Spect	No pref	Don't know	CL	Spect	No pref	Don't know
Offduty	96%	2%	2%	0	96%	0	4%	0
Garrison	91%	3%	6%	0	96%	0	4%	0
Field trng	76%	19%	4%	0	92%	4%	4%	0
Deployment	72%	9%	4%	15%	84%	4%	4%	8%
Airborne ops	30%	5%	10%	55%	40%	8%	4%	48%
Air assault	34%	5%	10%	51%	40%	12%	4%	44%
Special ops	42%	5%	7%	46%	48%	4%	4%	44%
Combat ops	48%	9%	7%	35%	56%	4%	8%	32%

Conclusions

As the Army's first major evaluation of the performance impact of contact lenses, this study provides substantive findings relevant to the potential use of contact lenses among combat troops. Together with the study's data on ocular physiology and wear and care problems, the results reported in this paper form the cornerstone of the database needed to support establishment of comprehensive Army policy. The combined findings indicate the need for additional studies to address new types of contact lenses and specialized operational settings.

Based on the subjective performance findings obtained in the armor environment of this study, we conclude the following:

1. Contact lenses equalled spectacles in terms of participants' confidence in their ability to see adequately.

2. Contact lenses were judged superior to spectacles in visual ability afforded.

3. Contact lenses improved overall job performance for the great majority of CL wearers.

4. Difficulties related to wearing corrective lenses in the field were relatively common among both CL wearers and spectacle wearers.

5. Difficulties related to performing military tasks while wearing corrective lenses were infrequent among CL wearers, but frequent among spectacle wearers when equipment compatibility or physical exercise was involved.

6. Among CL wearers, lens-related environmental difficulties were infrequent except for conditions involving dust, wind, and smoke. Spectacle wearers more frequently reported environmental difficulties, especially for rain, dust, hot weather, and high humidity.

7. CL wearers expressed strong preferences in favor of contact lenses for performing military tasks, for most operational settings, and for diverse environmental conditions except dust.

References

1. Borish, I. M. 1970. Clinical refraction. Chicago, IL: The Professional Press.

2. McGraw, J. L., and Enoch, J. M. 1952. Contact lenses: An evaluation study. Fort Knox, KY: U. S. Army Medical Research Laboratory. USAMRL Report No. 99.

3. Crosley, J. K., Braun, E. G., and Bailey, R. W. 1974. Soft (hydrophylic) contact lenses in U. S. Army aviation: An investigative study of the Bausch and Lomb Soflens. Fort Rucker, AL: U. S. Army Aeromedical Research Laboratory. USAARL Report No. 74-10.

4. Van Norren, D. 1984. Contact lenses in the military service. American Journal of Optometry and Physiological Optics. 61:441-447.

5. Braithwaite, M. G. 1983. The use of contact lenses by Army aircrew. Journal of the Royal Army Medical Corps (British). 129:43-45.

6. Bachman, W. G., Leibrecht, B. C., Crosley, J. K., Price, D. R., Bentley, G. A., and Leas, P. M. 1987. An operational evaluation of extended-wear soft contact lenses in an armored division. Fort Rucker, AL: U. S. Army Aeromedical Research Laboratory. USAARL Report No. 87-12.

7. TRADOC Combined Arms Test Activity. 1986. Controlled investigation of contact lenses and operational performance (CICLOPS). Fort Hood, TX: TRADOC Combined Arms Test Activity. TCATA Test Report FT 484.

8. Kok-van-Aalphen, C., Van der Linden, J., Visser, R., and Bol, A. 1985. Protection of the police against tear gas with soft lenses. Military Medicine. 150:451-454.

A Helical Free-Electron Laser Structure for Hyperpower Devices (U)

*Herbert A. Leupold, Dr., and Ernest Potenziani II, Dr.
US Army Electronics Technology and Devices Laboratory (LABCOM)
Fort Monmouth, New Jersey, 07703-5000

INTRODUCTION

There has been considerable recent interest in mm sources and amplifiers for surveillance, communication, and weapon systems to penetrate dust, fog, smoke, and other battlefield obfuscation. The exigencies of more global military, aerospace and naval efforts, such as Army 21, SDI, Scott-Terminal, MILSTAR, and NETS have intensified further an already strong demand for powerful, broad-band wave sources and amplifiers. Free electron laser (FEL) devices would seem to be particularly attractive as possible fulfillments of that demand. Not only are their potential power levels virtually unlimited, but their relativistic electron beams "shrink" the period of the wiggler structure so that, in the electron reference frame, their longitudinal dimensions are reduced from what they are in the laboratory frame. This frees a wiggler design from the tyranny of size in that its dimensions need not be commensurate with the wave length of the desired radiation, as in more conventional wave sources. This characteristic greatly simplifies the manufacture of sources of short wave length, as their structural dimensions can be made orders of magnitude larger than the wave length generated, thereby obviating the need for difficult and expensive micromachining.

An example of a simple wiggler array is shown in Fig. 1. The alternating orientation of the component magnets (small arrows) generates an approximately sinusoidal field (large arrows) transverse to an axial electron beam. The alternating field causes the beam to oscillate in the plane of the beam, in the direction normal to the field and to the axial velocity of the electrons. This oscillation constitutes an acceleration and, since accelerated charges radiate, a beam of radiation results. If the product of the magnetic field in tesla and the wiggler period in centimeters is equal to one or less, the radiation from all parts of the beam interfere constructively and laser action results. The generated radiation is then confined to a narrow beam in the axial direction. Under such circumstances, the

wiggler is called an undulator.

Wiggler radiation is polarized in the plane of electron oscillation. For some purposes, circularly polarized radiation is preferable. For example, in radars circularly polarized radiation affords the advantage of easier discrimination from background noise. A source of such radiation would be a helical analogue of a wiggler. A mildly relativistic version of such a device was invented and built by R. M. Philips[1] ; it was named an ubitron. Figure 2 compares the power versus frequency potential of the ubitron with those of various high-power radiation sources[2]. In the militarily significant millimeter and near-millimeterwave region, the ubitron affords an approximate order of magnitude power advantage over the other sources.

The ubitron requires a constant transverse magnetic field of which the azimuthal orientation varies linearly with progression along the axis of electron motion. Such a field will accelerate the electrons into helical paths, rather than sinusoidal paths, and causes them to emit circularly polarized radiation by laser action if the field strength and twist rate have the required relationship. Since the magnitude of the ubitron field is constant, and since the power output varies as the square of the field, a ubitron will produce twice the power of a similar sinusoidal wiggler with the same field amplitude.

In view of these considerations, a team at the Naval Research Laboratory[2] designed and built a prototypical ubitron in which the field is provided by a bifilar electrical solenoid twisted to produce the desired wave lengths. The solenoid produces a field of 0.5 kOe at a twist period of 2.5 cm, with a coil current of 200 amperes. The coil and its associated power supply have a combined weight of several hundred pounds, depend on a source of electrical current, lack flexibility with regard to adjustment of twist, continually expend energy during operation, and are generally cumbersome. Accordingly, NRL sought the expertise of the magnetics group of the U.S. Army Electronics Technology and Devices Laboratory to design and assemble a permanent-magnet structure that has none of the aforementioned drawbacks. The permanent-magnet structure, or twister, was also to have the capability to produce a higher field than the solenoid at any given twist rate and, if possible, to extend operational viability to shorter periods.

DESIGN OF THE TWISTER

A natural starting point in twister design is the consideration of existing cylindrical magnet structures that produce high transverse magnetic fields. Then, the most promising of these structures are appropriately modified to produce the desired helical field. The three such base structures that were considered are pictured in the first column of Fig. 3. Configuration (a) is the simplest of the structures conceived at ETDL, consisting only of a transversely magnetized bar

magnet with a longitudinal hole through it. Configuration (b) is an octagonal approximation to a "magic-ring" structure that can produce as large a transverse field as desired in its interior, if the outer radius is large enough. Such structures have been used extensively by Halbach[3] in particle accelerators, and at ETDL[2] in the design of various electronic devices. The magic ring also has the advantage of confinement of magnetic flux to its interior cavity, so that it produces no stray fields that would disturb sensitive equipment in its vicinity. Confinement, of course, also implies design efficiency, since all of the flux produced goes into the interior working space, where it is needed. It is, obviously, more complex than structure (a). Structure (c) was designed at ETDL as a field source for NMR imagers. With its fourteen pieces, it is the most complex of the three alternatives, and is limited in its working field to about one-half the remanence of the magnetic material used. Like structure (c), it confines the flux it produces and has the added advantage of iron pole pieces that help smoothen irregularities or defects in the magnets to produce better field uniformity.

In the following analyses and comparisons, it is assumed that, in each structure, the interior working space to which field is to be applied is just large enough to circumscribe a right circular cylinder that is 1.7 cm in diameter. We also assume that rare earth permanent magnets are used, which are of material with perfect magnetic rigidity, i.e., $H_c = B_r = 10$ kG, where H_c is the magnetic coercivity and B_r the magnetic remanence. Together with the basic structures, a, b, c, shown in Fig 3, are the respective twisted structures d, e, f and the structures g, h, i that, in practice, must approximate d, e, f, respectively.

THE TWISTED BAR MAGNET

A simple twister structure results when a laterally magnetized bar with a coaxial cylindrical hole is twisted about the axis. (See Fig. 3a. and 3d.) We determine the on-axis transverse field by calculating the pole density distribution σ on the surfaces with the with the expression

$$\sigma = \hat{n} \cdot \vec{M} \qquad (1)$$

and then inserting σ into Coulomb's law and integrating over the surfaces.

The bar lengths considered are infinite, the heights 0.4 cm greater than the diameter of the cylindrical hole, or 2.1 cm, and the widths are varied. In the limit of zero twist and infinite width, magnetization perpendicular to the planar surfaces gives rise to a field $4\pi M$ due to poles on the parallel outer surfaces, and one of

$-2\pi M$ due to the poles on the surface of the cylindrical hole, resulting in a net field of $2\pi M$. If the direction of magnetization is parallel to the planar surfaces, there are no planar charges, and only the cylindrical surfaces contribute to a field of $-2\pi M$. These are the maximum field strengths obtainable with this structure. The field is reduced both by twisting and by the reduction of the structure to finite width. The reduction in field is less if the magnetization is parallel to the planar surfaces; only that arrangement is considered here.

The results of the calculations are plotted in Fig. 4 for two different values of twist designated by the ratio N of period length to hole radius b. The graph shows the effect of an increase of magnet width on the transverse field. As the width is increased, the planar poles recede from the axis, therefore producing less field there to counteract the larger field due to the poles on the cylindrical boundary. Hence, the net field is increased with width until the counteracting field from the planar poles is so small that further reductions are negligible. The more technologically meaningful mass per unit length is used as the ordinate of Fig. 4, rather than the width with which it varies linearly.

THE OCTAGONAL DIPOLAR STRUCTURE

Since this structure is also composed entirely of perfectly rigid magnets, its field can be found with Eq. 1, and by integration over all the surfaces that form the boundaries of the trapezoidal segments. The mass per unit length is plotted against field in Fig. 4, and the resulting curves, while qualitatively similar to those for the bar magnet with a hole, show that this structure produces higher fields for the same twist and mass per unit length.

THE CLAD STRUCTURE

The basis of the clad structure is shown in Fig. 3c. The magnets M_m supply the flux and are of cross sectional area A_W just sufficient to provide enough flux to result in the chosen uniform field of H within the working space W

$$2A_m = A_W H_W / B_m \quad . \tag{2}$$

where B_m is determined by the demagnetization curve of the magnets

$$B_m = H_m + 4\pi M = H_m + B_r \quad . \tag{3}$$

Applying Maxwell's equation at the inner surface of M we find that H_m is equal to the negative of the desired field magnitude $-|H_W|$, so that

(3) becomes

$$B_m = -|H_w| + 10 \qquad (4)$$

and A_m is determined.

The cladding magnets M_c ensure that all of the flux produced by magnets M_m goes to the interior, so that Eqs. (2) through (4) are valid. The details of how appropriate thicknesses of M_c are determined are discussed elsewhere[4-6], so here we give only the result that the maximum thickness t_c is given by

$$t_c = |H_w| l_m / H_c \ ,$$

where l_m is the length of the supply magnet, 1.7 cm, and H_c is the coercivity of the cladding magnet, 10 kOe, so that

$$t_c = 0.17 |H_w| . \qquad (6)$$

The Permendur pole pieces p aid to ensure that the flux in W is distributed uniformly.

When the structure of Fig. 3c is twisted to form that of Fig. 3f, we keep the transverse fields parallel to the magnetization of m and equal to H_w, so that the required cladding thickness t_c remains the same. However, additional flux is needed to feed an axial component of the field H_z that arises when the structure is twisted. Such an axial field exists within the supply magnets M_m and in the working space W. The total flux per unit length ϕ_L needed to provide a transverse field H_c after twisting, can be determined by the method of estimation of permeances[7,8] or, more simply, by noting that the pole pieces are equipotential surfaces and that the field lines are normal to the surface. Therefore, the magnitudes of field components are in the ratio of the direction cosines of the surfaces. Since H_w is specified, the other two components are also known, and the total flux leaving either surface can be calculated. The total flux per unit length ϕ_L of the structure is given by

$$\phi_L = |H_w| \left[W_w + 2k^2 \, (t_m + W_w/2)^3/3 \right] \qquad (7)$$

$$\phi_L = |H_w| \left[1.70 + 0.667 \, (t_m + 0.85)^3 \, k^2 \right] \qquad (8)$$

where

$$k = 2\pi / \lambda \ , \qquad (9)$$

t_m is the thickness of a supply magnet M_m and W_w is the width of the working space between magnets M_m. The clad structure is compared with

the other two configurations in Fig. 4.

In Fig. 5 are plotted the maximum fields for the three structures as functions of N. In this regard, the ranking of these configurations is absolute, with the dipolar structure the most desirable and the clad structure the least desirable.

REALIZABLE APPROXIMATIONS TO THE IDEAL STRUCTURE

With present technology, it is not feasible to produce the structures shown in the second column of Fig. 3 with the desired twist rates, where N is of the order of two or three. They can, however, be approximated by the configurations shown in the third column of Fig. 3, which are formed by stacking discrete rectangular or octagonal sections of finite thickness. In principle, the approximation can be made as good as desired by increasing the number of slices per period. However, large numbers of thin sections would be prohibitively expensive, mechanically frail, and, therefore, generally impractical. Fortunately, mechanically and economically viable devices that give very good approximations to the ideal continuous structure can be obtained with relatively coarse cross sections. For example, a dipolar configuration of sixteen sections per period would have the rather large rotation of 22.5 degrees between successive sections, yet, as can be seen from Fig. 6, produces a transverse field that differs from the ideal by less than 1%.

SUMMARY OF ANALYSIS AND CHOICE OF PROTOTYPE

While inferior to the dipolar structure with regard to transverse field per unit mass, the bar magnet is far simpler and less expensive, has fewer parts and is more mechanically robust. The cladded structure produces the smallest maximum fields ,is by far, the most expensive and, because of its pole pieces, shorts out solenoidal focusing fields superimposed from the outside. The last mentioned feature would tend to make it practicable only for short structures, where the focusing produced by the axial component of the twister field would be sufficient, or perhaps, where iron-free focusing structures can be inserted.

From the foregoing considerations it was decided that even though it does not produce the largest fields, configuration (a) is the best candidate for prototype construction. The technology of magnet fabrication is not sufficiently advanced to produce structures (b) and (c) without the incurring of prohibitive expense and great difficulty in finding a willing manufacturer. Structure (a) is simple, affordable, relatively easy to implement, mechanically robust, and still capable of producing fields twice as great as those produced by the NRL solenoid.

Figure 7 shows an exploded view of the chosen twister embodiment. The twist is obtained by the progressive canting of successive rectangular slabs as shown in Fig. 3g. The slabs are mounted in brass disks for mechanical strength and protection. Holes are bored near the circumferences of the disks at 23 degree intervals to afford the versatility of an adjustable or even a progressive twist.

The resulting structure as delivered to the Navy is shown in Fig. 8. It is a meter long, 7 cm in diameter, weighs about 18 kg and is composed of magnet slabs 2.5 mm in thickness. The structure was designed to produce transverse fields of 1.1 kOe and 2.3 kOe at twist lengths of 2.5 and 5.0 cm, respectively. The experimental values obtained from Hall probe measurements on the actual structure were 1.13 ± 0.05 kOe and 2.30 ± 0.10 kOe respectively. The former value is more than twice the 0.5 kOe produced by the NRL electrical solenoid with the same percent precision. The precision can doubtless be improved by selective replacement of aberrant magnets as indicated by axial field measurements along the entire length of the structure. In any case, experiments with the solenoid have shown the present precision to be adequate for effective functioning of the ubitron under current operating conditions.

FUTURE PLANS AND IMPROVEMENTS

For some twist rates, especially slow ones, the field strength can be augmented by circumscription of each magnet slab in an iron ring. For an untwisted structure, this increases the field from 2.5 kOe to 4.2 kOe, an increase of almost 70%. If the iron frame is square rather than circular, the increase is to 4.5 kOe or 80%. Such large fractional increases would not be realized at all twist rates because the decrease of field due to poles on the inner cylindrical surface of the twister and that due to those on the outer rectangular surfaces depend differently on twist rate. Therefore, effective removal of the outer poles by the iron frames may be of less value than in the untwisted structure.

After experimental testing on an operating ubitron shows the permanent magnets to be practical, an octagonal structure will be built for operation at quicker rates of twist for which the present transverse fields would be inadequate. Figure 4 clearly shows the superiority of the octagonal structure in this regard.

A particularly exciting development at NRL is the idea of operating at a twister harmonic rather than at its fundamental frequency. Since the required electron beam voltage varies as the reciprocal of the square of the harmonic number (N), great reductions in operating voltage are forded by an increase in that number. Even use of the lowest harmonic, N=3 (only odd harmonics are present), yields a voltage decrease by a factor of $1/N^2=1/9$ or almost an order of

magnitude. The mass, bulk, and complexity of a ubitron electron source varies approximately as the reciprocal of some power n of the electron voltage between 3 and 4. Therefore, if we assume the most pessimistic case of n=3 and N=3, the net mass reduction factor R is given by

$$R = 1/[(3)^2]^3 = 1/3^5 = 1/243 \quad .$$

If n=4,

$$R = 1/3^6 = 1/729 \quad .$$

or nearly three orders of magnitude. This principle, in conjunction with mass reductions already provided by the ETDL permanent magnet twister, affords possible extension of twister application from fixed stations or very large vehicles (ship, trucks, etc) to airborne and ballistic devices.

This work emphasizes the fruitfulness of cooperative ventures between different services of the DoD in bringing complex design problems to a successful conclusion.

REFERENCES

[1] R. M. Philips, IRE, **ED7**, p. 231, 1960

[2] R. H. Jackson and D. E. Pershing, Technical Digest, International Electron Devices Meeting (IEDEM),Los Angeles,CA, Dec 7-10, p. 342, 1986

[3] K. Halbach, in "Proceedings of the Eighth International Conference on Rare Earth Magnet Materials", University of Dayton, Dayton, OH, 1985, p. 123

[4] A. Tauber, H. A. Leupold, and F. Rothwarf, in "Proceedings of the Eighth International Conference on Rare Earth Magnet Materials", University of Dayton, Dayton, OH, 1985, p. 103

[5] J. P. Clarke and H. A. Leupold, IEEE Trans. Magn. **MAG-22**, p. 1063, 1986

[6] E. Potenziani II and H. A. Leupold, IEEE Trans. Magn. **MAG-22**, p. 1078, 1986

[7] H. C. Roters, in "Electromagnetic Devices", John Wiley and Sons, Inc., NY, NY, 1941

[8] H. A. Leupold, in "Proceedings of the Fifth International Workshop on Rare Earth Cobalt Permanent Magnets",Roanoke,Virginia,1981, p. 270

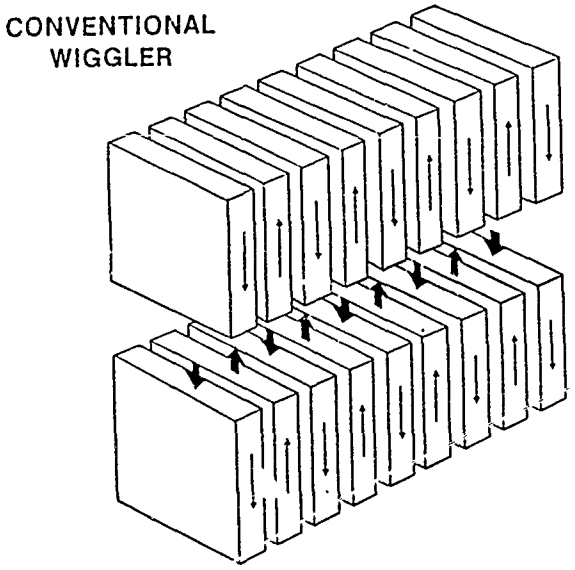


Figure 1. Simple wiggler array. The magnets are oriented so as to generate a sinusoidal field transverse to the electron beam.

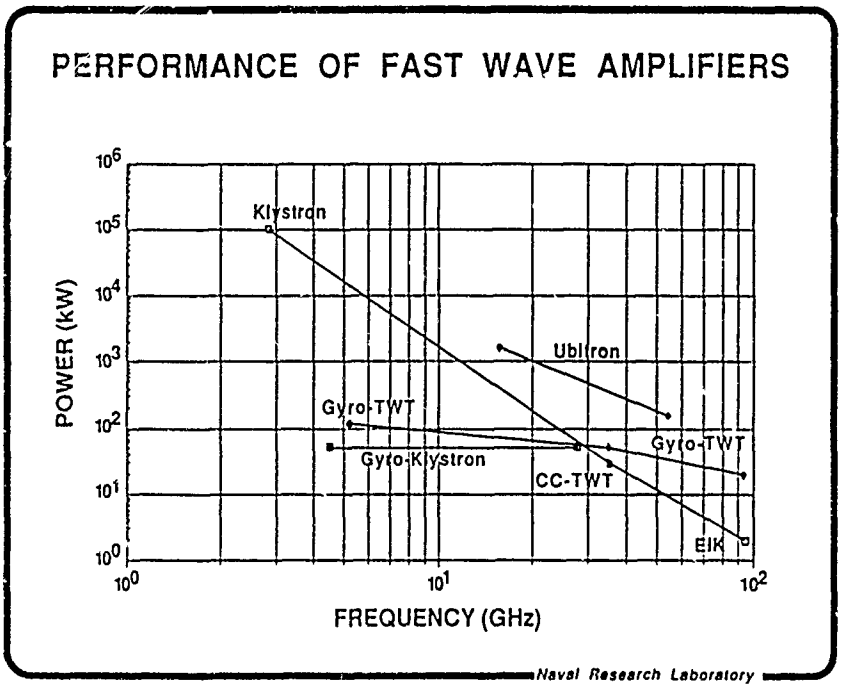


Figure 2. Comparison of power versus frequency for various high-power radiation sources

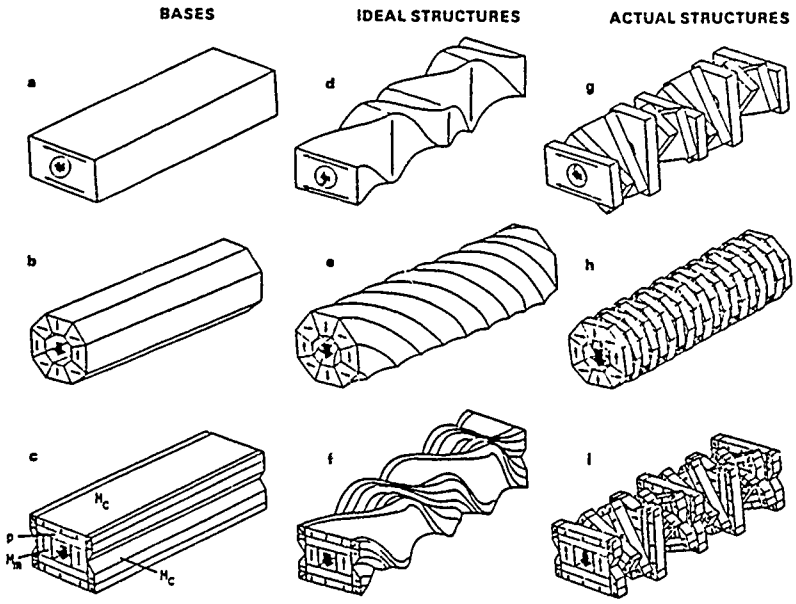


Figure 3. (a, b, c) are the three twister bases; (d, e, f) respective ideal twister structures; (g, h, i) respective approximations to the ideals. Heavy arrows show working space magnetic field direction while small arrows show magnet orientations.

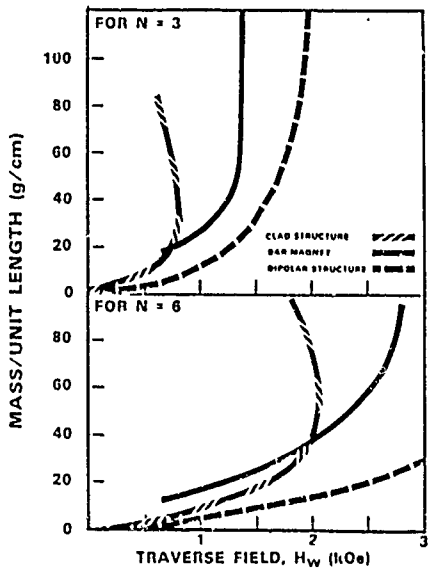


Figure 4. Mass of material needed to produce a given transverse field for the three structures with twists of N=3 and N=6.

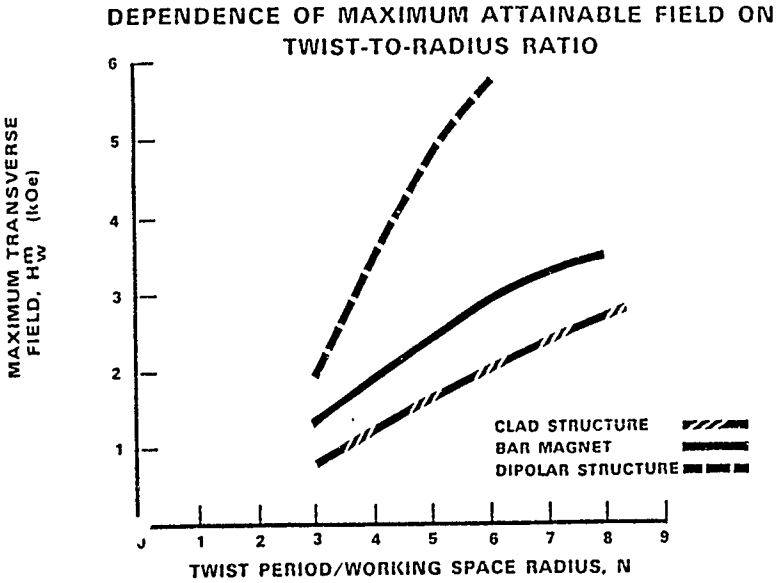


Figure 5.

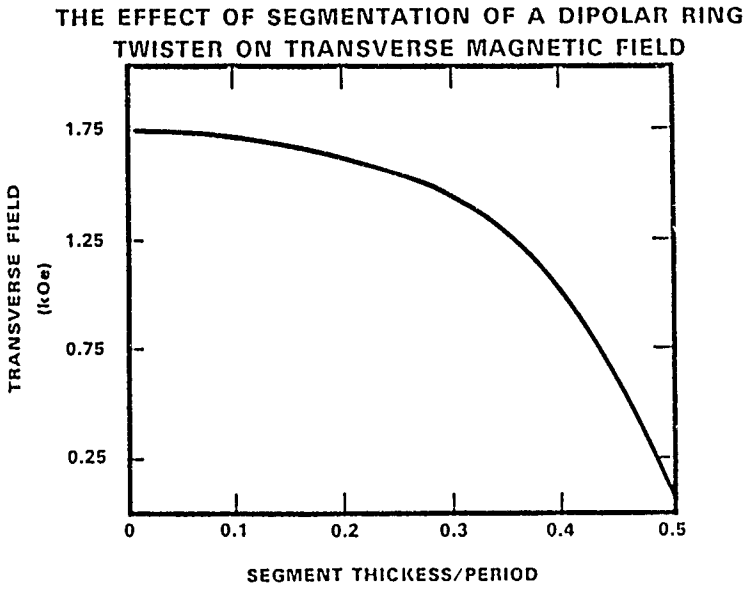


Figure 6.

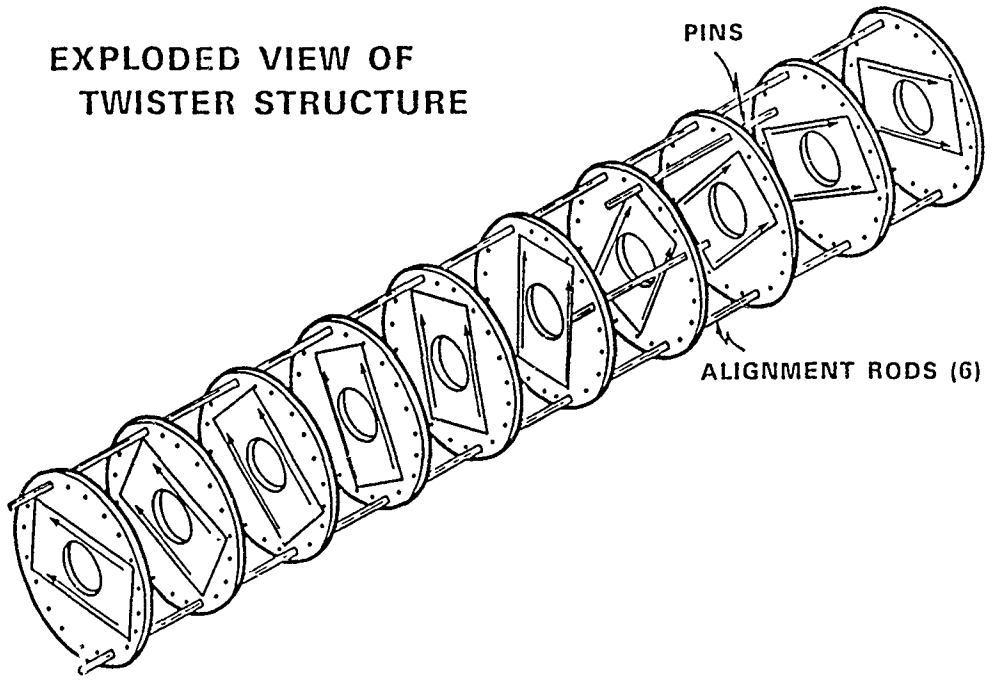


Figure 7. Exploded view of twister prototype. The periodic holes in the brass retaining disks afford easy adjustment of twist rate.

Figure 8. Prototype ubitron twister structure. Length is 1 meter, diameter is 7 cm, weight is approximately 18 kg and slab thickness is 2.5 mm. The structure consists of 400 individual magnets.

Use of Remote Sensing to Predict the Threat of Vector-Borne Diseases to Military Deployment (U)

*Kenneth J. Linthicum, MAJ, Charles L. Bailey, LTC, F. G. Davies, Dr., and C. J. Tucker, Dr.
U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, MD 21701-5011
Veterinary Research Laboratory, P.O. Kabete, Kenya and NASA/Goddard Space Flight Center, Greenbelt, MD 20771

Vector-borne infectious diseases have and will continue to pose substantial threats to military troop deployment throughout the world. One such disease, Rift Valley fever (RVF), is an arthropod-borne viral disease that produces devastating outbreaks in humans and domestic animals throughout much of Africa. As evidenced by the 1977-1978 epidemic in the Nile Delta and Valley in Egypt, the disease causes severe clinical manifestations and heavy mortality in immunologically naive human populations, such as U.S. military personnel.[1] Estimates of morbidity ranged up to 200,000 cases during this outbreak, with a mortality rate in hospitalized patients of 14%. Furthermore, the ecological areas most affected by RVF are likely sites of military troop concentrations.

Preventive and control measures up to now have been impossible to implement because occurrence of disease could not be predicted. Our recent field ecology studies clearly correlated RVF outbreaks with flooding, caused by heavy rainfall, of specific geographic formations [dambos][2] in which virus-infected mosquito vectors breed.[3-10] Figure 1 illustrates the life cycle of RVF virus in nature. When the transovarially infected, dormant eggs of Aedes mcintoshi mosquitoes, lying in the soil of dambos, are flooded, they hatch and produce mosquitoes capable of introducing the virus to susceptible hosts and maintaining the endemic cycle by the vertical transmission of the virus to the next generation of eggs. If the rains are sufficiently heavy and widespread to flood the mosquito eggs and permit standing water to remain for several weeks, a tremendous increase in the numbers of secondary vectors (Culex species) can create an epidemic of this disease. The epidemic cycle can be maintained by biological, mechanical, or aerosol transmission. Both transovarially and horizontally infected, primary vectors (Aedes mcintoshi) can continue to transmit the virus to their eggs. Secondary vectors pick up the virus from infected vertebrate

hosts and transmit the virus until the number of susceptible hosts are depleted or the vectors themselves die off, resulting in the termination of the epidemic.

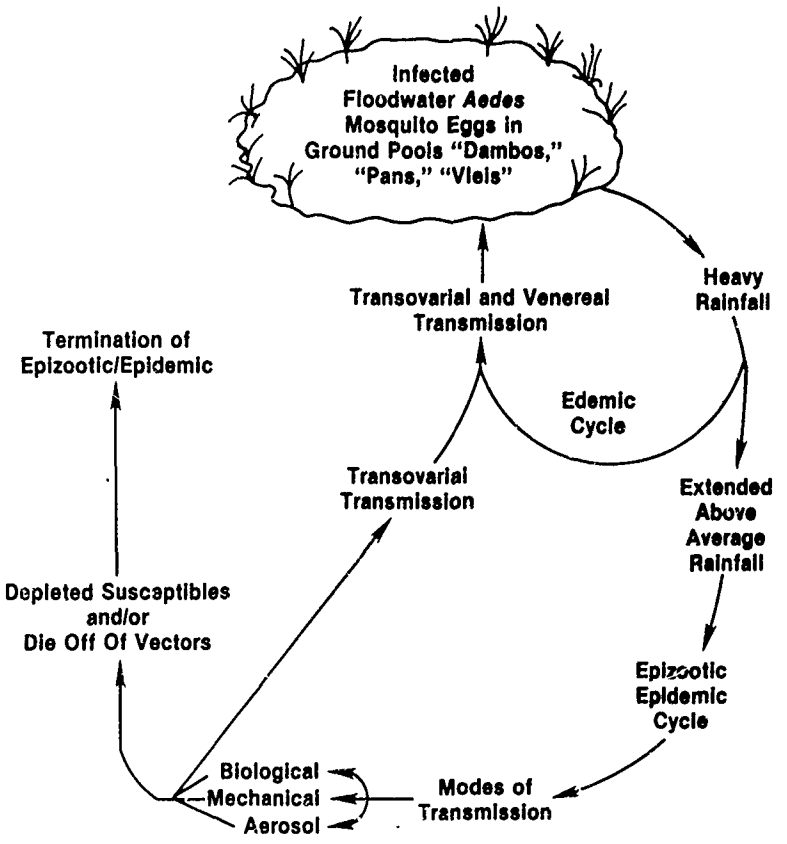


Fig. 1. Rift Valley fever virus life cycle.

Figure 2 depicts the rainfall at five sites in Kenya from 1950-1982 where epizootics occurred. It illustrates how RVF epizootics are closely correlated with the periods of very heavy rainfall that flood mosquito breeding habitats.[3] The epizootics of 1951-53, 1961-63, 1967-68, and 1977-79 are clearly and consistently associated with periods of high, positive surplus rainfall. Only a few clinical cases of RVF were observed late in 1957 and 1982, correlating with the lower positive values of the composite statistic at these times. No viral activity was recorded during periods of negative surplus rainfall.

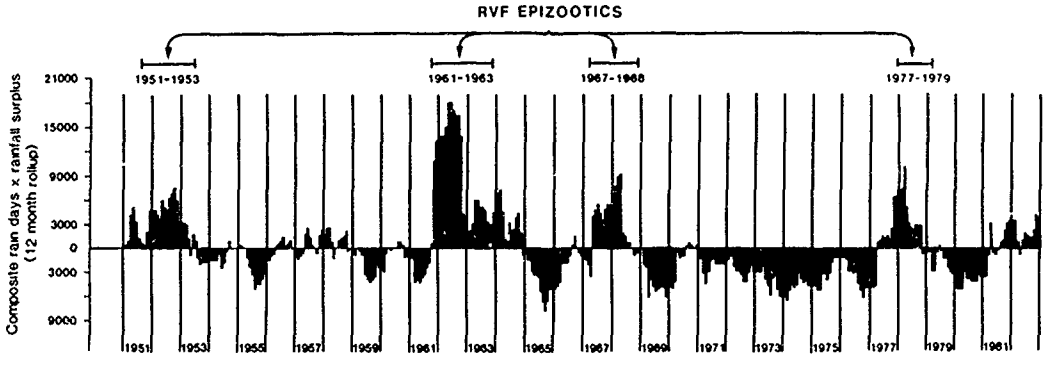


Fig. 2. The relationship between Rift Valley fever (RVF) epizootics and rainfall in Kenya for the years 1951-1982. The graph depicts a composite statistic based upon number of rainy days and rainfall for each month at five sites in Kenya where RVF epizootics occur. The zero line represents the 33-year mean rainfall for each month. Values above the zero line represent periods of positive, surplus rainfall.

The best strategy to prevent or lessen RVF and other vector-borne diseases is to monitor known parameters of epidemic activity so that control efforts can be implemented. Ideally, a predictive tool could permit maximum utilization of disease prevention and control programs. Satellite remote sensing technology is the newest and possibly the only method available to conduct surveillance activities over very large and diversified areas. Remote sensing of green vegetation dynamics is a well-developed technique that can serve as a sensitive indicator of rainfall patterns; several satellites collect these data worldwide.[11,12] In the past, satellite data have been used to survey ecological conditions predisposed to desert-locust activity. In this study, we analyzed factors related to RVF viral activity and green leaf vegetation dynamics, with the latter derived from satellite data as they relate directly to rainfall and subsequent vegetation development, during an interepizootic period in a restricted area in Kenya where epizootics occur. We describe a quantifiable prediction methodology that can be applied on a continental basis with data from polar-orbiting meteorological satellites.

The two areas in Kenya where RVF virus studies were conducted are described on the basis of the classification of Kenyan ecoclimatic zones[13] one through six and are defined in terms of climate and vegetation. The study areas, typical of ecoclimatic zones II and III, are illustrated in Figure 3 and described in Table 1.[14]

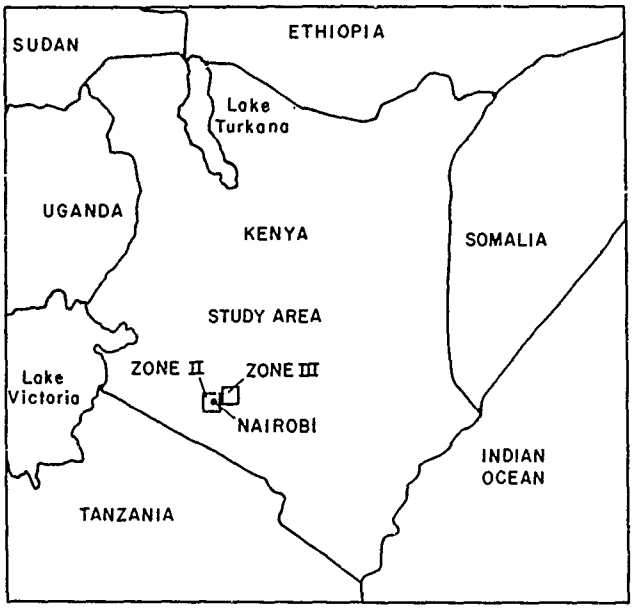


Fig. 3. Location of study areas in Kenya.

Table 1. Statistics for study areas in Kenya.

Ecological zone	Lat (S)	Long (E)	Elevation (M)	Total rainfall (MM)	Average monthly rainfall (range)	Average NDVI[a] (range)
II	1°20'	36°44'	1780	2201.8	71.02 (0.1-317.9)	0.1609 (0.0007-0.310)
III	1°12'	37°	1500	1864.2	60.14 (0.9-195.9)	0.2484 (0.0222-0.4007)

[a]NDVI = Normalized difference vegetation index.

Monthly rainfall data were obtained from the Kenya Meteorological Department. Adult mosquito population levels and RVF viral activity were continuously monitored from September 1982 to March 1985. Observations of the flooding of potential mosquito vector habitats were made concurrently while monitoring the adult mosquito population.

Digital remote sensing data were produced by the advanced very high resolution radiometer (AVHRR) sensor on polar-orbiting meteorological satellites operated by the National Oceanic and Atmospheric Administration (NOAA). The AVHRR records visible, near infrared, and thermal channels of the electromagnetic spectrum.[15,16] The characteristics of the NOAA/AVHRR systems used in this study are listed in Table 2.

Table 2. Characteristics of the NOAA/AVHRR systems. The NOAA-7 satellite was launched in June 1981 and taken out of operation 28 January 1985. NOAA-9 was launched in December 1984 and is still operational.

Characteristic	Measure
Coverage cycle	9 days
Scan angle range	±56°
Ground coverage	2700 km
Orbit inclination	98.9°
Orbital height	833 km
Orbital period	102 min
Ground resolution	1.1 km (nadir); 2.4 km (maximum off-angle along track)
	6.9 km (maximum off-angle across track)
Equatorial crossing	Descending mode, 1430 hours
	Ascending mode, 0230 hours

The normalized difference vegetation index (NDVI) analyzed in this study is a transformation between data from the visible channel (Ch1, 0.58 to 0.68 μm) and near-infrared channel (Ch2, 0.725 to 1.1 μm), and is expressed by

$$NDVI = (Ch2 - Ch1)/(Ch2 + Ch1)$$

The NDVI data have been shown to be highly correlated with such vegetation parameters as green-leaf biomass, atmospheric CO_2 draw down, and seasonal rainfall,[16,17] and represent the photosynthetic capacity of the area measured.[18]

Weekly AVHRR data were derived from the global area coverage data produced by the on-board processing of large-area coverage data (1.1 km by 1.1 km) and subsequently transmitted to receiving stations in Virginia or Alaska.[19] The satellite data have a grid cell (pixel) size of approximately 15 km by 15 km. Weekly composite data were formed by selecting the highest NDVI for each grid cell location from the daily data for that week. Weekly NDVI data were calculated and mapped to a Mercator projection. The highest value over 3 weeks was selected to represent the 3 week composite for each grid cell location. Data selected in this manner tended to occur near nadir, were largely cloud-free, and were characterized by low-aerosol conditions, which minimized atmospheric scattering.[20] Compositing was performed on the Hewlett-Packard 1000 and Ramtek image-processing system.

The potential viral activity factor (PVAF) statistic was developed from a methodology described and tested by Hielkema[21] to provide a single indicator of potential desert locust breeding activity. The PVAF progressively weights NDVI classes consistent with field observations that increasing vegetation densities are positively correlated with rainfall, a critical condition to RVF epizootics. The PVAF is calculated from NDVI data by the equation

$$PVAF = \frac{A \times 10^0 + B \times 10^1 + C \times 10^2 + D \times 10^3}{T}$$

where A is the number of grid cells with $0.05 < NDVI < 0.20$, B is the number of grid cells with $0.20 < NDVI < 0.29$, C is the number of grid cells with $0.29 < NDVI < 0.37$, D is the number of grid cells with $0.37 < NDVI < 0.65$ (0.65 is the maximum NDVI), and $T = 4$.

The relation among NDVI data, PVAF values and RVF ecological parameters are illustrated in Figures 4 and 5. During the 31-month study period, rainfall induced the flooding of dambo mosquito breeding habitats once in November and December 1982. This flooding corresponded to the isolation of RVF virus from mosquitoes on 11 occasions (six in zone II, and five in zone III).

In both areas, the maximum NDVI and PVAF values were recorded at the period starting 27 December 1982 (point C, Fig. 4 and 5), just following the most consistently rainy 3-month period (October, November, and December 1982) of the 31-month study period. These high values either coincided with (zone III, Fig. 5) or just followed (zone II, Fig. 4) the period of maximum mosquito vector populations and RVF viral activity.

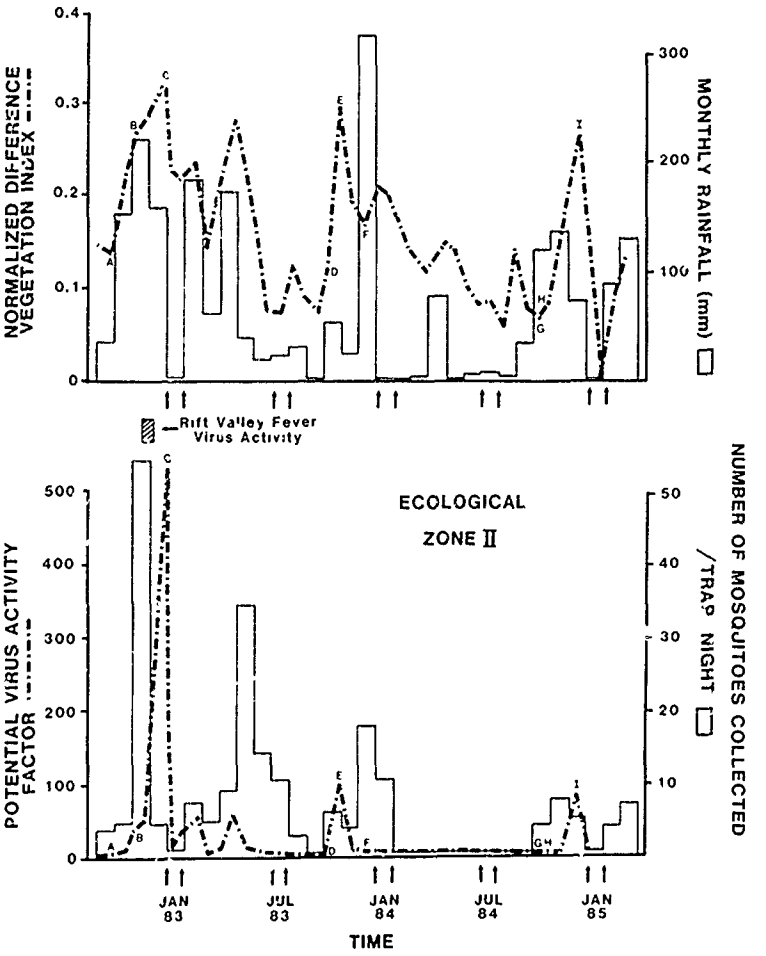


Fig. 4. Plots of AVHRR/NDVI and PVAF values, monthly rainfall, and mosquito collections from September 1982 to March 1985 at ecological zone II. Period of RVF viral activity shown by shaded box.

Figure 6 presents NDVI imag of Kenya for October, November, and December 1982 through 1984. Since NDVI represents a bounded ratio, values can range from -1 to 1. We have assigned brown, gold, green, red, pink, blue, and purple colors to increasing NDVI values from 0.0 to 0.6 as indicated by the color coding bar on Fig. 6. The letters on each of the images correspond to letters indicated on Fig. 4 and 5. The intense pink

and blue colors in the December 1982 (C, Fig. 6) image in the central portion of Kenya clearly stand out as the time of most intense green vegetation development, substantially higher than the corresponding period in the two subsequent years (F and I, Fig. 6). The December 1982 visual image correlates with the highest NDVI and PVAF values seen at point C (Fig. 4 and 5) and coincides with the only RVF viral activity detected.

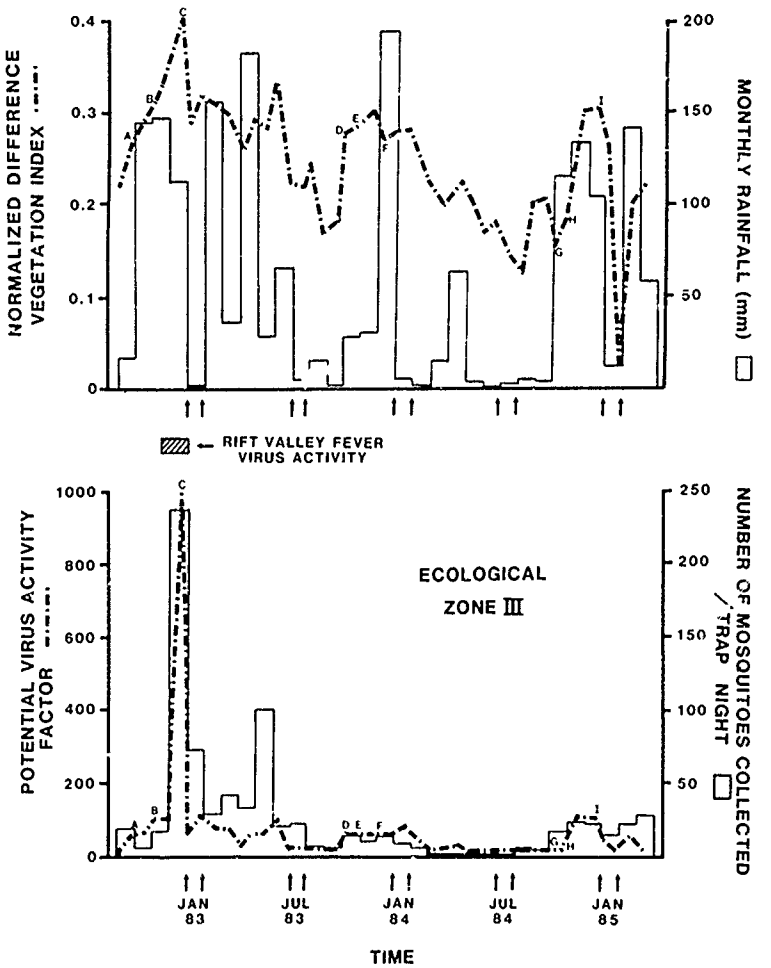


Fig 5. Plots of AVHRR/NDVI and PVAF values, monthly rainfall, and mosquito collections from September 1982 to March 1985 at ecological zone III. Period of RVF viral activity shown by shaded box.

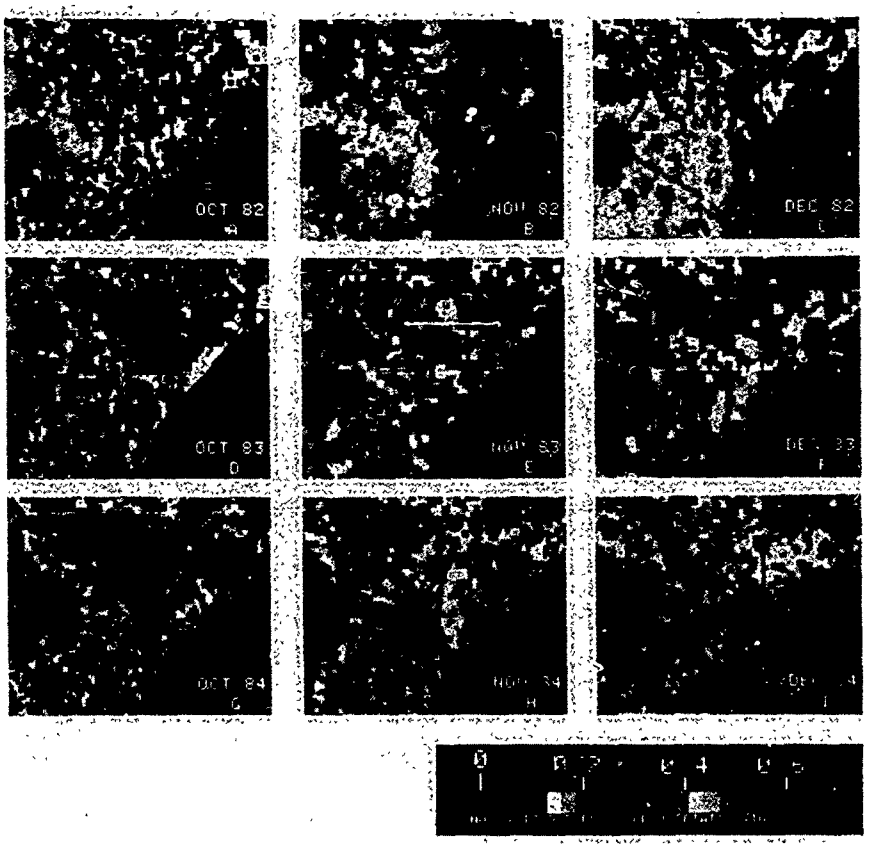


Fig 6. AVHRR/NDVI composite images of Kenya, October to December 1982, 1983 and 1984.

The positive correlation between NDVI and rainfall is evident in both ecological zones examined, even though other parameters controlling vegetation production, such as temperature and evapotranspiration, were not considered. In addition, the correlation between NDVI and mosquito populations, as reflected by our collection data, further corroborates the relation between NDVI and rainfall, as mosquito population levels are known to be highly related to rainfall patterns.

The ability of the PVAF to detect, with precision, RVF viral activity in our study areas during an interepizootic period indicates that RVF epizootics can be predicted reliably from this statistic. The PVAF generates rapid knowledge about potential viral activity conditions in ecologically equivalent areas and consolidates NDVI data, effectively reducing the operational decision-making process as it relates to control strategies. Potential viral activity factor data can be applied equally well to the entire geographic range of RVF in Africa (Fig. 7). Continental AVHRR data are currently available and operationally guaranteed on NOAA satellites through the 1990's.

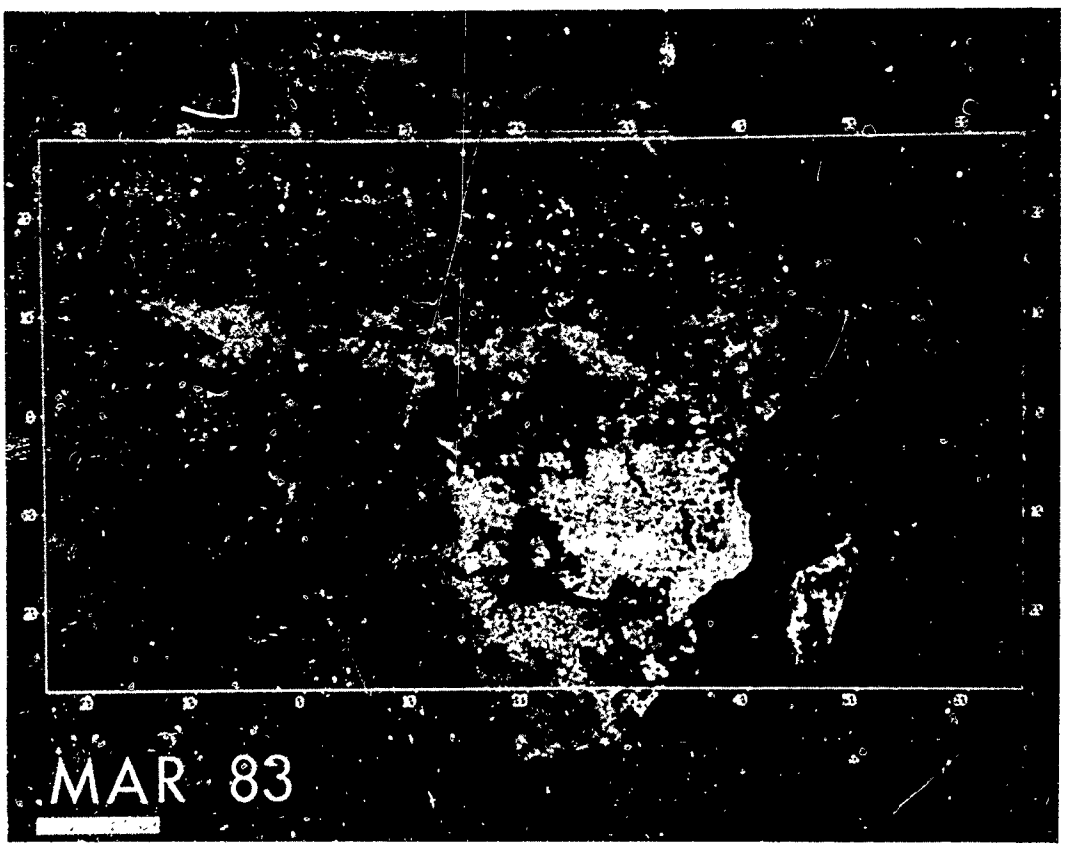


Fig. 7. African continental AVHRR/NDVI composite image of 21 Feb 1983 - 13 Mar 1983.

The isolation of virus in mosquitoes represents the earliest stages in a RVF epidemic. Detection of viral activity at this point in the RVF viral life cycle would allow time for specific control operations before an epidemic occurs. If the specific location of flooded mosquito breeding habitats were known, especially proximal to troop encampments, the efficiency of control operations could be greatly improved.

SUMMARY

These studies, for the first time, have established a mechanism whereby a disease threat to military troops can be continuously quantified and predicted on a real-time basis by analysis of satellite remote sensing imagery. The technology discussed here for RVF virus will certainly have an application to other militarily important diseases that decimate the continent and are ecologically linked, either directly or through transmission vectors.

Military medical intelligence can now be assisted in assessing the threat of vector-borne diseases. Figure 8 illustrates the worldwide availability of the remote sensing data to provide global predictive capability on a short-term notice. Other diseases, which could affect the success of military deployment operations, might also be monitored and eventually predicted by the methodology described here. Malaria, scrub typhus, leishmaniasis, Korean hemorrhagic fever, leptospirosis, schistosomiasis, and other diseases affected by rainfall patterns and vegetation changes are likely candidates for remote sensing monitoring.

In addition, naturally occurring disease outbreaks could be potentially distinguished from outbreaks caused by biological warfare and hence provide an important and previously non-existent method to monitor biological treaty compliance on a real-time basis.

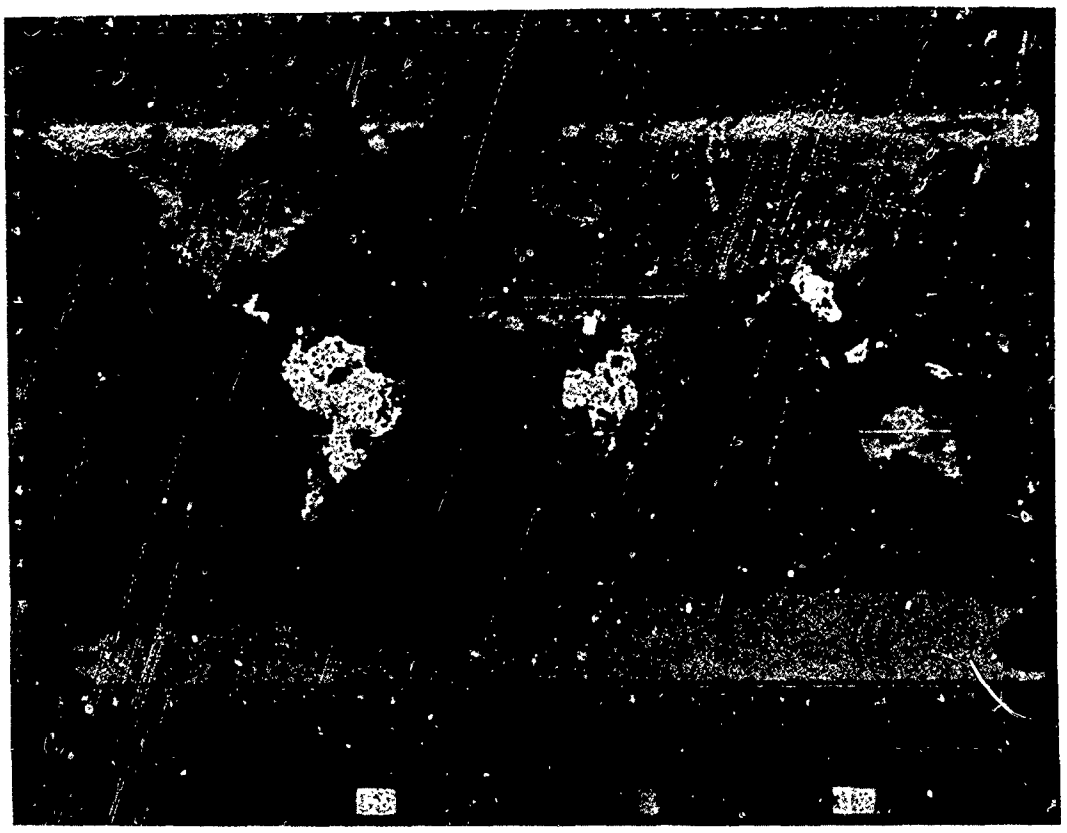


Fig. 8. Global view of NDVI data for December 1982.

REFERENCES

1. L. W. Laughlin, J. M. Meegan, L. J. Strausbaugh, D. M. Morens, and R. H. Watten, Trans. R. Soc. Trop. Med. Hyg. 73, 630 (1979).
2. E. Ackermann, Wiss. Veroff. Mus. Landerk. Leipzig 4, 147 (1936).
3. F. G. Davies, K. J. Linthicum, and A. D. James, Bull. WHO 63, 941 (1985).

4. K. J. Linthicum, F. G. Davies, A. Kairo, C. L. Bailey, H. F. Kaburia, K. J. Lindquist, in Advances in the Diagnosis, Treatment and Prevention of Immunizable Diseases in Africa, S. N. Kinoti, D. K. Koech, P. M. Tukei, Eds. (Kemri/Ketri, Nairobi, 1987), pp. 97-108.
5. K. J. Linthicum, F. G. Davies, C. L. Bailey, and A. Kairo, Mosquito News 43, 464 (1983).
6. K. J. Linthicum, F. G. Davies, C. L. Bailey, and A. Kairo, Mosquito News 44, 228 (1984).
7. K. J. Linthicum, F. G. Davies, and A. Kairo, Mosquito News 44, 595 (1984).
8. K. J. Linthicum, H. F. A. Kaburia, F. G. Davies, and K. J. Lindquist, J. Am. Mosq. Cont. Assoc. 1, 93 (1985).
9. K. J. Linthicum, C. L. Bailey, F. G. Davies, and A. Kairo, Bull. Entomol. Res. 75, 661 (1985).
10. K. J. Linthicum, F. G. Davies, A. Kairo, and C. L. Bailey, J. Hyg. 95, 197 (1985).
11. C. J. Tucker, J. U. Hielkema, and J. Roffey, Int. J. Remote Sens. 6, 127 (1985).
12. J. U. Hielkema, J. Roffey and C. J. Tucker, ibid. 7, 1609 (1986).
13. D. J. Pratt, P. J. Greenway, and M. D. Gwynne, J. Appl. Ecol. 3, 369 (1966).
14. K. J. Linthicum, C. L. Bailey, F. G. Davies, and C. J. Tucker, Science 7, 1656 (1987).
15. H. Schwalb, Nat. Oceanic Atmos. Adm. Tech. Memo NESS 95, (1982).
16. C. O. Justice, J. R. G. Townshend, B. N. Holben, and C. J. Tucker, Int. J. Remote Sens. 6, 1271 (1985).
17. C. J. Tucker, J. R. G. Townshend, and T. E. Goff, Science 227, 369 (1985).
18. P. J. Sellers, Int. J. Remote Sens. 6, 1335 (1985).
19. J. D. Tarpley, S. R. Schneider, and R. L. Money, J. Climate Appl. Meterol. 23, 491 (1984).
20. B. N. Holben, and R. S. Fraser, Int. J. Remote Sens. 5, 145 (1984).
21. J. U. Hielkema, FAO Tech. Rep. CGP/INT/349/USA (1980).

Ultraviolet Laser-Based Detection and Identification of Chemical Agent Simulants (U)

*S. Randolph Long, Dr.
Steven D. Christesen, Dr.
U.S. Army Chemical Research, Development and Engineering Center
Aberdeen Proving Ground, MD 21010-5423

I. Introduction.

Spectroscopic characteristics of chemical agents in the infrared spectral region have been exploited in the development of techniques for stand-off detection in both passive (XM21) and active (IR laser Lidar, DIAL) modes. Application of spectroscopic techniques outside the infrared has lagged due to the absence of characteristic spectral features of agent molecules in the visible and ultraviolet. However, our recent basic research on the interaction of focused ultraviolet (UV) laser radiation with agent simulants demonstrates significant potential for enhancement of point detection capabilities.

The action of focused UV laser radiation on target molecules results in the formation of neutral fragments and ions characteristic of the parent molecules. Several detection techniques can take advantage of the neutral fragment and ion formation. In this report, we present data on three applications of UV laser fragmentation/ionization to detection and identificiation of organophosphonates, which are simulants to the chemical nerve agents.

Neutral fragments typically have readily identifiable spectral features which may be utilized as markers for the presence of the parent molecule. In the case of organophosphonate simulants, the characteristic PO radical is of specific interest since the -P=O moiety is common to all organophosphonates. Section II discusses the salient features of our study of the PO radical generated by excimer laser photofragmentation of organophosphonates.

Ionization by a UV laser holds promise for improvements in mass spectrometric and ion mobility spectrometric detection of agent molecules. Section III presents results of our study of mass spectra of organo-

phosphonates generated by focused excimer laser radiation. The potential advantages of this ionization technique by comparison to the conventional electron impact ionization technique will be evident.

Ion mobility spectrometry, unlike mass spectrometry, is a technique which is applied at atmospheric pressure. Ions are identified according to their mobility in the atmospheric pressure gas. Most implementations of this technique, such as the ACADA chemical agent alarm in advanced development, employ chemical ionization techniques to generate ions of the analyte. Laser ionization, a direct process, offers a number of potential advantages in comparison with chemical ionization as applied to ion mobility spectrometry. This comparison, and our studies to date on laser ionization/ion mobility spectrometry, form Section IV.

II. Excimer Laser Photodissociation of Organophosphonates.

In applying focused ultraviolet laser photodissociaiton to detection of organophosphonates, we take advantage of the fact that, in an intense radiation field, molecules which have little or no absorption at the single photon level can be induced to absorb two or more photons simultaneously. The input of this amount of energy, using UV photons, is sufficient to decompose the target molecule into smaller fragments. These fragments typically have readily identifiable spectra in a convenient wavelength region that facilitate their detection. and thus provide a marker for the presence of the spectroscopically uncharacteristic parent.

The scheme we apply in the study of organosphonate photodissociation using excimer lasers[1,2] (ArF: 193 nm, KrF: 248 nm) is shown in Figure 1. Here a focused krypton fluoride (KrF) laser beam is used to induce the organophosphonate (dimethyl methylphosphonate, for example) to absorb several photons. The result is that the organic portions of the molecule are stripped away, leaving the bare PO radical. The PO radical may then be detected by laser-induced fluorescence (LIF) in its $A^2\Sigma^+ - X^2\Pi$ transition which has (0,0) bandheads at 247.7 and 246.3 nm. While either the KrF (248 nm) or argon fluoride (ArF, 193 nm) excimer may be used as the photolysis laser, the KrF laser is potentially more attractive for implementation since its gain curve extens to the PO A-X (0,0) bandhead at 247.7 nm. A single tunable KrF laser operating at this wavelength would be useful for both photolysis and probing for the presence of PO.

In experiments on the excimer laser photodissociation of organophosphonates, we have employed a two-laser arrangement, with the KrF or ArF laser used for photolysis and the radiation from a tunable Nd:YAG pumped dye laser system used for excitation of the PO A-X LIF. Studies were conducted using a vacuum system in flowing mode, with either the organophosphonate alone at pressures up to several tenths of a torr or

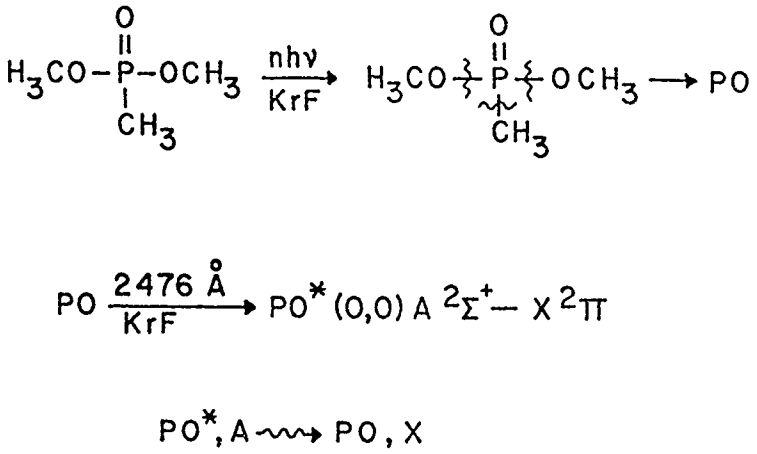


Figure 1. Photofragmentation/LIF detection scheme

buffered with nitrogen at pressures up to atmospheric passing through the flow cell containing four optical ports at right angles. The laser beams passed through the cell collinearly from opposite sides, while fluorescence was viewed perpendicularly to the laser propagation axis. The fluorescence was collected by a lens system and dispersed in a monochromator. Output of the photomultiplier detector was monitored by a boxcar averager (for spectral recording or reaction rate studies) or transient digitizer (for lifetime measurements). The excimer laser pulse (20 nsec width, energies up to 50 mJ/pulse used) was focused into the cell with a 200 mm focal length lens. The dye laser probe beam (5 nsec width, energy typically 0.4 mJ/pulse) was used unfocused. A master trigger system controlled the time delay between the photolysis and probe lasers.

Figure 2 displays the spectrum of fluorescence excited by 247.7 nm radiation 250 nsec after a focused KrF laser pulse acted upon approximately 0.7 torr dimethyl methylphosphonate (DMMP). The same LIF spectrum is obtained when the ArF laser is used for photolysis. The LIF spectrum consists of several sets of apparent doublets. These spectral features are confirmed as bandheads in the PO A-X electronic transition by the correspondence of our observed wavelengths with those given by PO A-X spectroscopic data in the literature.[3] The fluorescence excitation spectrum of the A-X (0,0) band recorded by setting the monochromator at a (v',v") = (0,1) bandhead and scanning the dye laser from 246 to 248 nm yielded bandheads whose wavelengths again agreed with literature values. (The distinctive and characteristic doubling of the fluorescence bands is due to the spin-orbit splitting in the $X\,^2\Pi$ ground state.) Taken together, the fluorescence and excitation spectra show unequivocally that the PO radical is formed during ArF and KrF laser photolysis of DMMP.

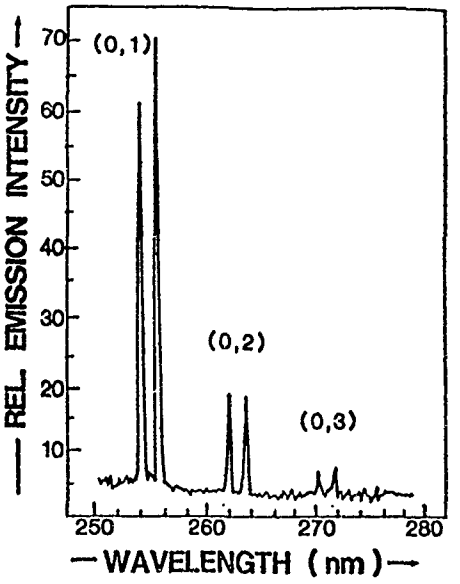


Figure 2. Laser-induced A-X fluorescence of PO

The radiative lifetime of the $A^2\Sigma^+$ state of PO was measured using the transient digitizer with the excitation laser (delayed by 350 nsec after the photolysis laser pulse) set at 247.7 nm (0,0 bandhead) and the monochromator set to pass a (0,1) bandhead using 0.5 torr DMMP. The observed A—X fluorescence decay lifetime yielded a radiative lifetime of 9 $\pm$ 2 nsec,[2] which has since been verified by Anderson and coworkers.[4] This short lifetime is indicative of a high oscillator strength for the A - X transition, demonstrating that this transition is a particularly sensitive vehicle for PO detection and encouraging our interest in incorporating it into a scheme for phosphonate detection.

In order to determine fluorescence quantum yields for PO A— X at elevated pressures, it is necessary to know rates of quenching of the A state by the atmospheric constituents nitrogen and oxygen. These have been obtained from Stern-Volmer plots of the inverse lifetime of the A state versus pressure in the 1 - 10 torr range for N_2 and O_2 as collision partners.[2] The slopes of these plots yield identical quenching rate constants $k_Q = 1.8 \times 10^{-10}$ cc/molecule/sec (5 microsec^{-1} torr^{-1}) for N_2 and O_2. This rate is near collisional, as the hard-sphere collision rate for PO on N_2 or O_2 is about 12 microsec^{-1} torr^{-1}. These data, coupled with the A state radiative lifetime, indicate that a 3% fluorescence quantum yield in the PO A-X transition may be expected at atmospheric pressure.

The actual demonstration that detection of organophosphonates at elevated pressures by this technique is feasible has been accomplished by measuring the PO LIF intensity as a function of nitrogen pressure up to atmospheric (760 torr) for the simulants DMMP, disopropylmethylphosphonate (DIMP), and diethyl ethylphosphonate (DEEP).[5] Figure 3 displays the results for DIMP using the ArF and KrF lasers for photolysis and includes the pressure dependence of the quantum yield calculated using the A state radiative lifetime and quenching rate by N_2 discussed earlier. Using the KrF laser, all three simulants are detectable at atmospheric pressure. As seen in Figure 3 for DIMP, the dependence of LIF intensity on pressure for KrF photolysis follows the quantum yield pressure dependence, indicating that collisions do not substantially affect the photolytic generation of PO using the KrF laser on organosphonates. The same is not true for ArF photolysis, for which the PO LIF intensity falls off rapidly with increasing pressure. This observation provides another motivation for choosing the KrF laser for the photofragmentation/LIF detection scheme.

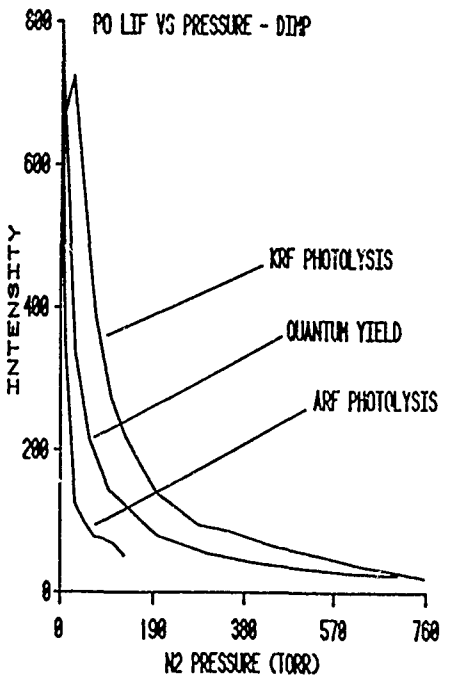


Figure 3. Pressure dependence of PO A-X fluorescence

Under ambient conditions, oxygen is present and one needs to be concerned about reaction of PO with O_2. We have monitored using LIF the rate of loss of PO by reaction with O_2 under pseudo-first order reaction conditions over an oxygen pressure range of 0.25 - 10 torr using KrF laser photolysis of DMMP[2] as well as CO_2 laser infrared multiphoton dissociation of DMMP.[6] The bimolecular rate constants extracted for the PO + O_2 reaction using these two different photolysis techniques are nearly identical: 1.2×10^{-11} cc/molecule/sec and 1.4×10^{-11} cc/molecule/sec, respectively. About one in forty collisions of PO with O_2 results in reactive loss of PO. The implication is that at about 20 nsec after the photolysis laser pulse, the PO concentration will be 1/e of its initial value. If a two-laser approach were used in this detection scheme, the probe laser delay should be less than 20 nsec to be effective.

An abundance of other data has been collected in these studies, such as the laser power dependence for the generation of the PO radical and internal state population distributions.[2] Space considerations prevent us from discussing those results, except to note that only ground vibrational state v"=0 PO is produced in the KrF laser photolysis of DMMP; probing by LIF via the A-X (0,0) band is thus the appropriate approach.

We conclude this section with the suggestion that the KrF excimer laser photofragmentation/LIF scheme is a viable and valid concept for the detection and/or monitoring of organosphonates under atmospheric pressure conditions. Since focusing of the excimer laser is necessary, point detection application is mandated. A clear potential advantage of this technique over other point detection systems is that information about the sample is transmitted by light: thus the sample need not enter the detection system, eliminating the memory effects many point detection systems experience.

III. UV LaserIonization Mass Spectrometry of Organosphonates.

The generation of neutral radical fragments as discussed in the previous section is one type of process occurring upon focused UV laser irradiation of a target gas. The second dominant process is the generation of ionic species. Thus, if a focused UV laser beam interacts with a beam of molecules in the ionization source region of a mass spectrometer, a mass spectrum of the sample may be recorded. Since mass spectrometry is the central technology for several agent detection systems at various stages of development, it is important to examine laser ionization of agent simulants to ascertain whether this ionization technique holds potential advantages for application to mass spectrometric detection of agent species.

Laser ionization mass spectra of DMMP and DIMP have been generated by focusing (typically with a 10 cm focal length lens) the output of a KrF or ArF excimer laser into a beam of the target molecules. The sample is injected into the vacuum chamber through a 0.5 mm diameter orifice along the axis of a quadrupole mass spectrometer. Pressure during mass spectral recording is 7×10^{-6} torr. The laser is operated typically at 20 or 50 Hz and the mass spectrometer is scanned at 0.2 amu/sec. The electron multiplier output of the quadrupole is amplified and broadened. The amplified current pulse is averaged in a boxcar integrator, whose output is recorded by computer. For power dependence measurements, dielectric filters are interposed in the laser beam to provide attenuation. Mass spectra generated by the standard electron impact (EI) ionization technique are recorded with the same mass spectrometer operating under the same conditions for laser ionization.

Figure 4 displays for comparison the mass spectra generated by 70 eV EI ionization (Figure 4a) and 2.5 mJ KrF (Figure 4b) laser radiation on DMMP. Prominent ions in the laser ionization spectra are at m/e = 124, corresponding to DMMP molecular ion, and at 109, 94, and 79, all corresponding to fragment ions attributable to DMMP. For KrF laser ionization, these ions, plus the less abundant fragments at m/e = 93, 92, and 91, are the only ions observed. In the case of ArF laser ionization, additional smaller fragments at m/e = 12 (C+), 15 (CH_3+), 31 (CH_3O+), 47 (PO+), and 63 (PO_2+) are observed. These smaller fragments appear as the result of greater photochemical activity at the shorter wavelength of the ArF laser and become relatively more prominent as the laser energy is increased. (Figure 4c will be referred to later.)

For both KrF and ArF laser ionization the dependence of ion current signal for the principal high mass ions (79, 94, 109, 124) on laser power is near linear and nearly the same for all ($n \simeq 0.95$ for KrF and $n \simeq 0.85$ for ArF, where n is defined by $S \propto I^n$, with I = laser energy, S = signal). This result is interesting, because since DMMP absorbs only weakly or not at all at 248 nm or 193 nm, the initial event is likely to be a simultaneous, two-photon absorption, leading to the anticipation of at least a quadratic power dependence for parent ionization. The observed reduced power dependence must result from laser power-dependent fragmentation of the parent (molecular) ion. From a practical standpoint, it is useful that all four characteristic high mass ions have nearly the same power dependence, since this means their relative abundances are not strongly dependent on laser power.

Comparison of the laser ionization mass spectra (Fig. 4b) with the EI mass spectrum (Fig. 4a) for DMMP shows no particualr advantage for laser ionization. The molecualr ion, which is the most significant for molecular identification, is prominent in all cases, as are the other high mass characteristic ions.

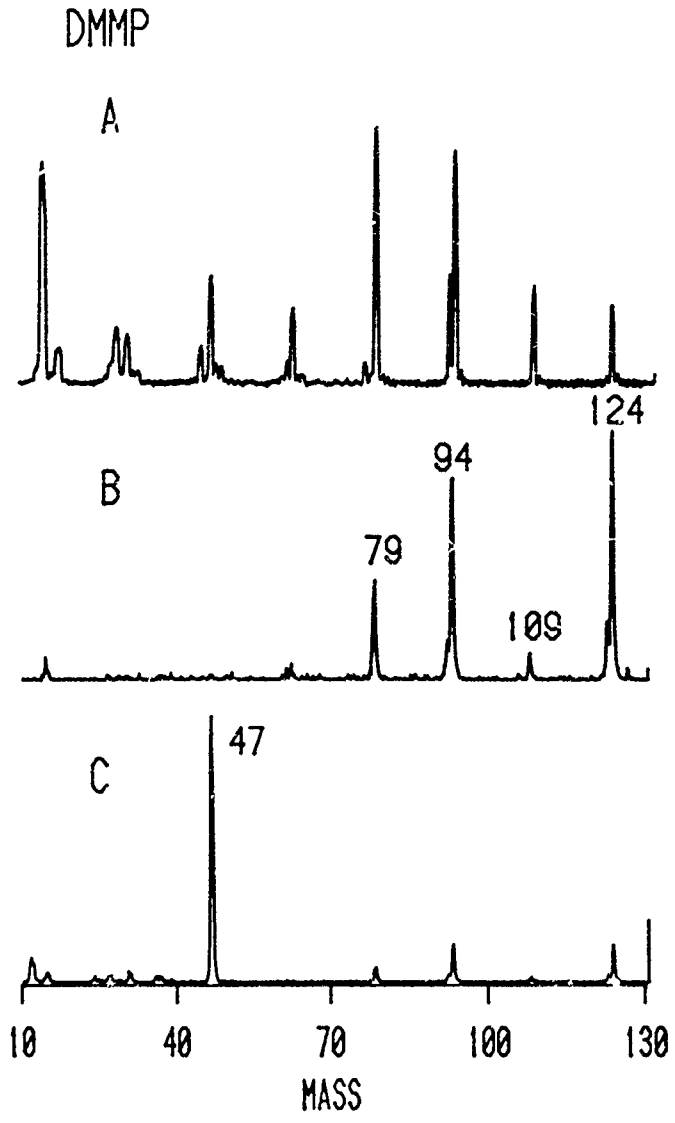


Figure 4. Mass spectra of DMMP. A:EI, B: 2.5 mJ KrF, C: 0.5 mJ, 246.3 nm dye laser

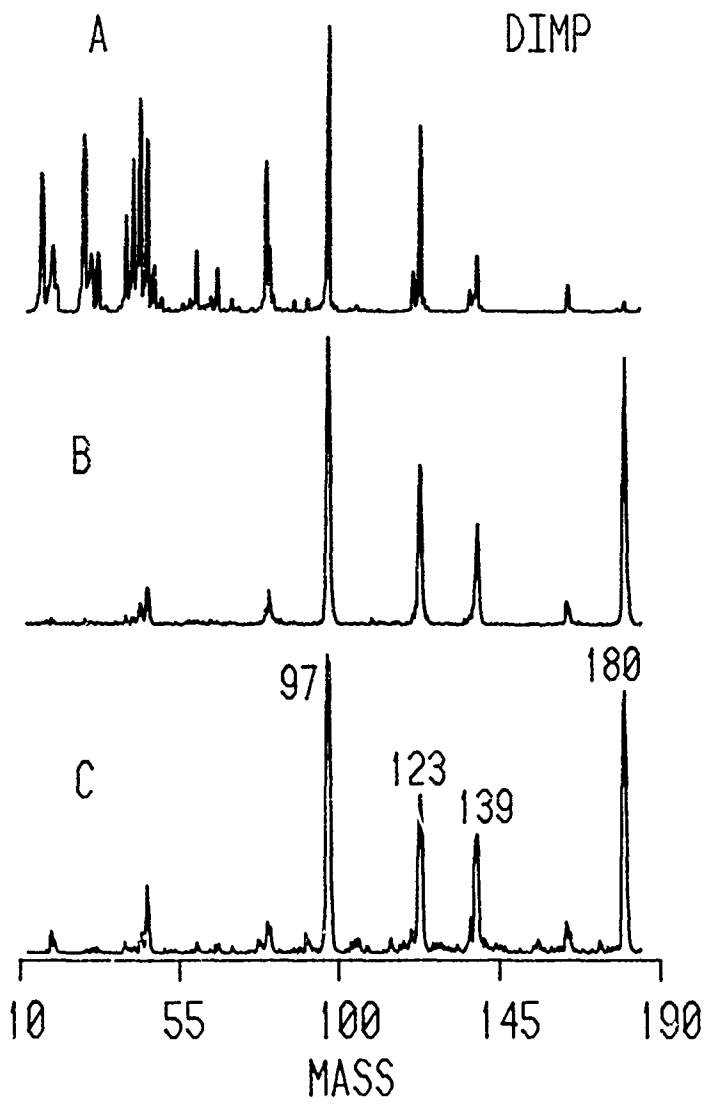


Figure 5. Mass spectra of DIMP. A: EI, B: 0.04 mJ ArF, C: 0.6 mJ KrF.

By contrast, quite striking differences are observed between laser ionization mass spectra of DIMP and the EI mass spectrum of DIMP as displayed in Figure 5. The molecular ion at m/e = 180 is barely visible using EI (Figure 5a), while it is very prominent for both KrF and ArF laser ionization (Figures 5c and 5b, respectively). Furthermore, a considerably lower degree of fragmentation characterizes the laser ionization mass spectra in comparison with the EI spectrum. Aside from these differences, the same high mass fragment ions (at m/e = 97, 123, 138, and 165) are observed for EI and excimer laser ionization, with qualitatively similar relative abundances. It appears probable that, following ionization of the neutral parent, the DIMP ion absorbs an additional photon and fragments to produce the high mass fragment ions. The relative abundances reflect the relative stabilities of these ions. DIMP is somewhat easier to ionize than DMMP at both 248 nm and 193 nm due to higher absorption coefficients at these wavelengths. The laser pulse energy required for 193 nm ionization is so small that a resonant two-photon process must dominate.

The fact that laser ionization generates a very prominent molecular ion is important, since the molecular ion is the single most significant one for identification. These results, particularly for DIMP, suggest that UV laser ionization is potentially more advantageous for organophosphonate detection than the standard EI ionization source currently being incorporated into several mass spectrometer-based detection systems for such purposes.

An additional observation deserves expression at this point. The structurally significant mass 47 PO^+ ion was not generally observed in the above studies due to the fact that the ionizing laser beam was somewhat defocused in the ion lens region of the mass spectrometer. When the focusing lens is translated to a position such that the focus is on the quadrupole axis, mass 47 is observed. We have examined the wavelength dependence of the PO^+ ion in the range 243-253 nm with DMMP as the sample molecule and display this in Figure 6. The tunable dye laser system (5 nsec pulse width) was used to generate the laser radiation. In the 243-249 nm region, several distinct spectral features are observed. These peaks occur at wavelengths corresponding to bandheads of the PO A-X electronic transition, as denoted.

The appearance of these features verifies that the mass 47 signal generated by laser ionization with wavelengths in the PO radical A-X region is substantially due to laser photofragmentation of the parent organophosphonate to form the neutral PO radical followed by resonant two-photon ionization (R2PI) of the PO neutrals. R2PI of PO via the A-X transition is expected to be facile, since the transition is very strong as evidenced by the radiative lifetime of 9 nsec for the A state.[2]

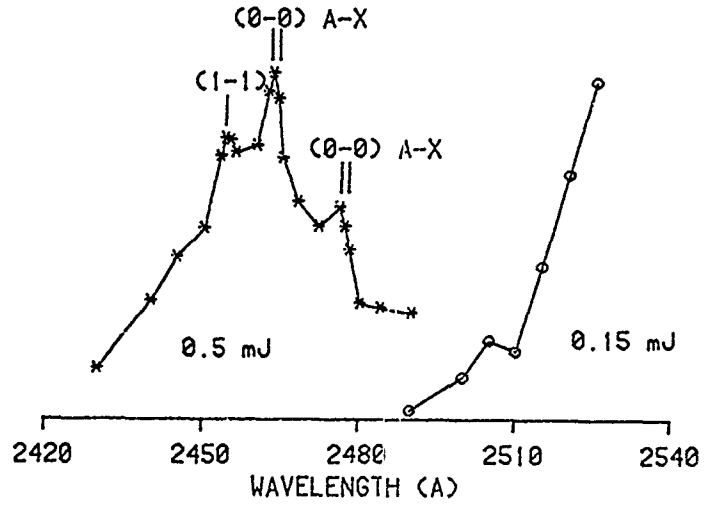


Figure 6. Wavelength dependence of mass 47 PO^+ ion.

Figure 4c is a mass spectrum recorded under the same conditions applied for Figure 6 and shows that at 246.3 nm, mass 47 is by far the strongest peak. Since nearly all the PO^+ signal is due to resonant ionization of PO neutrals formed by photofragmentation of the parent phosphonate, a substantial portion of the total ion intensity generated by the focused laser is due to the neutral photofragmentation/PO ionization process.

IV. Laser ionization/ion mobility spectrometry.

Ion mobility spectrometry (IMS) provides a relatively simple means of detecting and identifying chemical agent (and other) molecules. Ions characteristic of the analyte molecule are identified according to their mobility, as evidenced by their drift time over a known distance through atmospheric pressure gas under the influence of an applied electric field. In the usual implementation of this technique, ions are generated by ion-molecule reactions and the characteristic ions used for detection and identification are typically the protonated monomer ion $(M + H)^+$ or an adduct monomer ion $(M + R)^+$ where R = reagent ion.

The chemical nature of this ionization process leads to several potential disadvantages. The chemistry used for efficient ionization of one molecule may not be efficient for another class of molecules which one is interested in detecting. Furthermore, since the efficiency of chemical processes depends usually on temperature, the sensitivity of an IMS device based on chemical ionization may be dependent on temperature.

In addition, the signal due to a given analyte molecule will likely be apportioned among more than one peak, e.g., $(M + H)^+$, $(M + R)^+$, etc., which results in an effective reduction in signal amplitude.

Laser ionization is a potential means of circumventing some of the problems associated with chemical ionization because it is a direct ionization process. For organophosphonates, the possibility exists also to take advantage of the characteristic PO radical also generated by the ionizing laser pulse. With chemical ionization or with laser ionization at an arbitrary UV wavelength, in the ideal case, only a single peak due to the molecular ion can be expected. However, by using a laser beam with wavelength resonant with the PO A-X transition near 247 nm, as described in the previous section, it may be possible to add a peak due to PO^+ to the ion mobility spectrum. The addition of such a wavelength-dependent peak to the IMS of organophosphonates would more than double the information content of the spectrum.

The data of the Section II and III studies support a reasonable expectation that PO^+ can be generated by laser ionization at atmospheric pressure. The PO radical is formed in v"=0; thus a (0,0) A-X bandhead is most appropriate as the resonant intermediate transition. Resonant two-photon ionization is facile via the strong A-X transition. Collisions do not adversely effect the production of PO radicals in the laser photofragmentation of DIMP. The photofragmentation process leading to formation of PO radicals is of approximately the same importance as the process of ionization of the parent compound.

Laser ionization/IMS spectra were recorded using an ion mobility spectrometer built by PCP, Inc. Part-per-billion level concentrations of the analyte in zero air or UHP nitrogen are generated using the diffusion vial method in an AID standards generator. Approx. 15 mm downstream from the IMS sample inlet, the sample gas flow is crossed by a focused laser beam (10 cm focal length lens) at the central axis of the IMS. Ions generated by the laser drift against a counterflow of clean drift gas (zero air or ntirogen) under the influence of an electric field toward the collector 8.3 cm away. The current output of this collector is amplified, digitized, and accumulated in a signal averager or DEC MINC computer.

Figure 7 displays IMS spectra of DIMP at several concentrations in UHP N_2 using 246.3 nm laser radiation at an energy of 0.5 mJ focused by a 10 cm focal length lens. (Other conditions: IMS temperature 190°C, sample flow 128 cc/min, clean UHP N_2 drift flow 400 cc/min, electric field 144 V/cm). At all concentrations, only a single peak with K_o = 1.89 $cm^2/V/sec$ (at low concentrations) is observed. This peak is readily attributed to the DIMP monomer ion. Changes in all relevant

conditions, such as concentrations, IMS temperature and gas flows, laser wavelength, etc., did not allow observation of a mobility peak attributable to the PO^+ ion.

It is possible that collisions interfere with the resonant two-photon ionization of PO by quenching the real intermediate A state. A considerably higher laser intensity than that available with our dye laser system may be required to allow the ionization process to compete effectively with the collision quenching of the A state. Higher intensity laser radiation will be generated for continuing these studies.

The changes in peak widths for the DIMP monomer ion in Figure 7 deserves comment. The peak width increases from 750 to 950 microseconds with increase in concentration from 7 ppb to 50 ppb. There is no change in the drift time corresponding to the peak center. We attribute the increase in peak width to space-charge effects, i.e., when the number of ions occupying a given volume, such as the laser focal region, is increased, ion-ion repulsion is more substantial and results in peak broadening. When the concentration reaches 200 ppb, not only are space-charge effects intensified but also the formation of dimer ions occurs during the drift time, with the result that the peak is both broadened considerably and the centroid of the peak moves to longer drift time.

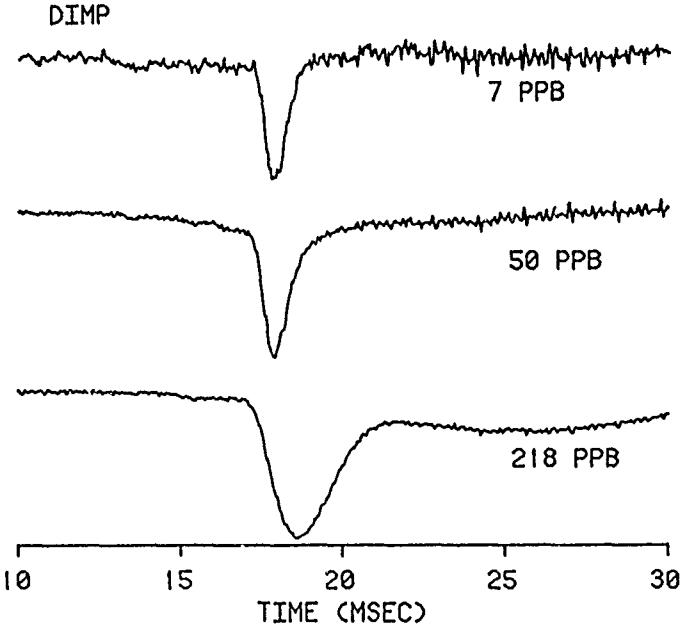


Figure 7. Laser ionization/IMS spectra of DIMP.

Due to space-charge effects and, at higher concentration, dimer formation, a simple plot of the peak intensity versus concentration is not linear. However, if one plots the integrated signal due to the DIMP monomer versus concentration, a linear relationship is indeed observed, suggesting that the laser generation of DIMP ions at atmospheric pressure is a simple, direct process.

V. Conclusion.

Several chemical agent detection systems may potentially benefit from the application of ultraviolet laser techniques. The concept of UV laser photofragmentation with laser-induced fluorescence detection of characteristic fragments is feasible under ambient conditions. The advantage of this scheme in a point detection role is that the sample under study need not contact any part of the detection system (since the information is transmitted by light), thus avoiding the sample memory effects which sometimes plague point detection systems which require intake of the sample.

Mass spectra of organophosphonates generated by laser ionization display substantial intensity in the molecular ion peak, the most significant for identification purposes. By contrast, the electron impact ionization technique, when applied to molecules even as small as DIMP, yields quite low intensity molecular ion peaks. Mass spectrometer-based detection systems may ultimately see enhanced identification capabilities by application of laser ionization. Laser ionization at atmospheric pressure in an ion mobility spectrometer appears capable of the same low ppb sensitivity of current chemical ionization/IMS devices such as the CAM and ACADA. The potential exists for substantially improving the identification capabilities of IMS-based devices through selective ionization of characteristic neutral fragments.

REFERENCES

1. S. R. Long, R. C. Sausa, and A. W. Miziolek, Chem. Phys. Lett., 117, 505 (1985).

2. R. C. Sausa, A. W. Miziolek, and S. R. Long, J. Phys. Chem., 90, 3994 (1986).

3. M. N. Dixit and N. A. Narasimham, Proc. Indian Acad. Sci., A68,1 (1968); B. Coiquart, C. Couet, N. Twan Arh, and H. Guenbant, J. Chem. Phys., 64, 1197 (1967).

4. K. N. Wong, W. R. Anderson, and A. J. Kotlar, J. Chem. Phys., 85, 2406 (1986).

5. A. W. Miziolek, R. C. Sausa, B. E. Forch, and S. R. Long, Ballistic Research Laboratory technical report, in press.

6. S. R. Long, S. D. Christesen, A. P. Force, and J. S. Bernstein, J. Chem. Phys., 84, 5965 (1986).

Computational Predictions of Close-In Blast Loading from Bare Spherical Charges (U)

Mr. Richard E. Lottero* and Mr. John D. Wortman

US Army Ballistic Research Laboratory
ATTN: SLCBR-TB-B
Aberdeen Proving Ground, MD 21005-5066

I. INTRODUCTION

Light combat vehicles such as the Bradley Fighting Vehicle (BFV) are being postulated as playing an increasingly important support role for both troops and other more heavily armored combat vehicles. As such, they have a much greater risk of coming under fire from anti-armor weapons. Accordingly, there is a need to quantify the blast loading on these vehicles so that they may be hardened against these blast loads, both for the survivability of the troops using the BFV, and for the vehicle itself. One of the survivability measures being considered is the exterior stowage of the TOW II anti-tank missiles that are carried by the BFV to eliminate the possibility of the accidental detonation of the warhead inside the BFV. However, stowage of the missiles on the outside of the BFV then leaves them in a more vulnerable position where a direct hit by an attacking shaped charge weapon could cause a high-order detonation of a warhead. Thus, the TOW II can itself become a threat to the BFV. Some work has already been done on developing buffering layers to be placed between the individual stowed missiles and between the missiles and the BFV hull to provide protection for the BFV.

The objective of this study is to make a first order estimate of the close-in blast loading on the BFV hull from the detonation of the warhead of a TOW II missile in a stowed position. The term "close-in blast" is used to describe situations wherein a condensed secondary explosive is about five or less charge radii from a surface and it is probable that the debris and explosive products will have a significant interaction with the target surface. Somewhat simply stated, materials subjected to this type of intense blast loading can fail locally through a sequence of high-intensity wave interactions which cause local shattering and spallation, or a broader impulse-driven failure of the structure through bending or breaking at welds. While the material response computations themselves are not the subject of this study, the nature of the computed loading can in-

dicate what type of failure, if any, might be expected. The addition of buffering material between the stowed missiles and the BFV hull, with no change in over-all standoff, can change the nature of the loading from a high-intensity wave interaction by reducing the peak pressures and spreading the waves out in time, but at the possible risk of actually increasing the impulse delivered.

Even in a controlled test situation, close-in blast produces a very severe environment. The transition of the detonation process from an initiator to an explosive charge and the subsequent buildup of a detonation front is complex, even for spherical charges. The pressures are extremely high and are not well known. Detonation front pressures for one-dimensional (1-D) laboratory reference experiments are inferred from detonation front speeds [1]. There is an intense debris and hot explosive products field in which gages with the required frequency response do not survive for a long enough period to give meaningful data. The validity of equation of state data is also in question for the first several volume expansions of the explosive products. It is also difficult to accurately predict the response of buffering materials. Because of the high stresses and strain rates, their material properties and equations of state are not well known. Their behavior is probably highly non-linear. Also, there may not be sufficient time for the buffer to dissipate enough energy or attenuate momentum, unless sophisticated techniques can be found to design multilayer buffers with impedance mismatches tuned to disrupt the principal anticipated loading frequencies.

This study was done using theoretical hydrodynamic computer codes (hydrocodes) and, where possible, comparing the computed loading to the total impulse estimates from recent experiments with a ballistic pendulum. A total of seven hydrocode computations are discussed here. Once the techniques for computing the loading are better developed, this loading can then be used as input to structural response codes to determine the response of the BFV hull. The primary emphasis of this paper is on seven hydrocode computations performed for simplified geometries. The initial work was done on the US Army Ballistic Research Laboratory (BRL) CDC 7600 (now removed) on which only limited resolution two-dimensional (2-D) problems could be run because of speed and storage limitations. Subsequent work has been performed on the BRL CRAY X-MP/48, which has allowed finer finite difference grid resolution for 2-D problems because of its greater speed and memory.

II. CONFIGURATIONS: ACTUAL AND SIMULATED

One proposed stowage configuration for the TOW II missiles is to mount them externally, typically in pairs head to tail with a buffer between them and with their

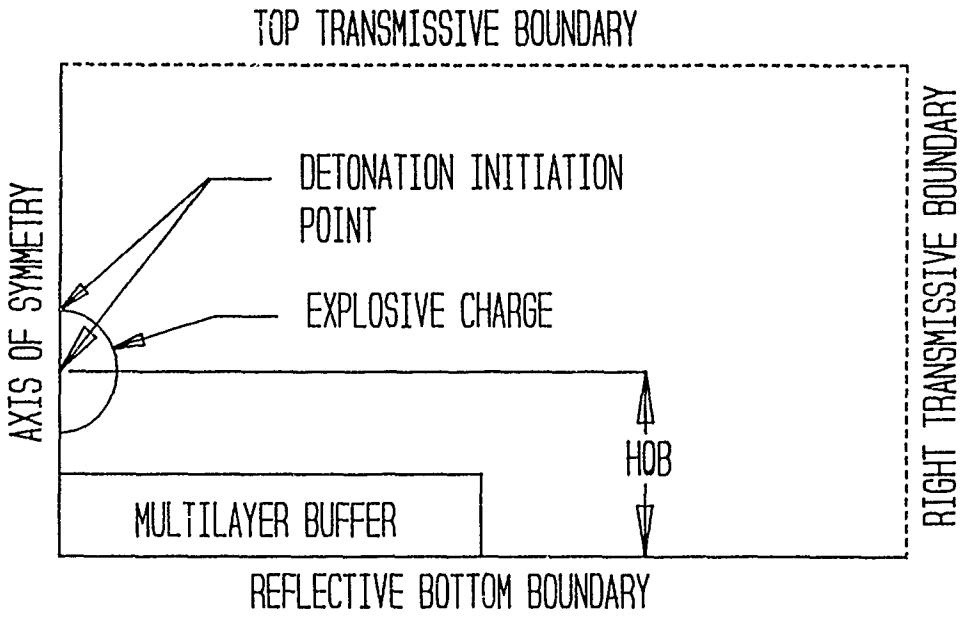


Figure 1: The Idealized Bare Spherical Charge Over a Reflecting Plane.

axes mutually parallel and parallel to the BFV hull. Several possible stand-off distances, measured from the hull surface to the missile axis, are considered here, including 15.24 *cm*, 17.78 *cm*, 18.18 *cm*, and 33.02 *cm*. Various multilayer buffers have been considered, which are to be placed between the missiles and the hull with the bottom layer typically in contact with the hull. A simplified model of one of these buffers is considered here. The TOW II warhead is a shaped charge with 3.09 *kg* of LX-14 coaxial with the missile axis. This warhead has been idealized for the hydrocode computations discussed herein as a bare spherical charge with its height of burst (HOB), measured from the BFV outer surface to the charge center, equal to the stowed missile axis height (see Figure 1). The BFV hull is modeled as a rigid, non-responding plane. Depending on the problem modeled, the detonation initiation point is either at the charge center or the top of the charge.

The actual warhead contains approximately 3.09 *kg* of LX-14, a mixture of 95.5 percent HMX and 4.5 percent estane. The warhead is located symmetrically along the axis of the missile, and includes a shaped charge liner in its hollowed forward section, a light casing around the outside of the explosive, the missile casing itself, and an initiator at the rear. This was simplified into a sphere of LX-14 of equal mass for five of the hydrocode computations, thereby allowing the use of 2-D cylindrical coordinates rather than three-dimensional (3-D) Cartesian. Two computations were done for 3.90 *kg* 50/50

pentolite (hereinafter referred to as "pentolite") spheres. This is 8.2 percent more than a mass with the same detonation energy as the 3.09 *kg* of LX-14. Hydrocode computations for seven separate configurations, all using 2-D cylindrical coordinates, are summarized in Table 1. The logic of the notation used for the problem names is as follows. The first letter stands for the explosive, "P" for pentolite and "L" for LX-14. The second letter stands for the detonation initiation point, "C" for center and "T" for top. When a third letter "B" appears, it indicates that a buffer was modeled. The trailing number indicates the simplified problem number in that particular sequence. The computations will be discussed individually later.

Table 1: Summary of Computations

Problem Number	PC1	PC2	LC1	LT1	LT2	LTB1	LTB2
Expl.	Pent	Pent	LX-14	LX-14	LX-14	LX-14	LX-14
M (kg)	3.90	3.90	3.09	3.09	3.09	3.09	3.09
Det Pt	Center	Center	Center	Top	Top	Top	Top
HOB(cm)	33.02	15.24	15.24	15.24	18.18	18.18	18.18
Time (ms)	1.072	0.876	0.400	0.369	0.360	1.200	1.200
Buffer	No	No	No	No	No	Yes	Yes
Buffer Model Layers	N/A	N/A	N/A	N/A	N/A	Dense Air Layers	Dense Air + RHA
Det Modeled	SAP Prog	SAP Prog	HULL F.Fire	HULL F.Fire	HULL F.Fire	HULL F.Fire	HULL F.Fire
Radial Cells	HULL 148	HULL 148	HULL 612	HULL 612	HULL 614	HULL 614	HULL 614
DX (cm) at Charge	0.500	0.500	0.246	0.246	0.246	0.246	0.246
Axial Cells	HULL 140	HULL 140	HULL 475	HULL 475	HULL 486	HULL 486	HULL 486
DY (cm) at Charge	0.500	0.500	0.246	0.246	0.246	0.246	0.246

III. HYDRODYNAMIC COMPUTER CODES

All hydrocode blast loading computations were done using versions of the HULL hydrodynamic computer code. The two computations for the pentolite spheres, PC1 and PC2, used an airblast version of HULL [2,3] that has been at the BRL for several years. This version evolved from an earlier version of HULL acquired from the US Air Force Weapons Laboratory. The modeling of the detonation process for these two computations was done using the 1-D Lagrangian hydrocode SAP [4]. SAP uses a programmed burn routine where the detonation front progresses at a prescribed Chapman-Jouguet (C-J) velocity [1]. The explosive burn routine in SAP incorrectly added the detonation energy to the unburned explosive ambient energy instead of replacing it, resulting in a slightly more energetic explosion than theory predicts. This error was not discovered until after the computations were completed. The pentolite explosive products equation of state in SAP consists of a polynomial curve fit based on density and internal energy [unknown source]. SAP was run until the pentolite had fully detonated and the shock front in the air had reached a radius approximately equal to the planned ground plane position for the given problem. The SAP flow field was then mapped into the 2-D HULL grid, where HULL used a similar equation of state for the pentolite explosive products.

The other five hydrocode computations, LC1, LT1, LT2, LTB1, and LTB2, used the most recent version of HULL [5] at the BRL. HULL solves the inviscid Euler equations for fluids, and includes a computation of principal plus deviatoric stresses for materials with shear strength. It uses a modified predictor-corrector central differencing scheme in a fixed Eulerian grid. The detonation of the LX-14 spheres was performed in the HULL code itself, using a forest fire burn technique where the detonation wave front was moved at the local speed of sound in the detonation products immediately behind the front. This is the most stable condition for a detonation wave without an added forcing function, it being neither underdriven nor overdriven [1]. The JWL [6] equation of state was used to describe the LX-14 explosive products. HULL was reprogrammed to replace the unburned explosive ambient energy with the C-J detonation energy when it was detonated. Also, a technique was employed to restore energy to the detonation process that was previously mathematically diffused forward of the detonation wave front because of HULL's pressure iteration method for mixed cells containing both burned and unburned explosive, and then later lost in the C-J energy replacement when that explosive in the cell was burned. Before this technique was employed, as much as 15 percent of the available detonation energy was being mathematically dissipated.

All HULL computations were performed using the multimaterial 2-D axisymmetric cylindrical coordinates version. The projection of the HULL grid on a plane such as shown in Figure 1 has the left boundary as an axis of symmetry, the bottom boundary as

a perfect reflecting plane representing the BFV hull, and the top and right computational boundaries as transmissive. All of the HULL computations used this same geometry and boundary definitions. Options are also available for 2-D and 3-D Cartesian coordinates, but these were not used for this study.

IV. HYDROCODE COMPUTATIONS

Computation PC1, performed on the CDC 7600, is for a center-detonated 3.90 kg sphere of pentolite at an HOB of 33.02 cm with no intervening buffer. The pentolite sphere had an undetonated radius of 8.247 cm. The detonation and initial expansion were computed using the 1-D spherically symmetric SAP hydrocode, with initial cell sizes of 0.250 cm. At a time of 21 μs after initiation, the SAP flow field was mapped into the 2-D cylindrical HULL grid, which had cell sizes in the area of interest of 0.50 cm. The computation was run for a total real time simulation of 1.072 ms. As an example of the loading on the ground plane, Figure 2 shows the overpressure vs time at the point on the ground plane directly below the charge (hereinafter referred to as "ground zero"). Here, the peak overpressure is 110.5 MPa, and the pulse duration is approximately 300 μs. It should be noted here that while pressure peaks are of interest in computations such as these, there is often much uncertainty as to their correct values and validity. Another often more useful measure of the blast loading on the ground plane is the total impulse, integrated over both time and a specified area. The total impulse at the end of the computation time on a 38.10 cm radius circle on the ground plane and centered at ground zero (hereinafter referred to as "the reference area") for computation PC1 is 2.03 kN-s. This radius was chosen because the ballistic pendulum, described later, has a face area nominally equal to that of a 38.10 cm radius circle.

Computation PC2, also performed on the CDC 7600, is identical to PC1 except that the HOB is 15.24 cm, and the SAP computation modeling the same center-detonated pentolite sphere is mapped into the HULL grid at 16 μs after charge initiation. It was run for a total real time simulation of 876 μs. Figure 3 shows the overpressure vs time at ground zero for PC2, which has a peak of 895.4 MPa, 8.10 times greater than that for PC1 even though its HOB is only reduced to 46.2 percent of that for PC1. The total impulse for PC2 on the reference area is 4.64 kN-s at 876 μs, 2.29 times that for PC1.

Computation LC1 is for a center-detonated 3.09 kg LX-14 sphere at an HOB of 15.24 cm performed on the CRAY X-MP/48, as were all succeeding computations. This is the correct TOW II warhead mass and explosive. The detonation of the LX-14 was modeled in the 2-D HULL computation using a forest fire burn routine. As the detonation

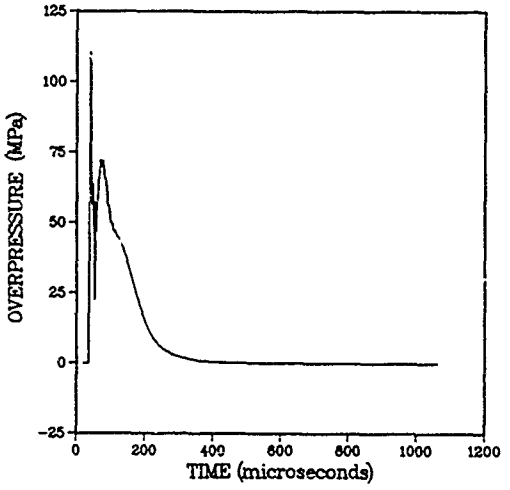


Figure 2: Overpressure vs Time at Ground Zero, Problem PC1.

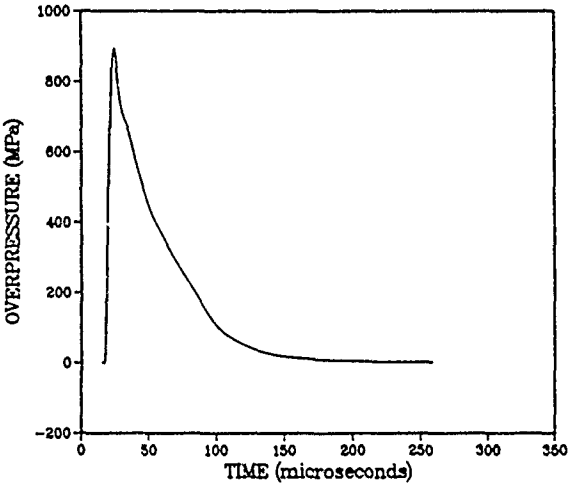


Figure 3: Overpressure vs Time at Ground Zero, Problem PC2.

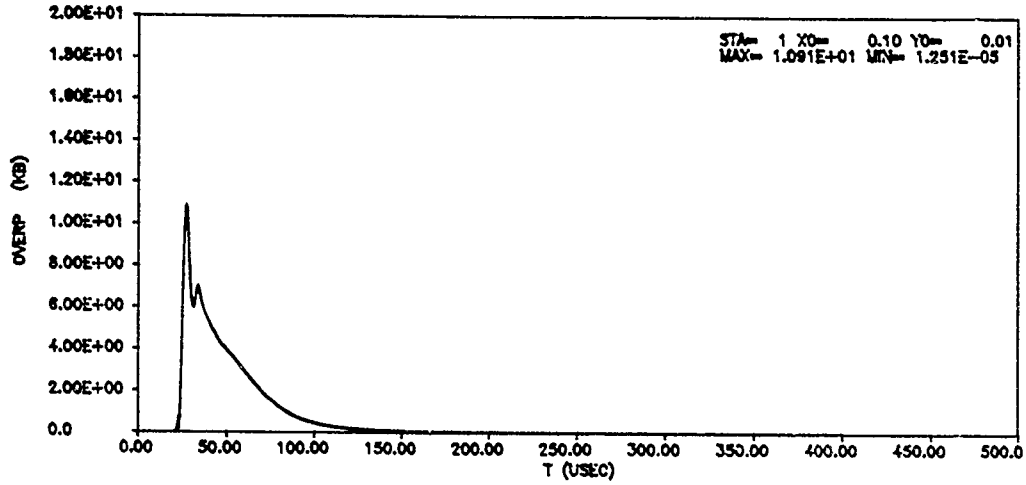


Figure 4: Overpressure vs Time at Ground Zero, Problem LC1.

proceeded, the initial ambient energy of the undetonated LX-14 was replaced by the C-J detonation energy as described previously. The explosive products were modeled using the JWL equation of state and the corresponding published constants [6]. It was run for a total real time simulation of 400 μs, and required 19.75 hours of Central Processor Unit (CPU) time. The flow field cell sizes in the region of interest were approximately 0.246 cm axially and radially, as was the case for the remainder of the computations. Figure 4 shows the overpressure vs time at ground zero for LC1, which has a peak of 1.091 GPa, 21.7 percent greater than that for the same point in PC2. The total impulse on the reference area for LC1 is shown in Figure 5. It has a value of 1.94 kN-s at the 400 μs ending time, 58.2 percent less than that for PC2. This is part of a consistent trend in the computations, where the computations with LX-14 spheres with comparable total energy release to pentolite spheres have significantly lower computed total impulse. This is discussed more fully in the section "Comparisons to Ballistic Pendulum Experiments." While there are no experimental shots for comparable bare LX-14 spheres, the other pendulum results do imply that the computed total impulse values for the bare LX-14 spheres in this and the following computations appear low for reasons stated in that section and in the Introduction.

In a more realistic attack situation, the detonation process in the stowed TOW II warhead would probably begin near the top, the more likely first contact point. This can result in a more intense local loading on the hull directly below the warhead because the detonation has a longer path through explosive material, even though the total mass of explosive is the same. The flow field grid, HOB, and LX-14 sphere for computation LT1 are identical to those for LC1. The only difference is that LT1 has the detonation

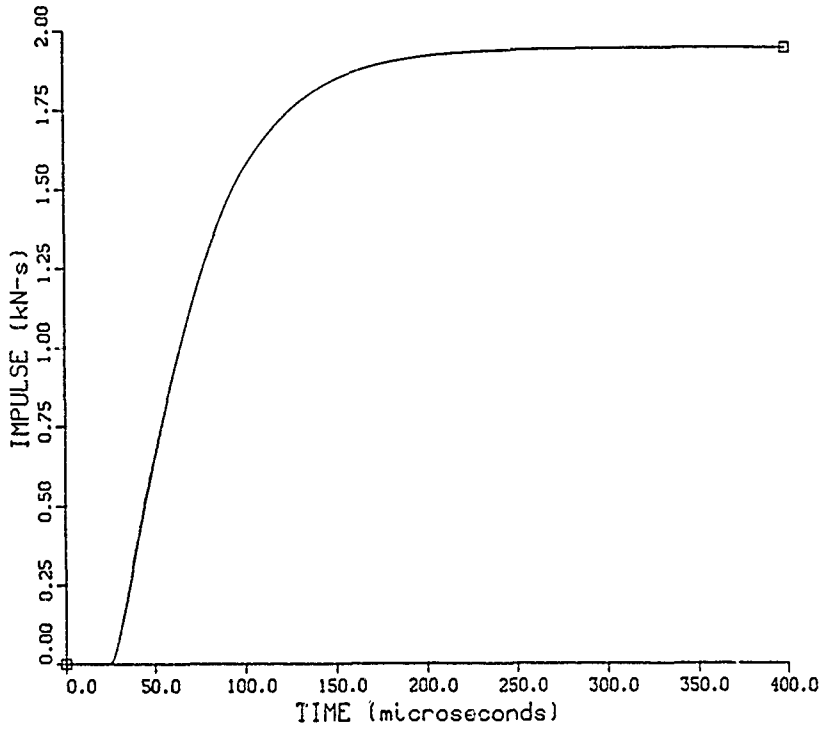


Figure 5: Total Overpressure Impulse On a 33.10 *cm* Radius Circle, Problem LC1.

initiated at the top of the sphere. This computation was run for a total real time of 369.29 μs, requiring 18.37 CPU hours. As may be seen in the overpressure vs time plot for ground zero in Figure 6, the peak overpressure for LT1 is 1.764 *GPa*, 61.7 percent greater than that for LC1. Figure 7 shows the total overpressure impulse on the reference area for LT1, having a value of 2.11 *kN-s* at the ending time of 369.29 μs, 8.8 percent greater than that for LC1 at its ending time.

Because a set of computations involving multilayer buffer simulations was planned where the charge HOB would be at 18.18 *cm*, computation LT2 was run without buffer layers to logically connect the set of computations with buffers to those without. It has nearly the same flow field grid as that for LT1 except for small modifications in the grid below the charge to match the axial and radial edges of the individual layers in the buffer. Like LT1, it has a 3.09 *kg* LX-14 sphere, but at a slightly higher HOB, with the detonation initiated at the top. It was run to a total real time simulation of 359.63 μs, requiring 18.47 CPU hours. Figure 8 shows the overpressure vs time at ground zero for LT2, which has a peak value of 1.082 *GPa*, 38.7 percent less than that for LT1 because of the increase in the HOB. Figure 9 shows the total overpressure impulse on reference area for LT2. It has a value of 1.97 *kN-s* at the ending time of 359.63 μs, 6.6 percent

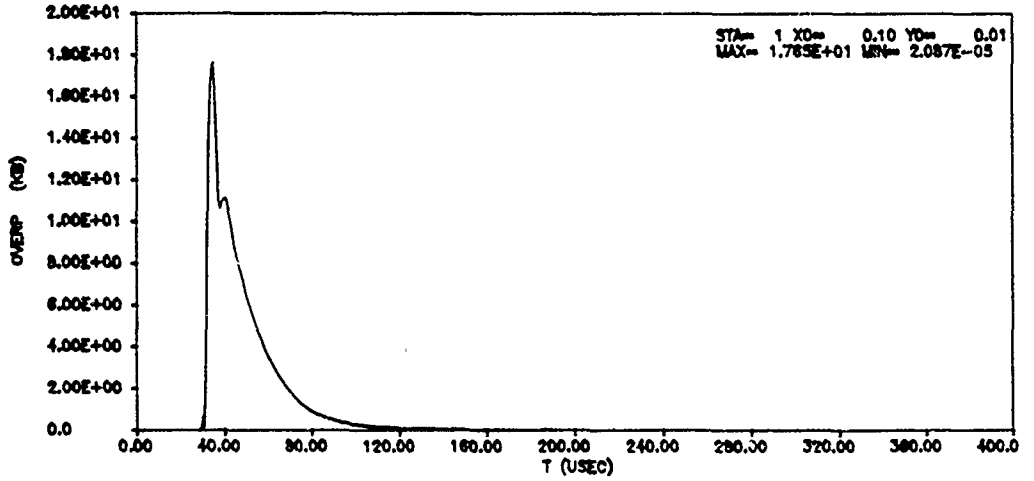


Figure 6: Overpressure vs Time at Ground Zero, Problem LT1.

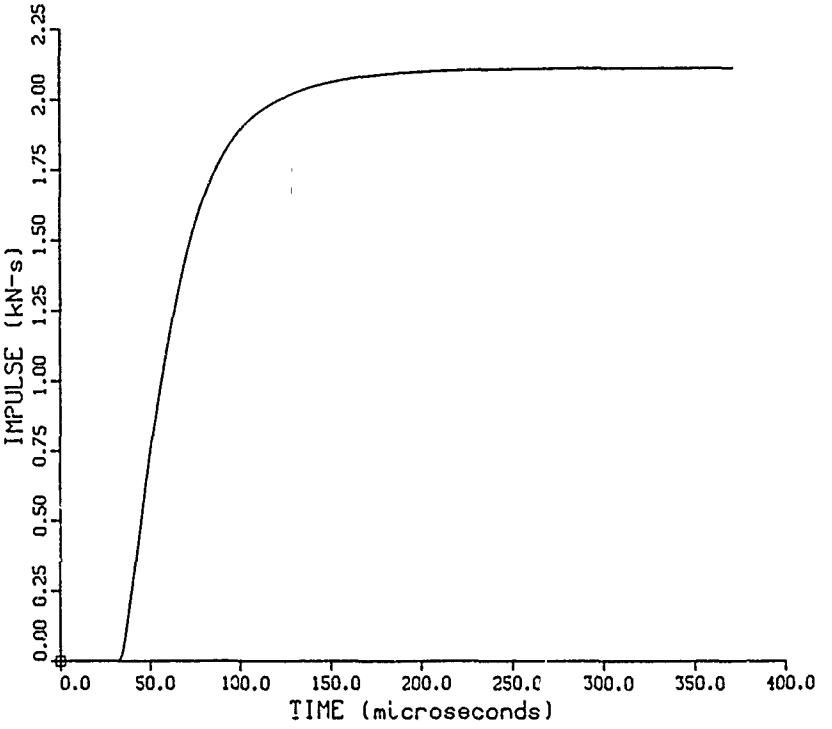


Figure 7: Total Overpressure Impulse On a 38.10 *cm* Radius Circle, Problem LT1.

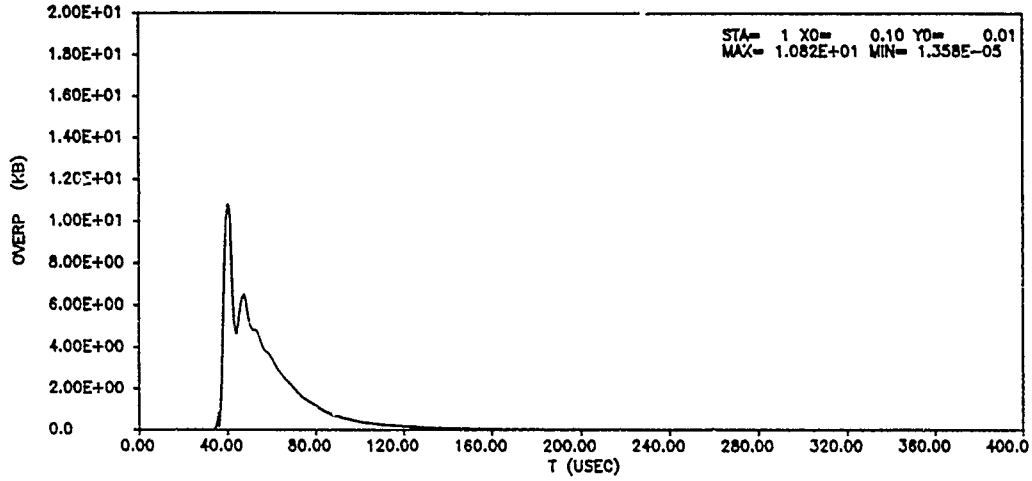


Figure 8: Overpressure vs Time at Ground Zero, Problem LT2.

less than that for LT1 at its ending time.

Computation LTB1 is identical to LT2 except that a simplified version of a proposed multilayer buffer is positioned between the charge and the BFV hull. The actual buffer consists of multiple layers of three different materials, all in contact with each other, and with the bottom layer in contact with the hull. It has a 1.27 *cm* thick top layer of 1.3 g/cm^3 Isodamp, a 1.905 *cm* thick layer of 0.13 g/cm^3 Kevlar-Nomex honeycomb, a 1.27 *cm* thick layer of 7.85 g/cm^3 Rolled Homogeneous Armor (RHA), another 1.905 *cm* thick layer of the Kevlar-Nomex honeycomb, another 1.27 *cm* thick layer of the RHA, and a final 1.27 *cm* thick layer of the Isodamp in contact with the hull. As a first approximation, these layers were modeled as air at ambient pressure with the same dimensions and density of each of the actual materials. This was done in order to preserve the mass of the buffer, on the assumption that in such a probable overmatch of the strength of the buffer materials, the masses and density differences at the interfaces might be the dominant factors in the modification of the loading. The radius of the buffer in the computation was 30.48 *cm*. It was assumed that this artificially dense air can be described at the initial conditions by the polytropic equation of state, $p = \rho I(\gamma - 1)$, where p is the pressure, ρ is the density, I is the specific internal energy, and $\gamma(= 1.4)$ is the ratio of specific heats. (Once the HULL computation began, the dense air comprising the buffer was allowed to behave as a real gas.) In order to start the computation with this artificial buffer at ambient pressure, an artificially low specific internal energy had to be specified for this dense air. The sound speed equation for a polytropic gas, $c = [\gamma(\gamma - 1)I]^{0.5}$, thus indicates that the simplified dense air layers will have artificially low sound speeds. Therefore, the number of materials and types of materials are simplified.

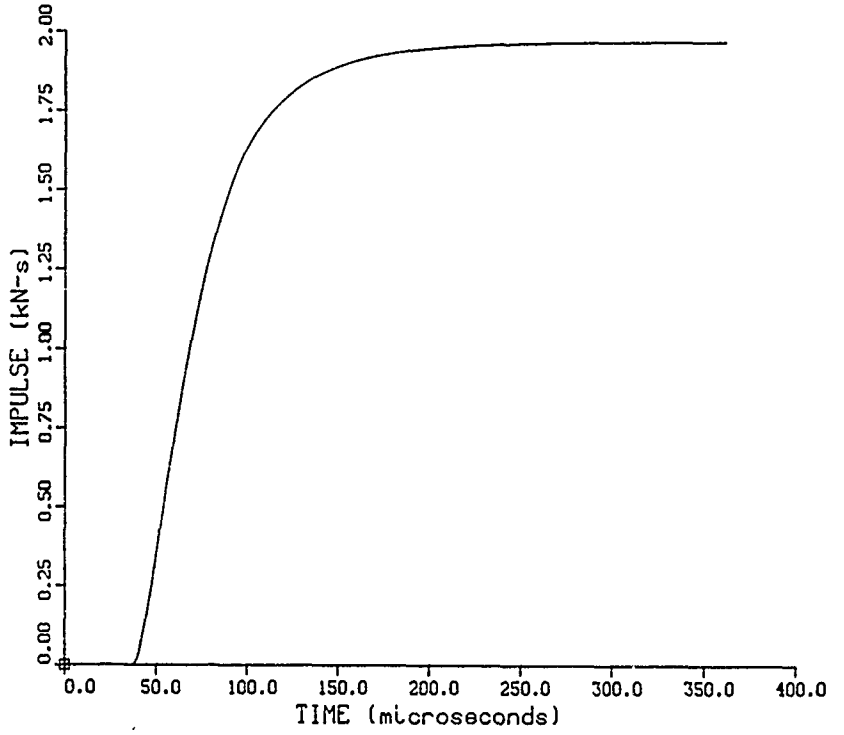


Figure 9: Total Overpressure Impulse On a 38.10 *cm* Radius Circle, Problem LT2.

and the individual masses and densities in the buffer are preserved, but at the expense of incorrectly modeling the impedance mismatches and wave transmission rates in the buffer layers, artificially stretching out the timing in the problem. This computation, LTB1, was run for a total simulated real time of 1.200 *ms*, requiring 48.50 CPU hours. Figure 10 shows the overpressure vs time at ground zero for LTB1. It has a peak value of 1.353 *GPa*, 25.0 percent greater than that for LT2. This increase in peak pressure is contrary to indications from other experiments, and is probably caused by the incident shock wave reflecting too strongly from the buffer because of the incorrect impedance mismatch caused by the combined effects of having a correct density but a low sound speed in the buffer layers. This can send a strong shock into the buffer. Figure 11 shows the total overpressure impulse on the reference area, which has a value of 7.86 *kN-s* at the ending time of 1.200 *ms* and is still increasing rapidly. This total impulse value is 3.99 times that for LT2 at its ending time. While some of this additional impulse may be due to the impedance mismatch discussed above, it is also due in part to the existence of the buffer itself, an effect also seen in other ballistic pendulum experiments not discussed in detail here. Briefiy, the addition of a buffer effectively lowers the apparent HOB, exposing the top layers to a much stronger blast. This effect was demonstrated in the comparison of computations PC1 and PC2. If the buffer does not have a layering with

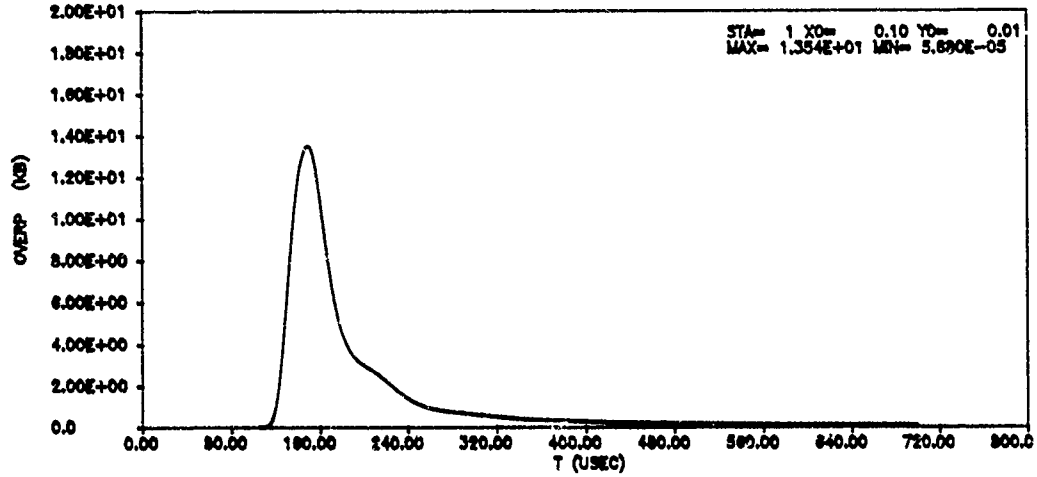


Figure 10: Overpressure vs Time at Ground Zero, Problem LTB1.

impedance mismatching designed to defeat the anticipated blast, it may be that some layers can act as energy-gathering devices that can themselves become attacking weapons when they drive into the surface they were meant to protect.

Computation LTB2 is identical to LTB1 except that in the buffer the two RHA layers are actually modeled as RHA. Computation LTB2 was also run to a total simulated real time of 1.200 ms, but because of the effect of adding to the problem another material which also has a high sound speed, it required 127.95 CPU hours. This was a factor of 2.64 times the CPU time required for LTB1. This graphically illustrates the reasons for attempting to find a simple way of modeling the complex materials in the buffer. Figure 12 shows the overpressure vs time at ground zero for LTB2, with a peak value of 2.878 GPa, 2.13 times greater than that for LTB1, primarily because of the impact of the RHA plates. Unfortunately, the total impulse on the reference area cannot be presented because some of the data gathering points in that area stopped storing loading data for unknown reasons during the running of the problem. Ballistic pendulum tests with buffers having RHA plates not only typically showed increases in impulse over unbuffered shots, but also had the RHA plates strike the pendulum hard enough that the RHA plates rebounded as much as 10 m.

V. COMPARISONS TO BALLISTIC PENDULUM EXPERIMENTS

This section provides a brief summary of some of the ballistic pendulum experiments run previously [7] at the BRL. The purpose of these experiments was to provide

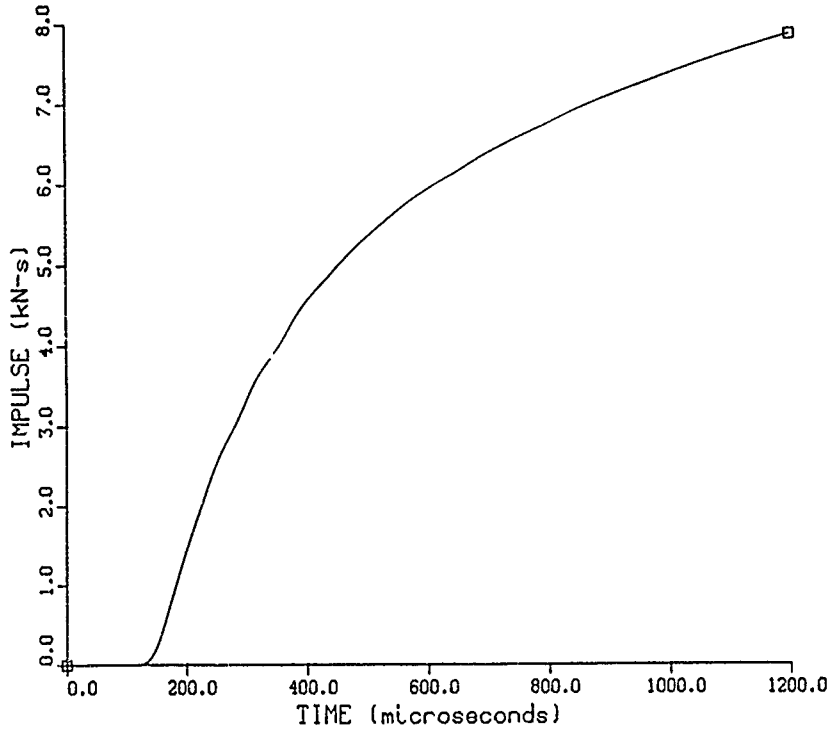


Figure 11: Total Overpressure Impulse On a 38.10 *cm* Radius Circle, Problem LTB1.

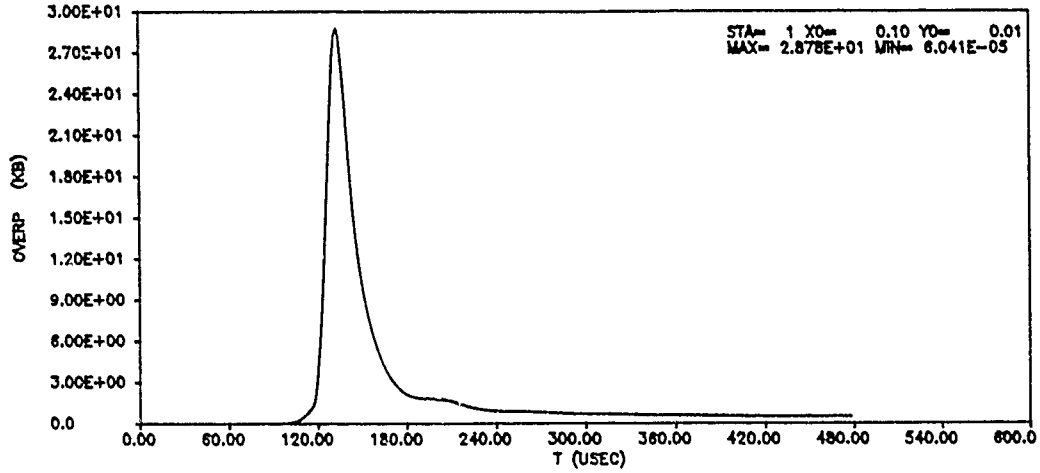


Figure 12: Overpressure vs Time at Ground Zero, Problem LTB2.

experimental reference data to estimate the total impulse delivered by bare spherical pentolite charges and various attack situations against stowed TOW II missiles. The charcteristics of the ballistic pendulum are as follows. It has a mass of 3.202 Mg, a length from its pivot point to its center of gravity of 2.18 m, a length to its center of face area of 2.44 m, a moment of inertia of 44.17 kg-m^2, a length to its center of percussion of 2.44 m, a period of 3.136 s, and a face area of 0.452 m^2. This face area is 99.0 percent of that intended for the reference 38.10 cm circle used for computing total impulse in the hydrocode computations. The impulse imparted to the ballistic pendulum from a given shot was found from first principles by measuring the swing amplitude and using the physical characteristics of the pendulum [7].

Comparisons of ballistic pendulum results with the hydrocode computations are informative. In Experiment E1, a 3.73 kg sphere of pentolite was center detonated at an HOB of 33.02 cm, giving a measured total impulse of 2.10 kN-s. Computation PC1 simulated a 3.90 kg sphere of center-detonated pentolite, a 4.6 percent greater mass, at the same HOB with a computed total impulse of 2.03 kN-s, a 3.3 percent lower value, at the ending time of 1.072 ms. This is quite good agreement, although it should be pointed out that computations PC1 and PC2 did overstate the detonation energy release as noted earlier. In another test, E2, a 3.83 kg pentolite sphere was center detonated at an HOB of 15.24 cm, giving a measured total impulse of 3.47 kN-s. Computation PC2 simulated a 3.90 kg sphere of center-detonated pentolite, a 1.8 percent greater mass, at the same HOB with a computed total impulse of 4.64 kN-s, a 33.7 percent greater value, at the ending time of 876 μs. This large discrepancy could be due to the uncertainty of the equation of state modeling and added energy discussed earlier. Computation LC1 modeled a 3.09 kg LX-14 center-detonated sphere at an HOB of 15.24 cm, somewhat similar to E2 but with a higher energy explosive. With the cautionary note that LX-14 releases energy at a larger amount per unit mass with a faster detonation wave velocity than pentolite, a rough comparison can be made based on the total theoretical energy contents of the two charges. Experiment E2 had a total theoretical energy release of 18.25 MJ, while computation LC1 had a 5.9 percent lower total theoretical energy release of 17.18 MJ. Computation LC1 had a total impulse of 1.94 kN-s, 44.1 percent less than that for E2. Computation LT1 for the same LX-14 sphere detonated from the top also had a correspondingly low impulse of 2.11 kN-s, 39.2 percent less than that for E2.

The impulse results above raise additional questions as to the validity of the JWL equation of state and its published constants for LX-14, at least within the context of its applicability to close-in blast within the framework that it operates in the HULL hydrocode. Two other ballistic pendulum experiments seem to confirm that concern, but these experiments do have the added complicating features of having an explosive initiator setting off a TOW II warhead mounted in the missile and loaded in its transportation case,

its axis parallel to the pendulum face. Thus, there is the unquantified contribution to the impulse of the initiator and the casing and missile debris. Experiment E3, with a TOW II warhead at an HOB of 33.02 *cm* set off by a small initiator, had a measured impulse of 3.01 *kN-s*, 43.3 percent greater than the bare pentolite sphere in E1 at the same HOB that had a 2.6 percent lower theoretical energy release of 17.77 *MJ*. Experiment E4, with a TOW II warhead at an HOB of 17.78 *cm* set off by a shaped charge device which itself has approximately 0.50 *kg* of LX-14, had a total theoretical energy of 19.96 *MJ* and delivered a total impulse of 4.26 *kN-s* to the pendulum. This is 2.16 times the impulse computed in LT2 for a top-detonated 3.09 *kg* sphere of LX-14 at a 2.2 percent higher HOB of 18.18 *cm* and with 8.6 percent less total energy than E4. A comparison between E4 and LT1, which the same top-detonated sphere of LX-14 as LT2 at a 14.3 percent lower HOB than E4, shows that E4 delivered 2.02 times the impulse of LT1.

VI. SUMMARY

Close-in blast can be an important damage mechanism for the current class of light armored vehicles. It is also a field of study that is relatively undeveloped, not receiving much attention until relatively recently. As demonstrated here, the means now exist for making reasonably good estimates of the close-in blast loading for highly idealized interactions of bare spherical charges over perfect reflecting planes with no intervening materials. Even so, there are obvious limitations in the ability of available equations of state, their related constants, and the hydrocodes in which they operate to model close-in blast to engineering accuracy even for the simple geometries discussed here. The addition of buffering materials not only complicates the analysis considerably, but can also actually increase the impulse on the vehicle by unintentionally acting as an energy gathering and delivering device. Much work remains to be done before theoretical computations can be used as design tools to either predict close-in blast loading and response, or assist designers in explaining the failure or success of a given configuration so that the designs can evolve toward providing proper protection at minimum weight and cost.

LIST OF REFERENCES

1. W. Fickett and W.C. Davis, "Detonation," University of California Press, Berkeley and Los Angeles, CA, Copyright 1979. ISBN 0-520-03587-9

2. J.A. Hasdal, B.S. Chambers, and R.W. Clemens, "Support to BRL: HULL Code Implementation on a CDC 7600," SAI-80-701-AQ, Science Applications, Inc., McLean, VA, August 1978.

3. B.S. Chambers, III, and J.D. Wortman, "Two-Dimensional Shore (Partial Island) Cells for BRL HULL," ARBRL-CR-00497, US Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, December 1982. (AD#A123357)

4. J.D. Beason, "A SAP Primer," AFWL-TR-80-3, US Air Force Weapons Laboratory, Kirtland Air Force Base, NM, September 1980.

5. D.A. Matuska, and J.J. Osborne, "HULL Technical Manual," Vol. 1, Orlando Technology, Inc., P.O. Box 815, Shalimar, FL, 1986.

6. B.M. Dobratz and P.C. Crawford, (Editors), "LLNL Explosives Handbook, Properties of Chemical Explosives and Explosive Simulants," UCRL-52997, Change 2, Lawrence Livermore National Laboratory, University of California, Livermore, CA, January 31, 1985.

7. B.P. Bertrand, Private communication documenting his ballistic pendulum experiments, February, 1987.

Evoked Potential Technique for Predicting Performance Under High Mental Workloads (U)

*Jeffrey H. Lukas, Dr.
Linda F. Mullins, Ms.
U.S. Army Human Engineering Laboratory
Aberdeen Proving Ground, MD 21001-5001

Stress is a chronic problem in the military. Soldiers must often perform their duties under severe hardships. Injury, loss of commrades, fatigue and a host of other problems plague the soldier yet he must cope if the mission is to succeed. The effects of combat on the psychological and physiological well being of the soldier are well known[1]. What is not known and needs to be explored is why some men faced with the same physical and psychological traumas manage to cope effectively while others succumb to the stress and become combat casualties. There are many factors influencing stress susceptibility including early environmental influences, education, training, morale, leadership, group cohesiveness etc. However, biological differences in excitability of the nervous system in any stress or challenge situation and its effect on cortical functioning may play a key role in how well a soldier will perform. How a soldier perceives and thinks about a situation will determine to a large part his reaction to that situation. Since thoughts and perceptions are cortical activities and since arousal modulates cortical functioning[2] it is likely that differences in excitability or arousability will determine how soldiers perceive a situation and therefore determine their ability to cope.

This research is part of a long term study of human and animal cortical functioning, which has developed under the experimental paradigm referred to as augmenting-reducing (A-R). The goals are to develop a technique to assess human cortical functioning in order to predict how well individual soldiers will cope with stress and to form a screening device to select those best able to cope to fill critical military positions. The A-R paradigm was developed by Buchsbaum and Silverman[3] using a response of cortical neurons referred to as evoked potentials to brief, repetitive sensory stimuli. These small electrical potentials can be recorded on the human scalp following signal averaging 50 to 100 repetitions to extract the signal (evoked potential) from the background

noise (EEG activity). Buchsbaum and Silverman measured the amplitude of the evoked potential recorded on the scalp over the vertex (C_z) to a series of light flashes that varied in intensity and calculated the linear slope of the best fit between evoked potential amplitude and intensity. Two main patterns of vertex slopes were reported: Augmenters have positive slopes indicating their evoked potentials increased in amplitude with stimulus intensity. Reducers show the opposite and rather surprising response of smaller evoked potentials as intensity increased. Reducers were thought to have hypersensitive nervous systems which require protective mechanisms at higher intensity stimulation.

The A-R measure of cortical functioning is related to personality and behavioral reactivity in both humans and animals. In particular, human augmenters appear to seek out higher levels of stimulation through experiencing new, and exciting experiences as measured by the Sensation Seeking Scale[4,5]. Animal augmenters also appear to be sensation seekers[6]. In addition, augmenter animals are better able to cope with aversive or stressful stimulation. Lukas and Siegel[7] found a close relationship between reduced cortical functioning (reducer) during intense unavoidable noise bursts and the inability to function effectively in a well learned food-reward task. In fact, one animal refused to make a response to receive food reinforcement for the first two days. Performance returned to normal only when cortical functioning returned to control levels after 7 days. Therefore, there appeared to be a tight and possibly causal relationship between cortical reducing and the failure to cope with aversive/stressful stimulation. In a similar manner, Lukas and Mullins[8] found that human reducers performed significantly worse than augmenters under a task producing high mental workloads. These results suggested that the augmenting measure of soldiers cortical functioning is related to an interest in exciting, novel and even dangerous experience and that augmenters perform better under high mental workloads and might even cope better under combat stress.

This study had 3 main objectives: (1) to measure cortical functioning using precisely controlled stimulus conditions[9] while maintaining the subject's attention focused on the visual stimuli[10,11]; (2) to determine what effect intensity range has on the correlation of A-R with personality since previous studies indicated that A-R to the most intense stimulus range correlated with personality[4,6], and (3) to extend the previous finding that auditory reducers perform significantly worse under high mental workloads to include a new subject sample and the visual modality to determine if the A-R phenomenon is replicable across modalities.

If this study supports previous evidence that auditory reducers cope significantly worse under high mental workloads, then greater confidence will be gained that A-R is a valid trait measure offering a window into the mind for predicting how well individual soldiers will cope with stress and high mental workloads.

Method

Subjects

Twenty male subjects volunteered to participate in this study and were paid for their efforts. Subjects ranged in age from 18 to 35 yr. All had normal vision and were not receiving drugs or medication at the time of testing.

Apparatus

Monocular visual stimuli were generated by a Maxwellian-view optical system. The light beam was supplied by a 150-W tungsten-halogen lamp powered by a regulated power supply to produce a continuous and highly reliable light source. Infra-red rays were removed from the light beam by a glass, heat-absorbing filter. White light flashes were produced by a rotating mirror positioned in the path of the light beam. Flashes were presented at the rate of 1/sec with a 100-msec duration. The optical system focused a 2-mm diameter image of the filament in the plane of the subject's pupil such that the light beam passed unhindered within the area of the constricted pupil. Therefore, changes in pupillary diameter could not affect the amount of light reaching the retina. The beam subtended a 22° visual angle. The maximum available retinal illuminance was 5.9×10^5 trolands which was adjusted in 0.5 log steps across a 3 log range using neutral density filters in conjunction with a balanced pair of neutral density optical wedges. A digital photometer was used to calibrate the neutral density wedges and filters. The shutter, lamp, and filters were secured on an optical bench located outside of a double-walled IAC chamber. The light beam projected through the glass windows of the chamber to prevent acoustic stimuli reaching the subject.

Procedures

Recording electrodes were located on O_z, P_z and C_z and the lower right eyelid for the ERG, all referenced to A_2. The ground electrode was placed on A_1. Electrodes impedance was maintained below 2 Kohms. Following electrode application, subjects were taken to an electrically shielded, sound-attenuated IAC chamber where they were positioned on a cot directly in line with the light beam such that the light beam passed

unhindered through the pupil of the right eye. The left eye was covered with an opaque patch. Subjects were instructed to relax, and maintain fixation on cross hairs in the center of the light flash. Subjects were shown how to maintain proper alignment and fixation and were allowed to practice until they were comfortable and could maintain fixation. Subjects were then adapted for 5 min to an ambient illumination of 34 ft L. Subjects received seven flash intensities in steps of 0.5 log unit in ascending order, with 100 light flashes presented at each intensity. Subjects counted the number of light flashes which occurred and reported this number at the end of each trial. The actual number of stimuli presented varied on each trial due to the number of artifacts rejected by the computer. Subjects initiated each trial when in proper alignment and ready to concentrate, and were given ample rest breaks throughout the experiment. Electrical activity from O_z, P_z and C_z were amplified 10,000 times with a bandpass of 1-100 Hz. The ERG had a bandpass of 1-300 Hz. The VEPs were computer averaged utilizing 1024 data points and a 300-msec sweep. Sweeps were rejected in excess of 90% of full-scale voltage (50 μV) to eliminate data contaminated with eye blinks or other muscular activity. Artifacts were relatively homogeneously distributed across intensity and rarely exceeded 5 per trial. Latency and amplitude measurements were obtained directly from the computer's oscilloscope screen by independently positioning two cursors along the VEP. Different components were measured from each electrode position as shown in Fig. 1. Latency functions were examined to insure that the appropriate component was measured at all intensities. The vertex positive potential (P_1) was often difficult to detect. If P_1 did not occur at the appropriate latency then the next earlier, positive peak was measured. Data from the P_z electrode were used for visual comparison only and were not measured.

Following the evoked potential procedures, subjects completed three personality inventories: Eysenck's Personality Questionnaire - Short Scale (EPQ-R) which is a 48-item questionnaire containing four scales of 12 items each: (1) Psychoticism, (2) Extraversion, (3) Neuroticism and (4) Lie. The second questionnaire was Form V of Zuckerman's Sensation Seeking Scale (SSS) comprised of 4 scales of 10 items each and a total score. The four scales are: (1) Thrill and Adventure Seeking, (2) Experience Seeking, (3) Disinhibition and (4) Boredom Susceptibility. The final personality inventory was Vando's Reducer-Augmenter Scale which consisted of 50 items comprising a total score.

Following completion of the personality questionnaires, subjects were given a break and then returned to complete a keeping track task[12]. Subjects were seated directly in front of a small screen (30 x 30 cm) containing a rear projection display in which 4 letters (P, Q, R, S) were presented in a random order with a duration of 100 msec and an ISI of 1.5 sec. Each trial consisted of 14 letters. In addition subjects received

40 white noise bursts thru a headset. The noise bursts were 80 dB SPL with a 30 msec duration and a 2 msec rise-fall. Two target noise bursts 10 dB louder were randomly presented in addition to the 40 nontargets in each trial. The visual and auditory stimuli were presented asychronously. The subjects primary task was to keep a separate mental count of each letter and report the number at the end of the trial. The secondary task was to react as quickly and accurately as possible with a button press to each target noise burst. The keeping track task was designed to manipulate mental workload by increasing the primary task difficulty from keeping track of one letter to keeping track of all 4 letters simultaneously while also reacting to the target noise bursts. Subjects were given 3 training trials each of the visual task, the auditory task and both tasks simultaneously. They then received 9 trials at each workload level in increasing difficulty and were provided verbal feedback following each trial. Keeping track of 4 categories, the most difficult task, was repeated an additional 9 trials to determine if group difference would persist with further practice. At the end of the study, subjects rated how difficult and interesting each task was as well as providing information on their strategy for keeping track.

RESULTS

Examination of the VEP data indicated 14 subjects had clear, easily identified components at all electrode positions for all 7 flash intensities. Calculation of the linear slope for C_z P_1N_1 amplitude across the 3 log intensity range produced 8 positive slopes (augmenters) ranging from .01 to 1.65 μV/log. Six subjects had negative C_z slopes (reducers) ranging from -.02 to -1.23 μV/log. Amplitude and slope data for these two groups (A-R) formed the Between-subjects effect in all MANOVAs reported in this paper.

Electroretinogram (ERG)

Under these stimulus conditions, the ERG was very clear and consistent for all subjects (Fig. 1). At low intensities the ERG was very simple, consisting of the photopic A and B waves generated primarily from the cones within the fovea. At 2 log attenuation, a small scotopic B wave began to emerge with a longer latency than the photopic B wave. This scotopic B wave grew in amplitude and decreased in latency until it merged with the earlier photopic B wave.

A MANOVA computed on the photopic A-B amplitude indicated the main effect for A-R and the A-R x Intensity interaction were not significant. As shown in Fig. 2, the two groups had very similar amplitude-intensity functions and similar slopes (augmenters:3.4 μV/log; Reducers:3.2 μV/log). The main effect for intensity was significant [$F(6,7) = 13.4$,

p<.002] with the ERG amplitude clearly increasing with intensity for both groups. In addition, all subjects had positive, augmenting slopes ranging from 1.5 to 6.3 μV/log.

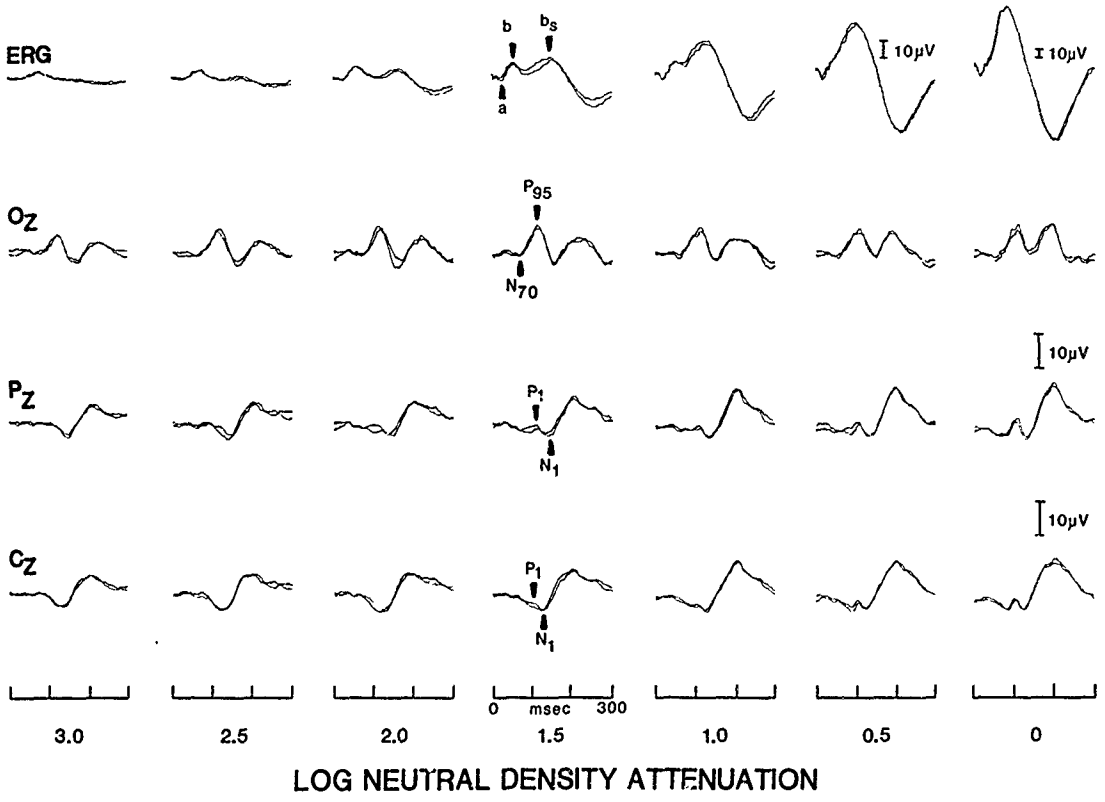


Figure 1. Grand average evoked potentials to 7 intensities ranging from dim (3 log attenuation) to bright (0 log). Two replications are shown. Both the photopic (a-b) and scotopic (bs) ERG components are identified.

Occipital VEP

The occipital potential was also very clear and consistent for all subjects. Both the N_{70} and P_{95} components decreased in latency as intensity increased. However, as can be seen in Fig. 1, the amplitude changed only about 4 μV (56%) across intensities with only half of the subjects achieving their maximum O_z amplitude at 0 log attenuation. In comparison the ERG increased over 500% in amplitude.

A MANOVA computed on the O_z amplitude demonstrated that neither the main effects (A-R or Intensity) or the A-R x Intensity interaction effect were significant. Examination of Fig. 2 indicates the groups had similar amplitude-intensity functions with similar slopes (augmenters:1.7 μV/log; reducers:1.1 μV/log). For the entire 3 log range, there were 3 reducers and 11 augmenters with slopes ranging from -1.5 to 2.8 μV/log.

Vertex VEP

Examination of the MANOVA for C_z amplitude showed that the main effects for A-R and Intensity were not significant. The A-R x Intensity interaction just failed to reach the .05 level of significance [$F(6,7) = 3.3$, $p<.07$)]. For all 7 intensities, there were 6 reducers and 8 augmenters with slopes ranging from -1.2 to 1.7 μV/log. Since the data were selected based on the C_z slopes, it is not surprising that the two groups had very dissimilar amplitude-intensity functions (Fig. 2) and different slopes (augmenters:1.1 μV/log; reducers:-0.9 μV/log).

Intensity Range

Slopes were computed for 3 partially overlapping intensity ranges: a low range from 3 log to 1 log attenuation (3.2 to 320 ft. cd) which roughly corresponds with the intensity range typically employed by Buchsbaum; a medium range from 2.5 to 0.5 log attenuation; and a high range from 2.0 to 0 log attenuation.

All subjects had positive ERG slopes for the 3 intensity ranges (Fig. 3) and the MANOVA indicated a significant main effect for Intensity [$F(2,11) = 4.1$, $p<.05$]. The groups did not differ nor was there a significant A-R x Range interaction. The ERG slope increased as intensity range increased for all subjects.

Although the occipital slope appeared to decrease with increasing intensity range (Fig. 3) there were no significant effects for A-R, Range or A-R x Range. There were 3 negative occipital slopes in the low range and all were vertex reducers. Two occipital reducers at the medium range were both vertex reducers and of the 7 occipital reducers at the high range, 4 were vertex reducers and 3 were augmenters.

For the Vertex slopes, only the main effect for A-R was significant [$F(1,12) = 13.8$, $p<.003$]. There were 5 negative slopes for the low range and 4 of these were reducers for the entire intensity range, 8 negative slopes at the medium range and 7 were reducers for the entire range, and 7 negative slopes at the high range all were reducers.

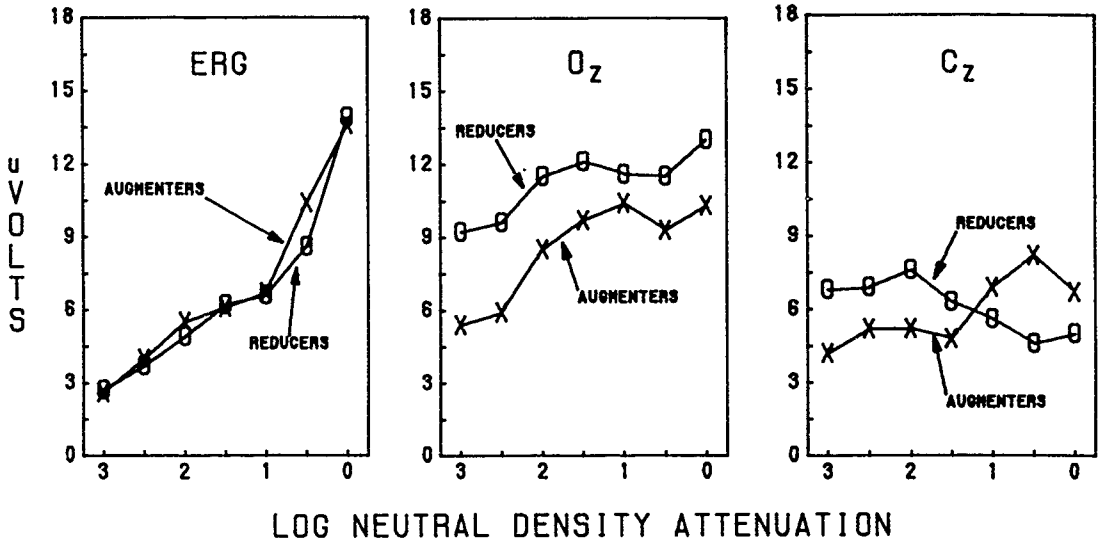


Figure 2. Comparison of the evoked potential amplitude for the augmenters and reducers at the 3 recording sites and 7 intensities.

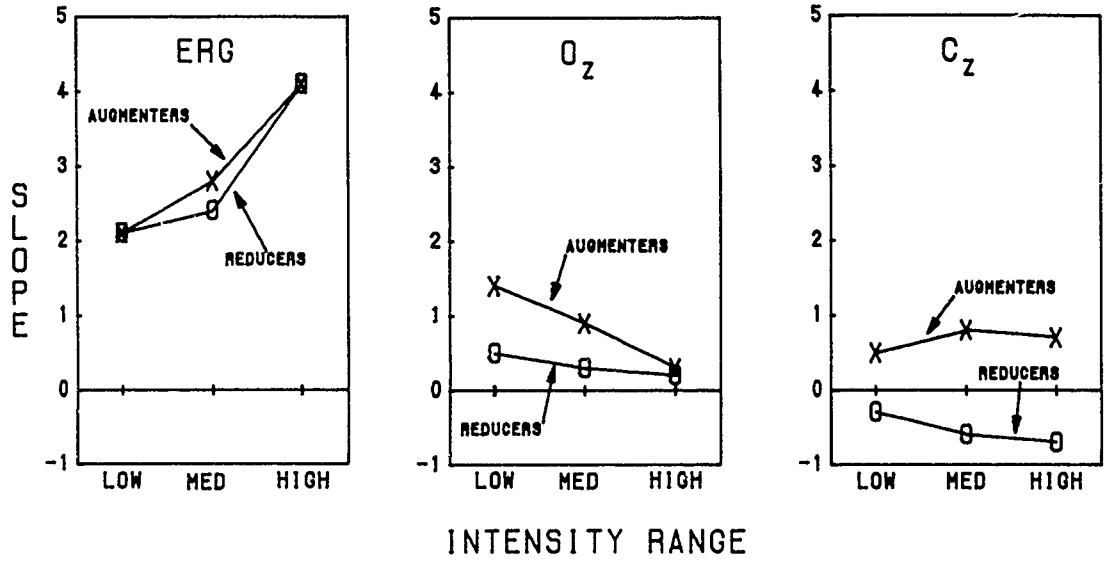


Figure 3. Comparison of the slopes for the augmenters and reducers at the 3 recording sites and 3 intensity ranges.

Personality Correlates

Pearson correlations between the personality scores and the occipital and vertex slopes for the low, medium and high ranges as well as for the entire range indicated only two significant correlations out of 80 possible. There was a cluster of positive but nonsignificant correlations between vertex slopes at the medium and high ranges and the SSS scales indicating a tendency for the vertex augmenter to be a Sensation Seeker.

Workload Performance

Performance on the Keeping Track task was analyzed with two separate MANOVAs. The first analysis was on total errors for the 9 trials completed at each category. As subjects kept track of more categories, the mental workload and the errors increased correspondingly [$F(4,9) = 21.2$, $p<.001$]. However, the reducers made significantly more errors [$F(1,12) = 10.2$, $p<.008$] especially at the higher workloads (Fig. 3). Both augmenter and reducers improved on the second set of 9 trials of keeping track of 4 categories but the reducers still averaged over twice as many errors as the augmenters.

Examination of the MANOVA for the log transformed reaction times indicated only the main effect for workload was significant [$F(4,9) = 4.6$, $p<.03$] with reaction times increasing as workload on the primary task increased. Although the reducers had longer reaction times at all workload levels (Fig. 4), the effect was not significant.

Conclusions

These results taken in conjunction with previous data[4,13] support vertex augmenting-reducing as a robust and viable phenomenon even when recorded under highly controlled stimulus and recording procedures[9]. Presentation of visual stimuli via Maxwellian-view offered several advantages over free viewing of a photostimulator typically employed in A-R studies. Since the light beam is focused to a 2mm point at the plane of the pupil, pupillary constrictions cannot decrease the amount of light reaching the retina. In addition, the flash produced by a photostimulator depends on the voltage which can vary. This system depends on a regulated power source to produce a constant light output which can be adjusted exactly using neutral density filters. Finally, unlike the photostimulator, there are no acoustic artifacts. Analysis of the retinal potential (ERG) demonstrated only positive, augmenting slopes for all subjects at all intensity ranges. This finding has two main implications for the A-R paradigm as employed in this study. First, considerable confidence can be placed on occipital and vertex slopes as

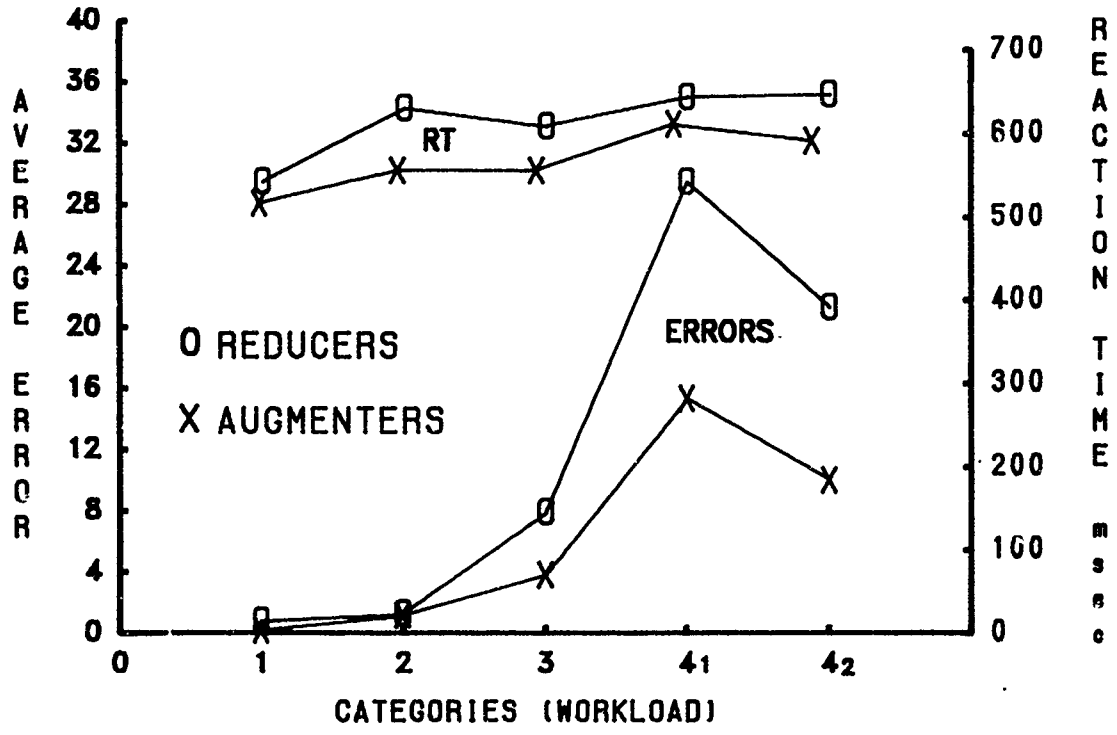


Figure 4. Average keeping track error for each workload level and average reaction time to the target noise bursts for the augmenters and reducers.

measures of human cortical functioning. Second, reducing is a central phenomenon since reducing was not seen at the retinal level. In the animal model of A-R, reducing appeared to be produced by cortical inhibitory loops actuated by intense visual stimuli or high cortical arousal[6,7]. With human subjects, occipital reducing was not observed at Buchsbaum's low intensity range[13]. However, 3 subjects in this study were occipital reducers at the low range; hence, it must be concluded that occipital reducing occurs at all intensity ranges. Still, there is no evidence from these data that occipital slopes correlate with personality or performance variables. The focus for future A-R research must therefore remain at the vertex.

In contrast to previous studies[4,5,8,10] few significant personality correlates were found. However, there was a clustering of positive correlations between vertex slopes to medium and high intensity ranges with Sensation Seeking Scales. Nine out of 10 correlations were positive

(range .02 to .41) indicated a tendency for the augmenter to be a sensation seeker. Neither the Vando R-A Scale or Eysencks Personality Questionnaire yielded any correlations with A-R. The lack of personality correlates is surprising especially since subjects' attention was directed towards the eliciting stimulus which was previously shown to be necessary for obtaining correlations with sensation seeking[10]. In the classic passive A-R paradigm, subjects are instructed simply to relax and observe the light flashes. However, there is no way of monitoring what the subject actually attends. This is a potential problem since attending the eliciting stimulus enhances N_{100} and P_{300} components especially under rapid stimulation rates[14]. In addition, the P_1N_1 slope is affected by what a subject is attending[15]. Mullins and Lukas compared a passive A-R paradigm to an active, attention condition where subjects reacted to target auditory stimuli with a button press such that their attention was controlled and could be monitored[10]. Attending the auditory stimuli enhanced P_1N_1 amplitude and provided a clustering of significant personality correlates with with A-R and the Vando and Sensation Seeking Scales. In contrast, slopes from the passive condition were randomly and nonsignificantly correlated with personality. Subjects attention was controlled in this study by having them count the light flashes and report the number following each trial. It is quite possible that this task was very easy allowing subjects to divide their attention. Future visual A-R studies will provide a target light flash in a similar manner to the target noise burst in order to control what the subject is attending.

The most important finding arising from this study was that the augmenters, as predicted, performed significantly better than the reducers under high mental workloads. This replicates previous results using an auditory A-R paradigm[8] and demonstrates that although both reducers and augmenters improve with additional practice with the most difficult task, the reducers performance is still significantly worse with double the error rate. This is a significant finding for at least 2 reasons. First, this adds credence to A-R as a measure of a biological trait[16] rather than as an unreliable, state measure[17]. Secondly, the replication of a significant relationship between a measure of cortical functioning and performance adds considerable confidence that the approach has merit and warrants further exploration. The goal tc achieve a biological screening device to select the best soldier to fill a critical military assignment has yet to be realized but at least the approach appears to be on the right track.

References

1. Grinker, R. R. & Speigel, J. P. (1945). Men under stress. New York: McGraw-Hill.

2. Steriade, M. (1970). Ascending control of thalamic and cortical responsiveness. International Review of Neurobiology, 12, 87-144.

3. Buchsbaum, M. S. (1976). Self-regulation of stimulus intensity: Augmenting/reducing and the average evoked response. In G. E. Schwartz & D. Shapiro (Eds.), Consciousness and self-regulation: Advance in research and theory (Vol. 1) (pp. 101-135). New York: Plenum Press.

4. Lukas, J. H. (1987). Visual evoked potential augmenting-reducing and personality: The vertex augmenter is a sensation seeker. Personality and Individual Differences, 8, 385-395.

5. Zuckerman, M., Murtaugh, T. M. & Siegel, J. (1974). Sensation seeking and cortical augmenting-reducing. Psychophysiology, 11, 535-542.

6. Lukas, J. H. & Siegel, J. (1977). Cortical mechanisms that augment or reduce evoked potentials in cats. Science, 198, 73-75.

7. Lukas, J. H. & Siegel, J. (1977). Aversive noise effects on performance and thalamocortical responsiveness in cats. Physiology & Behavior, 19, 555-559.

8. Lukas, J. H. & Mullins, L. F. (198). Auditory augmenters are sensation seekers and perform better under high workloads (Abstract). Psychophysiology, 22, 580-581.

9. Tepas, D. I., Guiteras, V. L. & Klingaman, R. L. (1974). Variability of the human average evoked brain response to visual stimulation: a warning! Electroencephalography and Clinical Neurophysiology, 36, 533-537.

10. Mullins, L. F. & Lukas, J. H. (1984) Auditory augmenters are sensation seekers - if they attend the stimuli (Abstract). Psychophysiology, 21, 589.

11. Mullins, L. F. & Lukas, J. H. (1987). Brain potentials and personality: A new look at stress susceptibility (Technical Memorandum 20-87). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory.

12. Monty, R. A. (1974). Keeping track of many things (Technical Memorandum No. 29-74). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory.

13. Lukas, J. H., Siegfried, J. B. & Mullins, L. F. (1987). Retinal and cortical augmenting-reducing to flash and pattern reversal stimuli. In R. Johnson, J. W. Rorrbaugh & R. Parasuraman (Eds.) Current Trends in Event-Related Potential Research (EEG Suppl. 40), (pp. 220-226). Amsterdam: Elsevier Science Publishers.

14. Hillyard, S. A., Hink, R. F., Schwent, V. L. & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. Science, 182, 177-180.

15. Schechter, G. & Buchsbaum, M. (1973). The effects of attention, stimulus intensity, and individual differences on the average evoked response. Psychophysiology, 10, 392-400.

16. Zuckerman, M. (1984). Sensation seeking: A comparative approach to a human trait. The Behavioral and Brain Sciences, 7, 413-471.

17. Raine, A., Mitchell, D. A. & Venables, P. H. (1981). Cortical augmenting-reducing - modality specific? Psychophysiology, 18, 700-708.

Initial Evaluation in Humans of a Live-Attenuated Vaccine Against Argentine Hemorrhagic Fever (U)

*C. MacDonald, MAJ, K. McKee, MAJ, J. Meegan, CDR, R. Kenyon, Dr., D. Parrish, Mr., J. Maiztegui, Dr., H. Lupton, COL, C. Peters, COL and J. Barrera Oro, Dr.
U.S. Army Medical Research Institute of Infectious Diseases, The Salk Institute, Government Services Division, Ft. Detrick, Frederick, MD 21701-5011
Instituto Nacional de Estudios sobre Virosis Hemorragicas, Pergamino, Argentina.

Historically, the military medical research establishment has been instrumental in the development, clinical evaluation, and mass utilization of new vaccines, chemoprophylactic agents, and therapeutic drugs of universal benefit. The impetus behind these efforts, of course, lies in the the legendary impact infectious diseases have had on the readiness and operation of military forces. The deployability of U.S. military personnel to virtually every corner of the globe imposes a heavy responsibility on the part of its medical support structure to address a wide variety of endemic disease threats of an unfamiliar, often hazardous nature. Moreover, the possible threat of biological weapons employed by hostile governments or terrorist groups mandates an aggressive program of prophylactic and therapeutic preparedness. Of notable concern is a group of exotic and frequently lethal viral hemorrhagic fevers found in many different parts of the world. Diseases in this group include Lassa fever, Bolivian hemorrhagic fever, Ebola and Marburg hemorrhagic fevers, Crimean-Congo hemorrhagic fever, and the hemorrhagic fevers with renal syndrome. Because of risks posed to individuals studying these often highly contagious diseases, limited effort has been devoted to their biology, epidemiology, and pathophysiology. The fact that many of these afflictions are found in relatively isolated, technologically depressed, areas of the world, compounds the problem, and genders little stimulus for sophisticated scientific investigation.

Argentine hemorrhagic fever (AHF) is one of this group of exotic and highly lethal disease threats. The etiologic agent is Junin virus (JV), a rodent-borne pathogen belonging to the Arenaviridae family. Viruses of this family, which include the Lassa and Bolivian hemorrhagic fever agents, are closely associated with various species of field-dwelling rodents (in the case of JV, predominantly Calomys musculinus), in which they cause chronic, asymptomatic infections. Under natural conditions, humans become infected through incidental contact with these chronically-infected animals

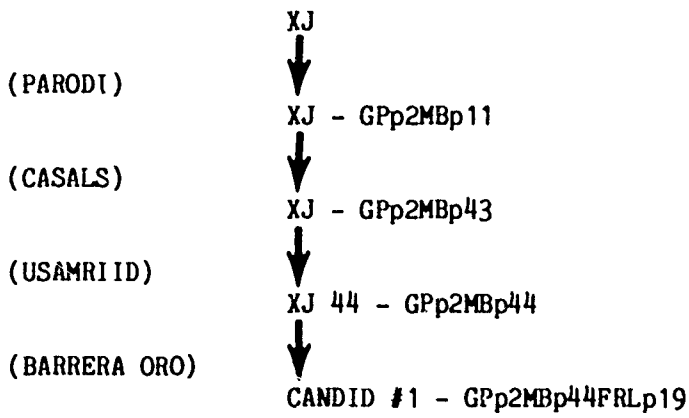
while they are shedding large quantities of virus in excreta and secreta.[1] Viral particles are highly infectious by the aerosol route, and stable over a range of environmental conditions, properties which help account for the relative frequency with which disease occurs among indigenous human populations in endemic areas.[1,2] Mortality in untreated human disease may be as high as 15 -30% under certain conditions.[2,3]

In contrast to some of these viral diseases, AHF is a disease of a developed nation - Argentina. Found exclusively in the fertile farmland regions around Buenos Aires, the endo-epidemic boundaries of AHF encompass population and economic centers of this nation. To Argentina, the disease is not just a major public health problem; it constitutes a significant economic concern as well. The development of preventive measures against AHF and it's relatives presents a major biomedical challenge. Efforts at eradication or containment of chronically-infected rodents have proved impractical, and attempts to develop effective methods for disease prevention have met with only limited success.[2] Candid #1 is a live-attenuated Junin virus vaccine developed to meet this challenge for AHF. This vaccine is the product of a collaborative effort between the Argentine Ministry of Health and the U.S. Army Medical Research Institute of Infectious Diseases (USAMRIID), and represents the first credible attempt to develop and extensively deploy a live, attenuated arenavirus vaccine for use in humans. Following on the heels of very promising preclinical evaluation, including protection against lethal challenge with virulent Junin and Machupo (Bolivian hemorrhagic fever) virus in animal models,[4,5] Phase I clinical testing of Candid #1 was initiated in October, 1985. The results of this Phase I study are presented.

VACCINE DEVELOPEMENT

Candid #1 is derived from the prototype Junin virus, XJ, a strain originally isolated by Parodi and coworkers from the blood of a human with clinical AHF.[6]

CANDID #1 PASSAGE HISTORY



Candid #1 was subjected to extensive preclinical testing to establish it's safety, immunogenicity, and protective efficacy prior to evaluation in humans.[7] Safety evaluation included determination of degree of attenuation, lack of adventitious agents, genetic stability, and inability to cause persistent or latent infection. Neurovirulence testing was performed in rhesus macaques, and persistence was studied in guinea pigs and rhesus macaques following subcutaneous and intramuscular inoculation.

Efficacy of Candid #1 was determined in preclinical studies by wild virus challenge of rhesus macaques 106 days following vaccination with Candid #1.[8] All animals vaccinated (20/20) resisted challenge, while 3/4 controls died within 4 weeks following challenge. The median protective dose of Candid #1 for rhesus macaques was determined to be < 16 plaque forming units (PFU). In recent studies, rhesus monkeys vaccinated with Candid #1 resisted challenge with Machupo (Bolivian hemorrhagic fever virus) virus as well. Similar results were observed in guinea pigs, with a dose of < 30 PFU of Candid #1 protecting against both virulent Junin and Machupo virus challenge.[4,5]

Clinical Design

Acutely conscious of the potentially disastrous effects which might occur during the initial human evaluation of a live virus vaccine derived from a highly hazardous pathogen, great efforts were taken in designing the trial to minimize the risks to participants as much as possible. In the specific case of this product, hypothetical concerns included possible development of genuine AHF (due to lack of adequate attenuation), an atypical infection (due to viral mutation or altered tissue tropism), or a persistent infection with resultant inflammatory or immunosuppressive disease. Consequently, testing of Candid #1 was initiated slowly and cautiously, with extensive precautions taken with the early volunteers. The first group of four healthy adult male volunteers, aged 19-28 years, was recruited from a pool of active duty U.S. Army medical research volunteers at USAMRIID. Following a thorough baseline physical examination and laboratory screen, subjects were hospitalized on a closed research ward and inoculated subcutaneously with 0.5 ml of Candid #1. Clinical evaluation utilizing a checklist based on an evaluation sheet used in assessing hospitalized AHF patients in Argentina, together with vital signs and physical examination, was performed daily for forty-two days, while volunteers were restricted to the ward. The same rigorous assessment was then performed weekly for eight weeks on these individuals as outpatients prior to scheduling additional cohorts of volunteers. In addition, an extensive range of clinical laboratory, serological, immunological, and virological parameters were monitored for more than 120 days following vaccination.

Satisfied with the absence of clinical and clinicopathological abnormalities in the initial group of volunteers, a second cohort consisting of twelve individuals (9 males and 3 females), aged 19-55 years, was recruited. These volunteers were divided in a randomized, double-blind

fashion into two vaccine groups and a control group. The two vaccine groups were inoculated intramuscularly with 1.0 ml of vaccine, using two different dilutions; the control group received saline diluent only. Subjects were examined daily as outpatients for 28 days, and serial laboratory specimens analogous to those in the first cohort were collected for >60 days. After prolonged observation of these initial two volunteer cohorts, and satisfied that no untoward short term effects had resulted from vaccination, immunization of a series of volunteer groups was initiated at USAMRIID. Several different vaccine dilutions were administered in an effort to examine dose-related differences in clinical and serological response (Table 1). Monitoring procedures in these later groups were modified as well, in order to avoid unnecessary testing of volunteers. At the same time, an initial group of 10 volunteers, 5 vaccinees and 5 doubleblinded placebo controls, was inoculated with Candid #1, in Pergamino, Argentina, by Argentine co-investigators.

Table 1. Volunteer groups.

Subjects were selected from volunteers from USAMRIID's medical volunteer research personnel, and staff members of USAMRIID and INEVH.

Dosage Groups Vaccinated 1985-1988

Group	No./GP	Mean Dose PFU/ml	Dose Range	Route of Inoculation	Vaccine Dilution
Undilute I (UI)	4	29,000	29,000	SC	1:1
Undilute II (UII)	53	44,700	30,000-58,000	IM	1:1
Dilute I (DI)	4	2,000	2,000	IM	1:100
Dilute II (DII)	22	11,000	7,600-14,000	IM	1:4
Argentine I (AI)	5	51,000	51,000	IM	1:1
Total Vaccinated	88		Dose Range 2,000-51,000		

Three subjects served as doubleblinded placebo controls in one of the earliest vaccine groups at USAMRIID and five were included as doubleblinded placebo controls in the Argentina group. Six additional (unblinded) subjects served as controls in the first USAMRIID groups for laboratory values.

Clinical and Laboratory Observations

Specific clinical and laboratory parameters monitored in vaccinees are listed in Table 2.

Table 2. Clinical and laboratory evaluation of Candid #1 vaccinees.

Physical Exam and Vital Signs:

Group	No. Tested	Clinical Evaluation:
UI	4/4	Examined daily for 42 days, and then weekly for 8 weeks
UII	5/53	Examined daily for 28 days
	7/53	Vital signs daily for 28 days, examined only if reporting a clinical complaint
	41/53	Examined only if reporting a clinical complaint
DI	4/4	Examined daily for 28 days as outpatients
DII	15/22	Vital signs daily for 28 days, examined only if reporting a clinical complaint
	7/22	Examined only if reporting a clinical complaint
AI	5/5	Examined daily for 28 days, and then weekly for 4 weeks

Clinical Laboratory Tests:

	Groups				
	UI	UII	DI	DII	AI
Hematology: White blood cells differential, platelet count, hematocrit, other red cell indices, reticulocyte count	4/4	31/53	4/4	22/22	5/5
Total T lymphocytes and T4/T8 subsets	4/4	31/53	4/4	22/22	0/5
Serum chemistries: renal and liver function tests, muscle enzymes, amylase	4/4	5/53	4/4	0/22	5/5
Coagulation studies: prothrombin time, partial thromboplastin time, fibrinogen and fibrin degradation products	4/4	5/53	4/4	0/22	0/5
Urinalysis	4/4	31/53	4/4	22/22	5/5
Serum immunolglobins, serum complement	4/4	5/53	4/4	0/22	0/5
Serum interferon	4/4	18/53	4/4	15/22	0/5

Adverse clinical effects of vaccination with Candid #1 have been minimal. Symptoms and signs observed among the first 41 volunteers during the initial 28 days post-vaccination are summarized in Table 3.

Table 3. Symptoms and signs observed in volunteers within 28 days of vaccination.

Observations	Groups* I	II	III	Total	C
	N=9	N=5	N=72	N=89	N=8
Characteristic finding in AHF					
Fever	0[1]	0	0	0	0
Malaise/Fatigue	5	0	3	8	2
Headache	5	4	1	10	5
Myalgias	0	2	2	4	2
Dizziness	1	3	0	4	0
Anorexia	4	0	1	5	0
Palatal Petechiae	5	3	0	8	4
Petechiae	8	1	0	9	5
Adenopathy	7	4	1	12	5
Facial Flushing	0	0	0	0	1
Other complaints					
Gastrointestinal C/O					
Gastroenteritis	2	0	0	2	0
Gastric Upset	1	2	2	5	1
Respiratory C/O					
URI	7	0	0	7	2
Allergic Rhinitis	3	0	0	3	1
Apthous Ulcers	1	0	0	1	0
Dermatological C/O					
Pruritis/rash	4	1	2	7	0
Laboratory changes					
Leucocytes < 3500	3	0	6	9	2
Platelets < 110k	0	0	1	1	0

[1] Fever of 100.0 degrees (F) observed for one day in a subject with gastroenteritis.

*Subject Groups:
I - First two vaccine groups immunized at USAMRIID
II - First group of Argentine volunteers
III - Remainder of vaccine groups
C - Double blinded controls (3 in second group at USAMRIID, 5 with first Argentine vaccine group)

The most common reported reactions were symptoms related to concurrent minor illnesses. The majority of findings were reported in groups where subjects were examined and questioned daily. Observed symptoms were mostly self-limiting and minimally debilitating. One vaccinee developed a fever of 100.0 F on day 4 following vaccination, concurrent with an apparently unrelated gastroenteritis. Another vaccinee complained of persistent headaches two weeks post-vaccination; an extensive neurological evaluation was normal, and the headaches responded well to medication for musculoskeletal pain. Findings corresponding to those in the AHF checklist (Table 4) were uncommon, mild, and isolated. Routine laboratory studies, including total white blood cell counts, platelets, serum chemistry values, and urinalyses, remained within normal limits. Those minor variations occasionally observed in total white blood cell or platelet counts did not differ significantly from individual baseline values, and no pattern of clinical or laboratory abnormality was observed in any subject that was suggestive of AHF (Fig. 1).

Table 4. Checklist used in evaluation of initial vaccinees, detailing clinical findings common in Argentine hemorrhagic fever.

Symptoms:	
Malaise, fatigue	Low back pain
Headache, retro-orbital pain	Pain at injection site
Dizziness	Other complaints:
Nausea	
Epigastric (upper abdominal) pain	
Arthralgias, myalgias (joint and muscle pain)	
Signs:	
Fever	Redness or swelling at injection site
Gingival congestion	
Petechiae	
Conjunctival injection	**Hemorrhagic signs:**
Palatal lesions - prominent palatal vasculature, palatal petechiae,	Epistaxis
	Hematemasis
Palatal vesicles	Hemoptysis
Muscle tenderness	Melena, hematochezia
	Vaginal hemorrhage
Neurological signs:	
Nuchal rigidity	Hyporeflexia
Signs of meningeal irritation	Areflexia
Cutaneous hyperesthesia	Hypotonia
Tongue tremors	Confusion
Hand tremors	Stupor
Convulsions	Delirium
Other finding:	Coma

Disturbances of immunological function are regular features of clinical AHF. Frequently observed abnormalities include lymphopenia, elevated alpha interferon levels,[8] decreased serum complement, suppression of normal immune responsiveness to ubiquitious antigens, and secondary bacterial infections. As a consequence, a variety of immunological parameters were monitored among the Candid #1 vaccinees. Studies performed included quantitative serum immunoglobulin levels, serum complement activity, erythrocyte sedimentation rate, serum alpha interferon levels, and serial determinations of total T lymphocyte and T-cell subset populations (Fig. 2). All parameters studied remained essentially unchanged from baseline values; differences related to vaccine dose, specific serologic response, or presence of circulating virus (see below) were not observed.

Virus Isolation

Extensive and prolonged sampling of various types of clinical specimens was undertaken to determine the presence, degree, and duration of recoverable JV following vaccination with Candid #1 (Table 5). Conventional virus isolation techniques were used to isolate virus from plasma, throat swabs, and urine, while co-cultivation on VERO cell monolayers and serial blind passage for 4 weeks was utilized for peripheral blood mononuclear cells (PBM) and semen specimens.

Junin virus was recovered from PBM cultures in seven of the 46 vaccinees tested. Virus was isolated only during a window extending from 7 to 13 days post-vaccination. The maximum duration of PBM viremia was three days; six of seven subjects were positive for only one day.

Immunologic Response

Initially, conventional plaque-reduction neutralization (PRN) and immunofluorescent (IF) antibody testing was employed to assess the serological response to vaccination. However, very low or absent antibody responses in the early volunteer groups (e.g., only two of the first four generated detectable antibody responses, and these only to low titer) prompted modification and expansion of serological testing procedures with the capability to detect seroconversion. Development of a complement-enhanced neutralization (C'PRN) test and JV-specific IgM and IgG ELISA assay markedly improved identification of positive responders. Additionally, a JV-specific lymphocyte transformation (LT) assay provided an assessment of specific cellular immune responsiveness to the vaccine. Serologic results in all volunteers were verified by duplicate assays performed at both institutions.

Table 5. Viral assays of Candid #1 vaccinees.

Specimens Assayed for Virus

Group	Samples studied	No. subjects studied	No. days cultured	Total Samples cultured
UI	Plasma, PBM*	4	38	152
	Urine, Throat	4	31	124
	Semen	4	12	48
UII	Plasma, PBM	18	7(5), 5(13)	100
	Semen	5	4	20
DI	Plasma, PBM	4	7	28
	Semen	4	4	16
	CSF*	1	1	1
DII	Plasma, PBM	15	5	75
AI	Plasma, PBM	5	7	35

*PBM - Peripheral Blood Mononuclear Cells
CSF - Cerebrospinal Fluid (from subject with recurrent headaches)

Virus Isolations in Candid #1 Vaccinees

Group	Sample	No. cultures positive	Day isolated post-inoculation	Vaccine dose received
UI	None Positive	-	-	-
UII	PBM	4[1]	7, 9	3.2×10^4 (2/4) 5.5×10^4
DI	PBM	1[2]	11	2.0×10^3
DII	PBM	2	9, 13	1.1×10^4
AI	PBM	1	9	5.1×10^4

[1] Virus was isolated on both day 7 and 9 in one subject.
[2] Culture of semen obtained from subject on same day was negative for virus

Rate of response to Candid #1, when judged by a positive response to any of four tests (LT, C'PRN, ELISA, or IF), was impressive, 100% in 76 vaccinees (Table 6). The sensitivity of detection varied considerably among individual tests; IF, the least sensitive, had a positive response rate of 35%, while 97% of vaccinees responded by LT. Junin virus-specific IgG was detected in 94% of volunteers following immunization, while only 69% had Junin positive IgM. Neutralizing antibody activity (C'PRN) was detectable in 91% of vaccinees.

The LT response was frequently positive by three weeks post-immunization, while antibodies could generally not be detected before 4 weeks. The magnitude and duration of antibody response was variable; titers were generally low, and declined to undetectable levels in 30-50% of responders by 1 year. In contrast, 100% of individuals initially positive by LT were still positive one year after immunization. (Table 7). There was no significant differences in rate and magnitude of immune response observed among the various vaccine dose groups.

Table 6. Immune Response in Candid #1 Volunteers.

Response Rates of Candid #1 Vaccinees

Group	Tests				Positive by any test
	LT	NT	IgG	IF	
UI	3/3* 100%	4/4 100%	3/4 75%	0/4 0%	4/4 100%
UII[1]	38/41 93%	48/51 94%	30/31 97%	15/48 31%	41/41 100%
DI	4/4 100%	3/4 75%	4/4 100%	3/4 75%	4/4 100%
DII	20/22 91%	18/22 82%	20/22 91%	8/22 36%	22/22 100%
AI	NA	5/5 100%	5/5 100%	3/5 60%	5/5 100%
TOTAL	65/70 93%	79/87 91%	62/66 94%	29/83 35%	76/76 100%

* No. positive/ No. tested
1Testing not yet completed in 2-22 subjects

Table 6. (continued)

Comparisons of Immune Response

Group	LT+/S-	LT-/S+	LT-/S-	IgG+/IgM-	IgG-/IgM+
UI	0/3 0%	0/3 0%	0/3 0%	0/4 0%	0/4 0%
UII	1/41 2%	2/41 5%	0/41 0%	6/31 19%	0/31 0%
DI	0/4 0%	0/4 0%	0/4 0%	2/4 50%	0/4 0%
DII	1/22 5%	1/22 5%	0/22 0%	7/22 32%	1/22 5%
AI	NA	NA	NA	0/5 0%	0/5 0%
TOTAL	2/70 3%	3/70 4%	0/70 0%	15/61 25%	1/61 1%

LT - Lymphocyte transformation assay
NT - Complement-enhanced plaque-reduction neutralization assay
IgG, IgM - ELISA for Junin-specific IgG or IgM antibody
IF - Immunofluorescent antibody assay
S - Serology (NT or IgG)

Table 7. Magnitude and Duration of Immune Response.

Maximum Serum Titers (ST) or Lymphocyte Transformation Stimulation Indexes (LTSI) to Junin Virus in Candid #1 Responders

Test	Maximum ST or LTSI Range	G. Mean
ELISA - IgM	1/20 - 1/1280	1/285
ELISA - IgG	1/20 - 1/1280	1/132
IF	1/16 - 1/1024	1/300
C'PRN	1/6 - 1/768	1/56
LTSI	4.5 - 71	13.1

Duration of Immunity to Junin Virus in Candid #1 Vaccinees

Test	Months after Vaccination <3	3	6	9	12
IgG	31/31* (100%)	30/31 (97%)	22/25 (88%)	7/10 (70%)	6/13 (46%)
C'PRN	29/29 (100%)	28/29 (97%)	28/29 (97%)	21/23 (91%)	7/10 (70%)
IF	11/11 (100%)	9/11 (82%)	7/11 (64%)	6/11 (55%)	0/11 (0%)
LT	19/19 (100%)	5/5 (100%)	22/22 (100%)	ND	10/10 (100%)

* Responders remaining/total responders tested

CONCLUSIONS

The safety, reactogenicity, and immunogenicity of a live attenuated AHF vaccine, Candid #1, have been evaluated in over 70 human volunteers in the United States and Argentina. During more than two years of study and observation, we can conclude that the vaccine is well tolerated, with no significant adverse clinical reactions observed to date. A low level of minor clinical reactions were reported following vaccination, however, none could be specifically attributed to the vaccine. Serial monitoring over extended periods of time have detected no hematological or biochemical abnormalities, alterations in immunologic function, or evidence of viral persistence resulting from vaccination. Candid #1 appears to be highly immunogenic, inducing antibodies to JV which are detectable using a variety of different types of testing in 91%-94% of vaccinees, and a JV-specific cellular immune response in 99% of subjects immunized. Preliminary findings suggest that immunogenicity and reactogenicity do not vary with dose or route of inoculation; however, additional studies, now in progress, are needed to validate this conclusion.

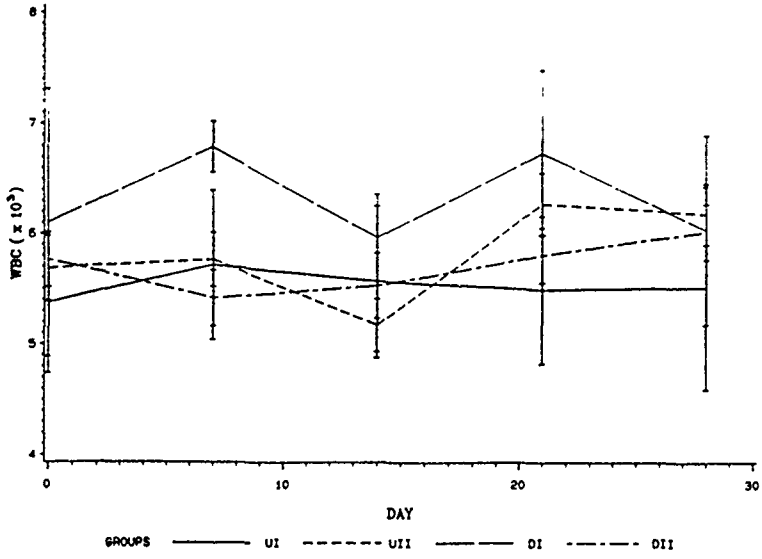
Preliminary assessment of Candid #1 indicates that this vaccine is suitably attenuated for use in humans. Previous studies using a highly realistic, albeit more sensitive primate model for human AHF have shown that Candid #1 protects against lethal challenge using virulent, wild-type JV. Moreover, cross-protection against a related and equally lethal disease, Bolivian hemorrhagic fever, has been demonstrated in non-human primates. Based on the encouraging safety and immunogenicity data outlined above, the ultimate test of vaccine efficacy, a controlled field trial, has been scheduled to be initiated in September 1988, in the southern portion of Santa Fe Province in Argentina. Based on our findings in animal model systems, there is every reason to anticipate a high degree of protection following immunization with this vaccine.

REFERENCES

1. C. J. Peters, in Textbook of Human Virology, R. H. Belshe, Ed. (Littleton, 1984), pp. 513-545.
2. C. J. Peters, P. B. Jahrling, C. T. Liu, R. H. Kenyon, K. T. McKee, Jr., and J. G. Barrera Oro, Curr. Top. Microbiol. Immunol. 134, 5 (1987).
3. J. I. Maiztegui, Bull. WHO 52, 567 (1975).
4. J. G. Barrera Oro, K. T. McKee, A. I. Kuehne, B. G. Mahlandt, J. Spisso, F. E. Cole, H. W. Lupton, presented 3rd Annual Meeting of the American Society for Virology, Wisconsin, July, 1984.
5. J. G. Barrera Oro, unpublished reports.
6. A. S. Parodi, et al., Dia Med. 30, 2300 (1958).
7. J. G. Barrera Oro, and G. A. Eddy, presented IV International Conference on Comparative Virology, Alberta, Canada, October, 1982.

8. K. T. McKee, Jr., B. G. Mahlandt, J. I. Maiztegui, G. A. Eddy, and C. J. Peters, J. Inf. Dis. 152, 218 (1985).
9. S. C. Levis, M. C. Saavedra, C. Ceccoli, E. Falcoff, M. R. Feuillade, D. A. M. Enria, J. I. Maiztegui, and R. Flacoff, R., J. Inf. Dis. 149, 428 (1984).

Figure 1. Mean total white blood cell counts (upper panel) and platelet counts (lower panel) in Candid #1 vaccine dose groups UI (N=4), UII(N=31), DI (N=4), and DII (N=22) weekly from Day 0 to Day 28 following immmunization.



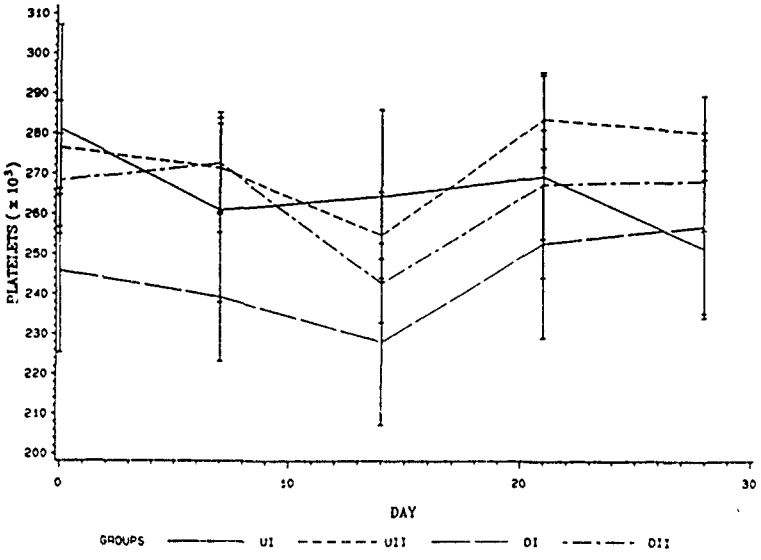
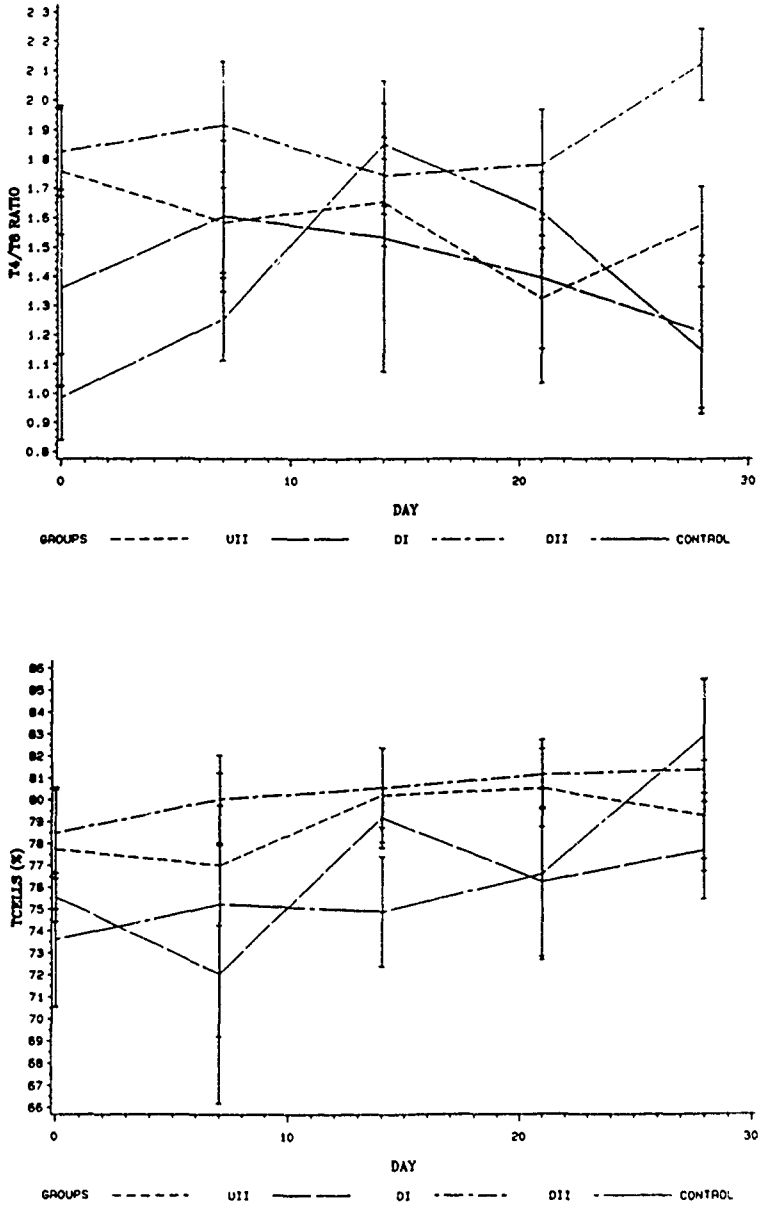


Figure 2. Mean Percentage of T lymphocytes (lower panel) and ratios of T4/T8 subsets (upper panel) in Candid #1 vaccine dose groups UII(N=12), DI (N=4), and DII (N=14) and a control group (N=5) weekly from Day 0 to Day 28 following immunization.



Hypertonic Saline-Dextran Solutions for the Far Forward Resuscitation of the Combat Casualty (U)

*Peter A. Maningas, MAJ, MC
Charles E. Wade, PhD
John P. Hannon, PhD
William G. Rodkey, LTC, VC
Letterman Army Institute of Research
Presidio of San Francisco, CA 94129-6800

INTRODUCTION

Acute hemorrhage is a major cause of death during conventional land warfare. During the Vietnam conflict, half of all soldiers killed exsanguinated in the field, prior to reaching a fixed medical treatment facility (1). Therefore, early adequate fluid resuscitation and definitive surgical care remain crucial to the ultimate survival of the severely hypovolemic combat casualty. However, the anticipated lack of air superiority in the next military conflict will delay evacuation to surgical facilities. Furthermore, supplying large quantities of crystalloid or blood to the field medic for treatment of the combat casualty is logistically impractical. Presently, the field medic carries only four 500 ml bags of standard crystalloid. This quantity of fluid is insufficient to treat effectively even one casualty sustaining a two unit hemorrhage (2). Therefore, the development of a more efficacious fluid for the field management of the combat casualty is necessary.

This paper describes the development of the new resuscitation fluid 7.5% NaCl in 6% Dextran 70 for the field management of hemorrhagic shock, Specifically, the non-clinical and clinical experimentation performed at the Letterman Army Institute of Research that has culminated in the FDA-approved clinical efficacy trials is discussed.

HYPERTONIC SALINE SOLUTIONS

In 1980, Velasco and associates (3) reported that the hypertonic saline (HS) solution 7.5% NaCl given in volumes equal to 10% of shed blood resulted in 100% long-term survival of dogs subjected to a fixed-pressure hemorrhage. This finding is in comparison to 0% survival in normal saline-

treated controls. Numerous investigators (4-7) have reported that small volume infusions of HS improve cardiovascular function following hemorrhage more effectively than an equivalent volume of standard crystalloids.

It appears, however, that the hemodynamic response to the hypertonic saline solutions is transient. Nakayama and coworkers (4) demonstrated in bled rats that infusion of HS dramatically improves mean arterial pressure, skeletal resting membrane potential, and intracellular water for a period between one and two hours. In sheep, the infusion of 4 ml/kg of HS increases cardiac output for a period lasting less than 45 minutes (5). We have demonstrated that mean arterial pressure falls 15 minutes after infusion of HS to levels that were not significantly different from those in normal saline-treated controls (8,9).

An explanation for this transient response in hemodynamics may be related to one of the mechanisms of action of hypertonic saline solutions. Hypertonic saline solutions cause a shift of intracellular water to the extracellular space (5). In 40-kg sheep, the infusion of 4 ml/kg of HS is estimated to move approximately 700 ml of water from the cells into the extracellular space, 360 ml of which is partitioned into the plasma volume because of the low hydrostatic pressure (5). However, sodium ions diffuse rapidly through the capillary endothelium. It is likely that after infusion of the hypertonic sodium chloride, water accompanies sodium down its concentration gradient out of the intravascular space. This may account for the deterioration in cardiac output and arterial pressure seen minutes to hours after infusion of hypertonic saline solutions.

HYPERTONIC SALINE-DEXTRAN SOLUTIONS - ANIMAL STUDIES

Recently we evaluated the use of 7.5% NaCl in 6% dextran 70 (HSD) in an attempt to prolong the beneficial hemodynamic effects and optimize the survival observed with 7.5% NaCl alone. The rationale for the addition of dextran solutions is based on the plasma volume expanding characteristics of these solutions. Dextran solutions increase the colloid osmotic pressure within the intravascular space and result in a plasma volume expansion of 24 to 28 ml/g dextran 70 (10). Given that a 4 ml/kg dose of 7.5% NaCl increases plasma volume approximately 8 ml/kg, a 4 ml/kg infusion of 7.5% NaCl prepared in 6% dextran 70 would then be expected to increase the plasma volume approximately 14 ml/kg.

The following non-clinical experiments were conducted at the Letterman Army Institute of Research to evaluate the efficacy of this new resuscitation solution: Study 1 - comparative survival of swine following a severe lethal hemorrhage after treatment with different hypertonic and hyperoncotic solutions; Study 2 - organ blood flow after infusion with HSD following a potentially lethal hemorrhage; and Study 3 - effects of HSD on body oxygen economy.

Study 1 - Survival

Previous experiments had demonstrated that a small volume infusion of HSD was superior to an equivalent volume of HS in sustaining cardiac output and plasma volume after a fixed-pressure non-lethal hemorrhage (11,12). In our initial study (8), we examined whether the beneficial hemodynamic effects observed in these non-lethal hemorrhage experiments would result in an improvement in survival following a severe potentially lethal hemorrhage. This information was critical in the developmental transition of this solution from non-clinical experiments to clinical trials.

Methods

Sixty chronically instrumented swine were randomized into one of four treatment groups: 0.9% NaCl (NS, n=15), HS (n=15), 6% dextran 70 (DEX, n=16), and HSD (n=14). Each animal was bled 46 ml/kg in 15 minutes. This volume of hemorrhage corresponds to 70% of the swine intravascular volume and is uniformly fatal within thirty minutes if untreated. Five minutes after the completion of hemorrhage, the animals were infused with their respective treatment in a volume (11.5 ml/kg) equal to 25% of the shed blood. The five-minute interval period was chosen because the WDMEV (Wound Data Munitions Effectiveness - Vietnam) data revealed that the average interval from the time a soldier was wounded and the time he was initially seen by a field medic was four minutes. The initial dose of 11.5 ml/kg was chosen because this volume represented the amount of fluid that could be reasonably expected to be infused in a field environment or during air evacuation. In a succeeding study, we compared decreasing doses of HSD using the identical hemorrhage model. This was done to establish the lower limits of therapeutic efficacy. Animals were infused with HSD in a quantity equal to 4 ml/kg, 2 ml/kg, and 1 ml/kg.

Results

In this randomized, blinded study, we observed that 100% of those animals infused with HSD survived until euthanized at 96 hours. In comparison, animals infused with NS, HS, and DEX had 96-hour survival values of 13%, 53%, and 69%, respectively. The survival in the HSD group was significantly better than those in the NS group ($P<0.001$) and the HS group ($P<0.01$) (Figure 1).

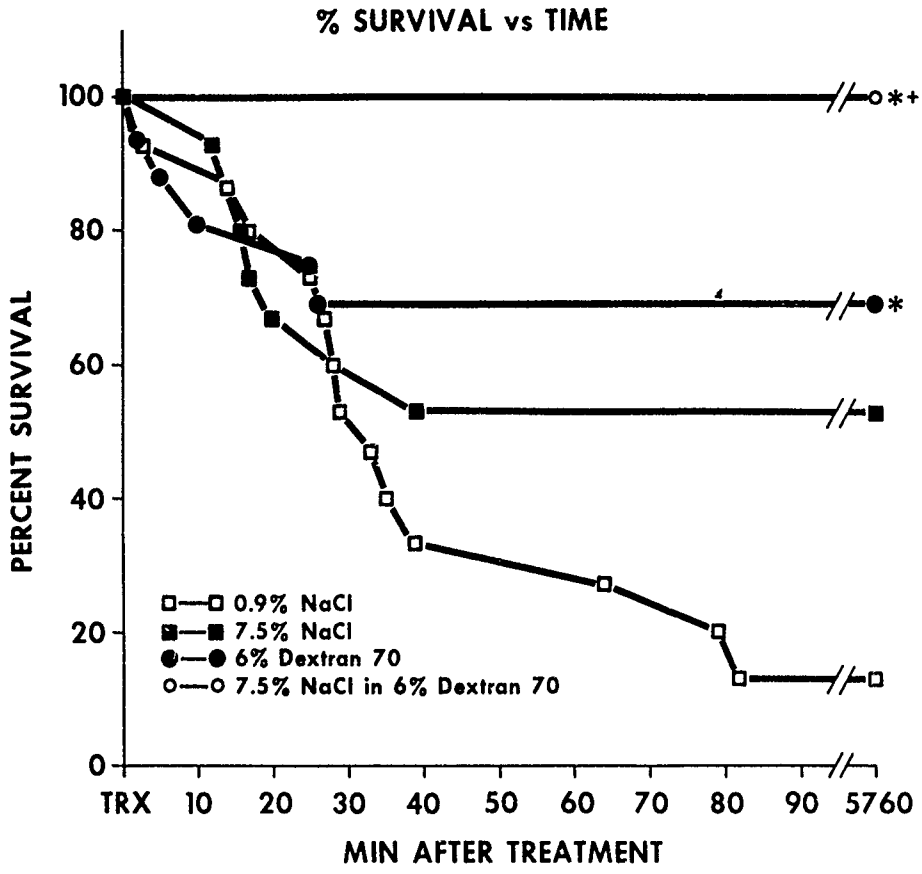


Figure 1. Survival outcome after a 11.5 ml/kg infusion of various treatment solutions.

All long term survivors exhibited normal behavior 3 hours after infusion. No post-treatment complications were noted in any group. Specifically, there were no clinically apparent manifestations (i.e., seizures) of the transient hypernatremia caused by the hypertonic solutions. Additionally, histologic studies performed after the animals were sacrificed at 96 hours showed no evidence of organ or cellular damage.

Immediately after treatment, the mean arterial pressure (MAP) in the HS and HSD groups increased to 80% and 86% of baseline values, respectively, while the MAP of NS animals increased to only 47% of prehemorrhage levels. Fifteen minutes after treatment, the MAP of the HSD group was 92% of baseline, while the MAP of HS animals decreased to levels not significantly different from those of the NS group.

In the succeeding experiment those of which evaluated the effectiveness of varying small doses of HSD, we found the 96-hour survival to be 83%, 53%, and 13% in those animals infused with 4 ml/kg, 2 ml/kg, and 1 ml/kg, respectively. The survival in those animals infused with the 2 ml/kg and 1 ml/kg dose was significantly less than that of the 11.5 ml/kg dose.

Study 2 - Organ Blood Flow

In terms of improving survival following potentially lethal hemorrhage, resuscitation with HSD was found to be significantly more effective than with HS alone, but the mechanisms responsible for this enhanced effectiveness remain uncertain. This experiment evaluated the effect that the combination of hypertonic saline in dextran has on organ blood flow (13). The importance of organ blood flow is that redistribution of the organ blood flow may be a critical determinant of survival following hemorrhagic shock (14).

Methods

Organ blood flow was measured by the radiomicrosphere technique (15) in unanesthetized swine bled 46 ml/kg in 15 min and subsequently infused with a volume of HSD (n=5) or NS (n=5) equal to 25% of the shed blood. Radiomicrospheres were injected before hemorrhage, immediately after hemorrhage, and 5 and 30 min after treatment. Organ blood flow was calculated by standard equations (15).

Results

After infusion of HSD, the cardiac output and MAP rose dramatically to levels greater than prehemorrhage values (Fig. 2). The cardiac output 30 min after treatment was 132% of baseline. In comparison, the cardiac output of NS-treated animals was 55% of baseline 5 min after the end of treatment and rose to only 70% of baseline in those animals surviving at 30 min. The increase in cardiac output in animals infused with HSD was the result of an increase in stroke volume to baseline values. In comparison, the stroke volume in the NS group remained approximately one third of baseline levels.

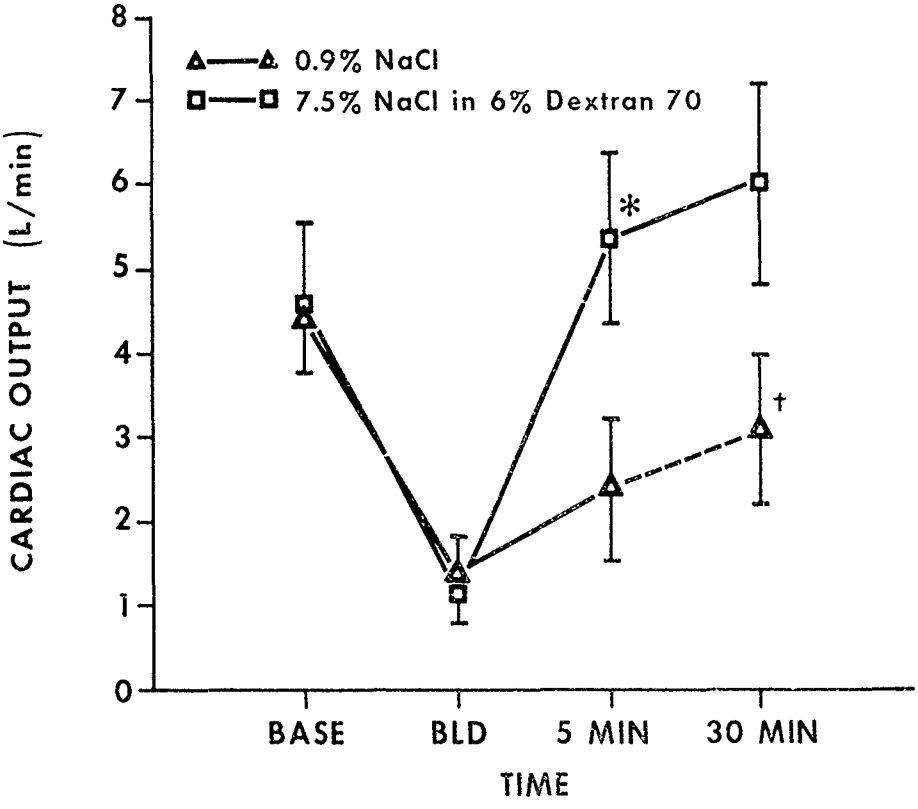


Figure 2. Cardiac output after hemorrhage and infusion of experimental fluids.

In general, infusion of HSD caused a dramatic increase in overall visceral blood flow as compared to NS. However, there was no difference between HSD and NS in the proportion of the cardiac output going to the vital organs or other functional compartments (i.e., kidney, splanchnic bed or carcass).

Infusion of HSD produced a dramatic increase in myocardial blood flow which was 3.5 times greater than baseline and more than twice that produced by NS.

Pancreatic and renal blood flow fell to 5 and 8% of baseline at the end of hemorrhage, respectively. Five minutes after infusion of HSD, blood flow to both organs increased to two thirds of baseline, whereas the blood flow in animals infused with NS remained at shock levels. Similar increases in organ blood flow after infusion with HSD were seen in the small intestine and liver.

Study 3 - Effects on body oxygen economy

The previous study suggested that resuscitation with HSD improved survival by increasing cardiac output and organ blood flow, and as a consequence, improved body oxygen economy. In our third animal study we specifically addressed the latter assumption, namely the effect of HSD on body O_2 economy.

Methods

Six chronically instrumented swine were studied. In contrast to study 1, these animals were splenectomized to allow fluid volume shifts to be determined. In addition, they were restrained in a Pavlov sling during the experimental period to facilitate measurements of energy metabolism and related functions, including O_2 consumption, tissue O_2 delivery and demand, cardiac output, hemoglobin level, acid-base status and the plasma level of various hormones. Following control measurements, the animals were bled 37.5 ml/kg over a 1-hr period and immediately thereafter given a 4 ml/kg bolus of HSD. Subsequent reovery was monitored for 4 hr.

Results

Over the course of hemorrhage total body O_2 consumption increased approximately 30%, an effect that was attributable to muscle activity and marked increases in sympathoadrenal activity. Cardiac output was reduced to 55% and the hemoglobin level to 71% of control values. Fluid transfer (12 ml/kg) from the intracellular to the intravascular space accounted for the latter effect. Arterial O_2 delivery to body tissues, the product of cardiac output and arterial O_2 content, was reduced by 56%, and plasma lactate levels rose from 0.6 to 13.6 mmol/L. As a consequence, tissue O_2 demand more than doubled, and the ratio of O_2 delivery to O_2 demand was reduced to less than 20% of control levels. Resuscitation with HSD, as in earlier studies, returned cardiac output to control levels. Resuscitation also caused a further fluid transfer to the vasculature (13 ml/kg). Recovery subsequent to resuscitation with HSD was characterized by a reversion of O_2 consumption toward control levels and a partial resolution of the O_2 debt and O_2 demand that accumulated during hemorrhage (Fig. 3). These effects appeared to be attributable to a reduction in muscle and sympathoadrenal activity.

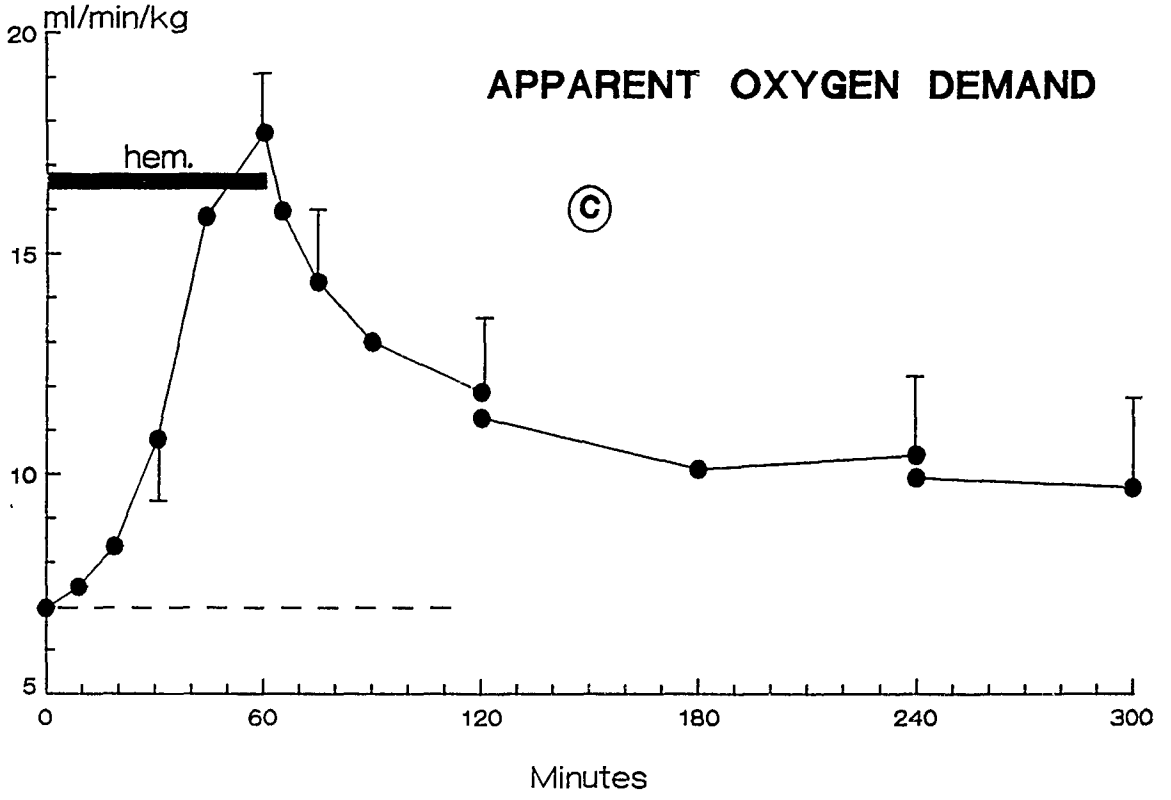


Figure 3. Oxygen demand during hemorrhage and after resuscitation with HSD.

HYPERTONIC SALINE-DEXTRAN SOLUTIONS - CLINICAL TRIALS

On the basis of the promising results of the previous animal experiments, one of the authors (PM) developed the protocol for the initial introduction of this new resuscitation solution in human trauma patients. The purpose of this preliminary study was to assess the feasibility of conducting a study on the use of HSD in the prehospital setting and specifically to examine: 1) the safety of infusing the hypertonic saline-dextran solution in trauma patients, 2) the ability of an Emergency Medical Service (EMS) system to execute successfully the study protocol, and 3) the logistical problems associated with infusion of the solution in the field. This preliminary study was conducted at Ben Taub General Hospital (BTGH), Houston, Texas in association with Kenneth L. Mattox, M.D., Professor of Surgery, Baylor College of Medicine. Information gathered from this study was intended to provide insight into the future planning of a larger prospective clinical efficacy trial.

Methods

Patients in the study were victims of penetrating injuries who had a systolic blood pressure of 90 mm Hg or less at the time of initial prehospital assessment by paramedics from the Houston Fire Department Emergency Medical Services (HFDEMS) Division. Patients were excluded for age less than 18 years, pregnancy, or if a previous history of seizures, coagulopathy, liver or renal disease had been obtained by the paramedics from the patient or next of kin.

Patients included in the study received one of two treatments: Group 1 received an initial bolus of 250 ml of the crystalloid Plasmalyte A (PA), and Group 2 received 250 ml of HSD. The treatment was started en route to BTGH or in the field if rapid evacuation was logistically delayed. All fluids were infused as quickly as possible through a peripheral intravenous site using a standard 14-gauge catheter. Once the patient arrived at the hospital, volume replacement included PA crystalloid therapy and whole blood administered as warranted by the patient's clinical condition.

Results

Over a four-month period (Dec 8, 1986 - April 8, 1987), forty-eight hypotensive patients who were victims of penetrating trauma were given 250 ml of HSD (n=23) or PA (n=25) as their initial resuscitation fluid in the field or during transport to BTGH. The two groups were comparable in age, initial blood pressure, total prehospital time, and injuries as determined by the initial field Trauma Score, Injury Severity Score, and anatomic site of injury.

Infusion of the hypertonic saline/dextran solution proved to be safe. There were no complications associated with the infusion of the HSD. Specifically, there were no seizures or anaphylactic reactions, or coagulopathies that could be attributed to the infusion of HSD. Also, there were no difficulties in crossmatching patients after infusion of HSD or subsequent evidence of phlebitis at the site of infusion.

Twenty of the twenty-three patients (87%) receiving HSD survived to be discharged from the hospital, whereas nineteen of the twenty-five patients (76%) receiving PA survived to discharge. One patient in the PA group remains in the hospital.

The mean systolic blood pressure increased 29 mm Hg after infusion of HSD, as opposed to 18 mm Hg in the PA group. Heart rate did not change significantly after infusion of either solution.

Twenty-two patients required blood transfusion (PA n-12, HSD n-10). Of those patients transfused, there was a trend toward the transfusion requirement being less in the HSD group (Fig.4). Although no estimates of blood loss were accumulated, the mean hematocrit in the HSD group was no different than that in the PA controls, suggesting that the trend toward a reduction in transfusion requirements was not due to less blood loss in the HSD group.

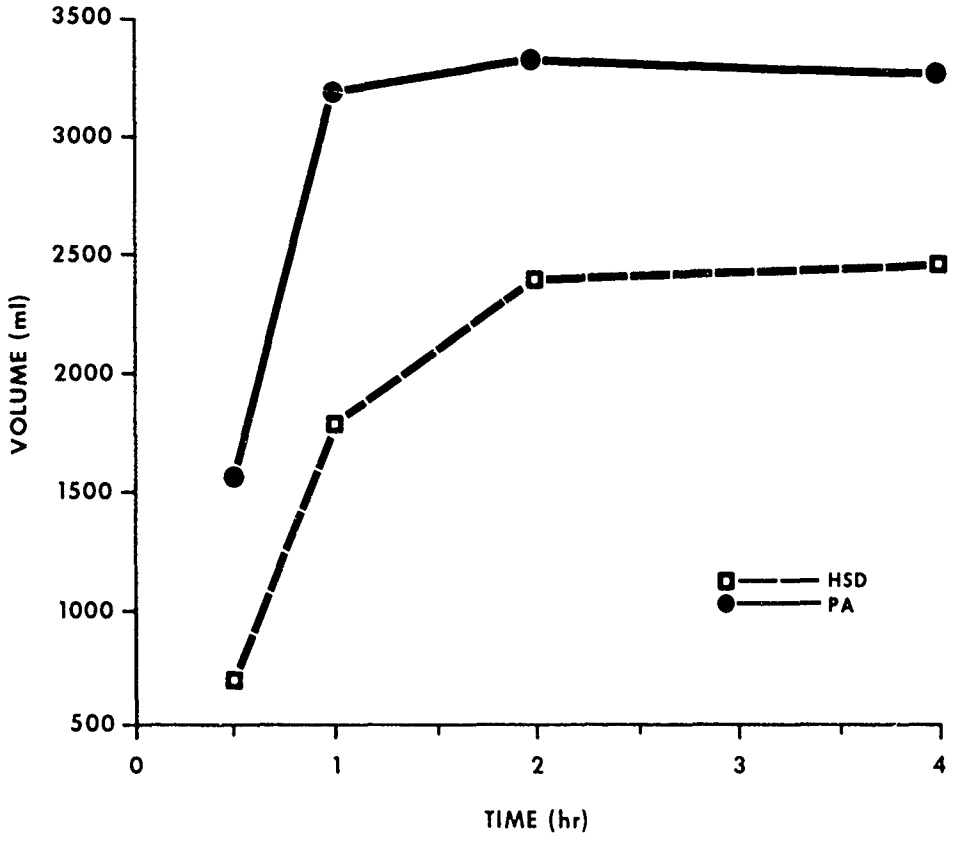


Figure 4. Blood requirements during the initial four hours after admission.

DISCUSSION

In these series of experiments, we have shown that a small volume infusion of 7.5% NaCl prepared in 6% dextran 70 is superior to an equivalent volume of standard crystalloid and 7.5% alone in its ability to treat hemorrhagic shock. In the first experiment, 100% survival was achieved in animals infused with 11.5 ml/kg of the HSD solution after an

otherwise lethal hemorrhage. A further reduction in the total amount of HSD infused to 4 ml/kg resulted in similar survival while reducing significantly the level of hypernatremia observed after infusion. The potential clinical and logistical impact of this new resuscitation solution is profound. Much less fluid may be required to resuscitate any given casualty. A combat casualty who has bled one liter of whole blood may require only 250 ml of the hypertonic saline-dextran solution, as compared to the 3 liters of lactated Ringer's solution currently recommended. For any given amount of lactated Ringer's solution, twelve times the number of casualties may be treated with the hypertonic saline-dextran solution. Therefore, this solution may alleviate the present logistical burden of having to supply large quantities of fluids to field units while still optimizing the survival of the injured soldier.

In the second and third experiments, emphasis was placed on further elucidating the mechanism of action of the HSD solutions. The effectiveness of this solution appears to be related to the synergistic effect between the hypertonicity of the sodium chloride and the hyperoncotic properties of dextran. Initially, the hypertonic saline shifts fluid from the intracellular space to the intravascular compartment. The addition of dextran to 7.5% NaCl, however, appears to prevent the sudden deterioration in plasma volume seen with 7.5% NaCl alone. Thus, that fluid which is shifted from the intracellular space as a result of the hypertonicity of 7.5% NaCl is retained within the intravascular compartment as a result of the hyperoncotic properties of dextran.
These effects were enhanced by dilutional and dextran-induced reductions in blood viscosity and by neurohumorally modulated decreases in total peripheral resistance. The net result was an improvement in venous return and cardiac filling. This augmentation in cardiac filling results in a substantial improvement in overall visceral blood flow which is associated with or results in an improvement in energy metabolism, most notably a reduction in tissue O_2 demand.

Finally, in the preliminary clinical trial, we demonstrated the feasibility and safety of conducting a study on the field administration of HSD in trauma patients. In those patients infused with HSD, there were trends towards improvement in survival, blood pressure and a reduction in resuscitation fluid requirements. Based in large part on the success of the previous non-clinical and clinical studies, a large multicenter trial on the efficacy of this solution in the field setting has been approved by the FDA and is currently in progress. This study is being conducted at Ben Taub General Hospital, Houston, Texas, Denver General Hospital, and Milwaukee County Medical Complex and is expected to be completed in 1989.

CONCLUSION

We have demonstrated in non-clinical and clinical studies that small volume infusion of the new resuscitation solution 7.5% NaCl prepared in 6% dextran 70 is superior to an equivalent volume of standard crystalloid in its ability to treat hemorrhagic shock. The hypertonic saline-dextran solution has the potential advantages of improving survival, producing a beneficial cardiovascular response with smaller fluid volumes, and reducing total fluid requirements during resuscitation. This new resuscitation solution may substantially improve the survival of combat casualties, thus optimizing the conservation of the fighting strength, thereby offering our Armed Forces a competitive edge.

REFERENCES

1. Bellamy RF: The causes of death in conventional land warfare: implications for combat casualty care research. Milit Med 1984;149: 55-62.

2. American College of Surgeons Committee on Trauma, Chicago, Ill. Advanced Trauma Life Support Course for Physicians, 1985.

3. Velasco IT, Pontieri V, Rocha-e-Silva M, et al: Hyperosmotic NaCl and severe hemorrhagic shock. Am J Physiol 1980;239:H664-H673.

4. Nakayama S, Kramer GC, Carlsen RC, et al: Infusion of very hypertonic saline to bled rats: membrane potentials and fluid shifts. J Surg Res 1985;38:180-186.

5. Nakayama S, Sibley L, Gunther RA, et al: Small volume resuscitation with hypertonic saline (2,400 mOsm/Liter) during hemorrhagic shock. Circ Shock 1984;13:149-159.

6. Traverso LW, Bellamy RF, Hollenbach SJ, et al: Hypertonic sodium chloride solutions: Effects on hemodynamics and survival after hemorrhage in swine. J Trauma 1987;27:32-39.

7. Prough DS, Johnson JC, Poole GV, et al: Effects on intracranial pressure of resuscitation from hemorrhagic shock with hypertonic saline versus lactated Ringer's solution. Crit Care Med 1985;13:407-411.

8. Maningas PA, DeGuzman LR, Tillman FJ, et al: Small-volume infusion of 7.5% NaCl in 6% Dextran 70 for the treatment of severe hemorrhagic shock in swine. Ann Emerg Med 1986;15:1131-1137.

9. Wade CE, Bossone C, Hunt M, et al: Cardiovascular, hormonal and metabolic responses to resuscitation with small volumes of hypertonic solutions following hemorrhage. Fed Proc 1987;46:805.

10. Gelin LE, Dawidson I: Plasma expanders and hemodilution in the treatment of hypovolemic shock, in Cowley RA, Trump BF (eds): Pathophysiology of shock, anoxia, and ischemia. Baltimore, Williams and Wilkins, 1982, p 454-464.

11. Smith GJ, Kramer GC, Perron P, et al: A comparison of several hypertonic solutions for resuscitation of bled sheep. J Surg Res 1985;39:517-528.

12. Kramer GC, Perron PR, Lindsey C, et al: Small volume resuscitation with hypertonic saline dextran solution. Surgery 1986;100:239-246.

13. Maningas PA: Resuscitation with 7.5% NaCl in 6% Dextran 70 during hemorrhagic shock in swine: Effect on organ blood flow. Crit Care Med 1987;15:1121-1126.

14. Bellamy RF, Pedersen DC, DeGuzman LR: Organ blood flow and the cause of death following massive hemorrhage. Circ Shock 1984;14:113-127.

15. Heymann MA, Payne BD, Hoffman JIE, et al: Blood flow measurements with radionuclide-labelled particles. Prog Cardiovasc Dis 1977;20:55.

Novel Derivatives of Human Hemoglobin Ao for Possible Use as Emergency Resuscitation Fluids. (U)

Dr. Mario A. Marini*, Col. Robert M. Winslow, Mr.Scott Christensen, Mr. Frank Medina, Col. Jovan Tasesky, Sgt. Shawn Snell and Mr. Angelo Zegna.

Letterman Army Institute of Research, San Francisco, CA 94129-6800

For some time, studies have been conducted on the feasibility of using hemoglobin or hemoglobin-based derivatives as emergency resuscitation fluids. The major disadvantages of cell-free hemoglobin include toxicity, high oxygen-affinity (P_{50} ca. 10 torr) and the short intravascular retention time ($t_{1/2}$ 1 to 2 hrs), but advances in the field have led to general optimism (1-4).

Removal of the bulk of the membrane material from a hemolysate (stroma-free hemoglobin, SFH) was shown to reduce toxicity (5), but traces of contaminating materials present in SFH may still cause serious toxicity (6). Many products currently under development have not been tested extensively in vivo. Some preparations have been reported to cause morphological and hemodynamic alterations in pigs (7), altered reticuloendothelial function (8) hypoxic and ischemic effects on retinal morphology (9) and possibly synergistic toxicity with endotoxins (10,11). Stroma-free hemoglobin and its derivatives cause coronary vasoconstriction in isolated rabbit hearts (12) and interfere with the autoregulatory mechanisms (13) although extensively purified hemoglobin (HbAo) and pasteurized cross-linked hemoglobin do not interfere (14).

Within the red cell, the binding of oxygen to hemoglobin is modulated by high hemoglobin concentration, the complex interaction of 2,3-diphosphoglyceric acid (15), CO_2, chloride and other anions (16) as well as pH (17) to give a P_{50} of 25 to 30 torr (18). In a cell-free solution, these modulators are weakly bound and are not sufficient to maintain the lower oxygen affinity required for tissue oxygen delivery.

The actions of chloride, 2,3-DPG and CO_2 are at least partially effective by introducing negative charges at the N-terminus of the α-chain and/or into the β-cleft. The P_{50} could be maintained at an appropriate level by the introduction of a covalently bound adduct at either or both of these sites (19). This finding explains the decrease in the oxygen affinity of hemoglobin by carboxymethylation (20-22) of the N-terminal amino groups. Also, a covalent bond can be formed with amino groups using pyridoxal phosphate (23,24) as an analog to 2,3-DPG.

It is generally understood that free hemoglobin is rapidly cleared from the circulation because of its dissociation into dimers which can be filtered through the glomerulus (25). As a means of preventing the dissociation, intratetrameric cross-linking was attempted using disubstituted aspirin derivatives (26) which were initially intended for use as anti-sickling agents (27). Reaction with oxyhemoglobin increased the oxygen binding but, when the reaction was conducted using deoxyhemoglobin (28), the P_{50} was elevated and the β-chains were said to be cross-linked, although this may be in error (27,29). Cross-linking and a simultaneous elevation of P_{50} can also be achieved with glycoaldehyde (20), dextran dialdehydes (30) and 2-nor-2-formylpyridoxal 5′-phosphate (31). A large number of other studies have used pyridoxylated hemoglobin which is polymerized with glutaraldehyde (32,33) to produce the desired P_{50} and a prolonged intravascular retention time.

The variable toxicity reported for various hemoglobins and derivatives is a cause for concern. Compositions of these solutions are quite variable and contain, in addition to the major component, HbAo, other minor hemoglobins as well as all the intracellular enzymatic components. When this mixture is subjected to the derivatization procedures, a further proliferation of molecular species occurs. In most instances, these solutions are used without further purification. It is therefore difficult to attribute any aberrant effect to any individual component in such a complex mixture. Attempts to purify these mixtures chromatographically have been largely unsuccessful and have considerably decreased the yield. To alleviate the heterogeneity, we have conducted our studies on the purified HbAo using mild reagents in low concentrations for brief times. Many reagents have been tried (Table 1), but for this paper we wish to report our progress with the reagents disuccinimidyloxalate (34), disuccinimidyltartartate (35), pyridoxal phosphate and erythrose-4-phosphate (36).

Table 1. Reagents Used with Stroma-free Hemoglobin Solutions

	Symbol
I. No Apparent Reaction:	
1) Succinimidyl 4-(N-maleimidomethyl)cyclohexyl carbonate	SMPB
2) Bis[2-(succinimidooxycarbonyloxy)ethyl] sulfone	SCS
3) N,N'-Dicyclohexylcarbodiimide	CDI
4) Dimethyladipimate	DMA
5) 2,3-diphosphoglycerate	2,3-DPG
II. Reaction with Decreased P_{50}:	
1) d-Erythrose	E
2) Glyceraldehyde 3-phosphate	G-3-P
III. Reaction with Increased P_{50}:	
1) Erythrose 4-phosphate	TP
2) Bis-dibromosalicylfumarate	DBBF
3) Disuccinimidyltartarate	DST
4) Disuccinimidyloxalate	DSO
5) Pyridoxylphosphate	PLP

Materials and Methods

Whole, outdated, packed red blood cells were washed three times with cold, isotonic, sterile saline. The buffy coat was removed, and the packed red cells were lysed with 2.5 volumes of cold water. The solutions were centrifuged at 20,000 g for 1 hr, and the upper 2/3 was passed through a mixed-bed ion exchange column (BioRex RGJ01-X8) to obtain solutions of approximately 5 g/dl with a conductivity of less than 10 μmhos. These preparations contain all the components of the red blood cell. These solutions containing 10-20 g Hb were then chromatographed on a Waters Delta-Prep 3000 using a 5.7 x 30 cm stainless steel column containing Waters QMA-Accell as previously described by Christensen et al. (37). The major component of SFH is the normal adult hemoglobin, HbAo, in about 80% yield.

Disuccinimidyloxalate (DSO) was purchased from Polysciences Inc., Warrington, PA. Erythrose-4-phosphate (TP) was obtained from Sigma and synthesized from glucose-6-phosphate by the procedure of Ballou and MacDonald (36). Disuccinimidyltartarate (DST) was purchased from Pierce and synthesized by the method of Smith et al. (38). Pyridoxal phosphate (PLP) was also obtained from Sigma. All other reagents were of the highest purity available.

Reaction Conditions:

Solutions of HbAo (3-6 mM in heme) were placed in a titration vessel and flushed with water-saturated air or nitrogen for two hours. The pH was adjusted to the desired end-point using a Radiometer TTT80 titrator and the ABU80 autoburette set in the pH-stat mode. All reactions reported here were run at room temperature. The reactions were initiated by the addition of the reagents in powder form (1-8 molar equivalents), and the pH was maintained at the end-point. Reactions were considered complete when no addition of acid (or base) occurred in 30 min. Generally they were completed within 2 hrs. These solutions were then passed through the mixed-bed ion exchange resin and subjected to analysis. Initially the reactions were performed with SFH to aid in the selection of reagents to be studied. The results of this screening are shown in Table 2.

Table 2. Stroma-Free Hemoglobin Derivatives

	P_{50}*	METHEMOGLOBIN (%)**		
		2 hrs	40 hrs	1 mo
Whole blood	22	--	--	--
SFH(deoxy)	12	2.9	4.1	14.2
Erythrose	8	0.8	0.9	--
CDI	12	0.6	0.8	4.5
SMPB	13	2.0	4.5	6.2
SCS	11	2.4	5.7	8.2
DMA	12	2.5	3.9	5.3
TP	25	0.6	0.8	4.1
DST	28	2.3	3.3	9.8
DBBF	28	2.4	3.8	5.9
DSO	27	3.2	4.0	4.5
PLP	29	0.9	--	--

* P_{50} readings by Hemox Analyzer after 2 hrs of reaction.
** HbO_2 deoxygenated 16 hrs at 22°C and reacted at pH 7.5 as deoxyhemoglobin with reagents indicated (4:1 molar ratio) at 22°C. One month samples stored at 4°C. The DSO and PLP were reacted with oxyhemoglobin.

Analytical Procedures:

Yields were estimated from the analytical chromatograms which were performed on a 0.5x5 cm column of MonoQ resin (Pharmacia, Piscataway, NJ). The samples (100 µl, 1 g/dl) were developed with a linear gradient of 20 mM Tris-HCl, pH 8.0 and 20mM Tris-HCl, pH 8.0 containing 0.4 M NaCl. Generally all the species were resolved at 35% Buffer B. The column was washed with 100% Buffer B to remove extensively modified protein. The extent of polymerization or cross-linking was estimated by

SDS-PAGE gel electrophoresis with the PHAST system of Pharmacia using 1 µl of 0.25 g/dl as previously reported (37). Oxygen equilibria were determined with the Hemox analyzer using Hemox buffer at 37°C. The P_{50} values were taken directly from the graphs and the Hill coefficients were calculated from a plot of the logarithmic values of the fractional saturation at 40-75% against the log of the oxygen tension. Hemoglobin concentration and methemoglobin content were determined by the procedures previously described (39).

Results and Discussion

Many bifunctional reagents have been used to modify and cross-link proteins. This approach must necessarily introduce another moiety into the protein which must eventually be metabolized. Considering the large amount of hemoglobin needed in an emergency situation (i.e. 4 mM), the cross-link should be innocuous. DST is one such reagent which could react with two available lysine residues positioned at approximately 6 A intervals to form a diamide of tartaric acid. When reacted with the oxygen-ligated hemoglobin, the derivatives are formed as seen by chromatography, but the P_{50} is unaltered. Reaction with deoxyHbAo, however, gives a mixture with a P_{50} of 20 to 28 torr. At pH 7.0, the maximun yield is 60% with a 5:1 molar ratio (DST:heme) and is not improved when excess reagent is added nor is the P_{50} increased; a typical chromatogram is shown in Fig. 1. It is apparent that a considerable heterogeneity exists which is not desirable.

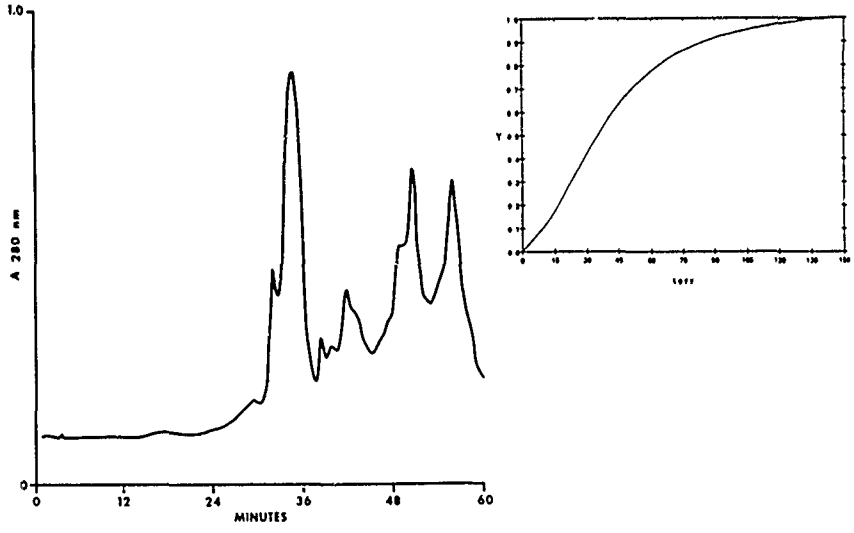


Figure 1. Analytical chromatogram of the products of DST and HbAo (4:1) at pH 7. The P_{50} (inset)= 34 torr with n = 1.9.

By analogy to the interaction of 2,3-DPG, we have considered that the negatively charged phosphates are directed into the β-cleft; therefore, a chemical analog of 2,3-DPG which would form a covalent bond would elevate the P_{50} and form a stable entity. This is the basis for the modifications performed with pyridoxal phosphate which could form a Schiff base. This derivative is not stable. Stability was achieved by reduction to the substituted amine (40), but this caused considerable heterogeneity (41). Though these derivatives have appropriate P_{50} values and an extended half-life (42), many investigators (32,33) have polymerized these mixtures with glutaraldehyde to yield materials with a large range of molecular weights, which cannot be resolved chromatographically.

We have reacted PLP with both oxy- and deoxyhemoglobin (4:1 PLP:heme) in the absence of Tris and without subsequent reduction. The solutions are passed through the mixed-bed ion exchange resin to remove any excess PLP. The P_{50} values were uniformly elevated to the range of 25-35 torr and the chromatograms showed one peak in the position of the HbAo. This peak has an oxygen equilibrium identical to the original hemoglobin and is devoid of phosphorus. Apparently the reaction was reversed on the anion exchange column. If these solutions, prior to chromatography, are incubated at 4°C for periods up to 1 month and then chromatographed, two peaks of approximately equal size appear (Fig. 2),

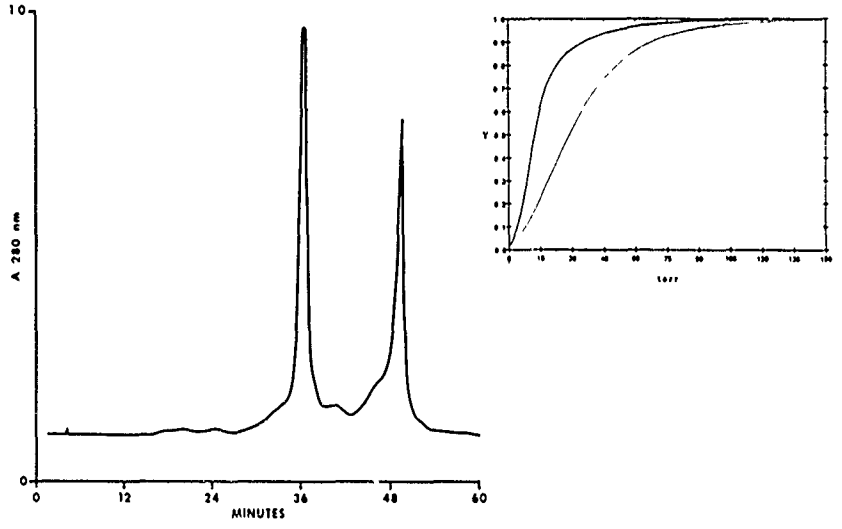
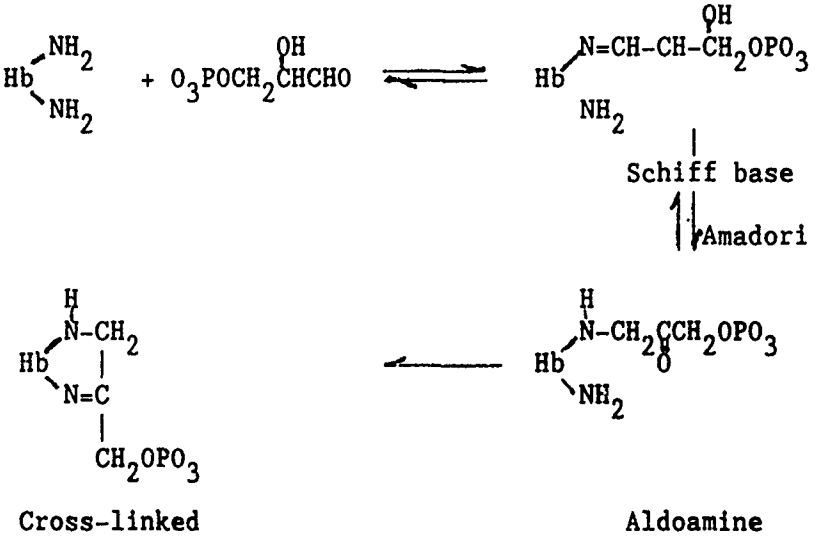


Figure 2. Analytical chromatogram of the products of PLP:HbAo (4:1) aged 3 wks at 4°C. The P_{50} for HbAo = 12 torr with n = 2.3; the P_{50} for the reaction mixture is 29 torr with n = 2.34.

one of which is in the original HbAo position and the other peak which contains 0.5 molar equivalents of phosphorus with a P_{50} of 25-30 torr. The second peak again gave two peaks when chromatographed, identical to that of the original chromatogram. Apparently, an equilibrium mixture exists which requires considerable time to be established. Efforts to stabilize this derivative by reduction gave the mixtures reported by McGarrity et al. (41).

A more stable derivative might be formed based on the glycolization of nemoglobin by carbohydrates. Originally we synthesized glyceraldehyde-3-phosphate (43) and reacted this with deoxyHbAo. It was thought that the molecule would react by means of Schiff base formation, which is a reversible reaction, but that, in time, the Amadori rearrangement would occur to form a stable derivative which might also be cross-linked according to the following sequence:



The results were disappointing, for in the time period of the reaction (2 hrs) no change in the P_{50} could be seen. This lack of interaction may have been the result of the enolization of the reagent. The next higher analog would be either threose-4-phosphate or erythrose-4-phosphate which exist as hydrated aldehydes (44). Reaction of TP with oxyHbAo gave a solution with a decreased P_{50} (7 torr), but with deoxyHbAo at a ratio of 4:1, the P_{50} increased to 25 torr (Fig. 3). The oxygen equilibrium may be further altered to P_{50} = 34 torr with a 8:1 molar ratio. Reaction must occur at a number of sites since at least 6 peaks are shown by chromatography (Fig 3). This reaction is quite dependent on the purity of the reagent which is commercially supplied (65-75%). TP made by oxidation is a cleaner product.

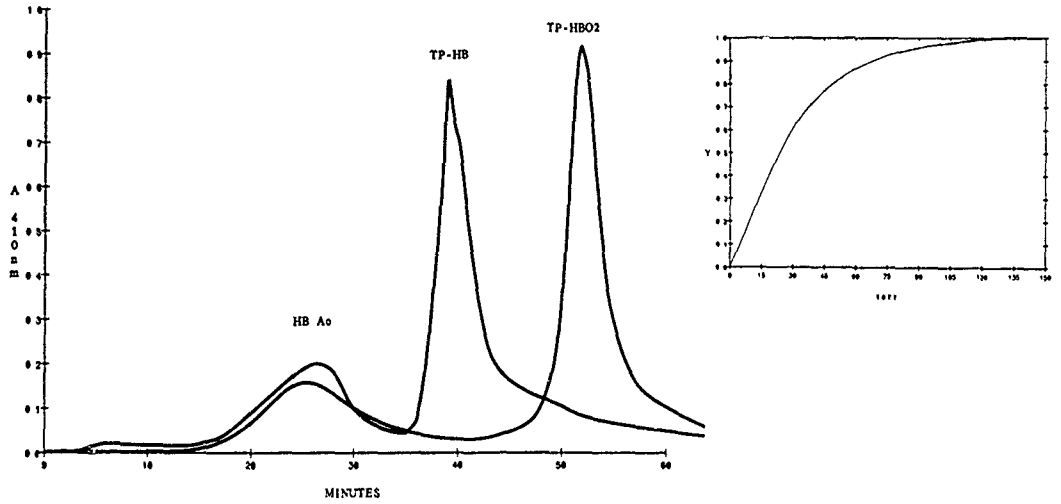


Figure 3. Reaction products of TP with both oxy- and deoxyHbAo (4:1) at pH 7.5. The peaks can be resolved into 6 components. P_{50} is 35 torr with n = 2.3 for the deoxyHbAo only.

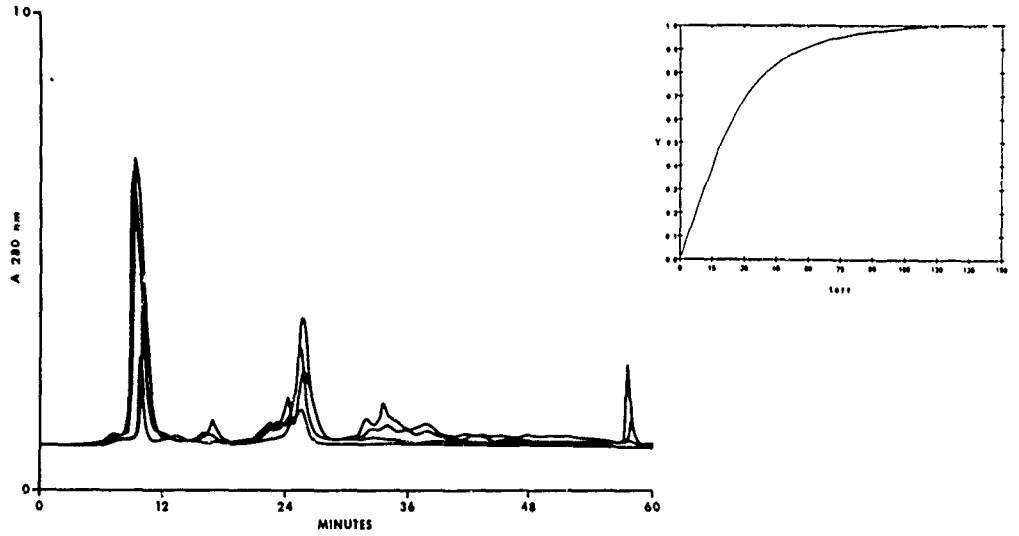


Figure 4. Products of DSO:oxyHbAo (molar ratios 1,2,4,6:1) at pH 7. The inset shows a P_{50} of 20 torr with n = 1.8.

An alternate procedure for cross-linking hemoglobin would be by the internal formation of amide bonds utilizing the free carboxyl and the amino groups of the protein. This would have the desirable effect of not introducing a foreign molecule but would leave the hemoglobin internally bonded. Initially the reaction was attempted with CDC, but at the concentrations used, no reaction was apparent (Table 2). DSO, however, was effective with both the oxy- and deoxy- forms. In either case, the P_{50} was elevated to desirable values, and extensive cross-linking was evident on SDS-PAGE gels (Fig 4). The analytical chromatograms of the reaction products using oxyHbAo were surprising in that a single major component was found for reactions using up to 5:1 molar ratios (Fig. 5). The P_{50} values reached a maximum of 28 torr at a molar ratio of 4:1 (Table 3). With the deoxy- form, apparently more derivatives are formed which are further altered with increasing concentrations of the reagent.

Table 3. Partial Analysis of HbAo Derivatives of DSO Formed at pH 7.0 and 22°C.

	DeoxyHbAo		OxyHbAo	
Molar Ratio (DSO/Heme)	MetHb (%)	P_{50} (torr)	MetHb (%)	P_{50} (torr)
0	1.1	11.0	0.8	11.4
1	4.9	13.3	1.4	15.2
2	5.1	16.0	1.7	17.5
3	6.1	18.5	2.3	20.2
4	7.2	21.5	3.1	21.1
5	7.9	24.1	4.1	24.0
6	8.0	24.9	4.6	24.1
7	8.2	28.1	5.9	24.0
8	8.5	28.0	6.2	24.9

In determining the most advantageous reaction conditions, for any particular reagent, a number of factors must be considered. Experience with the DSO reactions are generally applicable to all the reagents studied. The P_{50} appears to be higher at low pH (Table 4) but, methemoglobin is also elevated. This is more prominent with deoxyhemoglobin. As the pH is elevated, methemoglobin formation can be reduced, but the P_{50} is also lowered. Increasing the concentration of the reagent can partially correct the P_{50} values and increase the yield of derivative at the expense of the oxidation and the formation of a more heterogenous mixture (Fig 5).

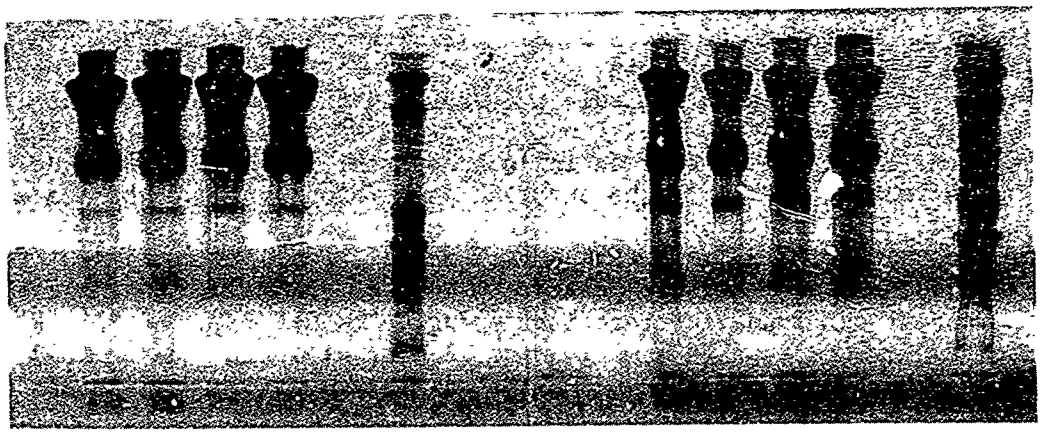


Figure 5. SDS-PAGE gels of the reaction mixtures of DSO with HbAo. Left gel: 1-4 molar equivalents DSO:HbAo. Right gel: 5-8 molar equivalents. The results indicate increasing polymerization with concentration. The standards are in the right hand lanes.

Table 4. Effect of pH on the Oxidation and Oxygen Equilibria of DSO (4:1) Treated HbAo.

	DeoxyHbAo		OxyHbAo	
pH	P_{50} (torr)	MetHb (%)	P_{50} (torr)	MetHb (%)
6.5	24.3	8.8	23.0	4.3
7.0	21.5	7.2	21.4	3.1
7.5	17.5	4.4	19.2	2.6
8.0	16.5	3.3	14.1	2.2

From a structural point of view, the more ideal reaction conditions would require a 4 or 5 molar excess with oxyhemoglobin at pH 6.5 or 7.0 which gives a greater than 50% yield of a product, which after chromatographic separation gives a P_{50} of 26 torr. It must be emphasized, however, that biological studies will ultimately dictate the more advantageous reaction conditions and that purity and yield are only two of many considerations.

Conclusion

Of the four derivatives of HbAo examined, all result in a yield of products which are greater than 50% and have P_{50} values in the range of 20-30 torr comparable to that of the red blood cells. These products can be purified chromatographically and exhibit excellent structural stability as well as resistance to oxidation with storage at $4^{o}C$. These derivatives, as well as those previously reported, are not to be considered the "ideal" emergency resusitation fluid if in fact an ideal solution exists. Many formulations meet the minimal requirements and are worthy of further study, but the number of derivatives possible with the large number of reactive residues of hemoglobin are limited only by our imagination and our understanding of this complex molecule.

References

1. Kothe, N., Eichentopf, B., & Bonhard, K. (1985) Surg. Gynecol. Obstet. 161, 563-569.

2. Odling-Smee, W., & McLarnon, C. (1984) Irish J. Med. Res. 153, 385-388.

3. Feola, M., Gonzales, H., Canizaro, P.C., Bingham, D., & Periman, F. (1983) Surg. Gynecol. Obstet. 157, 399-408.

4. Birkigt, H.G. (1983) Kunstliche Organe. 16, 103-108.

5. Rabiner, S.F., Helbert, J.R., Lopas, H., & Friedman, L.H. (1967) J. Exp. Med. 126, 1127-1142.

6. Feola, M., Simoni, J., Canizaro, P.C., Tran, R., Raschbaum, G., & Behal, F.J. (1988) Surg. Gynecol. Obstet. 166, 211-222.

7. Birkigt, H.G., Schultz, M., Sedlarik, K., Ebener, E., Riedel, E., & Freitzche, L. (1985) Kunstliche Organe. 18, 139-149.

8. Marks, D.H., Patressi, J., & Chaudry, I.H. (1985) Shock 16, 165-172.

9. Schuschereba, S.T., Friedman, H.I., DeVenuto, F., & Beatrice, E.S. (1983) Lab. Invest. 48, 339-352.

10. White, C.T., Murray, A.J., Smith, D.J., Greene, J.R., & Bolin, R.B. (1986) J. Lab. Clin. Med. 108, 132-137.

11. Dunn, D.L., Nelson, R.D., Conde, R.M., & Simmons, R.L. (1983) Surgery 83, 653-659.

12. Vogel, W.M., Dennis, R.C., Cassidy, G., Apstein, G.A., & Valeri, C.R. (1986) Am. J. Physiol. 251, 13-20.

13. Biro, G.P., Taichman, G.C., Lada, B., Keon, W.J., Rosen, A.L., & Sehgal, L.R. (1987) Artif. Organs 12, 40-50.

14. MacDonald, V.W., Marini, M., Klinker, M.T., & Christensen, S. (1988) FASEB Journal 2, A1711.

15. Benesch, R., & Benesch, R.E. (1967) Biochem. Biophys. Res. Commun. 26, 162-167.

16. Antonini, E., & Brunori, M. (1971) in "Hemoglobin and Myoglobin in their Reactions with Ligands." Elsevier, New York.

17. Kilmartin, J.V., Fogg, J., Luzzana, M., & Rossi-Bernardi L. (1973) J. Biol. Chem. 248, 7039-7043.

18. Winslow, R.M., Samaja, M., Winslow, N.J., Rossi-Bernardi, L., & Shrager, R.I. (1983) J. Appl. Physiol. 54, 524-529.

19. Fantl, W.J., Manning, L.R., Veno, H., DiDonato, A., & Manning, J.M. (1987) Biochemistry 26, 5755-5761.

20. Acharya, A.S., & Manning, J.M. (1983) Proc. Nat. Acad. Sci. USA 80, 3590-3594.

21. Hedlund, B., Carlson, J., Condie, A., & Drayton, C. (1983) in Advances in Blood Substitute Research." Liss, New York, pp. 71-77.

22. Saunders, S., & Hedlund, B.E. (1984) Biochemistry 23, 1457-1461.

23. Benesch, R., Benesch, R.E., Kwong, S., Acharya, A.S., & Manning, J.M. (1982) J. Biol. Chem. 257, 1320-1324.

24. Bolin, R., & DeVenuto, F. (1983) in "Advances in Blood Substitute Research." Liss, New York, pp. 1-8.

25. Bunn, H.F., & Jandl, J.H. (1968) Trans. Assoc. Am. Physicians 81, 47.

26. Walder, J.A., Zaugg, R.H., Walder, R.Y., Steele, J.M., & Klotz, I.M. (1979) Biochemistry 18, 4265-4270.

27. Walder, J.A., Walder, R.Y., & Arnone, A. (1980) J. Mol. Biol. 141, 195-216.

28. Tye, R.W., Medina, F., Bolin, R.B., Knopp, G.L., Irion, G.S., & McLaughlin, S.K. (1983) in "Advances in Blood Substitute Research." Liss, New York, pp. 41-49.

29. Chatterjee, R., Welty, E.V., Walder, R.Y., Pruitt, S.L., Rogers, P.H., Arnone, A.G., & Walder, J.A. (1986) J. Biol. Chem. 261, 9929-9939.

30. Tam, S.-C., Blumenstein, J., & Wong, J. (1976) Proc. Nat. Acad. Sci. USA 73, 2128-2131.

31. Benesch, R., Benesch, R.E., Yung, S., & Edalji, R. (1975) Biochem. Biophys. Res. Commun. 63, 1123-1129.

32. Sehgal, L.R., Rosen, A.L., Gould, S.A., Sehgal, H., Dalton, L., Mayoral, J., & Moss, G.S. (1980) Fed. Proc. 39, 718.

33. DeVenuto, F., & Zegna, A. (1983) J. Surg. Res. 34, 205-212.

34. Takeda, K., Sawada, I., Suzuki, A., & Ogura, H. (1983) Tetrahedron Lett. 24, 4451-4454.

35. Hill, M., Bechet, J.-J., & d'Albis, A. (1979) FEBS Lett. 102, 282-286.

36. Ballou, C.E., & MacDonald, D.L. (1963) in "Methods in Carbohydrate Chemistry." Academic Press, New York, pp. 293-295.

37. Christensen, S.M., Medina, F., Winslow, R.M., Snell, S., Zegna, A., & Marini, M.A. (1988) Biochem. Biophys. Methods, in press.

38. Smith, R.J., Capaldi, R.A., Muchmore, D., & Dahlquist, F. (1978) Biochemistry 17, 3719-3723.

39. Snell, S.M., & Marini, M.A. (1988) Biochem. Biophys. Methods, in press.

40. Benesch, R., Benesch, R.E., Renthal, R.D., & Maeda, N. (1972) Biochemistry 11, 3576-3582.

41. McGarrity, M.J., Er, S.S., & Hsia, J.C. (1987) J. Chromatog. 419, 1-14.

42. Greenburg, A.G., Schooley, M., & Peskin, G.W. (1977) J. Trauma 17, 501-504.

43. Ballou, C.E., & Fischer, H.O.L. (1955) J. Am. Chem. Soc. 77, 3329-3331.

44. Serianni, A.S., & Barker, R. (1984) J. Org. Chem. 49, 3292-3300.

An Experimental Investigation of the Aerodynamic Breakup Behavior of Heated Newtonian Liquids by Ambient Air. (U)

Joseph E. Matta, Ph.D.
Chemical Research, Development and Engineering Center
Aberdeen Proving Ground, Aberdeen, Md., 21010

1. INTRODUCTION

Various binary weapons designs rely on aerodynamic forces to atomize and disperse liquid agent. Typically, the air is at a much lower temperature than the liquid and cooling occurs during atomization. It is conceivable that when hot binary agent is atomized by ambient temperature air, evaporative cooling during atomization could significantly affect the resultant drop size. In a reported wind tunnel experiment by Weiss and Worsham (1) it states, "... in order to obtain meaningful results there must be no marked change in temperature until atomization is complete " They found, even with molten wax which has an estimated vapor pressure of 10^{-5} mm Hg, that the liquid's temperature drops and it begins to freeze while drop breakup is still occurring. The rapid liquid cooling during atomization leads to ambiguity in temperature and thus also in liquid physical properties. Since predictive models (2) used to evaluate the breakup performance of a disseminated liquid are very sensitive to the liquid physical properties, ambiguity in the model predictions exist. Therefore, in order to improve the predictive breakup modeling capability, an experimental investigation of the aerodynamic breakup behavior of heated Newtonian liquids by ambient air was conducted.

A 7.62 cm I.D. wind tunnel was constructed and used to atomize heated liquids and was designed for eventual use with binary agents. The resultant aerosol size (i.e. mass median diameter, MMD) was measured downstream using a Malvern 2600 Laser Particle Sizer which uses the principle of diffraction to measure particle size in a non-intrusive manner. Prior to the heated trials a series of dissemination tests were conducted with various Newtonian liquids in order to develop an empirical expression relating the resultant aerosol MMD to physical fluid properties and flow conditions. Although Weiss and Worsham (1) have reported an empirical breakup expression derived from data obtained using a similar dissemination device, it was deduced from liquids whose physical properties did not vary significantly. In their study, the liquid surface tension and fluid viscos-

ities varied only between 18 to 29 dyne/cm and 3.35 to 11.2 centipoise, respectively. However, since the surface tension of ambient and hot binary liquid can vary up to 50% and the viscosity by at least 2 orders of magnitude, development of an empirical expression from breakup data on fluids covering a much wider range of physical properties was deemed necessary.

2. EXPERIMENTAL

2.1 Wind Tunnel Facility

Figure 1 is a photograph of the wind tunnel test facility used for the atomization trials. Seen from left to right are the four major components of the system: blower, tunnel with liquid injection system, particle sizer, and aerosol scrubber, The Buffalo 5E blower (0.465 m diameter wheel), powered by a 2500 rpm variable speed D.C. motor, can generate wind speeds of 60 m/sec. Connected to the blower is a sheet metal transition (0.46 m long) that is used to connect the fan outlet (12.7 cm diameter) to a Plexiglas tube (10.2 cm i.d. and 0.46m long). Situated on the floor below the tubing is a stainless steel reservoir (8-liter capacity) used to hold and heat the test liquids. The thermostatically controlled, 1040-watt heating tape can condition the test liquids up to 150 °C. To the right of the reservoir is a regulated nitrogen gas cylinder that was used to force the liquid form the reservoir, through a flow meter, and out the L-shaped injector nozzle (1 cm i.d.) mounted with its longer dimension (12.7 cm) lying along the tunnel centerline. A thermocouple lead was inserted through an elbow into the nozzle and positioned in the center of the injector (1 cm from the downstream exit) to measure the liquid temperature just prior to costream injection into the tunnel.

Figure 2 is a close-up view of the wind tunnel and liquid injection system. Shown is the steel collar (22.9 cm long) used to mount the injector nozzle and to support the downstream end of the Plexiglas tunnel. Visible to the left of the collar is the pitot tube used to measure the centerline wind speed, and to the right of the collar is a portion of the Teflon sleeve that was inserted to reduce the tunnel diameter from 10.2 to 7.6 cm and achieve higher wind speeds. Attached to the downstream end of the sleeve is a section of Plexiglas tubing (20.3 cm long and 7.6 cm i.d.) where the fluid was atomized. Just visible to the left of the tube is the exit tip of the injector nozzl Also seen in the figure is the laser diffraction particle sizer and the aerosol scrubber inlet. The scrubber was used to avoid a large concentration of aerosol buildup in the test area. The laser particle sizer was positioned so the aerosol was measured 0.3m downstream from the injector nozzle. During the initial dissemination trials, high-speed, flash photography was used to confirm that complete atomization had occurred within this distance.

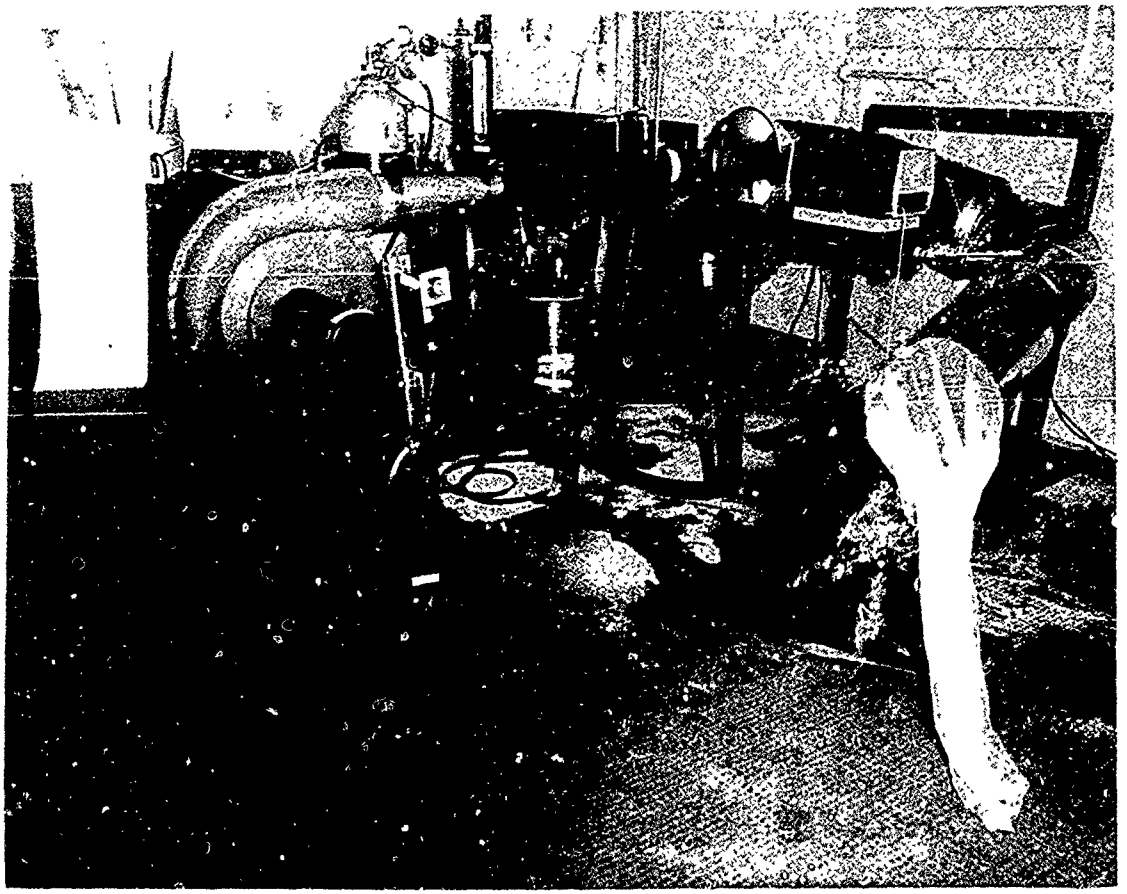


Figure 1. Wind Tunnel Facility.

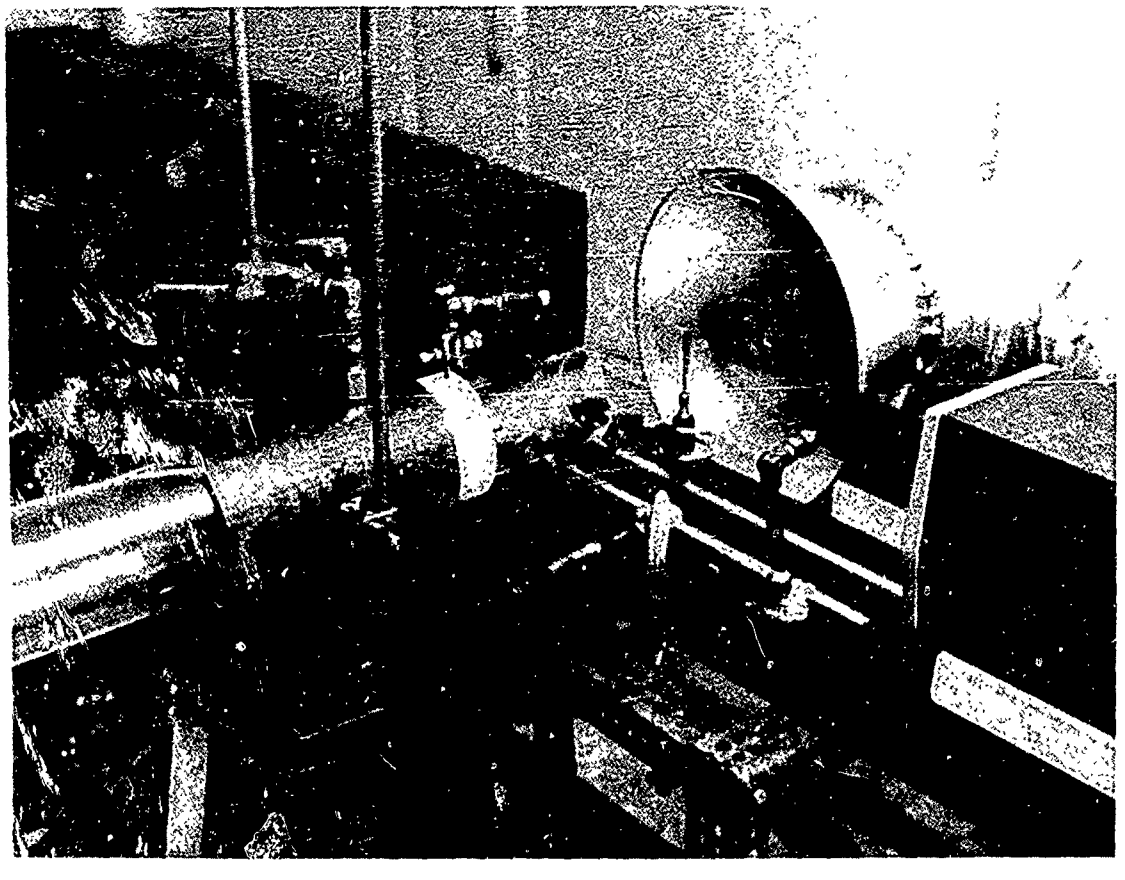


Figure 2. Close-up of Wind Tunnel and Injection System.

2.2 Laser Particle Sizer

A Malvern Instrument 2600 HSLBD particle sizer was used to measure the particle size distribution. The instrument consists of a 2 mW Helium Neon laser, which acts as a source of coherent and monochromatic light with a wavelength of 632.8 μm; and a multiconcentric ring diode detector, which measures the amount of laser light diffracted by various size particles. The laser light strikes a particle and is diffracted through an angle that is inversely proportional to the particle dimension. A lens focuses the diffracted light onto a multiring diode array positioned at the focal plane of the lens. The amount of light diffracted into each ring is measured; the diode output is digitized and converted into a particle distribution by the integral computer. For this study, a 1000-mm lens was used to collect the diffracted light and resulted in a 15-bin particle diameter distribution ranging from 19.4 to 1880 μm (see Table 3).

2.3 Test Procedure

Prior to atomizing heated liquids in the wind tunnel, ambient temperature testing was conducted. Bis(2-ethylhexyl) hydrogen phosphite (Bis-2), water, polyethylene glycol 200 (PEG 200), and glycerol were atomized to generate data for development of an empirical expression relating drop size to test conditions and physical properties of the liquid. The liquids were tested using the facility shown in Figures 1 and 2 in the following manner:

(a) Power to the laser particle sizer was switched on, and a 30-min warm-up period was allowed for the laser to stabilize. Beam adjustments were made to ensure proper alignment of the beam with the diode detector, and an aerosol background measurement was made to obtain a zero baseline.

(b) The blower motor was activated, and the direct current, variable control was adjusted for the desired wind speed.

(c) The fluid reservoir was pressurized, and the solenoid controlling liquid flow to the injector nozzle was opened.

(d) Once the desired flow rate was achieved by adjusting an in-line needle valve, the aerosol was measured for 10 sec using the laser particle sizer. Immediately afterward, the injector solenoid was closed, and the blower motor was switched off.

(e) After the integral computer analyzed the diffracted light and printed out the corresponding size distribution and mass median diameter (MMD), the wind tunnel was prepared for the next trial.

3. RESULTS AND DISSCUSSION

Figure 3 is a plot of ln(MMD) versus ln(wind velocity) for the four liquids disseminated. The water and Bis-2 trials were conducted at 25 °C, whereas the PEG 200 and glycerol were disseminated at 19 °C liquid temperature. The physical properties (i.e., surface tension, viscosity, and density) and the injection flow rate, Q, are summarized in Table 1. The

TABLE 1. Physical Liquid Properties and Injection Flow Rate

Liquid	Temp (°C)	Injection Rate (cm^3/sec)	Surface Tension (dyne/cm)	Viscosity (poise)	Density (g/cm^3)
Water*	25	25.0	71	0.009	1.0
Bis-2**	25	14.2	30	0.06	0.93
Peg 200***	19	20.6	45	0.63	1.12
Glycerol*	19	13.3	63	16.0	1.27

Physical properties provided by
* Handbook of Chemistry and Physics,
** U.S. Army Chemical Research, Development and Engineering Center (CRDEC), the Chemical Simulant Data Center,
*** Dow Chemical Data Sheet.

two solid lines shown in Figure 3 are drawn with a slope of -1.5, indicating that the MMD varies inversely with velocity, V, to the 1.5 power. The particle-size dependency on liquid flow rate was investigated using water at a fixed airspeed of 60 m/s. The water droplet size was found to increase with flow rate as, MMD ~ $Q^{0.1}$ (see Figure 4). Using the above velocity and flow rate dependency, a least squares power fit was made to the data in Figure 3 to determine how particle size varies with viscosity, η, and surface tension, σ. The two glycerol trials, however, were not included since they were conducted after generating the least square fit. The following expression was obtained:

$$\text{MMD (microns)} = 3.03 \times 10^6 \, (Q/\rho)^{0.1} \eta^{0.095} \, \sigma^{0.878} \, V^{-1.5} \qquad (1)$$

where ρ is the liquid density, and all parameters are in centimeter-gram-second units.

Figure 5 is a plot of the measured MMD versus the predicted values and indicates the goodness of fit. A standard error of estimate (10.2 μm) was calculated for this data. Using the glycerol parameters in Table 1 and a wind speed of 59 m/s, a 418 μm MMD is estimated. Even though the viscos-

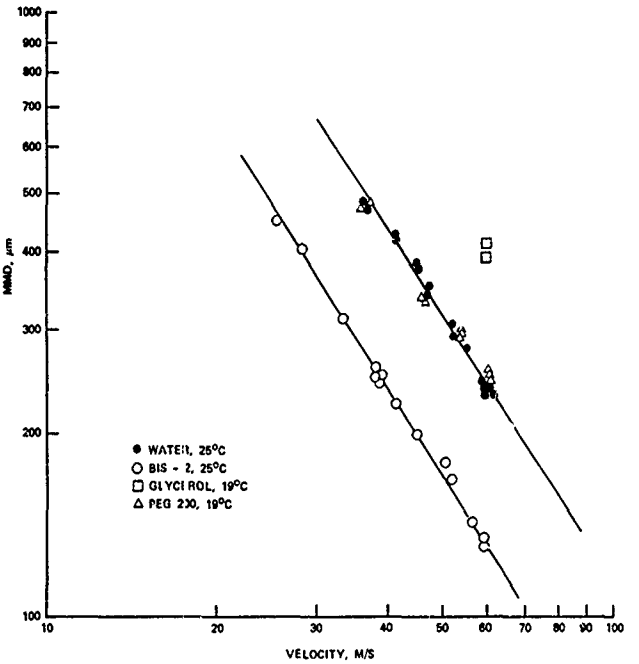


Figure 3. A ln-ln Plot of MMD versus Velocity for the Liquids Disseminated at Ambient Temperature.

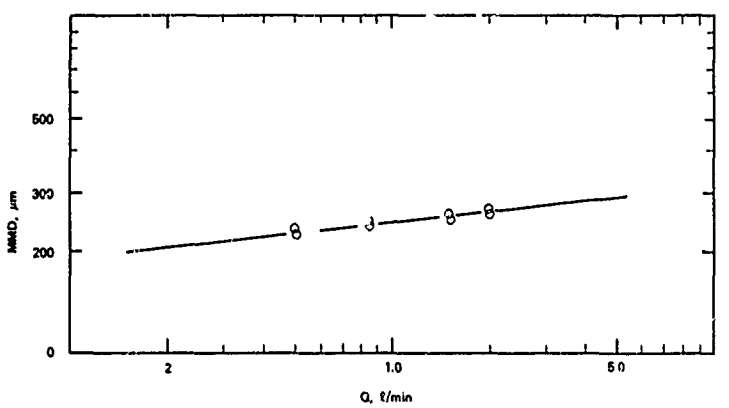


Figure 4. A ln-ln Plot of MMD Versus Injection Flow Rate for Water Disseminated at Ambient Temperature.

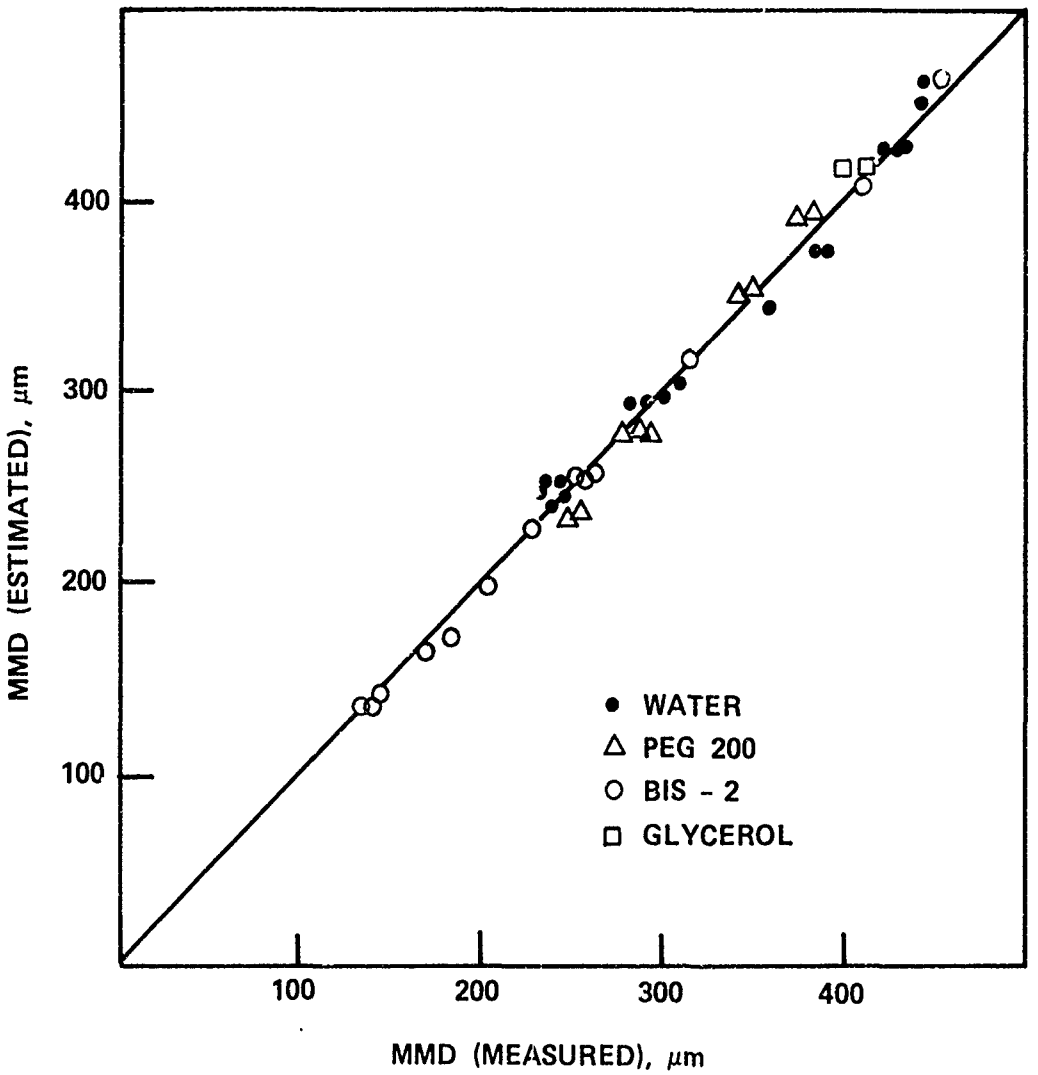


Figure 5. Measured versus Predicted Results Using Equation 1.

ity of glycerol is at least an order of magnitude larger than any of the fluids used to derive equation 1, the estimate agrees very well with the average of the two glycerol trials (i.e. 404 μm). Equation 1 differs considerably from that deduced by Weiss and Worsham. Not only did Weiss and Worsham report that drop size varied less with velocity, they also reported a larger viscosity and a much smaller surface tension effect, i.e. $MMD \sim (\eta\sigma)^{1/3}v^{-4/3}$. The observed surface tension and viscosity difference is not too surprising since these parameters were only slightly varied in the Weiss and Worsham study. The inappropriateness of the Weiss and Worsham expression is clearly evident by the results shown in Figure 3, because their expression predicts that under similar dissemination conditions Bis-2 should result in larger drop sizes than water, not smaller.

Tests also were conducted disseminating water and Bis-2 at temperatures up to around 100 °C, and the resultant MMDs measured are plotted in Figures 6 and 7. The wind velocities for the Bis-2 and water trials were 39 and 60 m/s, and the injection flow rates were 0.85 and 1.5 liters/min, respectively. The solid line in each figure is the predicted drop size using equation 1 and the temperature-dependent physical properties. The surface tension, density, and viscosity values used to derive the MMD temperature dependencies are listed in Table 2.

TABLE 2. Physcial Properties for Bis-2 and Water

	Temperature (°C)	Surface Tension (dyne/cm)	Viscosity (poise)	Density (g/cm^3)
Bis-2*				
	25	30.3	0.06	0.92
	45	28.5	0.039	0.91
	60	27.0	0.026	0.89
	150	20.0	0.011	0.81
	250	15.5	0.008	0.72
Water**				
	25	71	0.0096	1.0
	40	69	0.0065	1.0
	60	66	0.0047	0.98
	80	62	0.0036	0.97
	100	57	0.0029	0.90

*Provided by the U.S. Army Chemical Research, Development and Engineering Center (CRDEC), the Chemical Simulant Data Center.
**From Handbook of Chemistry and Physics.

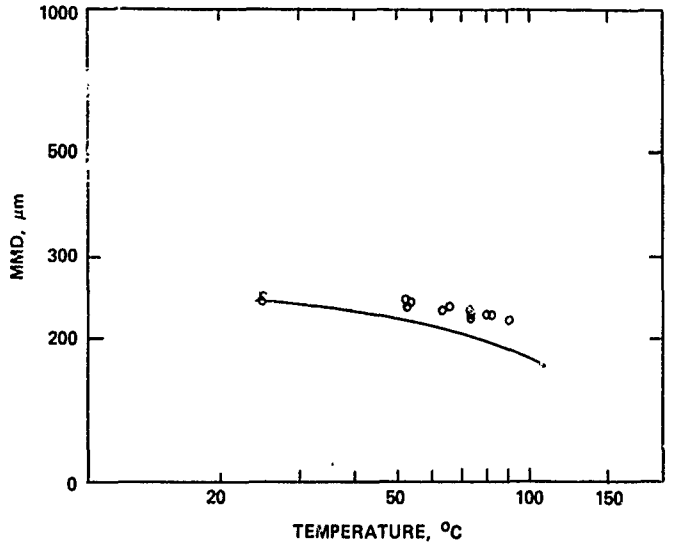


Figure 6. A ln-ln Plot of MMD versus Temperature for Water Trials at 60 m/s. Solid line is the predicted results of Equation 1.

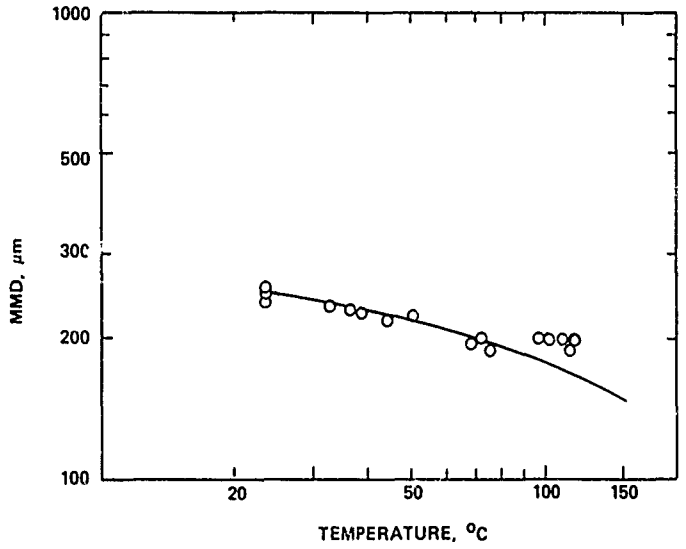


Figure 7. A ln-ln Plot of MMd Versus Temperature for Bis-2 Trials at 39 m/s. Solid Line is the predicted results of Equation 1.

For the heated trials it was necessary to eliminate the use of the innermost diode ring of the Malvern detector. The vapor from the evaporating drops refracted the incident laser beam and caused a portion of the beam to fall onto the innermost diode ring. This gave a false detector response for particle sizes between 872 and 1880 μm. Because for any ambient trial this bin size accounted for less than 1% of the total droplet mass, its elimination during the heated trials was considered to cause an insignificant effect on the results.

The MMD results for water versus temperature (Figure 6) demonstrate that significant cooling may occur during atomization. The measured MMD for the heated trials are larger than were predicted by using the physical fluid properties that correspond to the liquid temperature measured just prior to breakup. However, for Bis-2, whose vapor pressure is estimated about 10^{-5} times that of water, good agreement between prediction and experiment is observed for liquid dissemination temperature up to 80 °C (see Figure 7). This better agreement is expected because the vapor pressure for Bis-2 is much lower than the vapor pressure for water, and the heat of vaporization is an order of magnitude less. Thus, a much smaller cooling effect during atomization is anticipated. The deviation observed between prediction and experiment for liquid Bis-2 injection temperature greater than about 80 °C was not expected. Initially, it was thought that this deviation was caused by degradation of the Bis-2. However, a gas chromatographic (GC) analysis of the disseminated heated Bis-2 indicated no significant degradation. It was then considered that possibly the Bis-2 vapor was refracting a portion of the beam beyond the innermost diode and was being interpreted as aerosol diffracted light. If so, the measured MMD is larger than that of the actual aerosol, since the light refracted into the rings closest to the beam axis correspond to the largest particles. Although the innermost diode ring was eliminated, it was not possible to eliminate any additional rings without affecting the MMD measurement. The next closest ring to the beam axis corresponds to particle sizes within the expected range of the aerosol and thus, if eliminated, the measured MMDs would be biased towards smaller sizes. To substantiate this proposed effect it was decided to conduct additional heated Bis-2 trials at a higher wind speed, i.e. 60 m/s rather than 39 m/s. The velocity increase would reduce the particle size and shift the diffracted light distribution away from the inner rings which could then be eliminated without biasing the aerosol size measurement. Furthermore, with an increased airflow the vapor concentration is diluted with more air and the vapor effect itself is reduced.

Table 3 is a typical size distribution measured for an ambient temperature Bis-2 trial conducted at a 60 m/s wind speed. In this example no diode rings were eliminated and a MMD of 123.6 microns is measured. When the light scattered into the three innermost diode rings, corresponding to particles ranging from 1879 to 376 microns, is eliminated a MMD of 123.2

climate and also are strengthened by leaders and subordinates sharing of discomfort and sharing of training.

In combat, distinctions based on rank are blurred. Survival and victory depend on the intense cooperation of all ranks during combat. We must not make the fallacious assumption that an open climate is endemic exclusively to democratic societies. Even organizations which appear, at first glance, to be rigid and inflexible, such as the Wermacht in World War II[5] and the North Vietnamese Army[6], showed that in battlefield situation and in the small unit level criticisms and suggestions were a part of an open climate on the small unit level.

Even though British society is still somewhat stratified, British forces exhibited such an open climate. Argentine forces, with few exceptions, did not.

Recent U.S. Army Research Institute research indicates that an open command climate leads to high level of morale, cohesion and competence. However morale and competence are intertwined[7]. High performing groups have high morale and thus perform well. U.S. Army training should emphasize developing open command climates that produces free-wheeling criticism and, of course, good humor. Such an ambience requires patience and time.

Time

Time is an important factor for the development of cohesion. Military tradition grows out of years and years, if not centuries of military heritage, lore and myth. British soldiers and Argentine officers recount with pride their military traditions. Perhaps more than any army in the world, British soldiers and Ncos readily speak of their forebears and long-ago battles.

While the U.S. Army teaches military history to its officer cadets, there seems to be little emphasis on military history within the training cycles of the average soldier. There are recent attempt to resurrect traditions and historical lore, as for example with the 10th Mountain Division of the U.S. Army. The U.S. Army might model its training of history and military tradition by examining how British forces inculcate soldiers with a sense of military history. One factor in teaching soldiers their military tradition is time of service. British troops serve long tours which reduces personnel turbulence.

Close personal (horizontal and vertical) bonds and learning whom one can and cannot trust take months and probably years to come to full

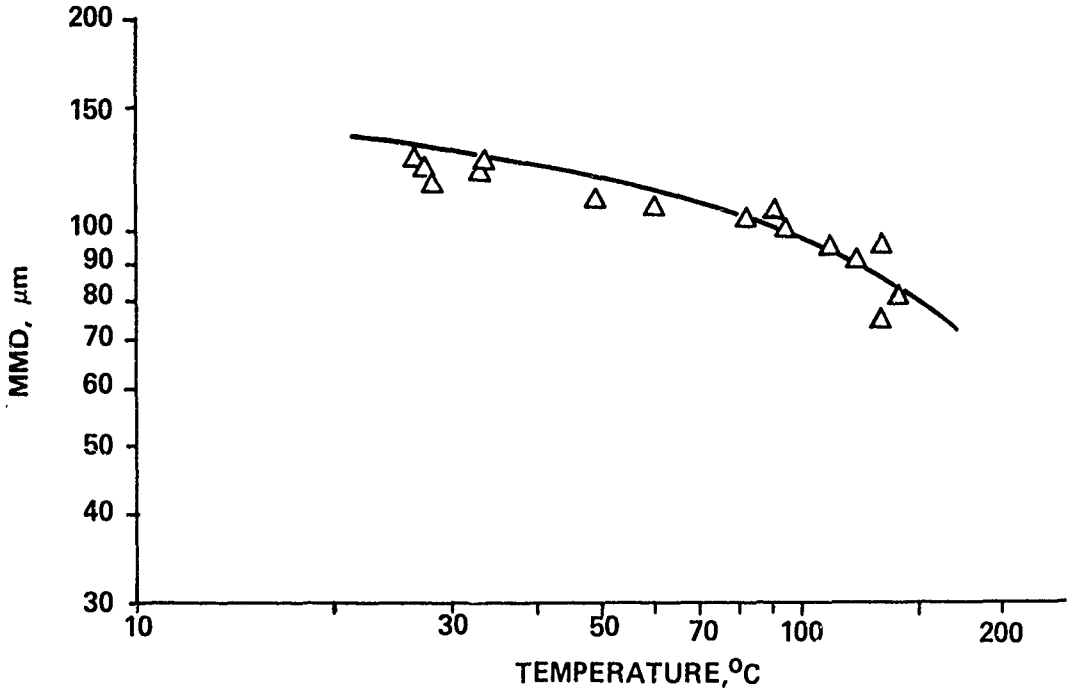


Figure 8. A ln-ln Plot of MMD Versus Temperature for Bis-2 Trials at 60 m/s. Solid line is the predicted results of Equation 1.

4. CONCLUSION

1. An empirical breakup expression was developed for dissemination modeling which can be used to predict the resultant drop size, i.e. MMD, of a Newtonian liquid atomized by a high velocity airstream.

2. The predictive expression developed is much improved over the often used Weiss and Worsham equation, since it was derived and experimentally verified over a broader range of physcial properties, and it indicates that,

   (a) Surface tension effects are much more significant in controlling drop size than previously expected, and

   (b) The viscosity effect is less important in controlling liquid breakup than reported by Weiss and Worsham.

3. With the developed predictive equation, the question of evaporative cooling was addressed. For heated water trials, evaporative cooling during atomization was found to occur and affect the resultant drop size. However, with heated Bis-2, a liquid with a latent heat of vaporization an order of magnitude less than water, evaporative cooling effects were not observed.

5. REFERENCES

(1) Weiss, M.A. and Worsham, C.H., Atomization in a High Velocity Airstream, ARS Journal 29, 252-259, 1959.

(2) Lapple, C.E., Henry, J.P., and Blake, D.E., Atomization - A Survey and Critique of the Literature. (AD 821314), SRI Technical Report No.6, April 1967.

The Exploitation of GPS to Satisfy Precise Platform Requirements

Joseph McGowan and *Kevin Schlosser
U. S. Army Aviation Systems Command
U. S. Army Avionics Research and Development Activity
Fort Monmouth, NJ 07703-5401

BACKGROUND

GPS is a space based radio navigation system which provides position, velocity, and time. The GPS system comprises three major segments: SPACE, CONTROL, and USER segment. The Space segment consists of a constellation of NAVSTAR satellites. Each satellite transmits specially coded signals that allow the range and rate of range change to the user to be measured. These fundemental measurements are known as pseudo-range and delta-range. As with all radio navigation systems, the geometric relationship between the transmitters and receiver antennas causes a Dilution Of Precision (DOP) in the navigation solution. Details on the operation of each segment and the effect of DOP can be found in reference 1.

Performance data collected in the past decade indicates that GPS readily meets the accepted specification of 15 meters (SEP) in position and 0.1 m/s (RMS per axis) in velocity. The fact that GPS performance often exceeds these levels led the technical staff of AVRADA to postulate that at a more fundamental level the system was performing well beyond the requirements of the GPS System Specifications. In order to further investigate this, a flight test was conducted with the objective of characterizing the quality of the GPS measurements, i.e. psuedo-range and delta range. The results of this test are documented in reference 2.

The test was conducted using a Rockwell/ Collins 5-channel receiver. The error in range measurement was found to be comprised of three significant components; a bias of 2 to 10 meters, a time varying term 3 to 5 meters in magnitude, and a random component with an RMS value of 0.5 to 1.0 meters. The delta range measurement was found to hold more promise for high accuracy applications. This is because its error contains no significant bias or systematic terms. When converted to its equivalent LOS velocity measurement, it was found to be comprised of a random component of about 1 cm/s (RMS). This level of error was consistent over

time and was independent of which satellite was being tracked[1].

The delta range random error far exceeds predictions based on carrier to noise ratio (C/N_0), and is consistent with the short term oscillator stability currently being used in DOD GPS User Equipment. The implications for high accuracy applications are:

1. Using standard GPS receiver components, delta range is a consistently high quality measurement.

2. The dominant error in delta range arises within the receiver itself, and seems to be directly related to the quality of the frequency standard.

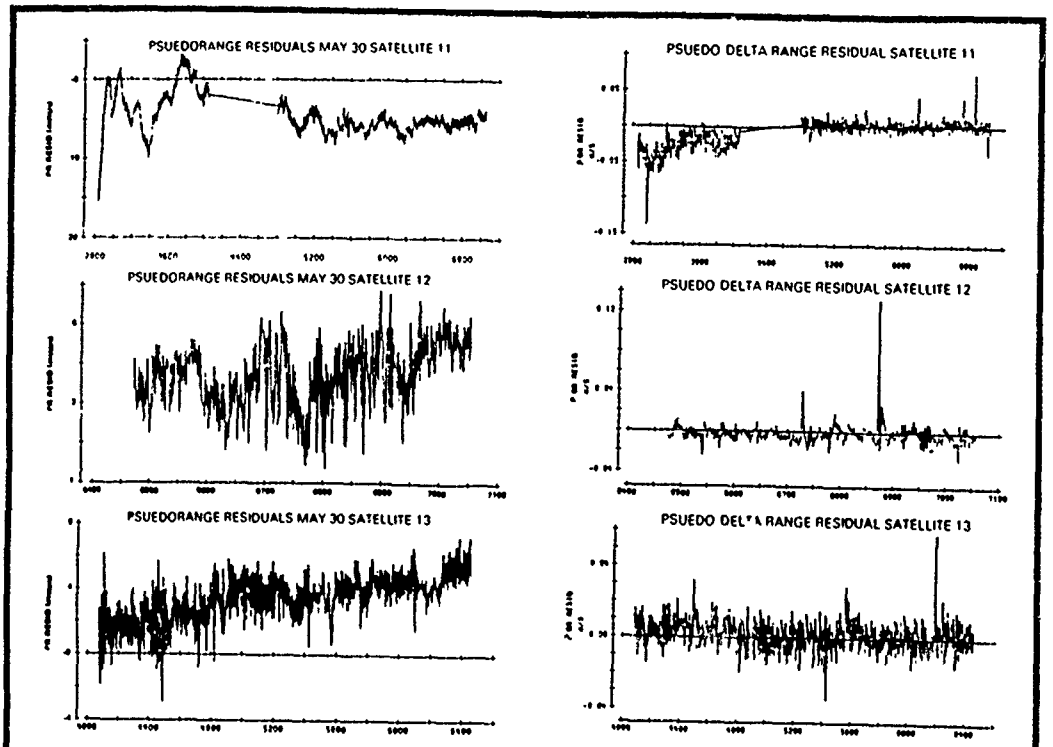


FIGURE 1. GPS MEASUREMENT ERRORS

1. This level of delta range error applies to periods of good constellation geometry. This is not to say that it isn't true at other times, just that the analytical technique was not as valid during poor geometry. See reference 2 for details.

Figure 1 shows the range and delta range errors for the 5 channel receiver.

Receiver measurements for a 2-channel receiver would differ somewhat from those described above. In these receivers, the hardware channels are time shared among the satellites and measurements are available roughly half as often. The range measurements would be of nearly equal quality. The delta range errors would differ appreciably in quality. This is because of the shorter dwell time on each satellite. In the 5-channel receiver, delta range is accumulated over a period of 780 msec while in the 2-channel this interval is 160 msec. From theoretical considerations, the difference in quality can be expected to be proportional to the square root of the ratio of these two intervals, or 2.2. In other words, the delta range error can be expected to be zero mean with an RMS value of 2.2 cm/s. This level of error has been observed in recent tests.

Hover Hold

The Army's next generation of light attack and utility rotorcraft (eg. Light Helicopter, Experimental - LHX) will be required to maintain an automatic, stabilized hover in and out of ground effect, in visual and instrument conditions, and in winds up to a maximum of 25 knots with a gust spread up to 15 knots. The position accuracy requirements, as shown in Table 1, are stringent. The horizontal and vertical hover hold accuracy requirements are the root-sum-square of the respective allocated hover sensor and flight control accuracies. Seventy percent of the error is allocated to the hover sensor system with the

	TOTAL	SENSOR	FLIGHT CONTROL
Max RMS Horizontal Error (meters)	1.0	0.84	0.55
Max RMS Vertical Error (meters)	0.3	0.25	0.17

TABLE 1. HELICOPTER AUTOMATIC HOVER HOLD REQUIREMENT

remaining thirty percent allocated to the flight control system.

The most appealing method for meeting the hover position sensor requirement is to utilize the LHX baseline navigation sensor suite to perform the hover position sensing function. This concept holds several advantages over a dedicted hover sensor. Most notable is the savings in space, weight, and power which accrue by utilizing sensors that are basic to future Army rotorcraft navigation suites.

Meeting the horizontal position measurement accuracy requirement poses the most risk to the successful achievement of the hover sensor requirements and necessitates use of an externally referenced sensor. Self-contained sensors available on Army aircraft (doppler and inertial) accumulate errors significantly larger than the total error budget for the hover positioning system. GPS provides the external reference that, when integrated with data from an inertial sensor, appears to meet the hover position sensing accuracy requirement. The hover position sensing system currently being tested by the Army consists of a 5-channel GPS receiver combined with an inertial navigation sensor (Figure 2).

Several modifications to the standard GPS/inertial integration technique are necessary to achieve the required horizontal hover position accuracy. A Kalman filter must be developed for operating in the low

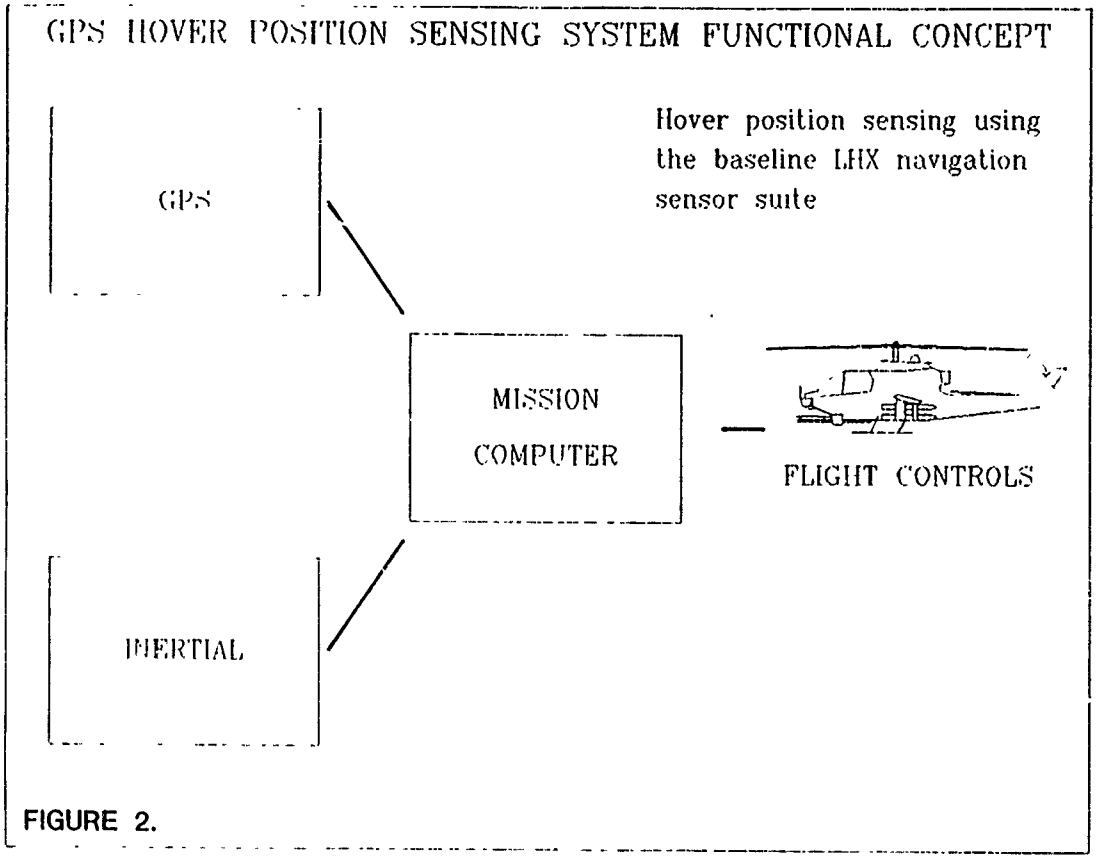


FIGURE 2.

dynamic environment characteristic of a helicopter at hover. In addition, the hover position sensing system must have the capability to lock-on to

a satellite constellation for the entire hover duration. Since the hover position sensing system measures position relative to an initial nominal point, the system must be inhibited from switching constellations during hover and creating a large bias error that was not present at the start of the hover. Figure 3 illustrates the effects of a satellite constellation switch during hover.

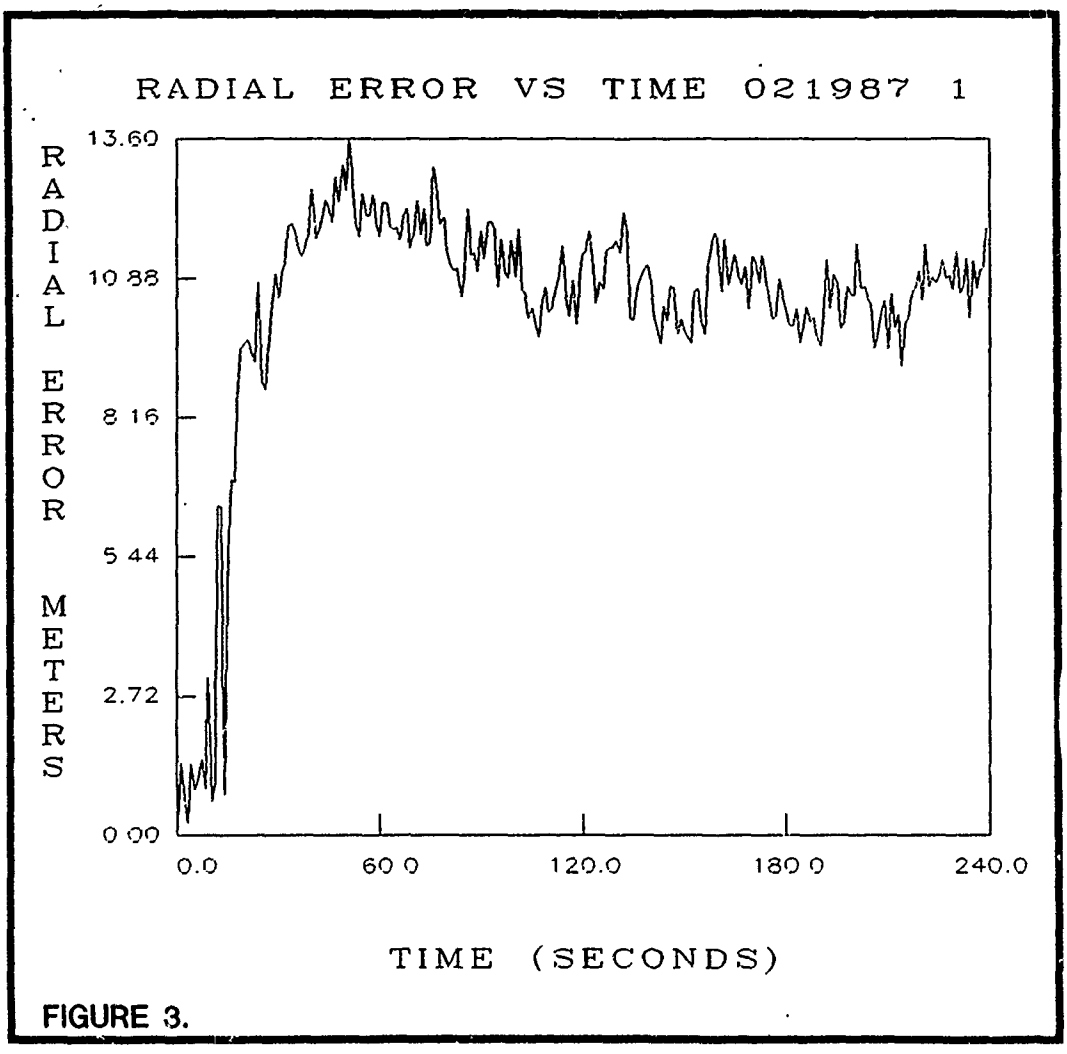


FIGURE 3.

Specific GPS error sources are classified as bias or random error sources. In general, only random errors will produce errors in the hover sensor application of GPS. Bias errors do not need to be addressed. As long as the hover filter does not attempt to eliminate a bias error, it

will be present with the same magnitude throughout the hover duration and is not involved with the relative solution.

The error sources with the greatest impact on hover sensor accuracy include GPS receiver measurement noise, mechanization errors, and quantization errors. These errors combine to form a total pseudorange random error on the order of 1.44 meters (evaluated at C/No = 30db for P-Code), and a total delta-range random error of .10 meters. The pseudorange error is a major contributor to hover position error and must be reduced through system mechanization improvements and inertial velocity information aiding.

The other major error which must be addressed is error due to the propagation gradient. The propagation gradient is the change or drift in ionospheric delay over time. Ionospheric delay is the delay of the GPS signal as it transmitted from the satellite to the user equipment through the ionosphere resulting in a pseudorange bias error. Ionospheric delay drift can be quite dynamic and can result in changes of pseudorange bias error as high as .46 meters per minute given a horizontal DOP of unity. For horizontal DOP greater than 1.0, this error is magnified. References 3 and 4 provide detailed discussion of this topic. The solution to ionospheric delay drift is to estimate the ionospheric delay error just prior to entering hover mode and then to monitor the ionospheric delay error throughout the hover and formulate corrections as necessary.

Initial hover flight tests were conducted in February 1987 at the Army Airborne Electronic Research Activity, Lakehurst, N.J. The resulting radial horizontal position error of the 5-channel unaided GPS receiver was on the order of 1.00 meter RMS for a three minute hover . The horizontal DOP during the hovers varied between 3.0 and 4.0. The carrier to noise ratio ranged between 40 and 45 db throughout the flight. Preliminary results combining the unaided GPS position data with inertial data indicate that accuracies on the order of .5 meters RMS are achievable for a three minute hover. Simulations show that the position accuracy is relatively insensitve to GPS pseudorange process noise estimates.

Conclusions for Hover Hold

The data collected from this initial flight test support the premise that the desired hover position sensor accuracy can be achieved using an integrated GPS/inertial system with modifications to the standard GPS navigation algorithms. The data indicates that modification of the GPS satellite selection algorithm to lock on to a constellation for the entire duration of the hover is essential. Error growth due to ionospheric drift was evident over the three to four minute hover duration. Ionospheric drift error correction will be investigated in the second phase of this program.

High Accuracy Velocity

The performance of certain advanced surveillance systems is greatly influenced by the accuracy of the velocity information available to it. It has been shown that there is a threshold value of accuracy above which overall system performance degrades rapidly, and below which little performance gain is achieved. It is encumbent on the system designer to identify this threshold so that the navigation system is neither under-specified, resulting in sub-optimal performance, nor over-specified, resulting in excessive cost.

Typical navigation accuracy requirements are given below in Table 2.

POSITION (m)	
HORIZONTAL	8.0
VERTICAL	40.0
VELOCITY (m/s)	
HORIZONTAL	0.02
VERTICAL	0.10

NAVSTAR GEOMETRIC CONDITIONS

HDOP: 1.58 VDOP: 2.71 TDOP: 1.68

TABLE 2. ACCURACY REQUIREMENT (RMS per axis)

In addition to the accuracy requirements reflected in the above table, a navigation solution is required at a rate of 5 Hz. These requirements indicate the need for a navigation suite comprised of an Inertial Measurement Unit (IMU) and an external reference to bound the IMU measurement growth.

The most stringent of the accuracy requirements is the horizontal velocity, where 0.02 m/s (RMS per axis) is required. There is a body of analytical evidence (Reference 2) that indicates that a good quality IMU coupled with a 2 channel Global Positioning System (GPS) receiver will provide the needed accuracies and data rates.

A block diagram of the primary navigation system is shown in Figure 4. The Inertial sytem is a Delco Carousel IVe, modified for this application to include a vertical channel output, a Rockwell/ Collins 2-

channel GPS receiver, the ASN-149 v3, and a barometric altimeter used to stabilize the IMU vertical channel and to aid the GPS receiver in the case of less than four visible satellites. There are also backup navigation modes that use TACAN or Data Link updates in the absence of GPS. These modes are not relavent to this paper and are not shown or discussed.

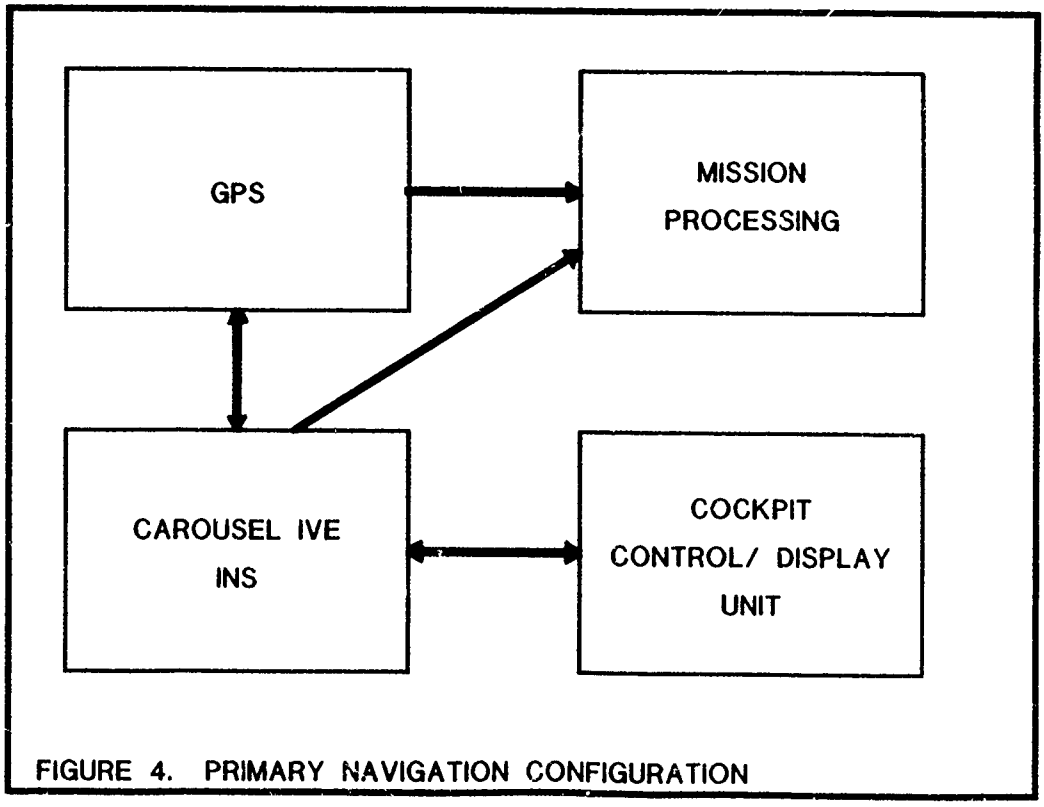


FIGURE 4. PRIMARY NAVIGATION CONFIGURATION

Flight tests were conducted to determine whether this configuration was meeting the required performance. Performance was scored against data obtained from the Aerial Profiling of Terrain System (APTS). APTS is an airborne surveying platform developed by The Charles Stark Draper Laboratories under the sponsorship of the U. S. Geological Survey. References 5 and 6 provide descriptions of APTS. It is sufficient here to state that it provides 1-sigma accuracies of 60 cm and 3 - 5 mm/s per axis in position and velocity, respectively.

Three hours of data was collected in two test flights . Data reduction showed that the real-time navigation solution was not meeting the requirement in horizontal velocity. Further analysis was conducted to

determine the modifications required to bring the system into compliance with the requirement. The analytical technique used in the 1985 tests was used to characterize the GPS measurements. The errors in the measurements were found to be consistent with the description given above in the Background section.

A GPS navigation solution that is not filtered in any way follows a simple prediction model, DOP times the measurement error. A filtered solution should always be at least this good. When the real time output was compared to this model, it was discovered that during times when DOP was low (3 - 5) the real time output was frequently worse than the model. It was evident that system performance could be improved by modifying the integration filter.

The filter can be modified in two ways. It can be reformulated based on different or enhanced models, or the original formulation can be retained and the filter re-tuned. The second method is by far the more desirable, as it entails no software re-design and is easily retrofitted into existing receivers.

The data recorded during the flights included the raw measurements of both the IMU and GPS. The approach used in re-tuning the filter was to develop an emulation of the real-time filter and to drive the filter with the recorded measurements.

The emulator was debugged and validated by comparing its output to the real-time navigation solution obtained in flight. The emulator provides a nearly identical time history[2], and more importantly, it provides an exact match for the statistical behavior of the real-time filter. With the emulator validated, a series of filter tuning iterations were run.

The integration filter is resident in the GPS Receiver/ Processor Unit (RPU) and is an error state mechanization. That is, the filter attempts to estimate the errors being committed by the IMU based on the psuedo-range and delta-range measurements made by the receiver.

The filter embodies a mathematical model of the way IMU errors propagate in time. The output from this model is used to predict the value of the next set of GPS measurements. When measurements arrive, a residual if formed by differencing the predicted and actual measurements. The IMU error estimate is then updated based on this residual. The update follows an equation of the form:

---

2. There are valid technical reasons for not obtaining an exact match. Reference 8 provides details.

$$X_{t+} = X_{t-} + (K) \times (\text{Residual})$$

where:

X_t - the error state vector (- and + indicate just before and just after a measurement, respectively)

K - a weighting factor between 0 and 1 (Kalman Gain)

In tuning the filter, the parameters which affect the value of the Kalman Gain are varied. These parameters can be lumped into two groups: those that reflect the fidelity of the model to the real world and those that reflect the quality of the measurements. The former group are commonly called "Plant Noise" or "Process Noise" and the latter are called "Measurement Noise".

In re-tuning this filter, great advantage was taken of apriori knowledge. Flight test data showed that the IMU velocity errors in the horizontal are well behaved. The model for IMU horizontal velocity error is therefor faithful, and a low plant noise value was used. We found the plant noise for vertical velocity in the original tuning to be optimistic, and it was raised somewhat. The behavior of the GPS oscillator frequency was found to be better than could be inferred from the original tuning, and the plant noise for this part of the model was modestly reduced. The benign dynamics of the aircraft mission profile also allowed a reduction in the g-sensitive terms of the oscillator error model.

The most significant tuning change made was in the delta-range measurement noise. Characterizations of this measurement from the flight data and the 1985 work showed it to be very precise. In fact, it is about 1000 times better than would be inferred from the original tuning. To be conservative, it was lowered by a factor of 100.

		BEFORE	AFTER
POSITION	N	3.2	2.5
	E	1.8	3.9
(meters)	D	5.4	2.1
VELOCITY	N	33	12
	E	107	15
(mm/s)	D	226	39

N - North E - East D - Down

TABLE 3. NAVIGATION ERRORS (RMS)

The before and after navigation errors for one segment of a flight

are summarized in Table 3.

Conclusions for Velocity Accuracy

The result of the re-tuning effort was a navigation system that fully complies with the position and velocity accuracy requirements. The improvement in performance was achieved with no additional hardware investment. Implementation of the revised tuning requires only the change of a table of values that is accessed by the RPU software.

There was another interesting effect of the revised tuning. Originally, the filter was not making much use of the delta range measurement. The bias shift in psuedo-range accompanying a satellite switch was therefor coupling into the velocity output in the form of a transient as high as 0.3 m/s. With the revised tuning, this effect was significantly reduced.

Overall Conclusions

GPS holds the promise to revolutionize navigation. The performance specifications for the general class of military users were designed to be applicable under a broad range of environmental and dynamic conditions. These performance levels will satisfy a great majority of user requirements.

Precise GPS fundemental measurements can provide much greater performance in special cases through integration with other navigation sensors and by tailoring the Kalman Filter design to the specific operational environment and sensor quality.

There is investigative work being done in other areas as well. Variations on the differential GPS concept are being utilized to provide reference trajectories in test environments. The high quality GPS measurements can be used to advantage for in-air alignment of inertial systems in short-notice takeoff situations. The high quality oscillators carried by the GPS satellites provide excellent time synchronization capability, and traceability to UTC allows global time transfer as well. Measurements of phase difference of the GPS carrier frequency made on two physically separate antennas can provide both baseline length and pointing information.

The underlying quality of GPS fundemental measurement data can be exploited in many circumstances where technical feasibility and/ or prohibitive cost have been a major constraint. This paper has described only two of the unforseeably many such applications.

REFERENCES

1. NAVSTAR GPS Joint Program Office, "Introduction to NAVSTAR GPS User Equipment", June 1987.

2. Greenspan, R. L., Donna, J. I., "APTS/ GPS Measurement Task Final Technical Report", CSDL-R-1845, February 1986.

3. Jorgenson P., "Ionospheric Measurements from NAVSTAR Satellites", presented at the ION Conference, Springfield, VA, 1979.

4. Martin, E. H., "GPS User Equipment Error Models", presented at the 33rd Meeting of the Institute of Navigation, 22-24 June 1977, Costa Mesa, CA.

5. Soltz, J. A., Mamon, G., Chapman, W., "Aerial Profiling of Terrain System Implementation", presented at Second International Symposium on Inertial Technology for Surveying and Geodesy, Banff, Alberta, Canada, June 1981

6. Hursh, John W., "Aerial Profiling of Terrain System (APTS), Journal of the Institute of Navigation, Volume 32, No. 3, Fall 1985

7. Donna, J., Greenspan, R. L., "GPS RCVR C4/ Carousel IV E Integrated Navigation System Flight Test Program Final Report", CSDL-R-1988, May 1987

8. Soltz, J. A., Donna, J. I., Greenspan, R. L., "GPS RCVR C4/ Carousel IV E Navigation Filter Tuning Refinement Final Report", CSDL-R-2016, November 1987

Niacinamide Prevents Sulfur Mustard (HD)-induced Pathology and Biochemical Changes in Human Lymphocyte and Mixed Human Leukocyte Preparations. (U)

*Henry L. Meier, Dr., John P. Petrali, Dr. and Clark L. Gross, Mr
United States Army Medical Research Institute of Chemical Defense
Aberdeen Proving Ground, MD 21010-5425

INTRODUCTION

Sulfur mustard gas (2,2'-dichlorodiethyl sulfide, HD) is a vesicant chemical warfare agent which produces incapacitating injuries to skin, eyes, and lungs (1-4). The use of HD in World War I had a devastating effect on exposed troops. In modern warfare, HD has again proven to be a potent tactical weapon as shown in the Iraq-Iran conflict. Despite 60 years of intensive research since HD was introduced as a weapon with massive incapacitating capabilities, the mechanism of HD-induced injury is only now being understood. We are attempting to develop an improved antidote, but until now there was no effective method for evaluating pharmacologic intervention in the prevention of HD injury to human tissue.

This paper discusses two in vitro models, the human leukocyte model (5) and the new human lymphocyte model, which will enable us to confirm the validity of our unique biochemical hypothesis for the HD-induced injury. Our hypothesis links DNA damage, alterations of metabolism, and pathology (5). HD extensively alkylates DNA in all cell types leading to considerable depurination of the DNA strands (6-12). Repair of apurinic sites in this DNA (13) results in activation of the chromosomal enzyme, poly (ADP-ribose) polymerase (14). Poly (ADP-ribose) polymerase uses NAD^+ as a substrate and polymerizes the ADP-ribose moiety from NAD^+ on different protein receptors (14). This polymerization would result in the depletion of the cofactor, NAD^+, which would cause an inhibition of glycolysis (14), activation of the hexose monophosphate shunt (15), secretion of active proteolytic enzymes (16), breakage of the anchoring filaments of basement membrane, and ultimately, blister formation (17).

The differences between humans and non-primate animal models in the development of the HD lesion include their cellular biochemistry (especially in synthesis of NAD^+)(18), proteases and inhibitors (19), and mechanisms of inflammation (20). Appropriate cell and tissue models which would exhibit typical human responses to HD and validate our biochemical hypothesis should also be useful in testing potential HD antidotes. Based on clinical evidence, the appropriate place for therapeutic intervention

would be during the latent period before overt pathology takes place. Therefore, our human tissue models were designed to measure relevant biochemical processes which are altered during the first 4-6 hours following HD exposure and correlating those changes with the development of pathology at 20-48 hours.

We selected human leukocytes (5,21) and human lymphocytes to study the effects of HD exposure. Human leukocytes and lymphocytes were employed to investigate changes in (1) biochemical parameters of the HD-induced injury and (2) changes in cellular viability. These cells are also easy to obtain, handle, and aliquot (multiple experiments can be performed from a single bleed from a single volunteer), as well as relevant to the type of injury. Clinical data from the Iran-Iraq war indicate that human leukocytes and lymphocytes appear to be primary targets in HD-exposed soldiers. Leukocytes allowed the study of HD on diverse cell types that can be made to behave similarly to both the basal cells of the epidermis and the inflammatory cells of the dermis.

In this paper, our models are described and discussed to demonstrate the following: [1] validation of part of our biochemical hypothesis, [2] their ability to correlate HD damage and biochemical, pathological, and morphological changes, and [3] the effectiveness of possible therapeutic regimens at preventing the development of HD-induced pathology. These models are the first in vitro tests developed where the effect of HD on human cells can be determined. Based on the results of these tests, we have been able to design therapeutic regimens and evaluate their potential in preventing HD-induced injury.

The mixed human leukocyte model shows that there is a HD-concentration-dependent decrease in NAD^+ levels which can be prevented by two types of compounds: NAD^+ synthesis substrates and inhibitors of poly (ADP-ribose) polymerase (5, 21). The human lymphocyte model is the first in vitro model where HD-induced pathology can be studied at a cellular level. Using this model, we can demonstrate that HD induces a concentration-dependent decrease in cellular viability at concentrations consistent with human exposure levels. The mechanism of lethality can be partially explained by membrane damage evident at the electron microscopic level. We are the first investigators to pharmacologically intervene and prevent HD-induced cellular pathology.

METHODS:

Human leukocyte model:

Blood from normal human volunteers was drawn and sedimented on dextran-EDTA mixture with dextrose added. Leukocytes were isolated and the contaminating red cells were lysed (22). Leukocytes were then resuspended in Tyrode's buffer or buffer containing the indicated drug, exposed to HD, and incubated at 37°C for the specified times.

Isolation of human mononuclear lymphocyte populations and polymorphonuclear granulocyte populations:

One hundred ml of human blood was diluted with saline (10 ml blood/25 ml saline) and centrifuged at 400g on a Percoll (Pharmacia Fine Chemicals, Piscataway, N.J.) cushion (1 ml cushion/1 ml blood; density = 1.080 g/ml) for 30 minutes at 20°C. The interface containing the mononuclear cells (about 90% lymphocytes) was removed and washed twice with Tyrode's buffer. The cells were resuspended and added to incubation mixtures containing various concentrations of mustard and/or antidotes. The red cell pellet under the Percoll cushion was lysed twice with cold distilled water for 30 seconds. The resulting polymorphonuclear leukocytes (99% granulocytes) were washed twice in Tyrode's buffer before they were added to reaction mixtures (22).

Assay of NAD^+/NADH:

NAD^+/NADH levels in the neutralized extracts were measured by the enzymatic cycling assay of Jacobson and Jacobson (23). NAD^+ levels were determined by measuring at least 3 separate samples (in duplicate) and the standard deviations were calculated and represented as error bars in the figures. Each experiment was performed in triplicate before being reported.

Exposure of human lymphocytes to HD for viability assay:

The lymphocytes were resuspended in the appropriate volume of RPMI 1640 medium, and aliquotted into 24 or 96-well tissue culture plates (Linbro®) on ice, containing RPMI 1640 medium, and the appropriate concentrations of niacinamide. The HD was freshly prepared in ice-cold RPMI 1640 medium (to decrease hydrolysis) and added directly to the wells. The plates remained in the hood for 1 hour at room temperature to insure safety from any HD vapor while the HD reacted.

Incubation and recovery of lymphocytes:

The tissue culture plates were transferred to a controlled environmental incubator where they remained for 16-48 hours at 37°C and 5% CO_2. Cell number and viability were determined by 2 methods: (1) a hemacytometer using trypan blue dye exclusion, and (2) an Epics C Coulter laser-based cell sorter using propidium iodide dye exclusion. In both methods, the mean viability was determined by counting at least 3 separate wells and the standard deviation was calculated and represented as error

bars in the figures. Each experiment was performed in triplicate before being reported.

Electron Microscopy:

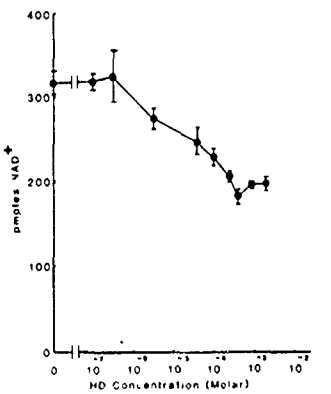
Samples of lymphocyte incubations were concentrated by centrifugation for 10 min at 300g. The resultant cell pellet was fixed for 1 hour by immersion at 20°C in a combined aldehyde fixative of 1.6% formaldehyde and 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer at pH 7.34 and 192 mOsm with respect to the buffer. Following primary fixation, the packed cells were washed 3 times for 15 min each in cacodylate buffer, post-fixed for 1 hour in buffered 1% osmium tetroxide, dehydrated in graded ethanols, recentrifuged to minimize losses and embedded in epoxy resin. Replicate thin-sections, differentiated with uranyl acetate and lead citrate, were analyzed for ultrastructural effects.

RESULTS:

HD concentration-dependent decrease in NAD^+ levels:

Figure 1 shows the concentration-dependent decrease of NAD^+ levels in HD-treated cells. NAD^+ levels began to decrease at concentrations greater than 10^{-5} M HD and at 10^{-4} M HD, dropped to 66% of control values and remained at this level even at higher concentrations of HD. The decrease in NAD^+ could not be explained by reduction of NAD^+ to NADH since the levels of NADH also decreased (approximately 25% decrease) over the same range. The possibility that the drop in total NAD^+ could be the result of cell lysis was rendered unlikely by an inability to demonstrate the presence of measurable levels of lactic dehydrogenase (a cytoplasmic enzyme), NAD^+ (mainly a nuclear constituent), and histamine (a granular component of human basophils) in the cell-free supernatant (21).

Figure 1. The HD concentration dependent decrease in NAD^+ levels in human leukocytes. The NAD^+ begins to decrease at 3×10^{-6} M and drops to 66% of control values at 10^{-4} M HD.



The time course of the decrease in NAD^+ levels in HD-exposed cells:

Leukocytes were exposed to 10^{-3} M HD and incubated at 37°C for various times (0, 15, 30, 60, 120, and 240 minutes) before being centrifuged, decanted, acidified, and assayed for NAD^+. The level of NAD^+

in control leukocytes steadily increased during 4 hours of incubation, indicating that there was a sufficiently large pool of substrates for continued NAD^+ synthesis. However, cells exposed to 10^{-3} M HD were able to maintain the increase in NAD^+ levels only during the first hour (Figure 2). During the ensuing 3 hours of incubation, these HD-exposed leukocytes demonstrated a consumption of NAD^+ when compared to the NAD^+ content of control leukocytes (Figure 2) (21).

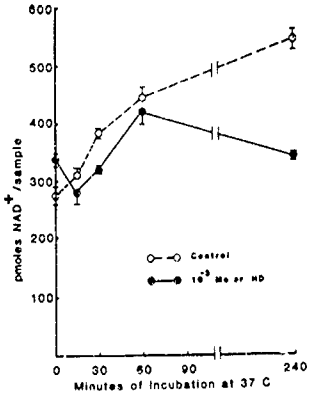


Figure 2. The time course of NAD^+ level decrease in mixed human leukocytes induced by HD exposure. There is no difference in NAD^+ levels of control cells and HD-exposed cells during the first hour of incubation. By the fourth hour of incubation, the NAD^+ content of the HD-exposed cells is 66% of control NAD^+ values.

Niacinamide effects on HD-suppresion of NAD^+ levels:

Niacinamide, an inhibitor of poly (ADP-ribose) polymerase and a precursor of NAD^+ synthesis, was chosen to study its effect on NAD^+ levels in HD-treated leukocytes. Leukocytes were incubated in various combinations of buffer, HD, and niacinamide. Niacinamide added 10 minutes before HD treatment both increased the level of NAD^+ at 4 hours in the control cells and blocked the HD-induced fall of NAD^+ levels (Figure 3). Maximal effects were observed at 10^{-4} M niacinamide, which caused a 196% increase in the amount of NAD^+ compared to HD-treated cells incubated in the absence of niacinamide. Niacinamide caused a "superinduction" of NAD^+ synthesis in the HD-treated cells (278% and 370% increases above those seen in untreated and HD-treated cells incubated in the absence of niacinamide respectively) (21).

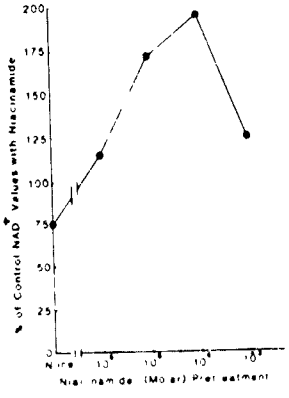


Figure 3. The effect of niacinamide on NAD^+ levels. The NAD^+ levels of HD- exposed leukocytes are compared to those of control cells. Niacinamide is added to each set of cells 10 minutes before HD or buffer treatment. HD appears to cause an increase in niacinamide stimulated NAD^+ synthesis, while niacinamide inhibits the consumption of NAD^+ by poly (ADP-ribose) polymerase. Niacinamide appears to have its maximum effect at 10^{-4} M and HD concentration was 10^{-3} M.

The kinetics of niacinamide protection:

The relationship between the time of niacinamide administration on its ability to prevent the HD-induced NAD^+ decrease was investigated. Cells were incubated with niacinamide for various times, both before and after exposure to HD (Figure 4). All groups of cells were incubated at 37°C for a total of 5 hours. After the first hour, 10^{-3} M HD was added to the appropriate cells. Ten ul of 10^{-2} M niacinamide in Tyrode's buffer was added at the indicated times to give a final concentration of 10^{-4} M. Niacinamide caused more than a doubling of NAD^+ levels in treated cells even when added 2 hours after incubation started. When the cells were pre-incubated for 1 hour with niacinamide, HD-treated cells demonstrated even higher levels of NAD^+ ("superinduction") which were approximately two-fold above that seen in niacinamide-incubated control cells. The "superinduction" in NAD^+ levels in HD-treated leukocytes (about 200% of untreated cells incubated in the presence of niacinamide) occurred even when niacinamide was added 2 hours after HD exposure (21).

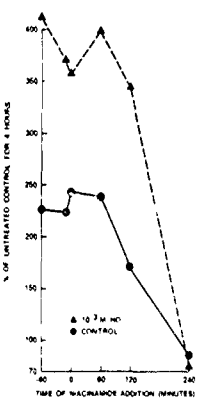


Figure 4. The time course of the ability of niacinamide to protect NAD^+ levels of the cells. Leukocytes were exposed to 10^{-3} M HD sixty minutes after the incubation was started (time zero). Niacinamide (10^{-3}M) was added at the indicated times during the 300 minute incubation. Niacinamide protected the NAD^+ levels when added at -60, 0, 30, 60, and 120 minutes.

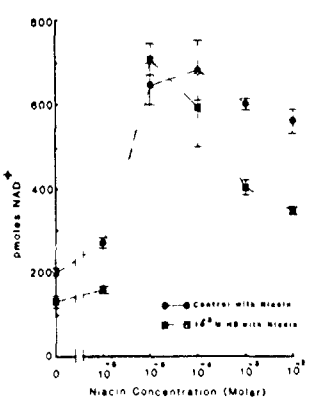


Figure 5. Protection of leukocyte NAD^+ levels by niacin. The NAD^+ levels of HD-exposed leukocytes were compared to NAD^+ levels of control cells when niacin was added to each set of cells 10 minutes before HD or buffer exposure. Niacin (a substrate for NAD^+ synthesis) causes a biphasic increase in NAD^+ levels which reaches a maximum value between 10^{-5} and 10^{-4} M niacin. In control lymphocytes, concentrations greater than 10^{-4} M niacin caused a gradual decrease in NAD^+ levels. Higher concentrations of niacin caused a precipitous decline in NAD^+ levels in HD-exposed cells.

The effect of niacin (nicotinic acid) on the NAD^+ level in leukocytes:

The ability of niacin, a substrate for NAD^+ synthesis but not an inhibitor of poly (ADP-ribose) polymerase, to affect the HD-induced NAD^+ depletion was investigated. Leukocytes from normal individuals were pre-incubated in buffer with niacin for 10 minutes before exposure to 10^{-3} M HD and their NAD^+ levels determined after a 4-hour incubation (Figure 5). Niacin caused a concentration-dependent increase in NAD^+ synthesis and prevented the loss of NAD^+ due to HD exposure. Maximum levels of niacin-induced synthesis and protection occurred at 10^{-5} M, with both untreated and HD-treated cells showing NAD^+ levels from 250-370% higher than those found in the absence of niacin (21).

The effect of HD on the different subpopulations of leukocytes:

The relationship of the HD-induced decrease in NAD^+ levels to poly (ADP-ribose) polymerase was further explored by dividing the leukocytes into polymorphonuclear granulocytes (a population of cells which do not contain poly (ADP-ribose) polymerase) and mononuclear lymphocytes (a population of cells in which most of the cells contain poly (ADP-ribose) polymerase). As shown in Figure 6, there is a HD-concentration dependent decrease in the NAD^+ levels of lymphocytes exposed to HD. The decrease in NAD^+ levels of the lymphocytes begins at HD concentrations as low as 10^{-5} M. However, there is no significant change in granulocyte NAD^+ levels even when they are exposed to concentrations of HD as high as $3X10^{-4}$ M (21).

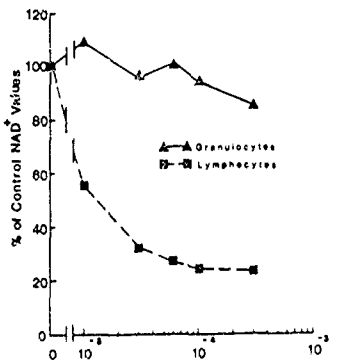


Figure 6. Comparison of the effects of HD on NAD^+ levels in human lymphocytes and granulocytes. Human lymphocytes and granulocytes were isolated from 100 mls of blood taken in a single bleed and separated on Percoll gradients. The cell numbers were adjusted to a constant value and exposed to the indicated concentrations of HD. NAD^+ levels in the lymphocytes decreased in a HD-concentration dependent manner while granulocytes showed almost no effect of HD on their NAD^+ levels.

The effect of HD on viability of lymphocytes:

A model was needed to study HD-induced pathology in human cells to determine whether prevention of NAD^+ level decreases initiated by HD exposure would influence the later development of the lesion. Using both trypan blue and propidium iodide exclusion techniques, no cell death was observed at 8 hours. The cell preparations were incubated with varying concentrations of HD (10^{-3} M to 10^{-5} M) for 24 hours. As shown in Figure 7, there was a steep dose-dependent decrease in cell viability after 24 hours. The cell viability began to show a decrease when the lymphocytes

were exposed to concentrations of HD at $3x10^{-5}$M. There appears to be an almost linear decrease in cell viability as the concentration of HD increases to 10^{-3}M.

Cells were isolated from 10 normal volunteers and exposed to 10^{-3} M HD to determine whether the effect of HD on cellular viability was dependent on individual variability. HD induced a 40 to 60% decrease in cellular viability in every individual studied. Each experiment was conducted with cells from a given individual and treated independently from other experiments.

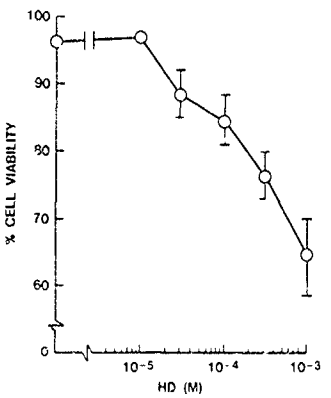


Figure 7. The effect of HD on cell viability. Human lymphocytes demonstrate an almost linear relationship between HD concentration and cell death. At 10^{-3} M HD, there was a 35% decrease in cellular viability as assayed by trypan blue exclusion 20 hours after exposure to HD.

Effect of the prophylactic use of niacin and niacinamide on the viability of cells treated with HD.

Niacin (a substrate for NAD^+ synthesis) and niacinamide (both a substrate for NAD^+ synthesis and an inhibitor of poly (ADP-ribose) polymerase) were shown to block HD-induced decreases of NAD^+ levels in human lymphocytes. They were studied to determine whether they could also prevent HD-initiated pathology as measured by cellular viability. The data in Figure 8 demonstrates that niacin, in spite of its ability to protect leukocyte NAD^+ levels in 4 hour cultures, failed to prevent the HD-induced decrease in lymphocyte viability. Niacinamide (Figure 8) not only prevented the loss in lymphocyte NAD^+ levels but also showed a concentration-dependent protection against HD-induced lymphocyte cell death.

Effect of niacinamide as a therapeutic treatment:

Since niacinamide was effective at protecting lymphocytes prophylactically, the effect of addition to HD-treated cells after exposure was also investigated. Niacinamide ($3x10^{-4}$M) was added to the incubation mixtures at various times after HD exposure. Figure 9 shows that niacinamide was as effective at protecting lymphocytes from HD-induced cell death when added 4 hours after exposure as when added simultaneously with HD. If the niacinamide was not added until 20 hours after HD exposure, no protection of viability was detected.

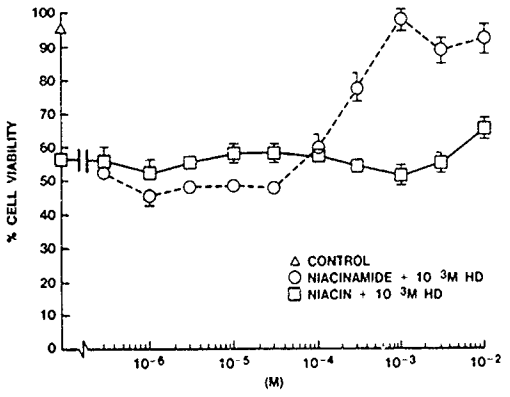


Figure 8. Niacinamide and niacin were studied to determine whether they were capable of preventing cellular toxicity by HD. Human lymphocytes were incubated with 10^{-3} M HD in the presence of the indicated concentrations of either niacin or niacinamide. Niacinamide began protecting against HD-induced cytotoxicity at concentrations as low as $3x10^{-4}$ M and completely protected the cells at 10^{-3}M. However, niacin conferred only marginal protection at its highest concentration, 10^{-2} M. Viability was measured by propidium iodide dye exclusion in the laser operated automated cell sorter 20 hours after HD exposure.

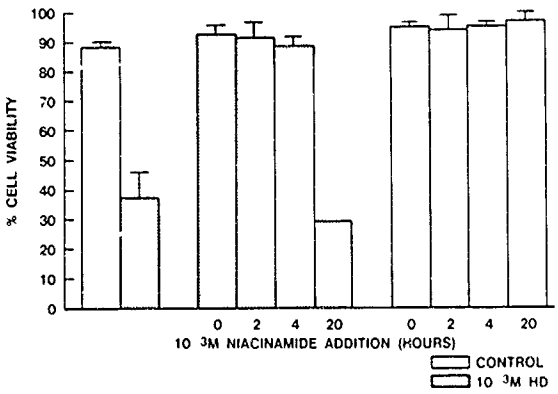


Figure 9. Niacinamide confers protection of human lymphocytes from HD-induced cytotoxicity even when added 4 hours after HD exposure. Cells were incubated with 10^{-3} M HD at time zero and 10^{-3} M niacinamide was added at the indicated times after HD addition. However, niacinamide failed to confer protection to the lymphocytes when added 20 hours after HD exposure. Viability was measured by propidium iodide dye exclusion in the laser-based automated cell sorter 24 hours after HD exposure.

The effect of 3-aminobenzamide on HD-initiated cell killing:

Known properties of niacinamide (a substrate for NAD^+ synthesis or an inhibitor of poly (ADP-ribose) polymerase, or both) could be responsible for protecting lymphocytes from HD-induced cell death. 3-aminobenzamide is a compound which inhibits poly (ADP-ribose) polymerase but cannot function as a substrate for NAD^+ synthesis. It was investigated to determine its ability in maintaining cellular viability in HD-exposed lymphocytes and exhibited (Figure 10) protection against HD-initiated cell death. At concentrations where 3-aminobenzamide confers complete protection against HD-induced cell death, it exhibits no toxic cellular effects.

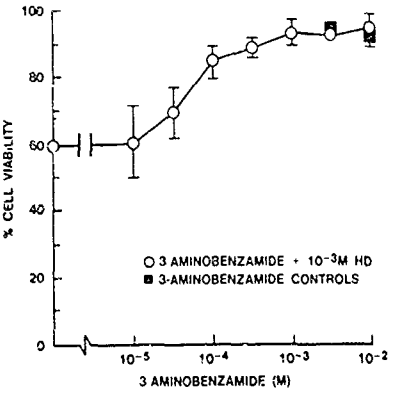


Figure 10. 3-aminobenzamide protects human lymphocytes from HD-initiated cell death. Lymphocytes were incubated with 10^{-3} M HD in the presence of the indicated concentrations of 3-aminobenzamide. At 10^{-4}M, 3-aminobenzamide protected the HD-exposed lymphocytes so that there was no difference in viability between them and control cells. Viability was measured by propidium iodide dye exclusion in the laser-operated automated cell sorter 24 hours after HD exposure.

Assessment of niacinamide protection at the electron microscopic level:

Lymphocytes were exposed to niacinamide ($3x10^{-4}$M)(Figure 11a), HD(10^{-3} M) (Figure 11b), or a combination of niacinamide ($3x10^{-4}$M) and HD (10^{-3} M) (Figure 11c) to insure that the protection of cell viability by niacinamide was a real phenomenon and not an artifact in which niacinamide limited dye migrated across the plasma membrane. The samples were examined after 24-hour incubation for pathology by electron microscopic techniques. Lymphocytes exposed to HD only presented fine-structural changes indicative of cytoxicity and cell death. These changes were observed in 7 out of 10 cells and included interrupted plasma membranes, loss of microvilli, rarefaction of large expanses of cytoplasm, condensation of nuclear chromatin, proliferation of cytoplasmic vacuoles, condensation and loss of mitochondria and other organelles, and extensive blebbing of the perinuclear envelope (Fig 11). HD-treated lymphocytes in the presence of niacinamide presented, with the exception of stunted microvilli, essentially normal fine-structure and appeared protected from cytotoxic effects of HD. These cells showed well-developed chromatin networks, prominent nucleoli, intact organelles, uninterrupted plasma membranes, and a cytoplasm rich in monoribosomes. The remaining cells presented to varying degrees those cytotoxic changes observed in the HD-treated population. Lymphocytes exposed to niacinamide alone were unremarkable.

Figure 11. Human lymphocytes exposed to niacinamide only, HD only, and HD in the presence of niacinamide for 20 hours at 37°C (magnification 7500X) (A). Representative lymphocyte from the niacinamide population. The cell demonstrates the typical ultrastructure of that of a small to medium sized lymphocyte. PM, plasma membrane; M, mitochondria; Mv, microvilli; Rb, free ribosomes. (B). Lymphocyte from that population exposed to HD only. Cell presents fine-structural changes indicative of cytotoxicity and cell death to include loss of microvilli and pyknotic nucleus. NB, nuclear membrane blebs; V, vacuoles; holes in plama membrane (arrows); N, nucleus. (c). Lymphocyte, representative of those treated with HD in the presence of niacinamide, apppears protected from the cytotoxic effects of HD.

DISCUSSION:

We proposed a working hypothesis for the biochemical mechanism (5) by which HD creates incapacitating lesions based on our knowledge of HD-induced pathology and other known effects of alkylating agents (10-12). This hypothesis was supported by our data indicating that HD alkylation of DNA caused enzymatic and spontaneous formation of DNA breaks (24). These findings were extended to activation of poly (ADP-ribose) polymerase, depletion of NAD^+, inhibition of glycolysis, activation of the hexose monophosphate shunt, and release of proteases. At the beginning of this research, there were no relevant models to study the validity of our hypothesis. Models in which to test possible therapeutic antidotes to HD also did not exist.

We have developed a new in vitro model (human lymphocyte viability model) which can quantitate the evolution of HD-induced pathology which may be observed at both the light and electron microscopic levels. Our other unique model (human mixed leukocyte biochemical marker model) (21) is capable of determining changes in biochemical markers after 4 hours. We have validated the activation of poly (ADP-ribose) polymerase and NAD^+ depletion portion of our hypothesis. We have been able to develop a logical approach to selecting candidate compounds to be studied as therapeutic regimens to prevent HD-generated pathology. More importantly, we have identified inhibitors of poly (ADP-ribose) polymerase as a promising class of compounds which should be further studied as antidotes to HD exposure.

The usual length of the latent period in HD-induced pathology is between 4-6 hours after HD exposure and we believe that the important biochemical changes involved in the HD lesion must occur before pathology is evident. We were able to demonstrate that 10^{-3} M HD caused a decrease in NAD^+ levels that was between 30-40% of control cells in 4 hours. NAD^+ levels at 30-40% of control values was also consistent with the hypothesis since only lymphocytes (comprising 30-40% of mixed leukocytes in normal individuals) contain the enzyme, poly (ADP-ribose) polymerase. This correlation was extended when the leukocyte populations were separated into granulocytes and lymphocytes. We were able to show that HD caused almost no change of NAD^+ levels in granulocytes but almost completely

depleted the NAD^+ levels in the lymphocyte population. The in vitro NAD^+ results with the leukocyte model were in good agreement with the in vivo NAD^+ results we obtained in HD-exposed human skin grafts on nude mice (25). It appears that the leukocyte model has a similar biochemical response to HD exposure as human skin. We were able to determine that the decrease in NAD^+ levels did not occur until after 1 hour post HD exposure and leveled off at 4 hours. This finding indicated that there was sufficient time to intervene in the progression of the HD lesion by interfering with the cause of the NAD^+ decrease.

According to our hypothesis, the NAD^+ was being consumed by the enzyme poly (ADP-ribose) polymerase. Therefore, we felt that two directions of action to prevent NAD^+ decrease were available for exploitation: 1) use of inhibitors of poly (ADP-ribose) polymerase such as niacinamide and 3-aminobenzamide to block NAD^+ decreases, and 2) use of substrates such as niacin and niacinamide to replace the lost NAD^+ through new NAD^+ synthesis. Both of these classes of compounds were effective in protecting leukocytes from HD-induced depletion of NAD^+. If the hypothesis was correct, then prevention of NAD^+ depletion should also hinder the development of HD-induced pathology. Like human skin, the leukocytes demonstrated no overt pathology in 4 hours.

A new model was needed in which pathology could be studied at 16-48 hours, the time when blistering occurred in human skin. Human lymphocytes were chosen as the appropriate model since they showed a HD-induced decrease in NAD^+ and could be kept in culture for extended periods of time. The lymphocyte model responded to HD damage with a similar time course as human skin (26) since it did not exhibit any overt pathology in the first 8 hours of culture but showed pathology at 16 hours which reached a plateau by 24 hours. Pathology in the lymphocyte was easy to assess by measuring the inability of the cells to exclude dye, an indication of metabolic cell death. Cellular viability decreased with increasing time and concentration of HD. Niacin (a substrate for NAD^+ synthesis) and niacinamide (both a substrate for NAD^+ synthesis and inhibitor of poly (ADP-ribose) polymerase) were studied to determine their ability in preventing HD-generated pathology. Niacin failed to offer any protection against HD-induced cell death (Figure 8). Niacinamide at 10^{-3} M conferred 97% protection in 24 hours and 88% protection in 48 hours against HD-induced cell death. Niacinamide was not only effective when added simultaneously with HD, but could also be added 4 hours after HD exposure and still prevent death of the lymphocytes. This finding indicates that niacinamide would not only be effective prophylactically, but also therapeutically. These results also indicate that niacinamide is not reacting with HD, since the HD (half-life of 8 min, unpublished data) would be hydrolyzed by the time niacinamide is added. Niacinamide must be interrupting the cascade responsible for generating HD-induced pathology, but it was still not known whether niacinamide was acting as a subtrate for NAD synthesis or as an inhibitor for the enzyme, poly (ADP-ribose) polymerase. 3-aminobenzamide, a specific inhibitor of poly (ADP-ribose)

polymerase but not a NAD^+ substrate, was shown to be as effective at blocking HD-induced cell death. Thus, it appears that the event responsible for HD lethality in lymphocytes is the inhibition of the enzyme, poly (ADP-ribose) polymerase.

HD-generated pathology in lymphocytes was correlated with pathology at the cellular level in human skin by comparing electron microscopic studies of HD-exposed lymphocytes with previous electron microscopic studies of human skin grafted on nude mice (17). In both types of tissue, there was the development of interrupted plasma membranes, loss of microvilli, rarefaction of large expanses of cytoplasm, condensation of nuclear chromatin, proliferation of cytoplasmic vacuoles, condensation and loss of mitochondria and other organelles, and extensive blebbing of the perinuclear envelope. Blebbing of the perinuclear membrane was of special interest. Similar blebbing occurred with regularity in the basal cells of the stratum germinativum of HD-exposed human skin grafts described in a previous study (27). The development of the lesions appear identical in both tissues. Lymphocytes exposed to HD in the presence of niacinamide show that niacinamide did not just prevent the dye from entering the cell but also prevented most of the structural damage in the lymphocytes, including interruptions in the plasma membrane.

We have now developed and validated two in vitro models using human tissue which is easy to obtain and responds like human skin. The data obtained from these models strongly support the first 4 parts of our hypothesis for development of HD-induced lesions. We have identified classes of compounds which should be investigated as potential antidotes. Using our models, we have shown that inhibitors of poly (ADP-ribose) polymerase are effective at preventing HD-initiated pathology. Specifically, we have identified that the FDA approved vitamin, niacinamide, is effective at preventing both significant cellular damage and changes in relevant biochemical markers.

References

1. S.W. Warthin and C.V. Weller, The Medical Aspects of Mustard Gas Poisoning, CV Mosby Co, St. Louis, 1919.

2. W.P. Anslow and C.R. Houck, Chemical Warfare Agents and Related Chemical Problems, NRDC Report, Division 9, Washington DC, 1946, 440-479.

3. B. Renshaw, Chemical Warfare Agents and Related Chemical Problems, NRDC Report, Division 9, Washington DC, 1946, 479-512.

4. H. Collumbine, Nature, 159 (1947) 151-153.

5. Meier, H.L., Gross, C.L., and Papirmeister, B. Army Science Conference 1984.

6. Papirmeister B, Davison CL, Gross CL: Army Science Conference Proc. 2; 217 (1968).

7. Papirmeister B, Davison, CL: Biochem. Biophys. Res. Comm. 17: 608, 1964

8. Papirmeister B, Davison, CL: Biochim. Biophys. Acta. 102: 70-92, 1965.

9. Papirmeister B, Gross CL, Petrali JP, Brinkley FB, Meier HL: Third Annual MRDC Chem Def Biosci Rev, 1983.

10. Fox M, Scott,D: Mutat. Res., 75, 131-168, 1980.

11. Lawley PD, Brookes P: Biochem. J. 109: 433-477, 1968.

12. Roberts JJ, Brent TP, Crathorn AR: Eur J Cancer 7: 515-524, 1971

13. Grossard F, Verly WG: Eur J Biochem 82: 321-332, 1978.

14. Hayaishi O, Veda K: ADP-Ribosylation Reactions, Biology and Medicine, Academic Press, NY, 1982.

15. Lehninger AL: Biochemistry, Chapter 16, Worth Publ. Inc., NY, 1979

16. Schnyder J, Baggiolini M: Proc Natl Acad Sci US 77: 414-417, 1980

17. Papirmeister B, Gross CL, Petrali JP, Meier HL: J Toxicol Cutaneous Ocul Toxicol 3: 395-410, 1984

18. Krehl WA: Fed Proced 40: 1527-1530, 1981

19. Quigley JP, Ossowski L, Reich E: J Biol Chem 13: 4306-4311, 1974

20. McAdams Jr J: J Invest Dermatol 26: 317-325, 1956

21. Meier HL, Gross CL, Papirmeister B: Tox Letters 39: 109-122, 1987

22. Lichtenstein LM: J Immunol 107: 1122-1130, 1971

23. Jacobson EL, Jacobson MK: Arch Biochem Biophys 175: 627-634, 1976

24. Papirmeister B, Gross CL, Meier HL, Petrali JP, Johnson JB: Fund Appl Toxicol 5: S134-S149, 1985

25. Gross CL, Meier HL, Papirmeister B, Brinkley FB, Johnson JB: Tox Appl Pharmacol 81: 85-90, 1985

26. Papirmeister B, Gross CL, Petrali JP, Hixson CJ: J Toxicol Cutaneous Ocul Toxicol 3: 371-393, 1984

Oxynitride Glasses and Glass Fibers

Donald R. Messier, Dr.
MTL, SLCMT-EMC
Watertown, MA 02172-0001

INTRODUCTION

Although earlier investigations[1,2] discussed the possibility of nitrogen substitution for oxygen in glasses, interest in oxynitride glasses as potential engineering materials dates from the time of a paper by Jack[3] entitled "Sialon Glasses," the "Sialon" in the title being an acronym for glass compositions containing various cations and the elements Si-Al-O-N. Although Jack's work was aimed at the understanding of the role of the glassy grain boundary phase in the sintering of silicon nitride, his results strongly suggested that the oxynitride glasses themselves were of great interest for further research and development. Work subsequently reported by Loehman[4,5] on the preparation and properties of Y-Si-Al-O-N glasses confirmed that these materials had some remarkable physical and chemical properties. Additional research on this system as well as on a number of other systems which form oxynitride glasses is discussed in a recent review by Messier.[6]

The findings discussed above, along with the interest at MTL in silicon nitride materials for high temperature structural applications such as heat engine components, led to our initial research program on oxynitride glasses in the system Y-Si-Al-O-N.[7,8] Those studies, which dealt with the preparation and characterization of small (10-30 g) specimens, confirmed and extended earlier work and reported the first elastic modulus measurements on these materials. Modulus and microhardness values were considerably higher than any previously measured for a glass, and subsequent work[9] showed that these glasses had outstanding hot water corrosion resistance. The remarkable properties of the glasses also stimulated our interest in the possibility of making high performance glass fibers, and attempts to make short (10-20 cm long) fibers were successful enough to warrant obtaining a patent[10] on oxynitride glass fibers. Our preliminary work as well as existing literature data indicated that glasses in the system Y-Si-Al-O-N as well as other oxynitride glass systems could have outstanding potential for two demanding Army structural applications: transparent armor, and reinforcing fibers in high-performance composites.

The work reported in this paper concerns techniques developed at MTL for the preparation of large (100 g) batches of Y-Si-Al-O-N glass, and for the drawing of fibers of the same glasses. Our recent research has dealt with glasses in several additional systems, some of which are similar to commercial oxide glass fiber compositions with nitrogen substituted for

oxygen in the glass structure. We have found, however, that the techniques for making any of the glasses and fibers are similar to those described herein for Y-Si-Al-O-N materials, and for clarity and brevity in this paper have emphasized our work on glasses in that system.

OXYNITRIDE GLASS PREPARATION

- STARTING MATERIALS
  - Al_2O_3, SiO_2, Si_3N_4, M_xO_y POWDERS
  - AlN, SiO_2, M_xO_y POWDERS
  - (M = Ca, Mg, Y, Zr, etc.)
- BATCH PREPARATION
  - MIX POWDERS AND ISOSTATICALLY PRESS
  - BATCH SIZES TO 100 g
- GLASS MELTING
  - BORON NITRIDE-LINED CRUCIBLES
  - TUNGSTEN ELEMENT FURNACE, N_2 AT 200 MPa
  - 1600°C, 2HR

Figure 1. PROCEDURE FOR THE PREPARATION OF OXYNITRIDE GLASS BATCHES OF VARIOUS COMPOSITIONS IN AMOUNTS UP TO 100 g.

EXPERIMENTAL PROCEDURES

The procedure for the preparation of oxynitride glasses is outlined in Figure 1. Starting materials for Y-Si-Al-O-N glasses are mixtures of powders consisting of either alumina, silica, silicon nitride, and yttria, or aluminum nitride, silica, and yttria. Some experimental evidence[11] indicates that glass quality is improved by using the latter mixture, but further work would be necessary to prove that point conclusively. At any rate, glass preparation is begun by weighing out the various powders in the desired amounts, mixing by tumbling in plastic containers with the aid of a

few alumina balls, and compacting the resulting mixture by isostatic pressing in a rubber bag. The compacted powder is the put into a boron nitride-coated graphite crucible and heated to the melting point in a cold-wall, tungsten element furnace under nitrogen at a pressure of 200 kPa (2 atmospheres). Heating is continued for 2 hours after which the glass is quickly cooled to room temperature. In order to be able to cut or machine large specimens (100 g) without shattering them, annealing is necessary; that is accomplished by reheating the specimens to 700°C for 2 hours, after which any additional machining that is needed may be easily accomplished. The glass preparation process is described in more detail in a recent publication by the author.[12]

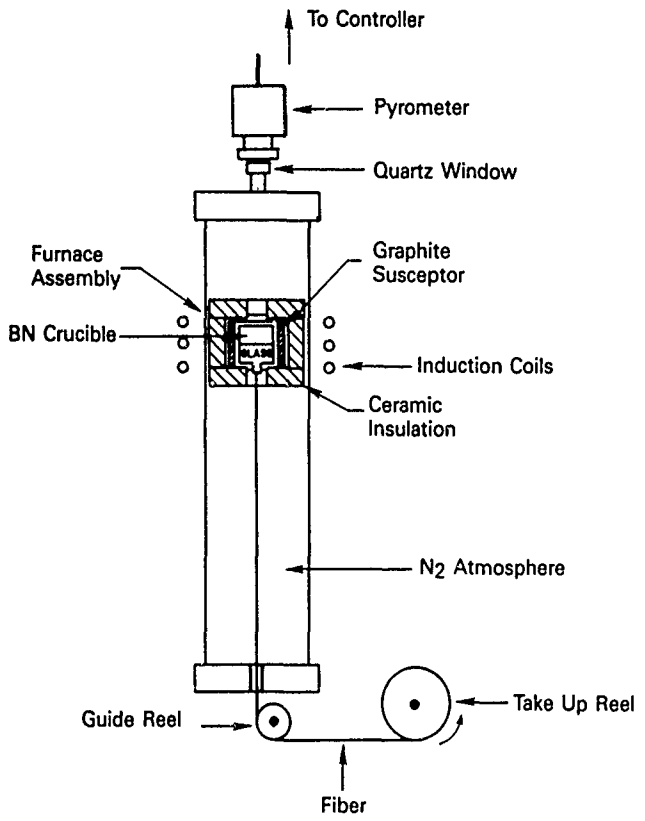


Figure 2. OXYNITRIDE GLASS FIBER-DRAWING APPARATUS.

The apparatus that was developed for drawing oxynitride glass fibers is illustrated schematically in Figure 2. A patent application is in progress on the process illustrated in the figure, this being the only known example of the preparation of continuous oxynitride glass fibers from the melt, or by any other means for that matter. The fiberizing process consists of drawing a fiber from molten glass contained in a boron nitride crucible having a centrally located 3 mm diameter orifice in its bottom. The technique is similar to that used commercially for making oxide glass fibers[13] except for the crucible material and use of a nitrogen rather than air atmosphere. As seen in Fig. 2, the furnace assembly is contained in a fused silica glass tube through which is maintained a flow of nitrogen gas. The furnace is heated by a radio frequency (400 kHz) induction unit that couples to the graphite susceptor via a water-cooled copper coil (Fig. 2). Precise control of heating rate and temperature are provided by a microprocessor programmer/controller (not shown in Fig. 2). Insulation is provided by easily replaceable stiff foam-type ceramic insulation that is formed using simple hand tools.

The fiber-drawing process is begun by heating previously fabricated oxynitride glass in the boron nitride fiber-drawing crucible in flowing nitrogen to the point where melting has occurred and a drop of molten glass has appeared at the outside of the orifice in the bottom of the crucible. At that point, a refractory ceramic (alumina) rod is inserted through the axially-located bottom fitting to touch the molten glass drop. The rod is then withdrawn, pulling a glass fiber out of the furnace enclosure through the bottom fitting. The end of the fiber is then attached to the takeup reel with a piece of tape. The reel is rotated by a variable-speed electric motor and the fiber is wound continuously for up to 5 to 10 minutes at which time upwards of 100 m of fiber has been obtained. A key feature of the process is that, while the fiber is drawn in a protective nitrogen atmosphere, it is wound outside of the apparatus under ambient conditions.

RESULTS AND DISCUSSION

The enhancement of the properties of an Y-Si-Al oxide glass by substitution of nitrogen for oxygen is indicated in Table I. That table compares Y-Si-Al-O-N glass containing 6.3 atomic % nitrogen made at MTL with a similar nitrogen-free oxide glass reported in the literature.[5,14,15] The property data in Table I follow trends similar to those observed for other oxynitride glass systems,[8] i.e., density, glass transition temperature, microhardness, elastic modulus, and fracture toughness all increase with increasing nitrogen substitution. The Young's elastic modulus value of 141 GPa (20.5 Mpsi) given for Y-Si-Al-O-N glass "a" is much higher than for the oxide glass, as well as far higher than typical values for silicate glasses (80 GPa). The glass is also nearly twice as hard as typical oxide glasses.

Table I. EFFECT OF NITROGEN ON PROPERTIES OF AN YTTRIUM-ALUMINUM-SILICATE GLASS

GLASS	Y-Si-Al-O-N "a"*	Y-Si-Al Oxide Glass**
NITROGEN CONTENT (Atomic %)	6.3	0.0
DENSITY (kg/m^3)	3720	3600
GLASS TRANSITION TEMPERATURE (°C)	944	918
MICROHARDNESS (GPa)	9.01	8
ELASTIC MODULUS (GPa)	141	115
FRACTURE TOUGHNESS (MPa $m^{1/2}$)	0.95	0.60

*MTL Data
**Literature Data from Several Sources

Figure 3 shows a number of Y-Si-Al-O-N glass disks (100 g each) that were fabricated at MTL for use in Hugoniot elastic property measurements to assess the potential of the material for ceramic armor applications. Nine of the disks contained 6 atomic % nitrogen, and the other five, 13 atomic % nitrogen. Fig. 3 demonstrates our ability to make the glasses reproducibly and in quantity. The disks shown in the figure were furnished to Israeli investigators[16] for machining and the Hugoniot property measurements. The results of those measurements (beyond the scope of this paper) indicate that the material looks attractive for use both by itself (e.g., for transparent armor), as well as for use as a high modulus matrix phase for low-cost, opaque ceramic armor. It is known, however, that the best criterion for evaluating ballistic performance is ballistic testing, and a contract effort is in progress to provide glass disks for such testing to be conducted in the near future at MTL.

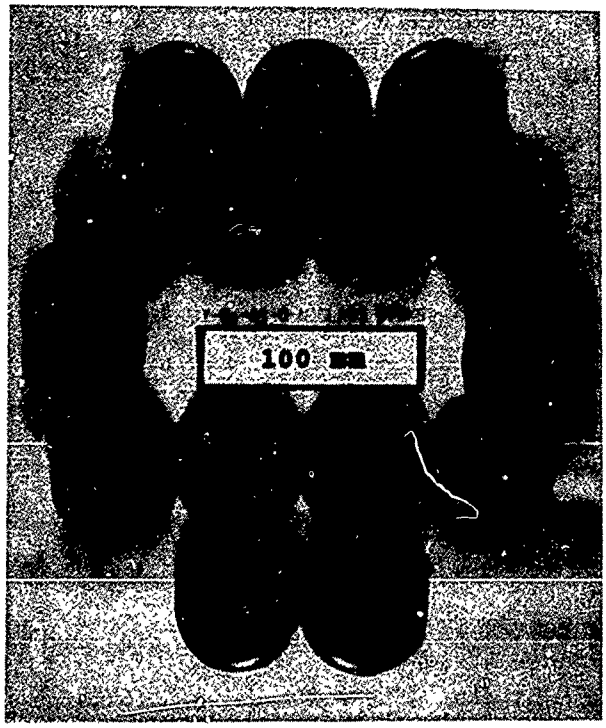


Figure 3. 100 g DISKS OF Y-Si-Al-O-N GLASS FABRICATED AT MTL.

The results of fiber-drawing experiments on Y-Si-Al-O-N glass composition "a" (6.3 atomic % nitrogen) are illustrated in Figures 4 and 5. As noted in the introduction, we have more recently been able to make oxynitride glass fibers of other compositions, but, for purposes of illustration and for brevity, have limited this paper to our results on one representative material.

Figure 4 is a transmitted light optical photomicrograph of portions of two melt-drawn glass fibers. Although these particular fibers were drawn discontinuously, they illustrate the potential for making good quality fibers from this composition, and the possibility of making very small diameter (5 μm) oxynitride glass fibers. Typical commercial oxide glass fibers are in the diameter range from 5 to 15 μm, and it is probable that applications for oxynitride fibers will require diameters in the same range.

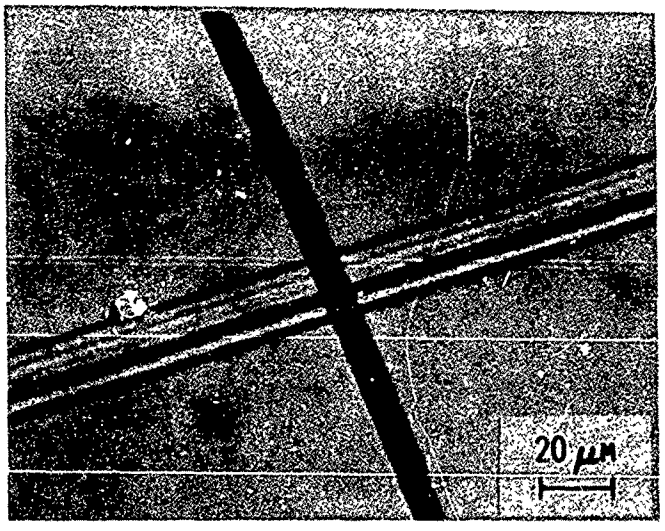


Figure 4. TRANSMITTED LIGHT OPTICAL PHOTOMICROGRAPH OF PARTS OF SHORT (100 cm LONG) Y-Si-Al-O-N FIBERS DRAWN FROM THE MELT. FIBER QUALITY AND UNIFORMITY ARE GOOD, AND THE SMALL DIAMETER (5 um) OF ONE FIBER INDICATES THAT FINE FIBERS CAN BE FORMED FROM THIS GLASS.

A bundle of Y-Si-Al-O-N glass fibers cut from a continuous fiber wound on a plastic spool is shown in Figure 5. This picture is included to demonstrate the feasibility of drawing long, continuous glass fibers. The length is chiefly limited by the amount of glass that can be loaded into the crucible. At present, however, fiber breakage occurs when the winding speed gets too high. With our initial unsophisticated setup, high winding speeds led to vibrations and oscillations that caused the fiber to break prematurely. Because fiber diameter is inversely proportional to winding speed, the breakage problem has thus far prevented us from obtaining fibers less than 20 μm in diameter. This is somewhat larger than desirable, and making finer diameter fibers requires improvements in our winding system, a task that is currently in progress.

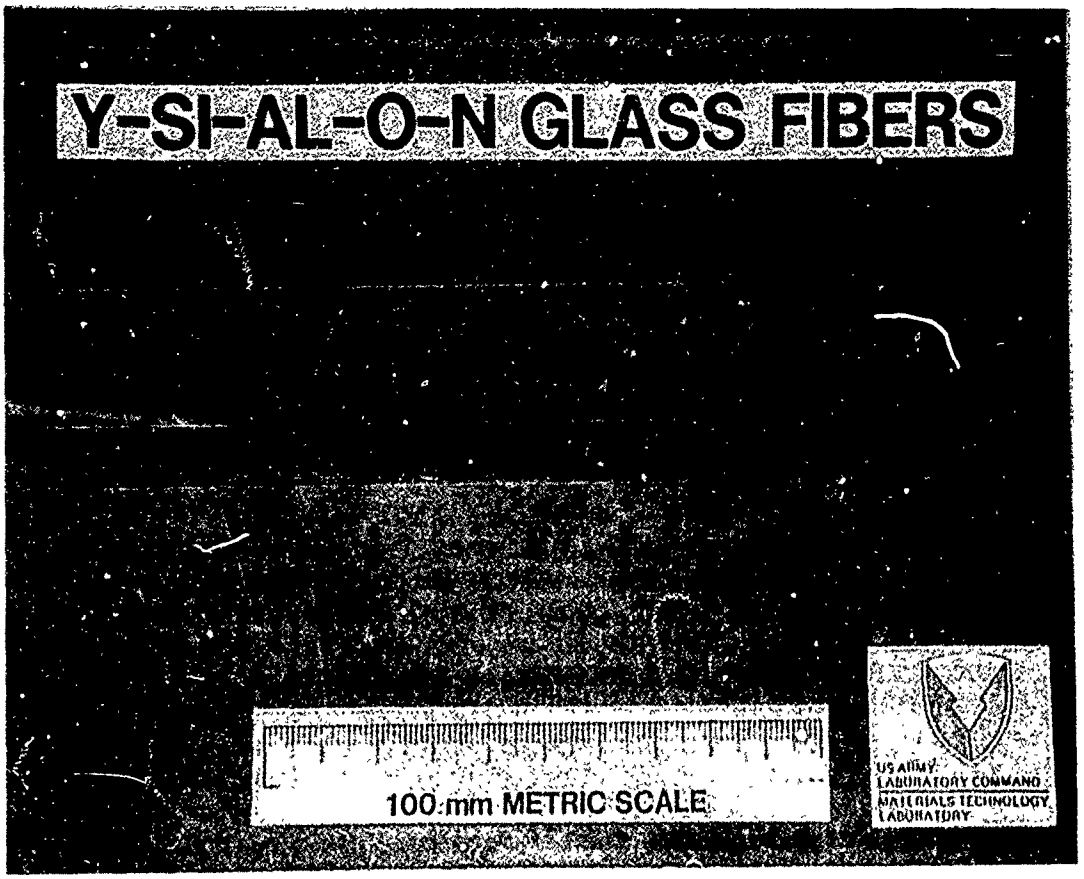


Figure 5. Y-Si-Al-O-N GLASS FIBERS CUT FROM A SPOOL OF CONTINUOUS FIBER DRAWN FROM MOLTEN GLASS.

One of the important questions in the fiber-drawing program was whether or not the chemistry and properties of the fibers were the same as those of the bulk glass, and that point is addressed in Table II which compares some key properties of Y-Si-Al-O-N glass and fibers of the same nominal composition. For convenience, the nitrogen contents in this table are given in weight % rather than atomic %, but the glass composition is the same as discussed in the rest of the paper.

Because of the small mass of fiber available, bulk chemical analysis and density measurements could done on the bulk glass, but not on the fibers. The table includes chemical analyses for nitrogen done on both glasses using wavelength dispersive analysis on the scanning electron microscope (S.E.M.), and those results agree well with each other as well as with the conventional chemical analysis within the experimental error of the techniques. Cation analyses (not shown for simplicity) also done with the S.E.M. were similarly in good agreement. Further results that we have obtained with a higher nitrogen content glass also gave similarly good agreement between the chemistry of the bulk glass and fiber lending further support to the finding that the two forms are chemically identical.

Table II. COMPARISON OF PROPERTIES OF Y-Si-Al-O-N "a" BULK GLASS AND GLASS FIBERS

FORM	BULK GLASS	GLASS FIBERS
NOMINAL WEIGHT % N	3.2	3.2
WEIGHT % N (CHEM. ANAL.)	3.3	-
WEIGHT % N (S.E.M.)	2.7	3.8
DENSITY (kg/m^3)	3720	-
MICROHARDNESS (GPa)	9.01	9.14
ELASTIC MODULUS (GPa)	141	135

Table II also shows that the hardness of the fibers is the same as that of the glass (the slightly higher microhardness value for the fibers is statistically insignificant). The elastic modulus value given for the bulk glass composition, initially determined at MTL, and independently confirmed by other investigators, is slightly higher than the value for the fiber (determined by a sonic velocity technique). Differences of that magnitude and direction are commonly found in comparing other bulk glasses and glass fibers, and it may be therefore concluded that the two forms are also the same in this respect.

CONCLUSIONS

The feasibility of making useful quantities (100 g) of Y-Si-Al-O-N and other oxynitride glasses has been demonstrated, and property data from this and previous investigations indicate that these glasses will be useful engineering materials. Possible Army applications for such glasses are transparent armor, and opaque ceramic armor using the glass as the bonding phase.

Our work has demonstrated for the first time anywhere the feasibility of making oxynitride glass fibers, and, even more significantly, that continuous fibers can be drawn from the melt. A melt-drawing process with several unique features has been developed, and it has been verified that the properties of the fibers are the same as those of the parent glass. The fibers have proven to be the highest elastic modulus glass fibers ever produced. Potential applications for such fibers include high performance resin matrix composites, and, possibly, reinforcements for optical fiber cables.

FUTURE WORK

Realization of the considerable potential of oxynitride glasses and fibers will require further research work. An area of glass research needing particular attention is better defining the composition and temperature ranges for glass formation in the Y-Si-Al-O-N system as well as in other systems of interest. The compositions that we have used so far, while demonstrating some impressive properties, have been selected rather arbitrarily, and even better properties would probably be found on materials with optimized compositions. The principal defects found in oxynitride glasses are metallic precipitates (silicon-rich) that result from thermal decomposition during melting, and their minimization or

elimination requires better understanding and control of the high temperature chemistry of the glass-melting process. Also needing attention is optimization of the entire heating-cooling-annealing cycles used for glass-making.

The minimum fiber diameter that we have been able to achieve thus far has been limited by the inadequacy of our fiber winding setup. Higher winding speeds to produce smaller diameter fibers should be easily achievable by straightforward improvements in the fiber-drawing apparatus. Another aspect of fiber making that remains to be looked at in detail is the identification of, and minimization or elimination of strength-controlling defects; if metallic particles such as those mentioned above are critical, the remarks on the need for understanding and control of high temperature chemistry also apply to fiber-drawing. Regardless of what remains to be done, however, it is certain that oxynitride glass fibers represent an exciting new technological opportunity.

ACKNOWLEDGMENTS

The capable assistance of R.R. Rich and R.P. Gleisner with all aspects of the experimental work on fiber drawing is greatly appreciated. Also appreciated is the scanning electron microscopy work done by J. Marzik, and the encouragement of R.N. Katz in the initial stages of this research.

REFERENCES

1. Mulfinger, H.O., "Physical and Chemical Solubility of Nitrogen in Glass Melts," J. Am. Ceram. Soc., 49 462-7 (1966).

2. Elmer, T.H., and Nordberg, M.E., "Effect of Nitriding on Electrolysis and Devitrification of High-Silica Glasses," J. Am. Ceram. Soc., 50 275-9 (1967).

3. Jack, K.H., "Sialon Glasses," in Nitrogen Ceramics, Ed. F.L. Riley, Noordhoff International, Reading, MA, 1978, 257-62.

4. Loehman, R.E., "Preparation and Properties of Yttrium-Silicon-Aluminum Oxynitride Glasses," J. Am. Ceram. Soc., 62 (9-10) 491-4 (1979).

5. Loehman, R.E., "Oxynitride Glasses," J. Non-Cryst. Solids, 42 433-46 (1980).

6. Messier, D.R., "Review of Oxynitride Glasses," Rev. Chim. Min., 22 518-33 (1985).

7. Messier, D.R., "Preparation and Crystallization of Si-Y-Al-O-N Glasses," in Ceramic Engineering and Science Proceedings, Vol. 3, No. 9-10, Sept.-Oct. 1982, The American Ceramic Society, Columbus, Ohio, 3-76.

8. Messier, D.R. and Broz, A., "Microhardness and Elastic Moduli of Si-Y-Al-O-N Glasses," J. Am. Ceram. Soc., 65 (8) C-123 (1982).

9. Wald, J.W., Messier, D.R., and DeGuire, E.J., "Leaching Behavior of Si-Y-Al-O-N Glasses," Int. J. High Tech. Ceramics, 2 65-72 (1986).

10. Messier, D.R., DeGuire, E.J., and Katz, R.N., "Oxynitride Glass Fibers," U.S. 4,609,631 (2 September, 1986).

11. Messier, D.R. and DeGuire, E.J., "Thermal Decomposition in the System Si-Y-Al-O-N," J. Am. Ceram. Soc., 67 (9) 602-5 (1984).

12. Messier, D.R., "Preparation and Properties of Y-Si-Al-O-N Glasses," Int. J. High Tech. Ceramics, 3 33-41 (1987).

13. Lowenstein, K.L., The Manufacturing Technology of Continuous Glass Fibres," 2nd Ed., Glass Science and Technology 6, Elsevier, New York, 1983.

14. Makishima, A., Tamura, Y. and Sakaino, T., "Elastic Moduli and Refractive Indices of Aluminosilicate Glasses containing Y_2O_3, La_2O_3 and TiO_2," J. Am. Ceram. Soc., 61 (5-6) 247-9 (1978)

15. Hampshire, S., Drew, R.A.L., and Jack, K.H., "Viscosities, Glass Transition Temperatures, and Microhardness of Y-Si-Al-O-N Glasses, J. Am. Ceram. Soc., 67 (3) C-46-7 (1984).

16. Brandon, D.G., Technion, Haifa, Israel, Personal Communication.

Expert Systems in the Tactical Environment

Robert F. Miller
Peter W. Dickson
Electronics Technology & Devices Laboratory
Ft. Monmouth, NJ 07703

INTRODUCTION

The task of an Explosive Ordnance Disposal (EOD) team is to disarm, detonate, or otherwise "render safe" bombs, rockets, projectiles, mines, or other ordnance. The ordnance fusing may have a motion sensitive and/or a time sensitive component, which requires the job to be done on site and as rapidly as possible. Although special tools and equipment have been developed to facilitate this process, the most critical element in the successful completion of the task is a detailed knowledge of the construction and operation of the explosive device.

Current Army tactics assign this job to a three man team, with five to seven teams forming a detachment. The data base of render-safe procedures is maintained at the detachment, partly on paper and partly on micro-fiche. The volume of the data base is approximately 15,000 pages of paper and 1,600 sheets of micro-fiche. The data base is made up of three types of publications; the Identification Guides, the Render Safe Procedures, and the List of Available Publications. The items in the data base are subject to periodic review and update, with changes to the data base occurring at a rate of about 10 items per week. It is anticipated that during the first days of an armed conflict the number of new items encountered would substantially increase, with a resulting increase in the volume of updates to the data base. Current Army staffing does not allow these detachments to have supply, maintenance, or clerical personnel; therefore all these functions must be performed by members of the three man team. A recent Army study of EOD units indicated that 16% of available man hours were spent in data base maintenance, a figure that would grow significantly under battle conditions.

On a typical mission, the EOD detachment will be informed of a situation requiring their expertise. Usually some indication of the ordnance to be encountered will be provided by the personnel on site

and/or by the method of ordnance delivery - air raid, rocket attack, artillery, etc. One or more three man teams will be dispatched for the mission. Depending on the accessibility of the site, they may travel by helicopter, by ground vehicle, or on foot. The amount of material they bring with them will depend on what they expect to find and on the mode of transportation, varying from as much as a truck and trailer of equipment down to as little as a 10 pound tool pack and an identification guide. The identification guides are printed documents containing the features and descriptions of the anticipated ordnance, along with an abbreviated render safe procedure. The ready availability of the complete render safe procedures data base is made difficult not only by its weight, volume, and the frequency with which it must be updated, but also by the security classification of its contents.

On arrival at the incident site, the EOD detachment will establish a "safe area" some distance away from the unexploded ordnance, taking advantage of whatever protection is offered by local terrain. At least one member of the team will remain in the safe area while data for the identification of the ordnance is being obtained, and during the render safe procedure. Access to whatever elements of the data base the team has brought with them is conducted in the safe area.

SYSTEM REQUIREMENTS

The EOD application requires an information retrieval system that is portable (preferably man-portable), will operate in a tactical environment, is capable of accessing an extremely large volume of data, and can display detailed engineering drawings. In order to qualify as man-portable, no single component of the system can weigh more than 37 pounds. The EOD community has expressed a strong preference that all components of the system be housed in one enclosure to minimize cabling and set-up time. It is also desired that the system be capable of battery operation for periods of two to four hours. The tactical environment covers extreme conditions in terms of temperature, shock, vibration, moisture, dust, etc. for operation and for transport.

The EOD information retrieval system requires active intervention by an EOD technician for data entry. resolution of uncertainty and ambiguity, and interpretation of results. The system can, at best, act as an assistant in the information retrieval process. In order to be able to use the system effectively, the operator must be able to easily enter data, and to understand the response of the system. The process of data entry may be impeded by a requirement for the operator to wear chemical protection gear. In addition to the requirements for ease of data entry and clear, unambiguous displays of query results and requests for information, an effective interface between man and machine requires that the

user have an accurate understanding of how the system operates. The EOD information retrieval system functions as an intelligent interface between the user and a large, complex data base. The system utilizes both data entered by the EOD technician and data not directly accessible to the user. In order to facilitate user understanding and acceptance of system conclusions, the EOD expert system allows the operator to review the data entered, to access certain elements of the data base, and to examine the rules that have been invoked in the decision making process.

The engineering drawings in the data base constitute the most demanding task for the system graphics processor and display. These drawings for the data base are currently generated electronically, using a cathode ray tube (CRT) display and a software zooming capability which allows portions of the drawing to be expanded to fill the entire screen. This system generates drawings with an addressable resolution of 10,000 x 13,000. These files are used to generate the paper drawings that constitute the current mode of distribution of the data base. These display requirements can be satisfied in the tactical environment through a combination of the use of flat panel displays utilizing thin film electroluminenscent (TFEL) technology, and a powerful graphic processing capability which will permit rapid panning and zooming of the graphic data.

The total volume of on-line storage required for the EOD application is on the order of hundreds of megabytes, with access time on the order of seconds. These requirements clearly indicate some form of optical storage as a solution. No other technology currently offers the large storage capacity, rapid access time, high data transfer rate, and low system volume obtainable with optical storage.

The functions of the EOD Information Retrieval System can be divided into three computational activities; the expert system software and data base management, the high resolution graphical data manipulation, and text processing. The expert system software and data base management task is not computationally intensive; the performance of this component will instead be limited by the access time and the information transfer rate of the mass storage media. The second system function, the manipulation of graphic information, consists of the display of engineering drawings, which typically consist of some 12,000 vectors. The third system function consists of the presentation of technical information and render safe procedures in the form of text files, with the ability to scroll smoothly, jump from page to page, access indices, and search for specific items. The response time for all these functions must be on the order of a few seconds or less in order to achieve a good man-machine in-

terface. The presentation of graphic and text data with this response time is computationally intensive, and will define the computing power required of the system.

PROOF OF PRINCIPLE SYSTEM

A proof of principle system has been constructed to investigate hardware and software alternatives, and to obtain feedback from the user community. The system uses a PC-AT architecture with an 80286 microprocessor, and a graphic display processor board containing a 68020 microprocessor and a Hitachi 63484 ARCTC graphic engine for parallel processing. Two sizes of TFEL flat panel displays, both with 640 x 512 resolution, are being tested in the proof of principle system. Operator input is entered through an LED touch panel on the display. Storage of the data base is on a WORM (Write Once Read Mostly optical media) with a capacity of 400 Mbyte per disk surface. The computer, power supply, and WORM drive are mounted in a case 7.5 x 14.5 x 16.5 inches, with a combined weight of 30 lb. The display is tethered to the computer on a 6 ft cable. Considerable reductions in weight, volume, and power are possible in the fielded system.

Viewing Area (inches)	Size (WxHxD) (inches)	Weight (lb.)	Power (W)	Resolution (lines/in)	Pixels
5.0x6.25	9.8x8.6x2.0	8	20	101.6	640x512
10.0x12.5	16.7x14.0x1.9	11	32	50.8	640x512

Table of dimensions; weight & power for display heads

The application program is based on PRIOR GKS/C, a portable, high performance implementation of the International Graphics Kernel System (GKS) standard. GKS provides graphics device independence, allowing the software to run with a wide variety of computers and displays. Many of the GKS lower level functions have been placed in the 68020 graphic controller's memory, thus balancing the computational load between the 80286 and the 68020. The system is able to process 1,000 to 2,000 vectors per second, allowing screen draw times of a few seconds for complex engineering drawings.

In order to access the proper render-safe procedure, the EOD technician enters the known characteristics of the unidentified ordnance. The ordnance is identified when a match between the characteristics entered and the attributes of an item in the data base has been achieved. As new characteristics are entered, the system monitors the number of items in the data base that match the user input. In the prototype system, data is entered through a touch panel. The preferred screen format for data entry consists of a menu with a limited number of choices. For

numerical or text data, the display is configured as a numeric or alphabetic keypad. The operator is given immediate feedback through an audio tone and/or highlighting of the item selected. The operator has access to on-screen help detailing what procedure is being performed, and what options are available in the identification process. In the event that the system will require more than a few seconds for search and calculation, a message is displayed informing the operator of the action taking place, with an estimate of the time required to complete it. When all known data has been entered, or when a unique match has been obtained. the technician is presented with an engineering drawing of the ordnance and asked to confirm the identification. If the characteristics available are not sufficient to uniquely identify the ordnance, a series of drawings will be presented, ordered according to the system's estimate of their probability, from the most likely to the least likely.

In the tactical environment, time and access to data are often restricted. The cost of the extended investigation required to obtain additional data is often an increased risk of detonating the ordnance. During the identification procedure, the system uses its estimates of the probability of each element in the data base to calculate the value of characteristics that have not yet been entered. The use of this information results in the most efficient collection of data during the identification process.

The algorithm used to estimate probability assumes that Bayesian statistics are valid; i.e., that the data base is complete and mutually exclusive. Operational data from the Viet Nam conflict shows that 75 to 80% of the items encountered by EOD units were in the data base. Lower figures can be expected during the first phases of an armed conflict, with later performance strongly dependent on the speed and efficiency of the updating process for the data base. Although it is possible for more than one item of ordnance to be found at an incident site, EOD technicians feel that there will be little difficulty in entering the characteristics of only one item at a time, satisfying the requirement for mutual exclusivity.

Given the validity of Bayesian statistics, then

$$P_i(H_j) = P_{i-1}(H_j)\ P(A/H_j)\ /\ P_{i-1}(A)$$

where

H_j is the hypothesis that the j*th* element of the data base matches the item being examined

$P_i(H_j)$ is the probability of H_j being correct after the *ith* attribute has been entered

$P(A/H_j)$ is the probability of H_j given attribute A

and

$$P_{i-1}(A) = \Sigma\ P_{i-1}(H_k)\ P(A/H_k)$$

$P_0(H_j)$ can be a uniform distribution over the entire data base, or it can reflect any *a priori* knowledge available to the EOD technician. $P(A/H_j)$ need not be a delta function, but could reflect the reliability of the data entered, giving some reduced value of probability to items in the data base with characteristics "close" to the attribute entered by the EOD technician. In order to calculate the value of characteristics as yet undetermined, the system must calculate the information content, or entropy associated with each of the undetermined characteristics. The characteristic with the minimum entropy will yield the maximum information. The entropy is given by

$$E(P_i) = -\Sigma\ P_i(H_j)\ \log_2 P_i(H_j)$$

The system must consider each as yet undetermined attribute A, and using each possible result of its determination $a_1,\ldots,a_m$ for the elements in the data base satisfying the characteristics entered to this point, compute $P_{i+1}(a_k)$ and the resulting entropy. The entropy associated with the determination of characteristic A is given by

$$E(A) = \Sigma\ p(a_k)\ E(P_{i+1}(a_k))$$

where $p(a_k)$ is the probability of occurrence of attribute value a_k.

The undetermined attribute resulting in the smallest entropy offers the highest information content for the identification procedure. If the EOD technician has determined that a particular characteristic is unobtainable, either because of the physical condition of the ordnance or because of increased risk, he may respond that the data is not available when that characteristic is requested by the system. Thereafter, this characteristic will not be considered in the entropy calculations.

The identification procedure is considered complete when only one item satisfies the characteristics entered, when no more data is available, or when one hypothesis is clearly superior. The criteria for clear superiority is an estimate of probability greater than .95.

As noted above, the array of initial probabilities allows for the consideration of any *a priori* information known to the EOD technician or to the system itself. This enables the incorporation of a "learning" feature. The systems for supply and for command and control anticipated on the battlefield make it likely that what was found in a particular area of operations yesterday will be found again today. Information about what was found by EOD units in the area will be filed in after-action reports, and therefore can be made available to each of the EOD information retrieval systems operating in the area. This information can be updated on a daily basis, through the process of transferring the current data on ordnance found in the area into the P_0 array as each team returns to its detachment headquarters. An additional mode of "learning" can be accomplished by adding characteristics of new elements to the data base through the incorporation of data from reports filed on "first seen" ordnance. Thus the system would have the ability to "learn" from its environment in terms of an optimized procedure for gathering information, and also in terms of a more complete data base.

In addition, when a particular hypothesis is confirmed, the system can obtain feedback about the validity of the characteristics entered by a particular user, and can be programmed to modify the $P_i(A/H_j)$ factors accordingly - e.g., operator A's estimates of diameter are accurate to within 1%, while operator B's estimates vary by 5%.

An unclassified data base was generated for the proof of principle system. The data base consists of engineering drawings and characteristics of 37 explosive devices in the projectile category, as well as data on possible fuses and text files describing simulated render safe procedures. In addition, a text file containing definitions of ordnance technical terms appearing in the drawings and render safe procedures was generated. Items of foreign manufacture were chosen for the data base to facilitate testing of the system through minimizing previous exposure of US Army test subjects to the possible test items. Diameters range from 47 to 122 mm, and length from 129 to 810 mm.

Identification parameters for each of the ordnance items consist of three types - DIMENSION, NUMBER, and STRING. DIMENSION items are numerical data, in millimeters, such as length and diameter. NUMBER items are positive integers representing some countable item, such as the number of fins, or the number of gas check bands. STRINGs are reserved for unique ASCII information such as markings, or for reference to other blocks of information in the data base. It is important to note that not all items in the database have elements for each data type or category.

Each ordnance description consists of many identification parameters. The structure of these parameters must be flexible so that various kinds of data traversing can be performed. Efficiency of traversal is particularly important for two operations: the examination of all identification parameters belonging to a particular ordnance, and the examination of all identification parameters of a given data type from all relevant elements of the ordnance data base. These kinds of data traversals can be efficiently implemented through the careful design of data fields and pointers or links, so that data can reference other fields. For large database applications, traversing multiple linked field structures becomes more efficient, as it can reduce the number of comparisons required for each search. Link fields or pointers can be used so that only the relevant data is referenced. As conditions change, the ability to redirect pointers to reflect the current configuration is vital.

The searching technique used in the prototype system reduces the number of comparisons necessary in the identification procedure by modifying the structure of the data links. For each identification parameter entered by the EOD technician, a traversal of the currently connected structure is performed in search of data agreement. When data agreement is not achieved for a particular ordnance, the entire set of structure elements for that ordnance is disconnected from the traversal mechanism. In the event that no element agreeing with the data entered can be found, or if further investigation requires the alteration of data, the EOD technician is able to examine and modify the data previously entered, and thus reconnect elements to the traversal mechanism.

The engineering drawings are stored on the optical disk in the form of individual files. These files contain vector information describing multiple views of a single ordnance. Typically these drawings may contain a dimensioned profile view, a dimensioned end view, an isometric view, and dimensioned cut-away views. The addressable resolution of the drawing space is 10,000 x 13,000.

Field Test

The EOD proof of principle system was scheduled for evaluation by EOD technicians in a tactical scenario during April 1988. Results of the tests will be given during the oral presentation of this paper.

The EOD technicians who participated in the tests were graduates of the EOD Advanced Individual Training course, augmented by personnel from operational EOD units to make up a total of 40 test subjects All subjects had approximately the same experience and training in EOD procedures. The most recent final test scores from the Advanced Individual Training course at the US Army Ordnance Missile & Munitions Center and

School (OMMCS) was used to rank the skill level of the subjects. Most of the test subjects had a Military Occupational Specialty (MOS) of 55D10 (EOD assistant with rank E3), although a few 55D20s (EOD specialist with rank E4 or E5) were included. In order to maximize the number of data points, the test subjects were organized as one-man teams.

Ten different projectiles were selected from the proof of principle system data base and placed by EOD personnel on a test range in a situation simulating an operational environment. Each of the projectiles was judged to be of approximately the same difficulty in identification. After one day's training on the proof of principle system, the test subjects were given a pretrial test to assure that all subjects had achieved the same skill level with the system. All participants then received operational briefings and conducted an identification for each of the ten items, one half with the EOD system and one half using the current paper manuals. EOD qualified personnel monitored the identification procedures. Exit interviews were conducted after each test. For each test, several items were measured - the time required to achieve an identification, any errors in procedure or in identification, the time required to access the render safe procedure, and the ratio of the time spent on the incident site to the time spent in the safe area.

EXPANSION OF SYSTEM FUNCTIONS

Several additional functions can be added to the system to reduce the clerical workload on the EOD technicians. These include prompts and automated formatting for the generation of after action reports, generation of requests for supplies consumed in the render safe procedures, and generation of reports for first seen ordnance. It is well within the state-of-the-art to include a video camera and an image capture device in the system, so that the first-seen ordnance report, which would be transmitted to other units electronically, can include a picture of the ordnance.

Since the EOD information retrieval system is able to compare the data entered with the characteristics of the item after the identification is confirmed, the system could generate some measure of performance for the tools and instruments used in the procedure, noting which elements require maintenance or calibration. The system could also generate a log of items used during the identification and render safe processes, and make recommendations for maintenance or resupply.

During the identification process, the EOD technician must assess the value of the information to be obtained versus the risk in obtaining that information. The EOD information retrieval system can be utilized in a training environment to provide the technician with feedback on the

utility of information and on the risks and benefits involved with alternate courses of action. In addition, feedback on the validity of data entered during actual operations could allow the system to suggest subject areas where the EOD technician should receive additional training.

In the identification procedure, the engineering drawings of the ordnance can be enhanced to provide greater spatial information. By digitizing the engineering drawings in three dimensions instead of two, the system will be able to generate a drawing of the ordnance from any viewing angle, utilizing a wire frame or shaded image with hidden lines removed. Additional data such as dimensions, call outs, cross sectional views and markings can be maintained in independent overlays. An order of magnitude increase in computing power will be required to maintain system response time during the generation of three dimensional drawings.

The conversion of the data base from paper to electronic format offers obvious advantages in terms of storage volume and ease of update. The realization of additional advantages, however, may require a restructuring of the data base. The paper data base is designed for linear access. This means that all the information required by the least qualified user, and by the most demanding situation, must be presented in a linear fashion to all users. The electronic data base is easily reconfigured for each user and for each situation, resulting in a more efficient transfer of information. Links to supplementary data are easily maintained so that, if necessary, additional data can be accessed.

EXTENSION OF PROOF OF PRINCIPLE SYSTEM TO FULL DATA BASE

The most difficult item involved in the extension of the proof of principle system software to cover the full EOD data base will be the definition of data structures and links for other categories of ordnance than projectiles. The definition of these structures for the projectiles in the data base required extensive participation and review by EOD specialists. An additional item of concern is that, when the data base management software must access several thousand items during a search, the seek time of the storage media will determine the response time of the system. The data structures and links will, therefore, have to be optimized in order to minimize this seek time. The algorithms for the determination of system entropy and for the presentation of the engineering drawings should present no difficulty in their application to the full data base.

ROBUSTNESS OF LEARNING FEATURE

In order to gauge the effectiveness of the learning function, several tests were conducted on the prototype system data base. The average number of characteristics required to uniquely identify an item was determined for both a uniform distribution and for several assumed initial distributions. It was assumed that all of the ordnance characteristics save one would be available during the identification process. Most ordnance items have unique markings, and if these markings can be read, they are the only characteristic needed to identify the ordnance. In order to avoid this trivial case, it was assumed that markings were not available in the identification procedure. This assumption is a reasonable approximation to the real world, since the markings are often obliterated in the firing process, or damaged or hidden on impact.

It was found that *a priori* information in the form of a non-uniform initial distribution could reduce the average number of inquiries required for unique identification by 15 to 20%. This technique did pose some danger in the form of an increased number of inquiries required to identify the outlying elements in the distribution.

ALTERNATE I/O DEVICES & TECHNIQUES

Several alternate input/output techniques are being investigated in the proof of principle system. The touch panel can easily be replaced by other pointing devices such as a mouse, trackball, joystick, or cursor control keys on a keyboard. The effectiveness of these devices depends strongly on the "computer literacy" of the user, and the touch panel seems to be the most effective for the broadest range of users. The use of voice for input and output frees the operator's hands for other tasks, and could be used as a supplemental technique for control. The use of color in the presentation of graphic and textual data offers considerable improvement in the highlighting of critical data, resulting in an increased speed for information transfer and a reduction in the number of errors.

AVAILABILITY of HARDWARE for FIELDABLE SYSTEM

The TFEL displays utilized in the proof of principle system are the result of a cooperative research and development effort by ETDL and private industry. These light weight, low power, militarized displays are commercially available as non-developmental items for incorporation into fieldable systems. Militarized CRTs are available with superior brightness, resolution, and full color capability; however the weight, power and volume required by these devices render them unsatisfactory for

man-portable applications. Work at ETDL continues on the development of TFEL displays with increased resolution, higher brightness and contrast, and full color capability.

The data volume and accessibility requirements for the EOD information retrieval system clearly indicate some form of optical storage as a solution. Although several manufacturers have announced optical drives meeting most of the military specifications required for this application, these optical drives are not yet readily available. The components and technology involved in magnetic drives are similar, and magnetic drives meeting the full environmental requirements of this application are available. The order of magnitude increase in the head to media spacing and the lower rotational speed in optical drives indicate that the development of full Mil-Spec optical drives should be possible. The temperature control of the laser diode over the full Mil-Spec range appears to be the most difficult task remaining in optical drive development.

The development of Mil-Spec optical media for these drives may be more difficult. Although the use of hermetically sealed disks or disk/drive units would ease the environmental requirements on the media, this could severely hamper the practicality of the device. The classified nature of the data base in the system requires removable media that can be easily secured and, if necessary, easily destroyed.

The choice of optical media will require tradeoffs in system functionality with respect to updates. If an erasable media were available, modified segments could simply be rewritten, with storage on the disk periodically rearranged to optimize file transfer rate by maximizing the number of contiguous segments. Optimum file structure could also be maintained through the use of CD-ROMS generated at a central processing facility, although this would restrict the frequency and availability of updates. A compromise solution would be the use of a write once media. Entire files could be rewritten, or post fields could be added to each sector. The first approach would require more space on the disk, while the second approach would degrade the data transfer rate for frequently updated records by increasing the requirements for head movement. Both approaches would decrease the storage capacity of the disk by requiring that the disks be fielded with sufficient blank space for updates. The ability to write on the disks in the field also imposes a penalty in size and power required for the drive. If a WORM media and format readable by a CD-ROM drive were developed, the teams would use the lighter, low power CD-ROM drives in the field, and the updating function could be accomplished by the team with a larger suite of equipment maintained at the detachment level.

The computing power necessary to accomplish the computational tasks, the graphic manipulations, and the text handling is available in Mil-Spec qualified components. Application specific circuits would have to be developed to permit timely access to the optical mass storage media, and to act as a high speed cache for that media.

OTHER APPLICATIONS

Army requirements for storage and retrieval of large volumes of information in a tactical environment are not limited to EOD. In order to make effective use of the Army's highly complex equipment on a rapidly changing battlefield, today's soldier requires timely access to data bases ranging in volume from megabytes to gigabytes. This information must be delivered in a form that is easily understood and quickly acted upon, preferably in a form that allows the use of computer technology to assist the soldier in the decision making process. This information must be delivered in a tactical environment by a system that is highly portable and capable of stand-alone operation.

The complexity of modern weapons systems and operational procedures has resulted in a tremendous growth in the amount of technical documentation required to support systems operation, training and maintenance. The current Army Materiel Command library of equipment publications consists of some 24,000 documents containing 2.7 million pages. These include technical manuals and technical bulletins detailing the operation and organizational use of the equipment, work requirements dealing with equipment upgrades and repair and maintenance procedures, and supply catalogs covering repair parts and special tool lists. An electronic format and delivery system for these elements offers significant advantages in the volume required for storage, in the ability to access the information required, and in the effort required for update of the information data base.

The application of paperless manuals and electronic assistance in the performance of maintenance tasks falls under the purview of the Militarized Electronic Information Delivery System (MEIDS) program, which is managed by the Project Manager, Training Devices (PM-TRADE). MEIDS will be required to be hand-portable and capable of stand-alone use. Maintenance personnel would load the system with the proper data base and diagnostic software using some high density form of storage, such as an optical disk. The diagnostic software will guide the technician in what data to obtain and what tests to run in order to facilitate the isolation of the cause of equipment failure. The hardware required for MEIDS for data storage and for the display of information, the software for the evaluation of the value of information, and the man-machine interface are very similar to those required for the EOD system. PM-TRADE has agreed

to accept the EOD system as a prototype for MEIDS, and will use the data obtained in the user evaluation of the EOD system as a guide for the specification of MEIDS requirements.

CONCLUSION

The EOD Information Retrieval System offers a significant improvement both in the efficiency and in the safety of the EOD technician in the identification of unexploded ordnance. In addition, the content of the system can be enhanced to provide improvements in the maintenance of the EOD data base, and in training and reporting procedures. Advances in the state-of-the-art in expert system software, optical devices for high-volume data storage, and flat panel displays for the presentation of high-resolution graphics and text have made it possible to effectively access large, complex data bases in the tactical environment. The hardware and software techniques used in this prototype system apply to a wide range of diagnostic and prognostic problems in the tactical environment.

3-D Model Based Machine Vision (U)

Vincent Mirelli
Center for Night Vision and Electro-Optics
Ft. Belvoir, VA 22060-5677

Introduction

The goal of model-based object recognition is to identify the presence of a given small number (e.g. two) of 3-D geometrical shapes in an image or geometrical configuration sampled from a probability space. An image is a discretization, $\{(i,j, a\,[\frac{1}{a}f(\delta i,\delta j)]) \mid (i,j) \in I^2\}$, where $I \equiv \{1,2,\ldots,n\}$ and $a,\delta \in R_+$, of a surface in R^3 represented by a function $f: R^2 \rightarrow R$ in a given cartesian coordinate frame.

This vision problem changes drastically depending on the complexity of the selected imagery database characterizing the above probability space. This complexity ranges from an imagery database obtained by discretization of all configurations of a thousand 3-D shapes confined to a 3-D box with a probability measure on the set of configurations invariant under the rotation group and the quotient group of translation modulo the group $(cZ)^3$, to some specific limit of this example in which the number of various 3-D shapes becomes infinite.

The main problems are:

1. The characterization of a given imagery database by an initial probability space (Ω,Σ,P)
2. The discovery of useful homomorphisms of this initial space into new probability spaces in order to obtain both reduction to useful information and suitable representation of this initial geometric information. More specifically, these homomorphisms are introduced in order to extract only useful geometrical information, to define the geometrical structure of the space in the most economical form, and to adopt for the vision problem the most useful geometrical representation for the configuration points of this space.

Note that while the characterization of the initial probability space is quite direct for an imagery database of low complexity, with increasing complexity of the imagery database

the mere possibility of constructing a probability space modelling the data depends critically on the above process of information reduction and economical characterization of the space geometry.

The other part of the 3-D model based vision problem is to obtain a classification map $f : \Omega \rightarrow D$, from the imagery or configuration space Ω to the classification space D, which minimizes the probability of error under the constraint that the time complexity of f is below a given value.

Research in machine vision, to date, has been mainly confined to construct these classification maps on the basis of a model for the given 3-D geometrical shapes which are required to be identified, and on the basis of a probability space for imagery of the following two extreme cases.

The case described above is typified as the probability space modelling all the configurations of a thousand 3-D shapes confined to a box. The other case, adopted for every instance of very complex imagery database, is typified by the trivial discrete probability space (I^{I^2},$P(I^{I^2})$, counting measure) carrying no structure or information at all, i.e., choosing not to make use of the specific geometrical information contained in the specific database because an effective methodology for modelling a complex database is not yet available.

The consequence of not confronting this difficult part of the machine vision problem is that the construction of classification maps f for more complex vision problems are of unacceptable performance.

There has been considerable research effort expended on developing: shape representation, invariant local features, multi-dimensional feature relational matching based only on the given shapes which are required to be identified, and etc. Our claim is that the rich geometrical structure that is present in a specific imagery database is going to transform the general machine vision problem into a specific one for which the above optimization problem is solvable.

We pause here to explain our perception of the essential difficulties in the probabilistic modelling of a specific complex imagery database. There are several steps in the construction of such imagery probabilistic models (IPM):

1. A purely phenomenological analysis of the empirical imagery data is required. Our experience, intuition and perception of the imagery phenomena must be articulated in a preliminary explicit phenomenological description. This expensive human resource step can be justified only when a certain level of theoretical capability in the second step has been demonstrated.

2. We need a formulation of a general framework for constructing geometrical spaces characterized by any consistent set of interacting properties of both local and increasingly more global kinds. The construction of shapes verifying global properties appear to be more naturally obtained by nonsmooth geometrical methods which build shapes from certain finite "bits" of space. Instead, the construction of shapes verifying local properties might be more natural in a different framework. This mismatch of frameworks requires a unifying general methodology.

3. Once a specific set of local and increasingly more global geometric properties have been identified as the relevant phenomenological ones for our vision problem, then a reformulation of this set of properties into a more economical mathematical structure is required. My guess is that the "economy" and power of the reformulation is obtained by a process of characterizing global properties by more local ones.

4. We need an effective method for constructing classification maps f which are both optimal in performance and low in complexity. The "economical" and implicit geometrical structure of a suitable image database model coupled with the methods of nonlinear functional and computational analysis gives us this methodology for constructing optimal classification maps. This methodology has already been successfully implemented in low level image processing. Here the modelling of imagery local properties is done in the Markov or Gibbs fields framework, see references below.

General Relationship Between Imagery Database & Given Targets and Imagery Model

The action of the family of observational transformations on the given set of shapes (i.e. targets), followed by the operations of cutting portions of these surfaces, generates a set A of geometrical configurations. In the construction of this set A, the observational noise has been neglected. We need an "economical" characterization of this set A. Because of the discrete nature of the observation, A is a large but finite set and the members of A are defined by a large but finite matrix of numbers from a finite set.

The role of approximation, by means of a metric or topology, in order to simplify the characterization of the set A is secondary. The essential role for an "economical" characterization of the set A is an "economical" structure β consisting of a system of relations and a small set of constants. This characterization is much more "economical" than a representation of the set A. This is analogous to a characterization of a set of functions by differential equations with initial/boundary value problems instead of representations in terms of special functions.

There is, of course, a large class of structures equivalent to β. The most suitable structure in this class with respect to a given imagery database (I.D.B.) is that structure admitting very simple substructures characterizing subsets R_i of our given imagery data such that the set R_i - A is statistically unlikely for our I.D.B. In the characterization of A by geometrical relations, those geometrical relations invariant under the observational transformations are most useful.

Given the level of observational noise, what are the length scales of the geometric structures of the given shapes which carry non-negligable amounts of information? Elsewhere, we will explain an information theoretical analysis for constructing a process to separate, from our given shapes, the various geometrical structures having different length scales and their associated information contents. Depending on the complexity of our given shapes (i.e. targets) and observational noise level, the model of our imagery database requires a geometrical structure organized on the basis of this set of length scales obtained by the above analysis.

As the number of the shapes and/or the size of a shape represented in our imagery database becomes unboundedly large, the introduction of probability on the geometrical structure modelling our I.D.B. becomes increasingly important. Thus, depending on the size of I.D.B., the associated vision problem lies somewhere on an axis, the two opposing poles of which are pattern recognition, entailing the role of probability, and structural analysis (i.e., morphology, artificial intelligence and knowledge-based techniques).

In a following section, we explain why it is necessary to introduce a probability structure in a model of a complex I.D.B. to further improve classification map performance while keeping its complexity low.

A Unifying Framework for a Complex Imagery Database

A cursory phenomenological analysis of our complex imagery database and our targets confronts us with a phenomenological description in terms of a consistent set of interdependent properties of both local and increasingly more global kinds.

Complex objects are described by local properties, and by increasingly more global properties. Examples of local properties are traditional geometric invariants (such as curvature, etc) or their discretizations. Partial differential equations or their discretizations are local descriptions. Examples of global properties might be relative position, inclusion of an object by another, set theoretic relations, contextual information (e.g., certain shapes are allowed).

A formal definition of the nature of a property with respect to its position on the continuous axis measuring the degree of locality or globality is the same as for differential equations: a property, defined by an equation $Gu = 0$, with G being a nonlinear operator, is said to have degree of globality δ whenever G verifies the following: for every $S \subset$ in the domain of functions belonging to the domain(G), and for every two functions u and v in the domain of G such that $u|_S = v|_S$, $Gu = Gv$ at every point which is at a distance $> \delta$ from the set S^c. If the space over which G is defined has a group structure, then an analogous definition is: there is a set S, such that $0 \in S$, and for every t, $u|_{S+t} = v|_{S+t} \Rightarrow Gu(t) = Gv(t)$. Informally, to verify that u is a solution of $Gu = 0$, we need to process all the data defining u, but the process can be reduced to a set of checks each entailing only a portion of data of a given fixed size δ, e.g. the function u restricted to a disk of ray δ.

Very global properties have a symbolic character because they are described by logic expressions or set theoretic relations. What is therefore needed, is the development of a mathematically consistent framework capable of formalizing structures combining these two interacting extreme elements (i.e., numerical and local versus logical and global). More precisely, what is the appropriate space of functions, which can be characterized by interdependent properties of increasingly more global nature? Once we understand such a framework we can view a complex object as a point sample in that space. By appropriate abstraction the underlying mathematical problem is very much related to constructing a unified theory for ordinary calculus and symbolic logic.

An example of how local and global properties can interact to produce efficient computation and dramatic complexity reduction is the well known multi-grid method for solution of partial differential equations. The finest grid represents the local description. The coarsest grid represents the global behavior. The interpolation and projection functions provide the linkage. The method provides memory savings and speed of computation. The speedy but inaccurate computations based on the coarse grid (global phenomena) actually speed up the more accurate computations of the fine grid (local phenomena). This is reminiscent of the interaction and transfer of control back and forth between bottom-up and top-down analysis in vision.

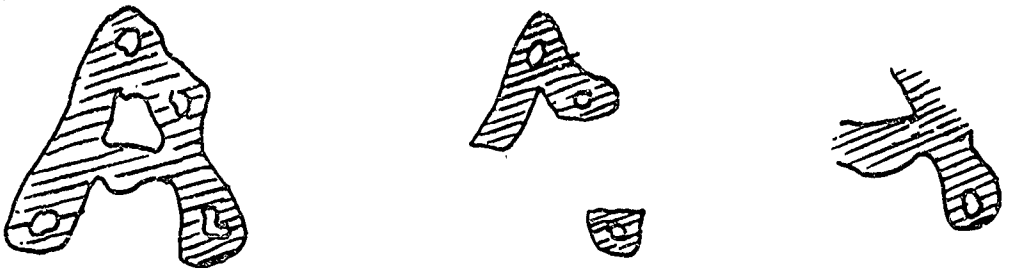
Another characteristic example is the analysis of nonlinear partial differential equations which exhibit soliton behavior. Here the local analysis, i.e. the solution of the p.d.e. is extremely complex. However, the description of the soliton (i.e. global behavior) is extremely simple. In addition, knowing that the nature of the global behavior can be described by a soliton, greatly simplifies the local analysis

In conclusion, we need to develop a coherent framework for the formulation of systems of relations or properties described above in order to characterize "economically" the principles and basic mechanisms governing image phenomena In the process of

developing such a framework, a useful initial step is to develop a framework for constructing models of a large set of idealized/simplier imagery databases. We give here just two examples of such imagery databases, but you can invent many others using your own imagination.

Examples:

1) First characterize the local and global properties, needed to describe the letter A as a two dimensional object, see figure below. Then construct the space of functions defined by the above properties. Also provide a framework that allows the description of all fragments of local and global information which can be thought of as representing a portion of a picture including the letter A . That is, the model should include in the same space:



In other words, provide a framework where subsets of local and global letter A information can be naturally described.

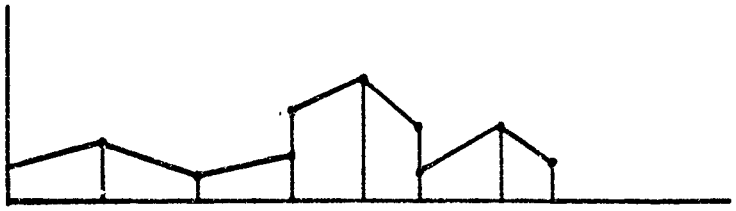
2) Describe a mathematical space of functions in R^2, which includes smooth Jordan closed curves, with the diameter of the set enclosed by the curve taking values clustering around three possible values. Additional properties of the space are the following. Each curve can contain at most two others, partitioning the plane into four connected components (generalization of Jordan property). Along any line intersecting all three, the curvature monotonically increases from the internal curve to the external one. Also between two curves, not including each other, the one with the larger diameter has everywhere larger curvature than the other. These last two relations between the local and global properties are verified with the exclusion of a small length for each curve.

Merging a Probability Structure with a Deterministic Model

We sketch an example showing how probability and geometrical structures can be co..sistently nerged in order to construct a model for a complex I.D.B. This example also shows how information can be reduced on the basis of our given targets.

The goal of model-based object recognition is to identify the presence of a given small number (e.g. two) of 3-D shapes h_θ, $\theta \in \{1,2\}$ in an image or geometrical configuration sampled from a probability space.

We can introduce a stochastic process to model imagery (i.e. a probabilistic space) in the following way. Consider a random tesselation in $[0,a]^2$. Label, with random marks, the vertices of the tesselation in such a way that only 3 vertices of each tile are marked. Marks represent heights. A piecewise linear interpolation will generate a surface. This is easily illustrated in one dimension by the following figure:



Jumps are obtained by introducing two random marks for each marked vertices of a tile. Texture or noise can be superimposed on this process. If the mark process takes value in $\overline{b} \equiv \{0,1,\ldots,b\}$, then the realization of this process is a point from the space $[0,b]^{[0,a]^2}$. A simple discretization transforms this space into the space $\Omega = \overline{b}^{\,\overline{a}^2}$, where $\overline{a} = \{0,1,\ldots,a\}$, thus obtaining the probability space $(\Omega,P(\Omega),P)$.

Let h_θ, $\theta \in \{1,2\}$, be two given closed surfaces in $[0,a] \times [0,a] \times [0,b]$ which we want to recognize.

A suitable approximating parametrization (i.e. representation problem) of these surfaces (i.e. h_θ) might be another way to construct this probability space as indicated in the following:

Let $\delta_n \equiv \delta(\alpha_1,\ldots,\alpha_n;n)$ be a parametrization of a subset H_n for approximating our surfaces h_θ. We have, of course,

$$\delta_n \to_n h_\theta$$

1. Estimation of

$$(\alpha,n) \equiv (\alpha_1,\ldots,\alpha_n,n) \in \cup_{k=1}^{\infty} A_k \times \{k\}, \text{e.g. } A_k \equiv (R^d)^k$$

by minimizing

$$\psi(d(f,\delta_n),n)$$

where $\mathbf{d}(\mathbf{f},\delta_n)$ is the distance between $\mathbf{f}$ and δ_n, and $\psi(\mathbf{d},n)$ is a function increasing w.r.t. both $\mathbf{d}$ and n. We want the parametrization of the surface to be local, with the degree of locality increasing with n. Local parametrization means:

a. A change in α_i controls a change only in a portion A_n^i of δ_n, such that $\lambda(A_n^i) \to 0$, where $\lambda(A_n^i)$ is the area of A_n^i.

b. There exists an F such that $F(\alpha_{i-1}, \alpha_i, \alpha_{i+1}, n; \delta_n)$ converges with n to the local geometric invariants of the surface.

2. If the surface is convex, then a parametrization linked with the supporting affine subspaces of the surface is a natural candidate.

3. A parametrization which is interpretable in terms of marked point processes has appeal because the structure of the problem is transferred into a well developed context.

Then, the probability space modelling images is obtained by introducing a probability distribution on the parameter space used for the approximation of the object surface. Associate with each α_i, $\alpha \in A_\kappa$, a statistical local feature or texture in terms of a mark from M. More precisely, this probability space is defined by a Markov field, conditioned for discontinuity at the boundary, having the parameter set $A \times M$ as state space, or by a conditioned-at-the-boundary Markov random measure on $A \times M$. Let us identify the space $\Omega_\infty \equiv (\overline{b} \cup \{\infty\})^{\overline{a}^2}$ with $\cup_{A \in P(\overline{a}^2)} \overline{b}^A$

Since we observe only aspects of the above given shapes $\mathbf{h}_\theta$, $\theta \in \{1,2\}$, we are going to construct the set S_θ of all possible realizations of these aspects for each given shape $\mathbf{h}_\theta$. Note that $S_\theta \subset \Omega_\infty$, and $S_\theta \cap S_{\theta'} = \varnothing$ for $\theta \neq \theta'$.

S_θ is constructed by the action of a subset of the Euclidean group G on each closed surface $\mathbf{h}_\theta$. Precisely,

$$S_\theta \equiv \{\Pi\, g\, \mathbf{h}_\theta | g \in G \wedge g\, \mathbf{h}_\theta \cap B = g\, \mathbf{h}_\theta\},$$

where, $B \equiv [0,a] \times [0,a] \times [0,b]$ is the box in which these shapes are confined, and

$$\Pi : P(B) \to \Omega_\infty$$

is a map defined by $\beta \in P(B)$, $(\Pi\beta)(i,j) = [\, inf\{K \in [0,b] | (i,j,K) \in \beta\} + \frac{1}{2}]$

We use the convention $\inf \varnothing = \infty$.

We make the simplification that images with high resolution and negligible sensor noise can model realizations of the real world, i.e. (Ω,D,P), with $D \subset P(\Omega)$, is a model of this real world. Observation is modeled by a family $\bar{\tau}_i$, $i \in I$, of transition kernels on $\Omega\times\Sigma_i$, from the measurable space (Ω,D) to the measurable space (Ω,Σ_i). Σ_i, $i \in I$, is a filtration of σ-algebras (i.e. $\Sigma_i \subset \Sigma_{i+1}$), with $\Sigma_i \subset P(\Omega)$, which models the observation information. The index I is a parametrization of a change in resolution or a global scale change of an image. If the space $(\Omega,P(\Omega))$ is sufficiently large, then a transition kernel τ_i on $\Omega\times(D\vee\Sigma_i)$ can be constructed from $\bar{\tau}_i$ such that it satisfies the properties: $\forall\ A \in D\vee\Sigma_i$, $\tau_i(\bullet;A) \in D$; $\forall\ \omega \in \Omega$, $\tau_i(\omega;\bullet)$ is a measure on $D\vee\Sigma_i$; $\forall\ \omega \in \Omega$, $\forall\ B \in D$, $\tau_i(\omega;B) = I_B\omega$, or equivalently $\tau_i(\omega;A(\omega)) = 1$, $\forall\ \omega \in \Omega$, where $A(\omega)$ is the Σ_i-atom of ω. If the space $(\Omega, P(\Omega))$ is not sufficiently large, then such a large space can always be constructed from it as shown in the next section.

The previous construction of P and S_θ characterizes an "interesting" vision problem if there is a relatively low computationally complex map $f : \Omega \equiv \overline{b^a}^2 \rightarrow N^{\{1,2\}}$, from the imagery space Ω to the classification space $N^{\{1,2\}}$ which minimizes the probability of recognition error, i.e.,

$$1 - P(\tau_i(\bullet;\cap_{\theta\in\{1,2\}}\{\omega\in\Omega \mid \#(\{\gamma\in S_\theta \mid \gamma\subset\omega\}) = f(\omega)(\theta)\})) = \varepsilon_i$$

where ε_i as function of i, is consistent with our experience about performance of human vision as function of the range or resolution.

We reemphasize, that the construction of an "interesting" P, given S_θ and τ_i, is the central vision problem.

Probability Structure of a Model Further Improves Classification Map Performance and Complexity

We simplify matters by assuming our real world is modeled by a probability space $(\Omega \equiv \{0,1,\ldots,n\}^{\{0,1,\ldots,n\}}, P(\Omega),\mu)$. We also assume that we have a single target. Thus Ω is partitioned into $\{\Omega_0,\Omega_1\}$, where Ω_1, and Ω_0 represent the sets of images with and without the target, respectively. Assume the obvious interpretation that a target is defined by a given string $a \equiv (a_1, a_2, \ldots, a_k)$ with $a_i \in \{0,1, \ldots,n\}$, and that $x \in \Omega_1$ means a is a substring of x, while $x \in \Omega_0$ means that a is not a substring of x. A formulation of the problem of optimizing the construction of a lowest complexity algorithm for deciding the membership of points of Ω to either Ω_0 or to Ω_1 could be the following. Consider a list of tests $T_i : \Omega \rightarrow \{0,1,2\}$ where $T_ix = j$ means $x \in \Omega_j$, $j \in \{0,1\}$ and $T_ix = 2$ means no decision can be made. The tests have the following properties:

1. If $j \in \{0,1\}$ then $\{x \mid T_i x = j\} \subset \{x \mid T_{i+1} x = j\}$.

   Note that 1. implies $\{x \mid T_i x = 2\} \supset \{x \mid T_{i+1} x = 2\}$.

2. The last test verifies: $T_n x \neq 2$ for all x, i.e. $\{T_n = 2\} = \varnothing$.

   Thus every T_i induces a partition of Ω into $T_i^{-1}\{0\}$, $T_i^{-1}\{1\}$, and $T_i^{-1}\{2\}$, with $T_i^{-1}\{0\} \subset \Omega_0$, $T_i^{-1}\{1\} \subset \Omega_1$.

   A special case of this scheme is obtained by adding the condition: $\forall x \in \Omega$, $T_i x \neq 1 \quad \forall i \leq n\text{-}1$.

We assume the T_{i+1} has greater complexity than T_i. The problem is to find the subsequence α_i, function of $T_k \quad k \leq \alpha_{i-1}$, which minimizes

$$\inf_\alpha \ \mu \ (\ \sum_{k=1}^{\tau} C_{\alpha_k} \ )$$

where $\tau(x) = \inf_i \{i \mid T_i x \neq 2\}$ and where C_k is the time taken to evaluate T_k at any given point $x \in \Omega$. This stochastic optimization problem is a standard one with a well-known solution. Assume now, that each point $x \in \Omega$ is partially observed, and possibly with some noise corruption. $y \in \bar{\Omega} \equiv \{0,1,\ldots,\bar{n}\}^{\{0,1,\ldots,\bar{n}\}}$ is the space of complex imagery at some level of abstraction. A transition Kernel $\tau \equiv \tau(dy \mid x)$ on $\Omega \times P(\bar{\Omega})$ models the observation of the real world. We define a new probability space $(\Omega \times \bar{\Omega}, P(\Omega \times \bar{\Omega}), P)$, where P is defined by $P(f \otimes g) \equiv \int f(x)\, \tau(g \mid x)\mu(dx)$. Define ν as the second marginal of P, i.e. $\nu(g) = P(1 \otimes g)$. Obviously, μ is the first marginal, i.e. $\mu(f) = P(f \otimes 1)$. The observation transformation τ could be $\tau(dy \mid x) = \delta_{F(x)}(dy)$, where $F : \Omega \to \bar{\Omega}$, then $\nu(g) = \mu(g \circ F)$. Since, F is highly noninjective (i.e. many points to one), the original partition of Ω into $\{\Omega_0,\Omega_1\}$ is lost; consequently, some image can be a representative for both cases, i.e. with or without the target being present. If probability information is available, even with partial observation, a most likely decision can be made. In the Bayesian framework with two hypotheses, an optimal decision exists in the family of deterministic decision rules. Thus, we generalize slightly the decision rules as follow: We introduce a filtration $(\sigma_i)_{i \in I}$, $\sigma_i \subset \{\Omega\} \times P(\bar{\Omega})$, a family D of finite stopping times with respect to this filtration, and a family $\varepsilon \equiv \prod_{k \in I} \varepsilon_k$ of statistical decisions, where ε_k is a set of measurable maps $Z_k : (\bar{\Omega}, \sigma_k) \to (\{0,1\}, P(\{0,1\}))$.

We have here the double choice of $\tau \in D$ and $(Z_k)_{k \in I} \equiv Z \in \varepsilon$. This generalized statistical decision (τ,Z) is trivially related to the previous case defined by a statistical

decision (or test) $T \equiv (T_k)_{k \in I}$ of the type: $T_k : (\bar{\Omega},\sigma_k) \rightarrow (\{0,1,2\},P(\{0,1,2\}))$; indeed : $T_k = 2I_{\{k<\tau\}} + I_{\{k \geq \tau\}}Z_k$ and $\tau(y) = \min \{k \mid T_k y \neq 2\}$. The optimization problem in this case is

$$\inf_{(\tau,Z) \in D \times \varepsilon} P(L(\theta,Z_\tau) + C_\tau),$$

where L is the loss function, and $\theta \equiv I_{\Omega_1 \times \bar{\Omega}}$. Thus, the statistical decision problem with problem predecomposition is a sequential inference problem. Sequential statistics takes into account the time required for the decision and the risk of errors. Thus, there are two types of optimal decisions to be taken:

1. the decision to terminate or continue the observation of the process at any stage based on the results of the observations made up to that stage, i.e., optimal stopping time.
2. the final decision of classification.

We have introduced here a predecomposition of the problem by means of our prechoice of a filtration. In the vision problem, the choice of filtration should be a decision which is part of the optimization problem.

It is important to note that the method of problem decomposition (or "divide and conquer") is much more complicated in the application to a statistical decision problem than to a corresponding deterministic decision problem, since, in the statistical setting, decomposition involves the new and difficult problem of sufficient statistics, i.e., let a problem A be decomposed into A_1 and A_2, and let t,t_1,t_2 be the respective sufficient statistics for these problems, then $\sigma(t_1,t_2) \subset \sigma(t)$, that is t has more information than those contained in t_1 and t_2. This is the reason the traditional decomposition of a vision problem for complex database into low level, middle level, and high level image processing is a poor approach.

By generalization of the circle of ideas underlying multigrid methods, K. Wilson's renormalization group, and nonlinear filtering with generalized filtrations, the complexity of machine vision can be systematically reduced by imbedding the probabilistic model of our image database in a family of simpler probabilistic models. Transition kernels provide the linkage among these models. A tractable iterative numerical algorithm for the optimization of our vision problem can be based on this network of interrelated I.D.B. models.

Suggestions for Asymptotic Analysis

1. Abrupt change detection:

   By exploiting the analogy to the subject of "Detection of abrupt changes in signals", search for relevant parameters for the asymptotic analysis of our laser radar problem.

2. Local asymptotic analysis:

   Choose a suitable Gaussian or Poisson type scaling for the relevant problem parameters so as to obtain the probability limit theorems for which the limiting errors of first and second type are non - degenerate.

   Obtain the asymptotic sufficient local statistics for the coordinates of the state space.

3. Large deviation asymptotic analysis:

   The performance of a machine vision algorithm is defined by the probability of error or the confusion matrix as a function of a few image features quantifying the degree of difficulty of a particular recognition problem. These image features can be derived by introducing several basic parameters, such as intensity of sensor noise, target range, and density of clutter, in a much simplified version of the imagery database model discussed here. Then, the probability of error, as a function of these parameters, can be derived for this simple model. The large deviation asymptotic analysis of this probability of error with respect to the parameters will give us the image features discussed above.

Some authors make use of smooth cost or penalization functionals to model imagery local properties instead of probability spaces. Probability measures cannot be derived from these functionals. We would like to suggest that these functionals might be interpreted as energy functionals or actions obtained by a large deviation analysis of probability measures modelling imagery database. This derivation is the analogous of asymptotic quantum mechanics when the Planck constant h tends to zero in the context of the W. K. B. theory.

Acknowledgement

Professor J. Baras for suggesting the soliton example.

Copy available to DTIC does not permit fully legible reproduction

References

(1) D. Stogan, W. S. Kendall, S. Mecke (1987) Stochastic Geometry and its Applications, John Wiley & Sons.

(2) Serra, J. (1982). Image Analysis and Mathematical Morphology. Academic Press, London.

(3) Besag, J. (1974). "Spatial Interaction and the Statistical Analysis of Lattice Systems (with Discussion)". J. Roy. Statist. Soc. Ser. B 36, 192 - 236.

(4) Besag, J. '1986). "On the Statistical Analysis of Dirty Pictures (with Discussion)". J. Roy. Statist. Soc. Ser. B 48, 259 - 302.

(5) Brown, T. C. & Silverman, B. W. (1987). "Edge Process Models for Regular and Irregular Pixels". Technical report no. 267. Stanford University.

(6) Geman, S. & Geman, D. (1984). "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". I.E.E.E. Trans. Pattern Anal. Machine Intell. 6, 721 - 741.

A Novel Laser Igniter Based on Resonant Multiphoton Excitation

*Andrzej W. Miziolek, Dr.; Brad E. Forch, Dr.; Rosario C. Sausa, Dr.
US Army Ballistic Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

INTRODUCTION: There has been a growing interest recently in the application of lasers which operate in the ultraviolet region to the ignition of reactive mixtures. One such study has described the use of the excimer lasers F_2 (157 nm) and ArF (193 nm) for the ignition of H_2/O_2 or H_2/air mixtures by photolysis of O_2 into O atoms.[1] More recently, experiments have been undertaken in which a KrF (248 nm) excimer laser has been used to photolyze O_3 into $O + O_2$ in the presence of CH_4 and H_2/O_2.[2] Both of these experiments have involved single photon photochemistry. Ignition of reactive gas mixtures which used multiphoton photochemistry was first demonstrated in our laboratory on simple hydrocarbon/air or N_2O mixtures using the ArF and KrF excimer lasers.[3] Interpretation of those results was based to a large extent on related experimental work in our and other laboratories which showed that uv laser (particularly the ArF excimer) multiphoton interaction with small carbon-containing fuels can be very extensive and can lead to substantial photofragmentation and fragment excitation. A particularly illustrative case involves the C_2H_2 molecule which upon irradiation by the ArF laser yields ground and excited state radicals such as C_2H, C_2, CH,[4] as well as the H and C atoms and C^+ ions.[5] On the basis of these results, it was not surprising to find that a C_2H_2/air mixture required only 0.25 mJ of incident laser radiation to ignite.

In a recent preliminary study we have observed a strong wavelength dependence in the amount of incident laser energy required to ignite a H_2/O_2 flowing mixture using a tunable laser system near 225.6 nm.[6] Specifically, we found that the most efficient ignition wavelength corresponded to the peak of the two-photon resonance excitation process for oxygen atoms in the J=2 ground spin-orbit state. The focused ultraviolet laser not only apparently caused photodissociation of O_2 into O atoms, but also, when on O-atom resonance, required the least amount of energy to ignite the gases. Furthermore, a plot of incident laser energy as a function of equivalence ratio yielded a minimum at 0.61, far from stoichiometry. This result further reinforced the conclusion that the

laser-oxidizer interaction is an important element in the ignition process of H_2/O_2 mixtures at 225.6 nm.

These preliminary experiments, however, were not detailed enough to identify the specific mechanism(s) involved in the ignition process. In particular, laser two-photon resonant population of the oxygen atom $3p\ ^3P$ states at 88,630 cm^{-1} (10.99 e.v.) can lead to a number of processes including excited state chemistry, heat deposition at the focal volume due to quenching collisions, and/or the absorption of a third photon leading to the formation of O^+ ions and free electrons.[7] It is the purpose of this paper to describe an experimental effort aimed at a much more comprehensive characterization of the multiphoton photochemical ignition phenomenon. The results presented here indicate that the ion formation channel is a key process since it represents an efficient and direct route for the production of the initial free electrons early enough in the laser pulse such that they become the seed material for the creation of a laser-produced spark, i.e., microplasma. The primary role of this short-lived microplasma (ca. 100 nsec) apparently is to be a localized source of highly reactive chemical intermediates at a very high temperature. If the spark is intense enough, then the resultant ignition kernel is sufficiently strong to permit transition into full combustion.

EXPERIMENTAL: The experimental schematic is given in Figure 1. Since it has been described in detail previously,[6,8] only the major points will be highlighted. Tunable uv laser radiation in the 225.6 nm region was focused with a 50 mm focal length lens at a position 1-2 mm above the burner surface. Typical laser energies up to 1 mJ/pulse yielded power densities around 10^{11} W/cm^2 in the focal volume. The water-jacketed H_2/O_2 burner was fabricated from a stainless steel Swagelock 0.25 in. terminator fitting through which a 0.9 mm hole was drilled. Matheson (Model 620) flowmeters were calibrated by a GCA Precision Scientific wet test meter for H_2, O_2, and N_2O flows up to 2 LPM. This resulted in orifice linear flow velocities in the 10^3 cm/sec range. The incident laser energies were always measured just before the focusing lens with a Scientech (Model 38-0103) disc calorimeter-power/energy meter. The emission signals were detected, averaged, and processed as described previously.[8] The excitation wavelength scans were performed manually, and each emission wavelength data point represented the average value for 512 laser shots.

Time-resolved emissions were digitized with a Tektronix 7912AD digitizer (7A24 amplifier and 7B90P timebase) and accumulated in a PDP-11/04 computer. The response time is ca. 25 nsec FWHM (see Figure 6) due to the relatively slow response of the EMI 9558QA photomultiplier detector tube. The power dependence of the O-atom emission intensities at 777.5 nm for O_2 and N_2O flows was measured using a 200 mm focal length lens to avoid the formation of microplasmas and only to measure the

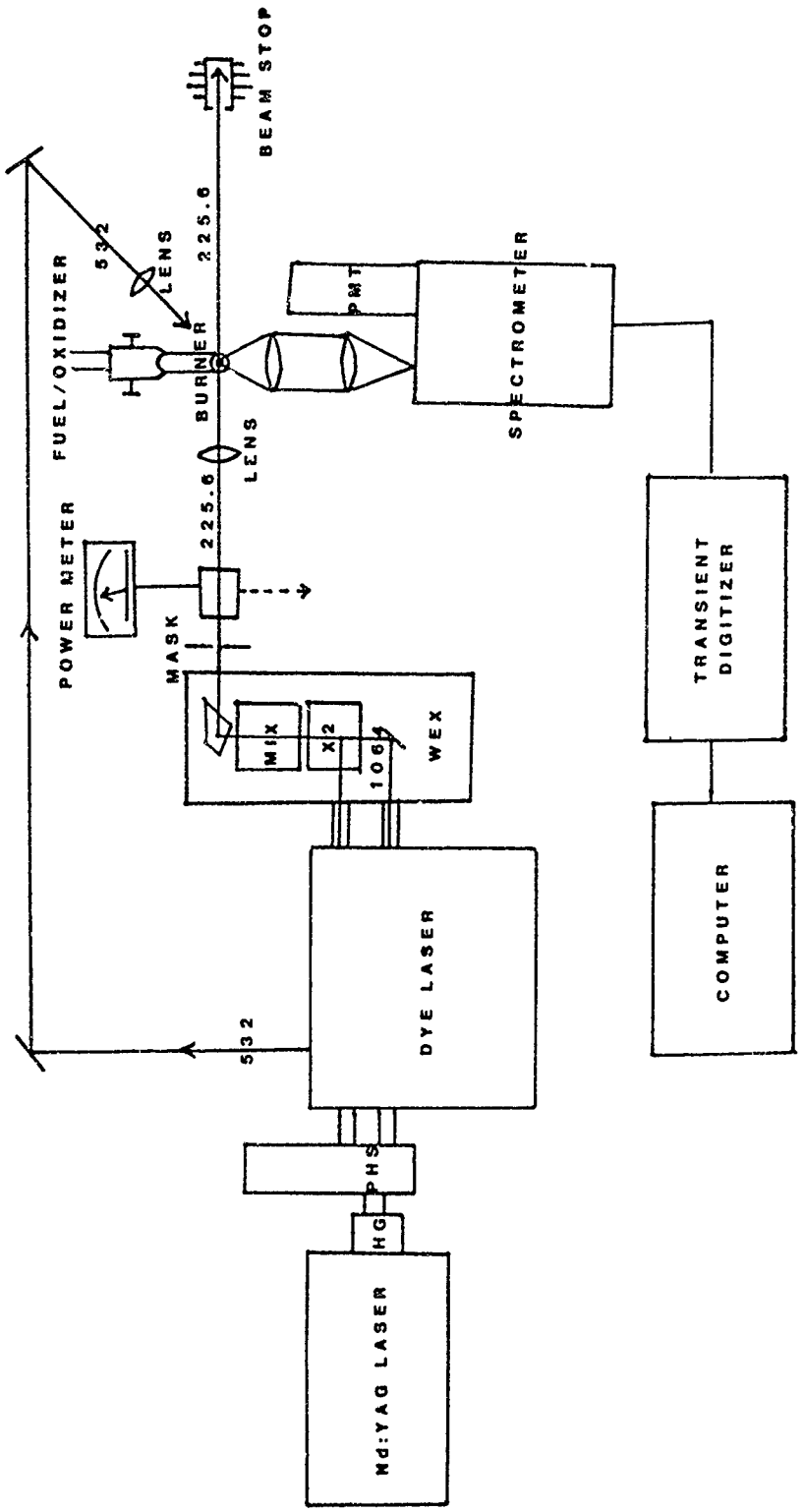


Figure 1. Experimental Schematic

photon dependence for the photolysis of those two molecules. For these experiments, the Nd:YAG laser amplifier flashlamp energy was varied, as before,[8] to change the output power at 225.6 nm.

RESULTS AND DISCUSSION:

A. Ignition: Figure 2 shows the wavelength dependence of the amount of incident laser energy necessary to ignite a premixed flow of H_2/N_2O. A similar type of behavior has been found for H_2/O_2 premixed flows. The curve clearly shows a strong dependence of the incident laser energy (ILE) on the laser wavelength with three prominent features around 225.6, 226.0, and 226.2 nm. The wavelengths of these three features are exactly the same as the fluorescence peaks which result from oxygen atom two-photon excitation of the ground electronic spin orbit split states J=2, J=1, and J=0, respectively.[7-8] This result unequivocally indicates that the electronic excitation of oxygen atoms is an important feature of the ignition mechanism. Also, the spectral widths of these features are considerably broader than those observed during the flame O-atom excitation scans. The reason for this difference will be discussed in the next section.

The dependence of the ILE on the equivalence ratio for H_2/O_2 flows is given in Figure 3. The lower trace results from the laser wavelength set at the peak of the O-atom two-photon excitation, while the upper trace is for the green laser beam, i.e., the second harmonic of the Nd:YAG laser (532 nm). Two points relevant to this figure should be discussed. The first is the observation that the minimum in the lower trace is far into the fuel-lean region. A similar type of behavior was noted for H_2/N_2O flows. The reason for this appears to be the same as before, i.e., the uv laser is clearly interacting significantly with the oxidizer (O_2) component of the reactive flow. Observation of the minimum so far from the stoichiometric point is, of course, in sharp contrast with the usual behavior found for spark ignition in a closed bomb. However, the recent report on the excimer laser ignition of H_2/O_2 mixtures also noted the most efficient ignition to be in the fuel-lean region.[1]

The second point relates to the upper trace of Figure 3 where ignition was caused by the green laser. Not only is the actual value of the ILE much higher at 532 nm than at 225.6 nm, but also the 532 nm dependence is virtually flat across a very wide range of equivalence ratios. This type of behavior has been observed previously in our initial ignition studies involving hydrocarbons[3] and can be explained by the inherent properties of the laser "spark" (gas breakdown) process. Specifically, due to the sharp threshold associated with the onset of absorption of laser energy, the spark (plasma) when produced, is typically much more energetic than the required critical ignition

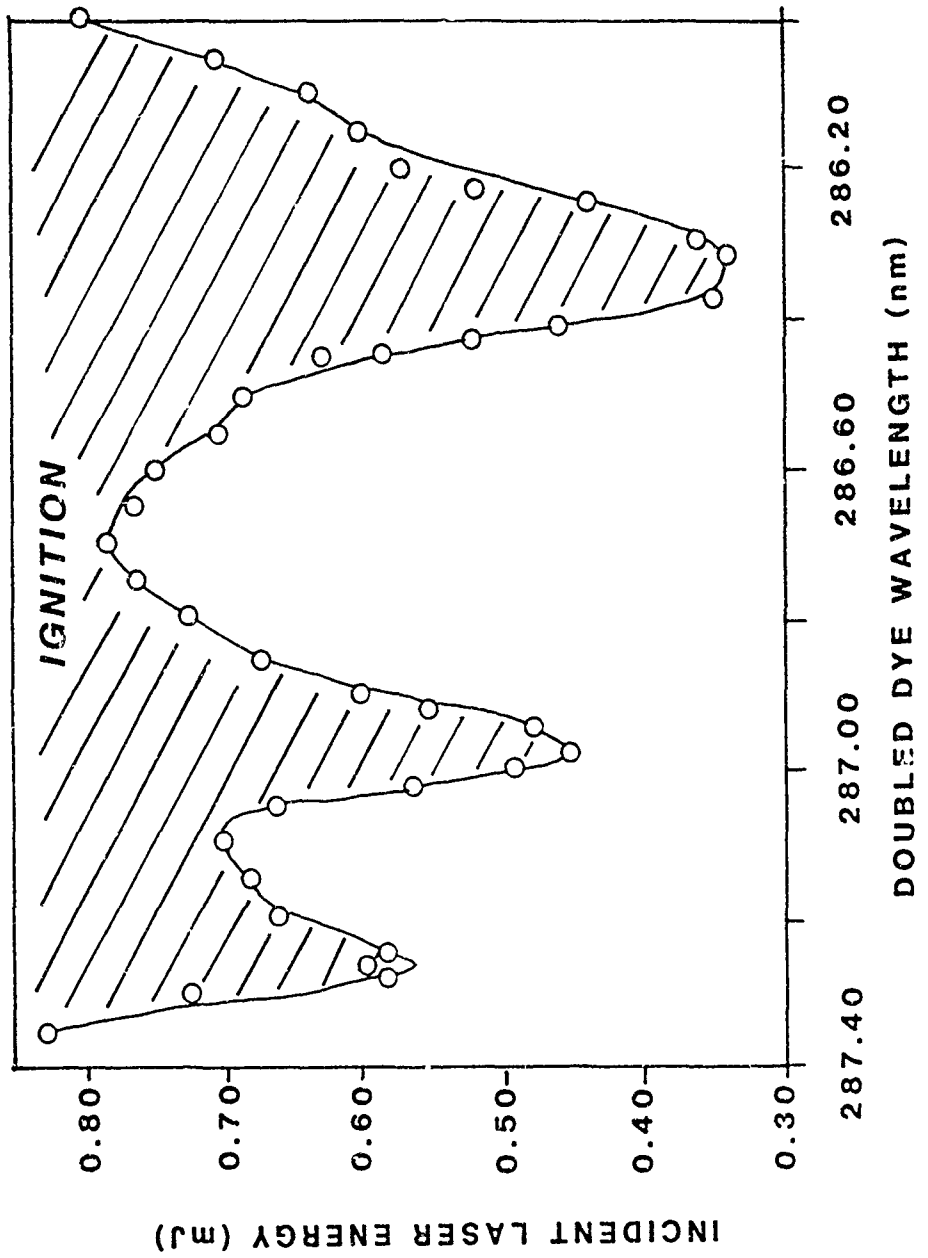


Figure 2. Incident Laser Energy Necessary to Ignite a Premixed Flow of H_2/N_2O as a Function of Laser Wavelength in the 225.6 nm (Doubled Dye +1.06 Micron) Region

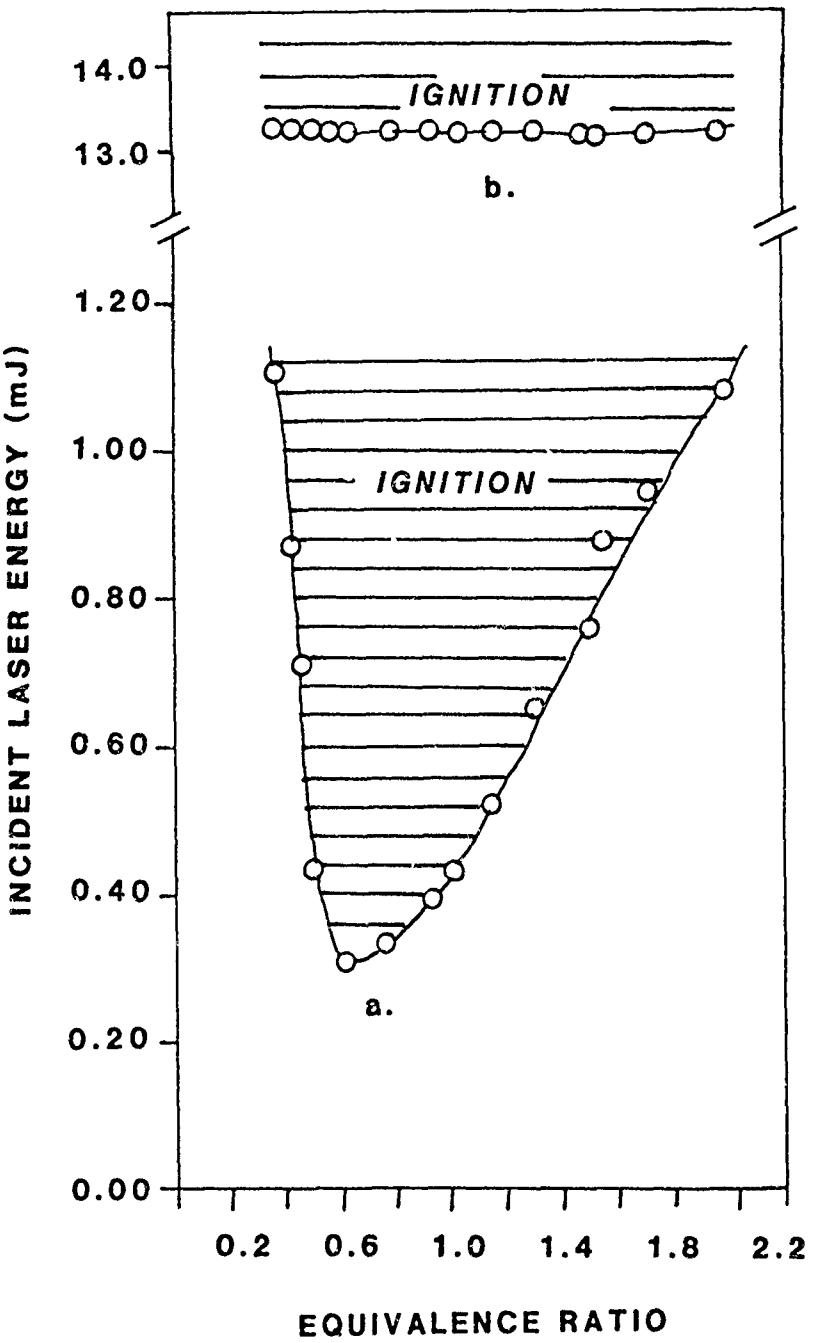


Figure 3. Dependence the Incident Laser Energy Required to Ignite a H_2/O_2 Mixture as a Function of Equivalence Ratio (a) For the Laser Wavelength Set at the Peak of the O-Atom Two-Photon Excitation, (b) For the Nd:YAG Second Harmonic (532 nm)

energy.[9] Furthermore, there is usually a sizable blast wave associated with the spark.[9] As indicated in Figure 3b, the spark intensity is sufficient to ignite mixtures at either extreme of equivalence ratios, and is clearly much greater than necessary for near stoichiometric mixtures.

Implicit in our discussions of multiphoton photochemical ignition of H_2/O_2 and H_2/N_2O is the fact that the process first has to start with the photoproduction of the oxygen atoms in the ground ^3P state. In order to study this process, we measured the laser power dependence for the production of the two-photon excited oxygen atoms whose fluorescence was detected at 777.5 nm.[8] The reason for doing this is that frequently such a power dependence study will indicate how many photons are involved in the process. Experiments were undertaken on flows of O_2 and N_2O respectively using a 200 mm focal length lens to avoid problems of microplasma formation. These measurements indicated that the photochemical formation of ground state oxygen atoms was a multiphoton process for both O_2 and N_2O requiring two photons in each case, i.e., for both cases we measured a four photon dependence for the O-atom emission. This is consistent with our previous measurement for N_2O,[8] whereas for O_2 this quantity had not been measured previously. The implication of these findings to our ignition studies will be discussed in the next section.

B. Microplasma Formation: During the course of our ignition experiments we began to take note of a faint source of white light that emanated from the laser focal volume region. The intensity of this light was clearly wavelength dependent with the brightest emission occuring at the wavelengths corresponding to the peaks of the O-atom two-photon excitation. In order to study this behavior in greater detail, we initiated both spectral and temporal studies of these microplasmas for O_2 and N_2O flows using the 50 mm focal length lens.

Figure 4a shows the excitation curve for O atom emission at 777.5 nm where the O atoms were themselves generated by the same laser focused into the O_2 flow. A similar plot (Figure 4b) is also included for nascent O atom two-photon excitation in a stoichiometric H_2/O_2 flame. Furthermore, two ignition plots for fuel-lean and stoichiometric H_2/O_2 mixtures are given in Figures 5a and b, respectively, with the differences in the absolute value of the ILE explained previously.[6] A number of similarities and differences can be seen and discussed. In particular, the spectral width for O-atoms produced in a O_2 flow is much greater than for the nascent flame O-atoms. The substantial spectral width in the O_2 flow is, however, quite comparable to the ignition profile in the fuel-lean case, while the spectral profile for the ignition of a stoichiometric mixture is somewhat narrower and less efficient than the fuel-lean case. The explanation for these

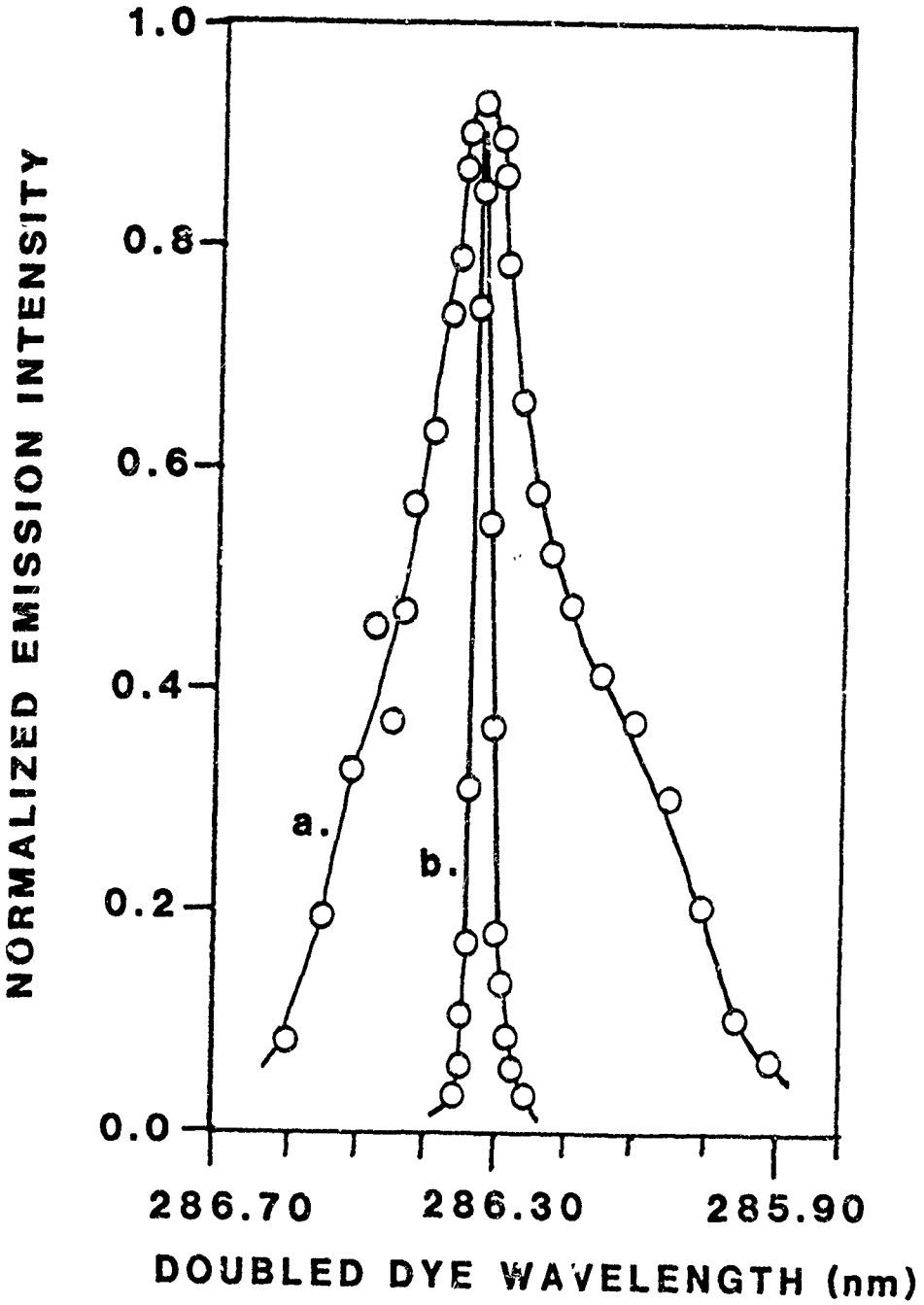


Figure 4. Excitation Curve for O-Atom Emission at 777.5 nm from (a) an O_2 Flow, (b) a Stoichiometric H_2/O_2 Flame

observations is that the spectral profile in the ignition case as well as in the case of the O_2 flow, is not really a representation of an atomic spectral property, but rather an indication of a much more complex process, i.e., the formation of a microplasma, which is inherently a highly nonlinear phenomenon. It is well-established that the laser-produced microplasma (spark) needs seed electrons in order to grow. In our case it finds them in the spectral wings of the two-photon resonant, three-photon ionization of oxygen atoms. When these free electrons are formed in the early part of the laser pulse, then the cascade plasma formation mechanism is initiated and the plasma is ultimately heated up to a very high temperature by the inverse brehmsstrahlung effect.[10-12] Thus, when this occurs, it is no longer valid to consider the O-atom emission at 777.5 nm as a simple two-photon laser induced fluorescence process. It is therefore not surprising that the spectral behavior of the ignition of premixed gases, a process sensitive to microplasma formation, should be similar to that of the microplasma producing precursor alone. This is, in fact, what is observed for the O_2 flow and the fuel-lean reactive mixture in Figures 4a and 5a. The fact that the stoichiometric ignition curve shows a narrower width and is less efficient as well, is explained by the fact that the H_2 hampers the growth of the microplasma, presumably due to its high ionization potential (I.P. for O_2 = 12.063 e.v. and for H_2 = 15.427 e.v.), and thus the plasma is relatively less intense. In search for other explanations for these wide spectral widths, a wavelength dependence of the initial step, i.e., multiphoton photochemical production of atoms, should be considered. It would be most unlikely, however, for such a dependence to yield the similar type of spectral profile which is found for all three spin-orbit components (Figure 2).

An additional parameter that was investigated is the temporal behavior of the O-atom emission at 777.5 nm. Figure 6 shows the time-resolved emission for scattered laser light (Figure 6a), flame O-atoms (Figure 6b), and O-atoms produced in the O_2 flow under conditions of microplasma formation (Figure 6c). Clearly the lifetime is much longer for the O_2 flow case since it actually is related to the lifetime of the plasma with direct laser produced signals from simple multiphoton photolysis and the two-photon excitation being a factor only in the leading edge of the trace.

The role of excited state O-atom chemistry has not been explicitly considered so far. The fact that a microplasma exists in multiphoton photochemical ignition implies that laser-populated excited state chemistry is probably not important since the kinetics of the microplasma process would appear to overtake other competing processes. Furthermore, we undertook a series of ignition studies of H_2/O_2 flows using longer focal length lenses (100 mm, 150 mm, 200 mm) and found that the ILE required to ignite increased considerably and ignition did not occ-

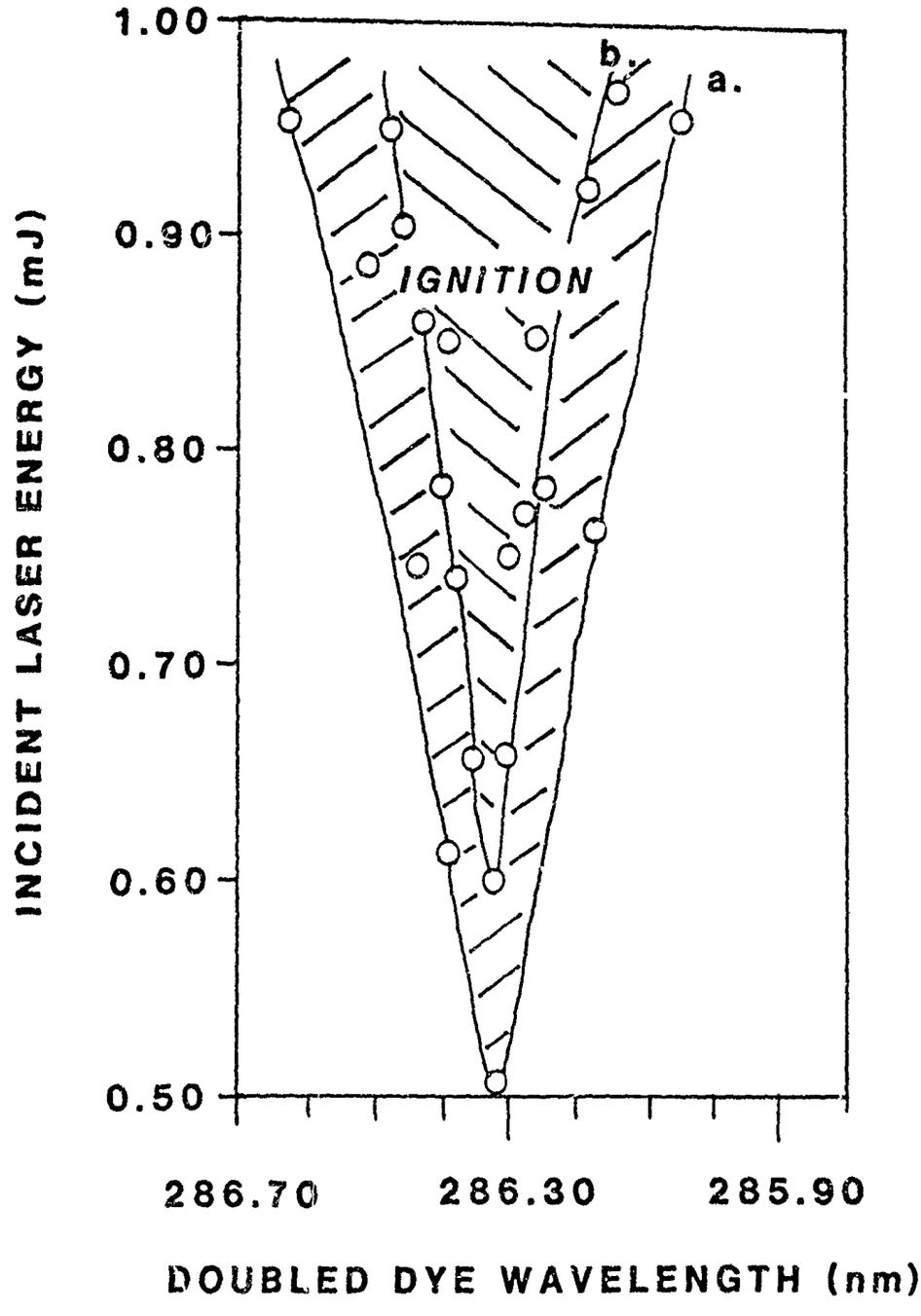


Figure 5. Incident Laser Energy Necessary to Ignite a Premixed Flow of H_2/O_2 Which is (a) Fuel-Lean (Equivalence Ratio = 0.6) and (b) Stoichiometric (Equivalence Ratio = 1)

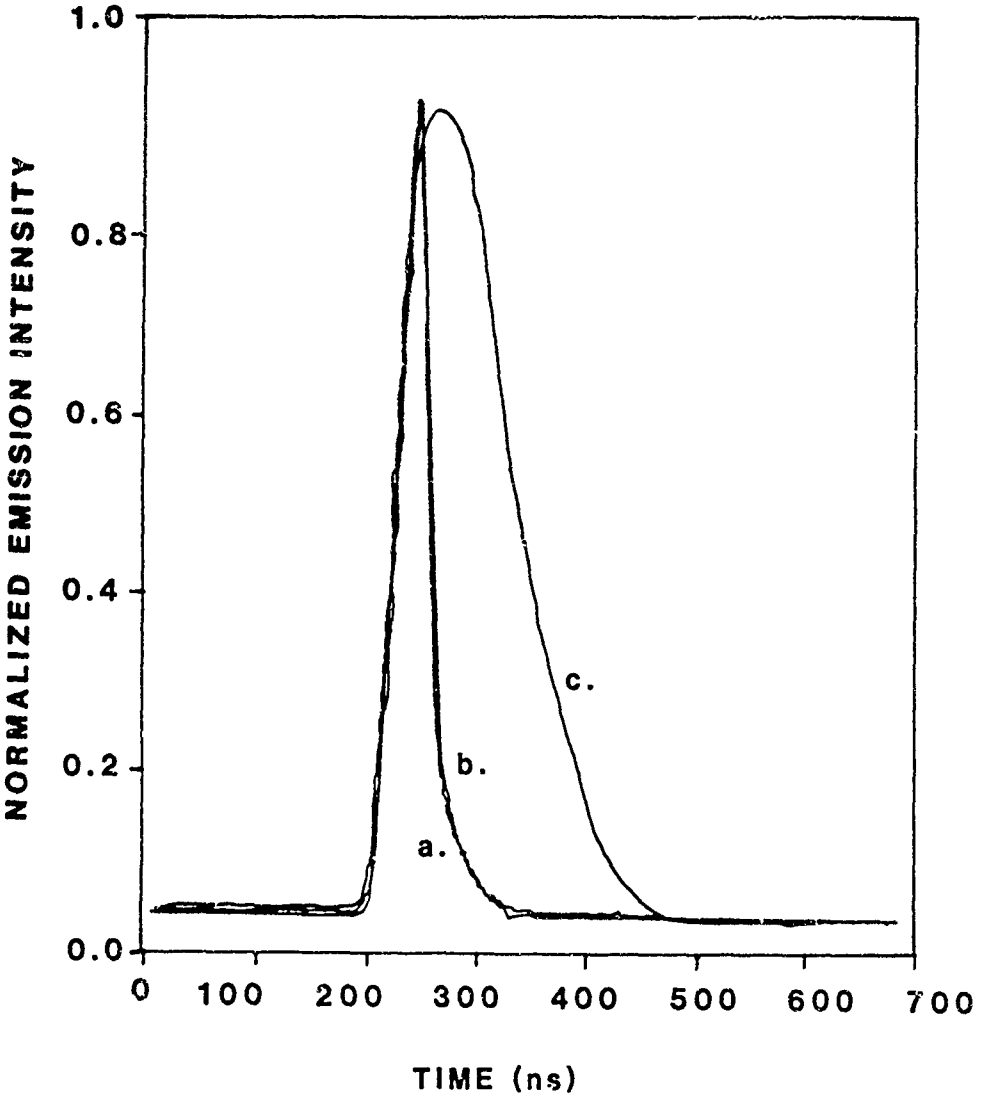


Figure 6. Time-Resolved Emission From (a) Scattered Laser Light, (b) Flame O-Atom Emission at 777.5 nm, (c) O-Atom Emission at 777.5 nm from Microplasmas Formed in O_2 Flows

without the formation of the microplasmas. On the other hand, even though our experiments strongly implicate the microplasma formation process as a key element in ignition, we cannot discount the possible importance of the photochemical formation of radicals in the converging laser beam near the microplasma. These radicals could very well be important in the early stages of ignition kernel growth, but after the microplasma had decayed. Clearly, future experiments with high speed photography, as well as in a closed bomb and with other appropriate optical diagnostic tools, would be very helpful in promoting further understanding of this phenomenon.

CONCLUSION: Focused ultraviolet laser radiation is capable of activating reactive gas mixtures through a new, previously unreported mechanism involving multiphoton photochemistry, ionization, and microplasma formation. The major difference between this work and previous work on laser spark formation is that multiphoton photochemical ignition provides a more efficient and controllable means for liberating the free electrons which then lead to the laser spark formation process. Due to these virtues, this laser ignition phenomenon should open up new opportunities in ignition studies. Also, since lasers possess certain attractive characteristics such as beam propagation through great distances, as well as excellent time-resolution, there are new opportunities for practical applications which require the activation of reactive systems. One such application is the use of this laser igniter as an ignition augmentation source for large caliber gun applications. In this situation, laser radiation would be distributed throughout the propellant bed via optical fibers to yield numerous ignition sites time-coordinated with a primary igniter. In this way, full-ignition time and location uncertainties that are frequently associated with low-vulnerability propellants would be avoided.

Another very important application is to DOD air breathing propulsion systems, particularly to "scramjet" (supersonic ramjet) engines. One of the technological barriers facing these devices is the relighting of the engines after flame-out at high Mach numbers. Photochemical laser igniters offer a solution since they are efficient and non-intrusive and therefore avoid flow-field perturbation problems posed by mechanical ignition devices.

ACKNOWLEDGEMENT: This research was supported in part by the US Army In-House Laboratory Independent Research (ILIR) program and by the US Air Force Office of Scientific Research, Contract #86-0008. The authors greatly regret the cancellation of the ILIR program in FY88 and hope that it will be restored in the near future.

REFERENCES:

1. M. Lavid and J.G. Stevens, "Photochemical Ignition of Premixed Hydrogen/Oxidizer Mixtures with Excimer Lasers," Comb. and Flame, Vol. 60, p. 195, 1985.

2. D. Lucas, Lawrence Berkeley Laboratory, private communication, 1986.

3. A.W. Miziolek and R.C. Sausa, "Photochemical Ignition Studies. I. Laser Ignition of Flowing Premixed Gases," BRL Technical Report BRL-TR-2644, 1985. Manuscript also submitted to Comb. and Flame.

4. J.R. McDonald, A.P. Baranovski, and V.M. Donnely, "Multiphoton Vacuum Ultraviolet Laser Photodissociation of Acetylene: Emission from Electronically Excited Fragments," Chem. Phys., Vol. 33, p. 161, 1978.

5. R.C. Sausa, A.J. Alfano, and A.W. Miziolek, "ArF Laser Photoproduction and Sensitive Detection of Carbon Atoms from Simple Fuel Molecules," ARBRL-TR-2798, April 1987.

6. B.E. Forch and A.W. Miziolek, "Oxygen-Atom Two-Photon Resonance Effects in Multiphoton Photochemical Ignition of Premixed H_2/O_2 Flows," Opt. Lett., Vol. 11, p. 129, 1986.

7. J.E.M. Goldsmith, "Resonant Multiphoton Optogalvanic Detection of Atomic Oxygen in Flames," J. Chem. Phys., Vol. 78, p. 1610, 1983.

8. A.W. Miziolek and M.A. DeWilde, "Multiphoton Photochemical and Collisional Effects During Oxygen-Atom Flame Detection," Opt. Lett., Vol. 9, p. 390, 1984.

9. F.J. Weinberg and J.R. Wilson, "A Preliminary Investigation of the Use of Focused Laser Beams for Minimum Ignition Energy Studies," Proc. Roy. Soc. Lond. A, Vol. 321, p. 41, 1971.

10. C.G. Morgan, "Laser-Induced Breakdown of Gases," Rep. Prog. Phys., Vol. 38, p. 621, 1975.

11. E. Yablonovich, "Similarity Principles for Laser-Induced Breakdown in Gases," Appl. Phys. Lett., Vol. 23, p. 121, 1973.

12. D.R. Cohn, M.P. Hacker, B. Lax, and W. Halverson, "Effects of Pressure and Magnetic Field Upon Physical Processes in Laser-Induced Gas Breakdown," J. Appl. Phys., Vol. 46, p. 668, 1975.

The Competitive Edge in Organic Synthesis: Improved Procedures for the Preparation of Important Compounds Containing Fluorine (U)

*August J. Muller, Dr.
David I. Rossman, Mr.
Chemical Research, Development and Engineering Center
Aberdeen Proving Ground, MD 21010-5423

INTRODUCTION

The Chemical Research, Development and Engineering Center (CRDEC) has an ongoing responsibility to prepare and purify samples of the alkyl phosphonofluoridates and other compounds containing fluorine for use in its CW detection, decontamination and physical protection research programs. Additionally, periodic routine laboratory syntheses are required in our research as an alternative to storing large quantities of highly toxic chemicals and managing the consequent safety and surety hazards associated with their storage and shipment. Consequently, simple yet safe, high yield, on-site preparations of these compounds are highly desirable.

The method that is usually employed to prepare the alkyl phosphonofluoridates is based upon the reaction of an alcohol with methylphosphonic difluoride (1) in the presence of an HF acceptor such as a tertiary amine (Eq. 1).[1] While this reaction provides good yields of the phos-

$$CH_3\text{-}\overset{O}{\overset{\|}{P}}\text{-}F_2 + ROH \xrightarrow{R'_3N} CH_3\text{-}\overset{O}{\overset{\|}{P}}(OR)\text{-}F + R'_3\overset{+}{N}H\,F^- \qquad \text{Eq. 1}$$

1

phonofluoridic esters, there are several required experimental steps that diminish the ease of performing this reaction. For example, the addition of the alcohol to 1 must be carefully controlled as the solution is cooled in order to moderate the heat from the reaction. Also, the ammonium fluoride salt that is formed necessitates a separate purification operation to effect its removal. It is evident that if this by-product were not

formed, these procedures would be facilitated since an extra purification operation would not be required.

Another reaction that is sometimes used to prepare the alkyl phosphonofluoridates involves the controlled addition of an alcohol to an equimolar mixture of 1 and methylphosphonic dichloride (2) dissolved in refluxing methylene chloride:[2]

$$\underset{\underline{1}}{CH_3\text{-}\overset{O}{\overset{\|}{P}}\text{-}F_2} + \underset{\underline{2}}{CH_3\text{-}\overset{O}{\overset{\|}{P}}\text{-}Cl_2} + ROH \xrightarrow{CH_2Cl_2} CH_3\text{-}\overset{O}{\overset{\|}{\underset{OR}{P}}}\text{-}F \qquad \text{Eq. 2}$$

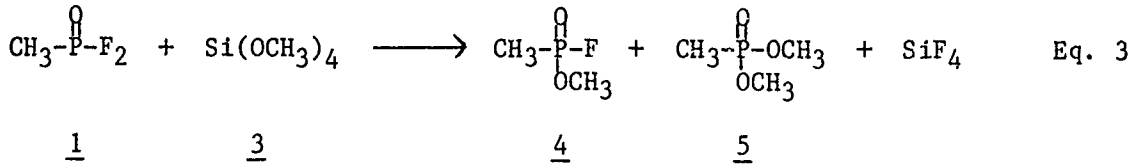
The dropwise addition of the alcohol and the necessity of removing the solvent from the fluoridic ester product are disadvantages which make this reaction less attractive as an expedient routine laboratory scale method.

These disadvantages inherent in the existing synthetic procedures, have led us to consider alternative approaches to the preparation of these compounds which minimize the difficulty of the reaction work-up scheme that is necessary to provide high yields of pure phosphonofluoridic ester products.

In the chemistry of organosilicon compounds, the reactions in which Si-F bonds are formed are usually highly exothermic. In reactions in which silicon tetrafluoride is formed, the Si-F bond energy[3] of 160 kcal/mol provides the driving force for these reactions. In view of the affinity of silicon for fluorine, and the lability of the fluorine atoms of 1, it seemed logical to investigate the potential reactivity of alkoxy-substituted silanes with 1. It seemed reasonable to suppose that a ligand exchange reaction with P-F systems might occur by a simple redistribution mechanism.

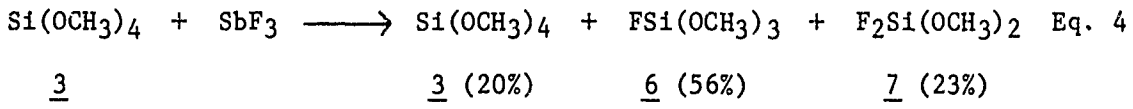
RESULTS AND DISCUSSION

Because methylphosphonic difluoride (1) and tetramethoxysilane (3) were available in our labs, these reagents were used to explore this postulated reaction. Thus, 1 and 3 were combined at room temperature to give approximately a 3:1 molar ratio of reactants in favor of the difluor 1. The reaction mixture remained visually unchanged for about fifteen minutes before a gaseous product evolved. A ^{31}P NMR spectrum of the resulting reaction mixture revealed the presence of methyl methylphosphonofluoridate (4) together with dimethyl methylphosphonate (5) in 79% and 19% yields, respectively. It is evident that all four of the methoxy groups of 3 are replaced by fluorine in this exchange reaction with 1 and that silicon tetrafluoride is the gaseous product evolved (Eq. 3).



Eq. 3

In contrast to these results in which a molar excess of 1 is employed, equimolar amounts of 1 and 3 react at a lower rate to give 4 and 5. For example, a 10:1 mixture of products in favor of 4 was observed after 2.5 hours, and a complete conversion to 5 was realized after 20 hours. These results may indicate that the fluorinated methoxy silane intermediates are more reactive toward an exchange reaction with the phosphoryl fluoride (1 or 4) than is 3. In order to test this hypothesis, the fluorinated methoxy silane derivatives were synthesized. A mixture consisting of 20% 3, 56% trimethoxyfluorosilane (6), and 23% dimethoxydifluorosilane (7) was obtained upon treating 3 with antimony trifluoride (Eq. 4). Trifluoromethoxysilane should also have been formed, but it



Eq. 4

reportedly[4] disproportionates to give 3 and silicon tetrafluoride.

A reaction was then conducted with 1 and a sufficient quantity of the mixture of silicon compounds to provide one exchangeable methoxy group for each molecule of 1. The resulting reaction employing the fluorinated methoxy silanes was indeed faster than that between 1 and 3, and a reduced reaction time of under four minutes was observed. In this case, the product mixture contained approximately a 4:1 mixture of 4 and the unreacted difluor 1, as evidenced by ^1H NMR spectroscopy.

As part of our attempts to determine the mechanism of this reaction, a number of potential catalysts were screened. Among the reagents tested for catalytic activity were potassium fluoride/18-crown-6, pyridine-HF, freshly distilled boron trifluoride etherate, triethylamine and methyl alcohol. In these trials a 4:1 molar ratio of 1 to 3 was employed, because the end of the reaction could be visualized by the emission of SiF_4. In all cased, instead of increasing the rate of reaction, the presence of the potential catalysts delayed the emission of SiF_4 (end of reaction) from 15-20 minutes (absence of added catalyst) to 25-40 minutes. In these experiments it was noted that a rise in the reaction temperature to 55-65 ^oC occurs simultaneously with the emission of SiF_4.

The presence of water, however, causes a marked acceleration in the rate of this reaction. For example, when a 4:1:1 molar ratio of 1:3:H_2O was combined, the emission of SiF_4 occurred almost instantaneously. The reaction temperature rose to just over 100 °C and within 15 seconds began to fall. The yield of 4, as determined by ^1H NMR spectroscopy, was over 90%. A possible mechanism that would account for these results is suggested in Equations 5-7:

$$\underset{\underline{3}}{Si(OCH_3)_4} + H_2O \longrightarrow \underset{\underline{8}}{(CH_3O)_3SiOH} + CH_3OH \qquad \text{Eq. 5}$$

$$\underset{\underline{1}}{CH_3-\overset{O}{\overset{\|}{P}}-F_2} + CH_3OH \longrightarrow \underset{\underline{4}}{CH_3-\overset{O}{\overset{\|}{\underset{F}{P}}}-OCH_3} + HF \qquad \text{Eq. 6}$$

$$HF + \underset{\underline{8}}{(CH_3O)_3SiOH} \longrightarrow \underset{\underline{6}}{(CH_3O)_3SiF} + H_2O \qquad \text{Eq. 7}$$

In this mechanism, methanol is produced by the hydrolysis of 3 (Eq. 5) and reacts with 1 to give 4 and HF (Eq. 6). The HF dehydrates trimethoxysilanol (8) that is formed in Eq. 5 resulting in the regeneration of H_2O and the formation of 6 (Eq. 7). The fluorosilane 6 then reacts with water and continues this cycle until SiF_4 is formed.

In contrast to the activating effect of silyl fluoro substituents, the presence of three methyl groups on silicon deactivates trimethylmethoxysilane (9) toward the exchange with 1 (Eq. 8). It was necessary

$$\underset{\underline{1}}{CH_3-\overset{O}{\overset{\|}{P}}-F_2} + \underset{\underline{9}}{(CH_3)_3SiOCH_3} \xrightarrow[18\text{-crown-6}]{KF} \underset{\underline{4}}{CH_3-\overset{O}{\overset{\|}{\underset{OCH_3}{P}}}-F} + \underset{\underline{5}}{CH_3-\overset{O}{\overset{\|}{\underset{OCH_3}{P}}}-OCH_3} \qquad \text{Eq. 8}$$

to utilize a catalytic amount of a KF saturated acetonitrile solution containing 1% (w/v) 18-crown-6 in order to obtain any evidence of a

reaction by ^1H NMR. In this experiment a 1:1 molar ratio of 1 and 9 resulted in an 86% yield of 4 after 6 days, together with a 5% yield of 5.

The reactivity of trimethoxysilane (10) with 1 is similar to that of the tetramethoxy silane 3. For example, when a 3:1 molar ratio of 1 to 10 was combined, the reaction mixture emitted a gaseous product (presumably $HSiF_3$) after 10 minutes. A ^1H NMR spectrum of the resulting solution revealed the formation of 4 in an 80% yield, together with a 10% yield of 5 and unreacted 1 (10%) (Eq. 9).

$$\underset{\underline{1}}{CH_3\text{-}\overset{O}{\overset{\|}{P}}\text{-}F_2} + \underset{\underline{10}}{HSi(OCH_3)_3} \longrightarrow \underset{\underline{4}}{CH_3\text{-}\overset{O}{\overset{\|}{P}}(F)(OCH_3)} + \underset{\underline{5}}{CH_3\text{-}\overset{O}{\overset{\|}{P}}\text{-}(OCH_3)_2} \qquad \text{Eq. 9}$$

The high yield of 4 obtained from this newly discovered reaction between 1 and 3 led us to attempt to apply the same novel synthetic methodology to a new synthesis of isopropyl methylphosphonofluoridate (11). For this phosphonofluoridic ester, the requisite silane, tetraisopropoxy silane (12), was employed. In contrast to the uncatalyzed reaction between 1 and 3, which was complete after 15 minutes, the reaction of a 4:1 molar ratio of 1 to 12 contained 74% of unreacted 1 even after 6 hours, as shown by ^1H NMR. A ^1H NMR spectrum of the reaction mixture after 3 days revealed that 11 was obtained in an 85% yield (Eq. 10).

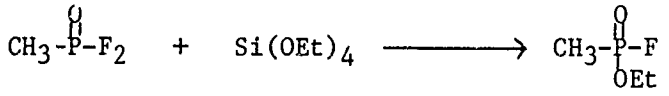
$$\underset{\underline{1}}{CH_3\text{-}\overset{O}{\overset{\|}{P}}\text{-}F_2} + Si(Oi\text{-}Pr)_4 \longrightarrow \underset{\underline{11}}{CH_3\text{-}\overset{O}{\overset{\|}{P}}(F)(Oi\text{-}Pr)} \qquad \text{Eq. 10}$$

The catalysis of this reaction was also studied. Initially, an amount of water equimolar to 1 was employed. Upon combining these reagents, the resulting solution immediately emitted a gaseous product in an exothermic reaction. However, the ^1H NMR spectrum of the solution revealed that the fluor-acid 13 had been formed in over 90% yield (Eq. 11).

$$\underset{\underline{1}}{CH_3\text{-}\overset{O}{\overset{\|}{P}}\text{-}F_2} + \underset{\underline{12}}{Si(Oi\text{-}Pr)_4} + H_2O \longrightarrow \underset{\underline{13}}{CH_3\text{-}\overset{O}{\overset{\|}{P}}(F)(OH)} + \underset{\underline{11}}{CH_3\text{-}\overset{O}{\overset{\|}{P}}(F)(Oi\text{-}Pr)} \qquad \text{Eq. 11}$$

Using less water, an amount equimolar with the silane 12, a reaction also ensued that was exothermic for a period of about 4 minutes. A ^{1}H NMR spectrum obtained 10 minutes after the reagents were mixed revealed that the reaction mixture was composed of approximately 60% of 11, 20% of unreacted 1 and 19% of the acid 13, i.e. the hydrolysis product of 1. Other attempts to catalyze this reaction using DAST (diethylaminosulfur-trifluoride), tetrabutylammonium hydroxide (40% aqueous solution), or isopropanol were unsuccessful. Therefore, a time period of a few days is necessary to utilize this reaction for a high yield synthesis of 11.

Ethyl methylphosphonofluoridate (14) was also prepared in high yield using this reaction. For example, 1 and tetraethoxysilane (15) were combined in a 4:1 molar ratio and allowed to react for 1.5 hours. At this time, the end of the reaction was indicated by a rise in the temperature of the reaction mixture together with the emission of silicon tetrafluoride as the solution darkened. Subsequent distillation afforded the desired product that was 94.5% pure by GLC (Eq. 12). The presence of 4.6%



Eq. 12

1 15 14

of unreacted 1 in the product suggests that an improved yield of 14 might be obtained if slightly more of the ethoxy silane 15 were used so that the molar ratio of 1 to 15 would be less than 4:1.

In summary, the uncatalyzed reactions of the three tetraalkoxy silanes (3, 12 and 15) with 1 proceed at different rates. Tetramethoxysilane (3) is the most reactive toward 1, whereas tetraisopropoxysilane (12) is the least reactive. Steric interactions, which develop in the transition state leading to the alkyl phosphonofluoridate products, can be invoked to explain the observed order of relative reactivity. The results of all the reactions of 1 with the silicon esters are summarized in Table 1.

Of the elements below silicon in group IVA of the periodic table, the organic compounds of tin have received the most attention. Thus, we decided to explore the potential use of organotin alkoxides in exchange reactions with compounds bearing labile fluorine atoms, such as the difluor 1. Upon mixing an equimolar amount of 1 and tributyltin methoxide (16), an exothermic reaction immediately ensued as a solid material formed in the reaction mixture. After an extraction of the mixture with carbon tetrachloride, a subsequent ^{31}P NMR analysis revealed

Table 1. Reactions of Methylphosphonic Difluoride (1) with Alkoxysilanes

Alkoxysilane	Ratio[a]	Products (% yield)[b]
3	1:3	4 (79), 5 (19)[c]
3	1:1	5 (>95)
3 + 6 + 7	1:1[d]	4 (80), 1 (20)
3 + H_2O[e]	1:4	4 (92), 5 (6)
9	1:1	4 (86), 5 (5)
10	1:3	4 (80), 5 (10), 1 (10)
12	1:4	11 (85)
12 + H_2O[f]	1:4	13 (90)
12 + H_2O[e]	1:4	11 (60), 13 (19), 1 (20)
15	1:4	14 (94.5)[g]

[a] Molar ratio of alkoxysilane to 1. [b] Product yields by ^1H NMR. [c] Product yield by ^{31}P NMR. [d] Equimolar ratio of total methoxy groups to 1. [e] Equimolar ratio of H_2O to the alkoxy silane. [f] Equimolar ratio of H_2O and 1. [g] Product purity by GC.

that the solution was composed of 67% of 5, 10% of 4 and 19% of unreacted 1. However, since the solid by-product (tributyltin fluoride) may contain some of the phosphorus compounds, this composition may not represent true product yields. In any event, it is apparent from the high percent yield of the diester formed in this reaction that 16 is much more reactive toward the exchange reaction with 1 and 4 than are the alkoxy silanes. Therefore, this reaction does not appear useful for the routine synthesis of the alkyl phosphonofluoridates because of the secondary reaction with the alkyl phosphonofluoridate to give the diester product.

We reasoned that perhaps the high reactivity of 16 would enable this compound to undergo an exchange reaction with less reactive molecules bearing labile fluorine atoms. One such compound that has received much attention[5] is perfluoroisobutene (17). The reactions of unsaturated fluorocarbons are very different from their hydrocarbon analogs. Whereas the chemistry of olefinic hydrocarbons is dominated by reactions with electrophiles, the perfluorinated olefins are reactive toward nucleophiles. There are two main types of reactions which perfluorinated olefins undergo in the presence of nucleophiles. One type is the addition-elimination reaction in which the nucleophile adds to the more electropositive carbon of the double bond to give an intermediate bearing a negative charge which eliminates fluoride to give the product bearing the nucleophile (Eq. 13). The other type involves a reaction with a protic nucleophile to give a dipolar intermediate which undergoes an internal proton transfer to afford the product resulting from the effective addition of HNu across the double bond (Eq. 14).

$$(CF_3)_2C{=}CF_2 \quad + \quad N\bar{u} \longrightarrow (CF_3)_2\bar{C}\text{-}CF_2Nu \xrightarrow[-F^-]{} (CF_3)_2C{=}CFNu \qquad \text{Eq. 13}$$

17 Nu = Nucleophile

$$(CF_3)_2C{=}CF_2 \quad + \quad HNu \longrightarrow (CF_3)_2\bar{C}\text{-}CF_2\overset{+}{N}\overset{+}{u}H \longrightarrow (CF_3)_2CH\text{-}CF_2Nu \qquad \text{Eq. 14}$$

17

We found that 16 did indeed react with 17 to give the product resulting from an addition-elimination reaction. The product 18 bearing the methoxy group was formed in a 62% isolated yield upon adding 16 to a solution of 17 in xylene at 0 - 5 °C followed by purification by distillation (Eq. 15).

$$(CF_3)_2C{=}CF_2 \quad + \quad n\text{-}Bu_3SnOCH_3 \longrightarrow (CF_3)_2C{=}CFOCH_3 \qquad \text{Eq. 15}$$

17 16 18

This reaction was also applied in a new synthesis of perfluoro-β,β-dimethylacrylonitrile (19) through the use of tributyltin cyanide (20) in a reaction with 17. The nitrile 19 is a difficult compound to synthesize, and it has been the subject of a recent synthetic study at CRDEC.[6] We were able to synthesize 19 in an isolated yield of 43% using reaction conditions similar to those employed in the synthesis of 18 (Eq. 16). It is not clear why the yield of 19 is lower than that of 18. Perhaps 20 is less reactive with 17 than is 16 because even when a 20% excess of 20 is employed in the reaction, a significant amount of 17 could be recovered from the reaction mixture.

$$(CF_3)_2C{=}CF_2 \quad + \quad n\text{-}Bu_3SnCN \longrightarrow (CF_3)_2C{=}CFCN \qquad \text{Eq. 16}$$

17 20 19

The vinyl ether 18 is a valuable intermediate leading to useful monomers in polymer chemistry. A recent report[7] describes the synthesis of 18 by a two step sequence involving the addition of methyl alcohol across the double bond of 17 to give 1-methoxy-2-hydroperfluoroisobutane (21) followed by the dehydrofluorination of this compound with sodium hydroxide to afford 18 (Eq. 17). It is apparent that the reaction shown in Eq. 15 is a more facile synthesis of 18 because it is an addition-

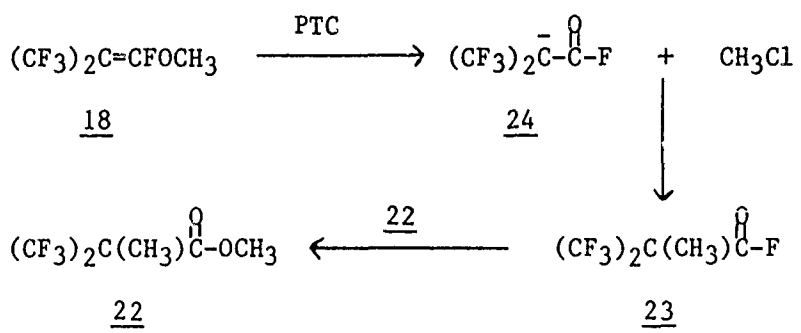
$$(CF_3)_2C{=}CF_2 \;+\; CH_3OH \longrightarrow (CF_3)_2CHCF_2OCH_3 \xrightarrow{NaOH} (CF_3)_2C{=}CFOCH_3 \qquad \text{Eq. 17}$$

17 21 18

elimination reaction that gives the product in one step.

The enhanced reactivity of organotin alkoxides over their silicon based analogs is also apparent from the reaction of 17 with methoxy trimethylsilane (9). In order to obtain products from the reaction of 17 with 9 it was necessary to heat the two compounds together in a sealed high pressure reactor at 180 °C. Although the presence of the vinyl ether 18 was confirmed in the product mixture, the major product was identified as methyl α-methylhexafluoroisobutyrate (22). This compound is likely formed from an exchange reaction of 9 with α-methylhexafluoroisobutyryl fluoride (23). The acid fluoride 23 has been reported[7] to result from the chloride ion induced rearrangement of 18. This rearrangement, induced by a phase transfer catalyst (PTC) such as trimethylbenzylammonium chloride, involves the demethylation of 18 to give methyl chloride followed by a methylation of the intermediate anionic specie 24 with methyl chloride or 18 to result in the formation of the acid fluoride 23 (Scheme I). In this system the catalyst could be the fluoride anion resulting from the decomposition of one of the fluorine bearing compounds in the reaction or product mixture.

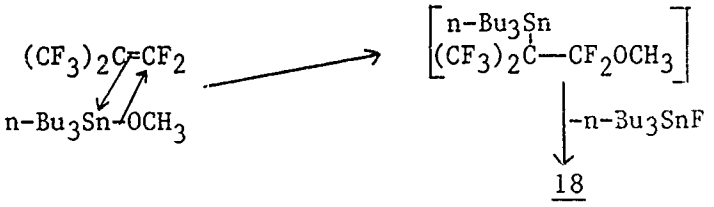
Scheme I



Although the reactions of 1 and 17 with the silicon and tin alkoxides appear similar, the mechanisms leading to the products are likely to be different. A reasonable mechanism to account for the reaction of 16 with 17 involves a nucleophilic attack by the methoxy group of 16 on the more electropositive carbon atom of 17 together with the addition

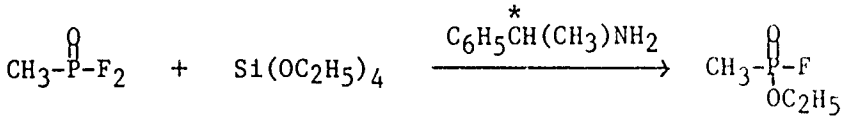
of the tributyltin moiety to give an intermediate which eliminates tributyltin fluoride and affords the product 18 (Scheme II). Low temperature NMR experiments are planned to attempt to observe possible intermediates in order to gain insight into the mechanism of this reaction.

Scheme II



In the case of the reaction of 1 with the silicon and tin alkoxides, however, we must account for the observed catalytic effect of water on the rate of the exchange reaction. It is important to note that alcohols do not react with 1 to form the alkyl phosphonofluoridates nearly as fast as the water catalyzed reaction between 1 and 3 takes place. Therefore, the enhanced rate observed in this system is most likely due either to the removal of HF by a dehydration reaction with a silanol, or to the formation of an activated complex between water and 1 which is highly reactive with an alcohol or, in the event of a direct substituent exchange, with the alkoxy silane.

We recognized that if 1 could form an activated complex with water then it should also form an activated complex with primary amines. We also hypothesized that since the fluorine atoms of 1 are enantiotopic, i.e. the phosphorus atom is a prochiral center, then employing a chiral primary amine as a catalyst instead of water might result in the formation of chiral alkyl phosphonofluoridates, as well as provide insight into the mechanism of this reaction. In order to ascertain the catalytic effect of a primary amine on this reaction, and to determine the influence of a chiral amine on the chirality of the alkyl phosphonofluoridate product, a reaction was conducted using 1 and tetraethoxysilane (15) together with (S)-(-)-α-methylbenzylamine as the chiral catalyst in a 4:1:1 molar ratio of 1:15:amine (Eq. 18). The



Eq. 18

1 15 14

amine did indeed serve as a catalyst in this reaction, since a reduced reaction time from 1.5 hours (absence of catalyst) to 10 minutes was noted. Unfortunately, after distilling the product, and dissolving a known amount in carbon tetrachloride, no optical activity could be measured.

Since no evidence for an activated complex intermediate could be obtained from this experiment, we can only account for the observed catalytic effect of the amine by assuming that a mechanism similar to the one proposed in Eqs. 5-7 is operative. Of course, in the case of amine catalysis, the hydrolysis step in Eq. 5 would be replaced by an aminolysis of the tetraalkoxysilane, and HF would cleave the Si-N bond to regenerate the catalyst and form the fluorinated silane. We hope to employ low temperature NMR spectroscopy to pursue these mechanistic studies in the future.

EXPERIMENTAL

General Procedures. Proton magnetic resonance spectra were obtained on a Varian EM-360A spectrometer, generally with carbon tetrachloride as the solvent, with tetramethylsilane as an external standard. ^{31}P and ^{19}F NMR spectra were recorded on a Varian FT-80A or a Varian XL-200 spectrometer. The external references for the ^{31}P and ^{19}F NMR spectra were 85% phosphoric acid and CCl_3F, respectively. A positive chemical shift value is taken downfield from the external reference. GLC separations were accomplished using a Hewlett Packard 5880A gas chromatograph equipped with an Analabs 6' x 1/8" OD 10% silastic LS-420 on Anakron Q (80/100 mesh) stainless steel column or a Varian 6000 gas chromatograph equipped with a 10% Silastic 420 (DC-LSX-3-0295) on Gas Chrom Q-II (80/100 mesh) 12' x 1/8" OD stainless steel column. Optical rotations were measured on a Perkin-Elmer 141 autopolarimeter. Tetramethoxysilane (3), trimethylmethoxysilane (9), trimethoxysilane (10) and tetraethoxysilane (15) were used as received from Petrarch Systems (Bristol, PA). Tri-n-butyltin methoxide (16) and tri-n-butyltin cyanide (20) were used as received from Aldrich. Perfluoroisobutylene was used as received from Flura Corporation (Durham, NC). Methylphosphonic difluoride (1) was distilled from dicyclohexylcarbodiimide prior to use.[8] Tetraisopropoxysilane (12) was prepared by Regis Chemical Company[9] according to the method of Sumrell and Ham.[10]

Preparation of the Fluorinated Methoxy Silanes (6 and 7). The procedure described by Noskov[4] was utilized to prepare the mixture of 3, 6 and 7. Into a two-necked 100mL round-bottomed flask was placed 30.3 g (0.17 mol) of antimony trifluoride (Aldrich). The flask was fitted with a pressure-equalizing dropping funnel and a distillation head. Into the dropping funnel was placed 20.6 g (0.135 mol) of tetramethoxy-

silane (3), and a nitrogen inlet was connected to the top of the funnel. The flask was purged with nitrogen and placed in an oil bath at a temperature of 85-95 °C. Tetramethoxysilane (3) was added dropwise as the products distilled into a cooled (dry-ice) receiving flask connected to the distillation head, and 11.0 g of distillate was collected. An analysis of the ^{1}H and ^{19}F NMR spectra revealed that the composition of the product mixture was 20% 3, 56% 6 and 23% 7. This mixture was used directly in the exchange reactions reported above. ^{19}F NMR: δ ($CFCl_3$)= 80.5 (doublet, J_{Si-F} = 196 Hz, $FSi(OCH_3)_3$), δ ($CFCl_3$) = 81.6 (doublet, J_{Si-F} = 196 Hz, $F_2Si(OCH_3)_2$).

Reactions of Methylphosphonic Difluoride with the Alkoxy Substituted Silanes. General Procedures. WARNING! The alkyl phosphonofluoridates are highly toxic acetylcholine esterase inhibitors; trained medical personnel should be available and the proper safety procedures should be followed while synthesizing or performing other experiments with these compounds.[11] These reactions were conducted using 1 to 2 gms (10 - 20 mmols) of 1 and an amount of the silane necessary to result in the molar ratios of reactants that have been described in the text. The neat reaction mixtures were stirred for the times specified (see text) in flasks fitted with calcium sulfate drying tubes. Reactions were monitored by a ^{1}H NMR analysis of a small aliquot of the reaction mixture dissolved in carbon tetrachloride. In some cases, the reaction temperature was measured with a thermocouple inserted into a septum-fitted two-necked flask. At the end of the reactions, the product composition was determined by ^{1}H NMR, ^{31}P NMR or by GC analysis. The spectral characteristics of these compounds were identical in all essential aspects to those of compounds prepared by the alternate methods described above.

Methyl Methylphosphonofluoridate (4). This compound was formed in the reaction of 1 with 3; 1 with the mixture of 3, 6 and 7; 1 with 9; and 1 with 10. The ^{31}P and ^{1}H NMR spectra are consistent with the assigned structure. ^{31}P NMR: δ 33.4 (d, J = 1039 Hz). ^{1}H NMR: δ 1.59 (doublet of doublets, J_{P-H} = 18.4 Hz, 3H, CH_3-P, $^{2}J_{P-H}$ = 6.0 Hz), δ 3.83 (d, J = 11.8 Hz, CH_3OP).

Dimethyl Methylphosphonate (5). This compound was formed in the reaction of 1 with 3; 1 with 9; and 1 with 10. The ^{31}P and ^{1}H NMR spectra are consistent with the assigned structure. ^{31}P NMR: δ 37.8, ^{1}H NMR: δ 1.37 (d, J = 17.8 Hz, 3H, CH_3-P), δ 3.66 (d, J = 10.9 Hz, 6H, OCH_3).

Isopropyl Methylphosphonofluoridate (11). This compound was formed in the reaction of 1 with 12. The ^{1}H NMR spectrum is consistent with the assigned structure. ^{1}H NMR: δ 1.21 (d, J = 6.3 Hz, 6H,

$OCH(CH_3)_2$), δ 1.41 (doublet of doublets), J_{P-H} = 71.8 Hz, 3H, CH_3-P, $^2J_{F-H}$ = 5.5 Hz, δ 4.74 (m, 1H, $OCH(CH_3)_2$).

Methylphosphonofluoridc Acid (13). This compound was formed in the attempted H_2O catalysis of the reaction between 1 and 12. The ^1H NMR spectrum is consistent with the assigned structure. ^1H NMR: δ 1.49 (doublets of doublets, J_{P-H} = 18.8 Hz, 3H, CH_3-P, $^2J_{F-H}$ = 6.0 Hz), δ 7.45 (s, 1H, P-OH).

Ethyl Methylphosphonofluoridate (14). A preparative scale synthesis of this compound was performed employing the reaction of 1 with 15. To 14.2 g (68 mmol) of the silane 15 was added 27.2 g (272 mmol) of 1. The temperature of the reaction mixture rose gradually from room temperature to 33 °C over a period of 50 minutes. During the next 5-6 minutes the reaction temperature rose to 70 °C as the solution darkened and SiF_4 was emitted. The resulting product was purified by a simple vacuum distillation (b.p. 53-54 °C/12mm) to afford 25.5 grams (84%) of the desired product 14. The distilled fluoridic ester product was 94.5% pure by GC and contained 4.6% of unreacted 1. The ^1H NMR spectrum is consistent with the assigned structure: δ 1.36 (t, J = 7.5 Hz, 3H, OCH_2CH_3), δ 1.54 (doublets of doublets, J_{P-H} = 19 Hz, 3H, CH_3-P, $^2J_{F-H}$ = 5.6 Hz), δ 4.19 (m, 2H, OCH_2CH_3).

1-Methoxy-perfluoroisobutene (18). To 20 g (0.10 mol) of 17 dissolved in 40 mL of anhydrous xylene was added dropwise with efficient mechanical stirring 32.1 g (0.10 mol) of tri-n-butyltin methoxide. The flask was cooled in an ice-water bath during this addition. After the addition was complete, the volatile liquid was removed under reduced pressure (1 mm) with the application of heat from a hot air gun. This material was fractionally distilled through a 6" packed column to give 14.0 g (62%) of the desired product 18: bp 101-103 °C.

Perfluoro-β,β-dimethylacrylonitrile (19). To 15.9 g (50.4 mmol) of tri-n-butyltin cyanide dissolved in 50 mL of frozen (dry ice/alcohol bath) anhydrous xylene was added 9.0 g (44.5 mmol) of perfluoroisobutene (17). The mechanically stirred slurry was allowed to warm to room temperature whereby the color of the reaction mixture changed from white to yellow to deep red. After stirring at room temperature for 2 hours, the volatile liquid was removed under reduced pressure. This material was fractionally distilled through a 6" packed column to give 5.0 g of the desired product that contained 11% 17 and 7.6% xylene. Considering the impurities present, the yield of 19 was 43%.

ACKNOWLEDGEMENT

Appreciation is expressed to Dr. H.S. Aaron, Dr. J.W. King, Mr.

L.J. Szafraniec and Mr. C.P. Ferguson for many helpful discussions. The authors are indebted to Ms. L.L. Szafraniec and Mr. W.T. Beaudry for performing the ^{31}P and ^{19}F NMR analyses, as well as Mr. D.K. Rohrbaugh and Mr. J.H. Buchanan for conducting the GC and GC-MS experiments.

REFERENCES AND NOTES

1. Perera, J., New Scientist, 11 April 1985 pp 10-11.

2. Bryant, P.J.R., Ford-Moore, A.H., Perry, B.J., Wardrop, A.W.H., and Watkins, T.F. J. Chem. Soc. 1553 (1960).

3. Walsh, R. Acc. Chem. Res. 14, 241 (1981).

4. Noskov, V.G., Kalinina, L.N., and Englin, M.A. Zh. Obshch. Khim. 42, 2028 (1972).

5. Zeifman, Yu.V., Ter-Gabrielyan, E.G., Gambaryan, N.P., and Knunyants, I.L. Uspekhi Khimii, 53, 431 (1984).

6. Hydro, W.R., and King, J.W. 1987 U.S. Army Chemical Research, Development and Engineering Center Scientific Conference on Chemical Defense Research, 17-20 November 1987, Abstract No. 51.

7. Misaki, S. J. Fluorine Chem. 29, 471 (1985).

8. This compound was obtained from the Chemical Process Laboratory, CRDEC. It is now commercially available from Alpha Chemicals, Danvers, MA.

9. Prepared under Synthesis Contract #DAAK11-82-C-0101.

10. Sumrell, G., and Ham, G.E. J. Am. Chem. Soc. 78, 5573 (1956).

11. Poziomek, E.J. Concepts Toxicol. 1, 243 (1984).

Supercomputer Applications in Projectile Aerodynamics

*Charles J. Nietubicz, Mr.
Jubaraj Sahu, Dr.
Karen R. Heavey, Mrs.
US Army Ballistic Research Laboratory
Aberdeen Proving Ground, Maryland 21005-5066

I. INTRODUCTION

The use of high speed scientific computers has always been an essential element in the computational fluid dynamics modeling of projectiles in flight. The ability to compute the aerodynamics of more complex projectiles, in both supersonic and transonic flow, has advanced due to the availability of computers with increased computational speed and memory. This is depicted in Figure 1 where the increase in predictive capability, indicated by the projectile sketches, has advanced due to the availability of new generations of supercomputers. The sketch with $M < 1$ indicates the ability to compute standard projectile shapes in transonic flow regimes. The physical nature of transonic flow requires the use of large computer memory and long run times. Until recently, the US Army Ballistic Research Laboratory (BRL), Aberdeen Proving Ground, Maryland, has had only limited access to supercomputers throughout the country and with that access, only limited memory was available. Figure 2 shows the computers which have been accessed in an attempt to determine computationally transonic aerodynamic coefficient data. Originally limited to 500,000 words of memory, the problem was intractable. With 1M (million) and then 2M words available, solutions have been obtained with only marginal results for the 3D transonic projectile calculations. The remote use of Cray supercomputers has helped provide the path for determining projectile aerodynamics in the transonic flight regime. The installation at BRL of the Cray X-MP/48 in December 1986 and the Cray 2 in July 1987 now provides the capability to perform timely calculations for projectile aerodynamics.

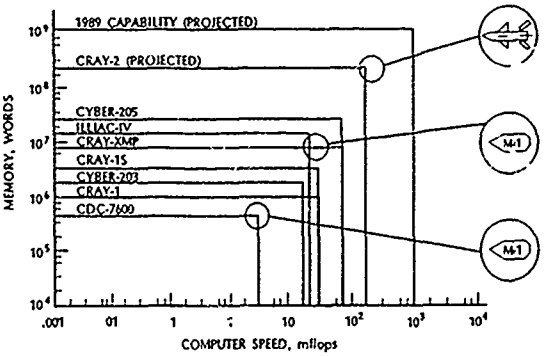


Figure 1. Computational Capability Relative to Available Computer Resources

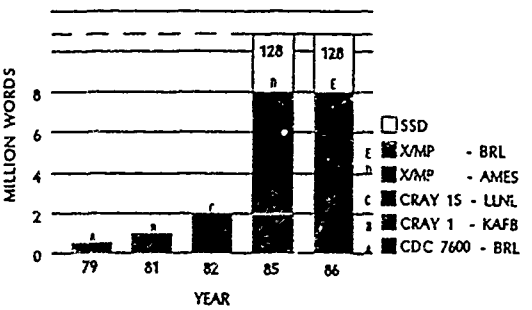


Figure 2. Remote Computing Facilities and Available Memory

This paper will discuss some of the current efforts in computational projectile aerodynamics and the use of color graphics workstations for the analysis of the computed data. The set of governing equations which have been solved and the solution technique utilized are discussed. Selected results will be presented for each of the following three areas: roll damping study, M825 projectile study, and M864 base bleed study.

II. GOVERNING EQUATIONS AND SOLUTION TECHNIQUE

1. GOVERNING EQUATIONS

The complete set of time-dependent thin-layer Navier-Stokes equations is solved numerically. The numerical technique used is an implicit finite difference scheme. Although time-dependent calculations are made, the transient flow is not of primary interest at the present time. The steady flow, which is the desired result, is obtained in a time asymptotic fashion.

The time-dependent thin-layer Navier-Stokes equations written in strong conservation law form for both the full 3D[1] and axisymmetric formulation[2] respectively are:

$$\frac{\delta \hat{q}}{\delta t} + \frac{\delta \hat{E}}{\delta \xi} + \frac{\delta \hat{F}}{\delta \eta} + \frac{\delta \hat{G}}{\delta \zeta} = \frac{1}{Re} \frac{\delta \hat{S}}{\delta \zeta} \tag{1}$$

$$\frac{\delta \hat{q}}{\delta \tau} + \frac{\delta \hat{E}}{\delta \xi} + \frac{\delta \hat{G}}{\delta \zeta} + \hat{H} = \frac{1}{Re} \frac{\delta \hat{S}}{\delta \zeta} \quad . \tag{2}$$

The general coordinate transformations are defined as:

$\xi = \xi(x,y,z,t)$	is the longitudinal coordinate
$\eta = \eta(x,y,z,t)$	is the circumferential coordinate
$\zeta = \zeta(x,y,z,t)$	is the near normal coordinate
$\tau = t$	is the time

The vector $\hat{q}$ contains the dependent variables $[\rho, \rho u, \rho v, \rho w, e]^T$ and the flux vectors $\hat{E}$, $\hat{F}$, $\hat{G}$ contain terms which arise from the conservation of mass, momentum and energy in the three coordinate directions. The source vector $\hat{H}$, in Equation (2), contains terms which result from an analytic determination of the flux vector $\hat{F}$, given the assumption of axisymmetric flow and constant angular velocity.[2] The viscous terms are contained in the vector $\hat{S}$ which is seen to have variation in the ζ direction only. This is representative of the thin-layer approximation.

2. SOLUTION TECHNIQUE

The numerical algorithm used for the solution of Equation (1) is a fully implicit, approximately-factored, finite-difference scheme that uses central differencing in η and ζ directions and flux-split[3] upwind differencing in the ξ direction. This code is relatively new and contains many of the improvements developed over the past few years. This includes: flux vector splitting, spacially varying time step and improved numerical smoothing methods. The flux-split upwind scheme can have several advantages over central finite difference schemes including natural numerical dissipation and better stability properties. The flux vector $\hat{E}$ has been split into $\hat{E}^+$ and $\hat{E}^-$, according to its eigenvalues. Three-point backward and forward differences are then used for the split fluxes $\hat{E}^+$ and $\hat{E}^-$, respectively.

Equation (2) is solved using the Beam and Warming[4] implicit approximately factored finite difference scheme which uses central differencing in both ξ and ζ directions. Code improvements have also been made to include a variable time step, numerical smoothing based on local solution gradients and code vectorization.[5] The Beam-Warming implicit algorithm is first-order accurate in time and second- or fourth-order accurate in space. The equations are factored (spatially split), which reduces the solution process to one-dimensional problems at a given time level. Central difference operators are employed and the algorithm produces block tridiagonal systems for each space coordinate. The main computational work is contained in the solution of these block tridiagonal systems of equations.

To suppress high frequency components that appear in regions containing severe pressure gradients e.g., shocks or stagnation points, artificial dissipation terms are added. In the present application, a switching dissipation model is used which is a blend of second- and fourth-order dissipation terms. This is similar to the model used by Pulliam[6] which uses a fourth-order dissipation in smooth regions and switches to a second-order dissipation in regions containing high pressure or density gradients. Incorporation of this dissipation model has resulted in an improvement in the quality of the results and has made the code more robust.

The axisymmetric code that solves Equation (2) uses a unique flow field segmentation procedure[7] to compute the full flow field over a projectile or a missile including the base region. A similar strategy[8] is used in the 3D code which solves Equation (1). Such procedures used for both the axisymmetric and 3D codes have the advantage of preserving the sharp corner at the base. The computations are for fully turbulent flows. A two-layer algebraic Baldwin-Lomax turbulence model[9] is used in both codes. In the axisymmetric code, a second algebraic turbulence model, based on a simple exchange-coefficient concept[10] was used in the base region.

III. ROLL DAMPING STUDY

The aerodynamic force and moment coefficients of interest for projectile design are shown in Figure 3. One of the goals of computational projectile aerodynamics is the ability to accurately determine these coefficients at subsonic, transonic and supersonic speeds. One measure of determining the adequacy of the computational approach is to compute the Magnus force and moment and make comparisons with experimental data. Good results for Magnus would generally indicate an adequately resolved flow field. More recently, however, it was learned that the roll damping coefficient is an even more sensitive parameter in determining adequacy of the viscous solution.

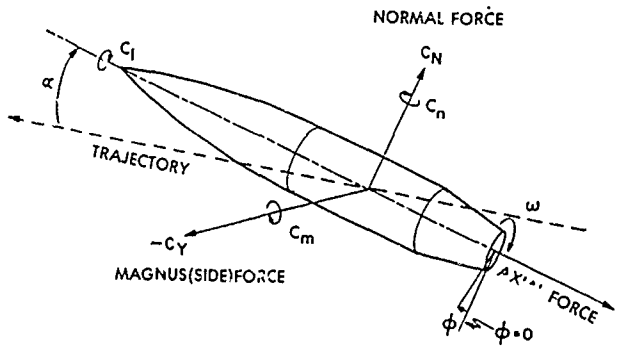


Figure 3. Aerodynamic Force and Moment Description

The Army-Navy spinner rocket configuration shown in Figure 4 is a simplified projectile shape which was used in the computational study to determine roll damping at transonic speeds. The computational results are compared to experimental data obtained from a BRL transonic range test and reported by L.E. Schmidt and C.H. Murphy.[11]

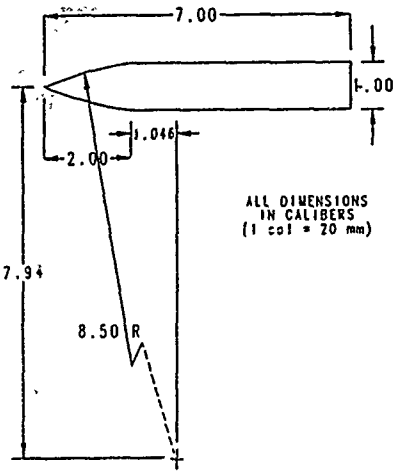


Figure 4. Army-Navy Spinner Rocket

The generalized axisymmetric code (Equation (2)) was used for these calculations. A computational grid consisting of 128 longitudinal points and 60 normal points was developed using a hyperbolic grid generator.[12] This results in a storage requirement of approximately 840,000 words and could not be conducted on the CDC 7600. Initial results were obtained remotely on the UIS Cray-1 computer in Kansas City, MO, and were completed on the BRL Cray X-MP/48.

The use of the BRL Cray resulted in the ability to obtain a significant number of solutions with almost real time turn around. Solutions were obtained from $M = 0.7$ to $M = 1.4$ for $\alpha = 0.0$. Mach contours are shown in Figure 5 for a free stream Mach number of 0.96. The coalescence of the Mach lines indicate the presence of a shock being formed at that location. A spark range picture was used to obtain the experimental shock location which was then superimposed on Figure 5. Both the experimental and computational shock position agree very well.

The results for the roll damping coefficient are shown in Figure 6 for a fully turbulent calculation. Most of the runs were made using a fully turbulent boundary layer and were found to overpredict the data by 20%. A few calculations were performed for laminar flow and indicated an under prediction of approximately 15%. From the experimental data the boundary layer was shown to be a mixture of both laminar and turbulent flow.

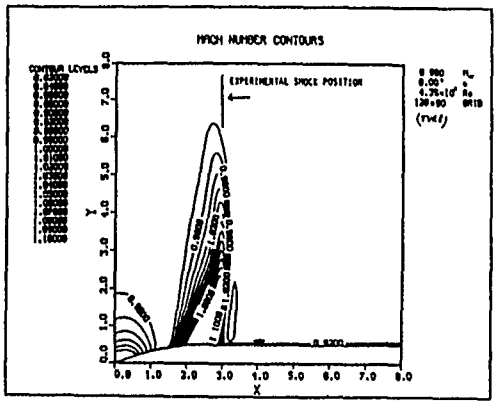


Figure 5. Mach Number Contours, $M_\infty = 0.96$

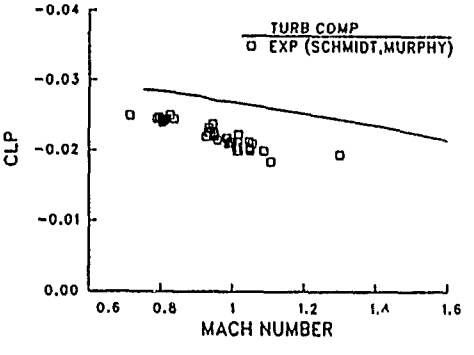


Figure 6. Roll Damping Coefficient vs. Mach Number

The combined use of a Silicon Graphics workstation and the availability of the BRL Cray resources has led to new a understanding in the convergence history of the Unsteady Navier-Stokes codes. During this study, files containing surface pressure as a function of convergence time were saved. These files were then analyzed on the color graphics workstation which presented a dynamic picture of the convergence history. Time dependent plots of surface pressure as a function of time were developed and the movement of pressure waves through the computational space were studied. The computed pressure waves were found to persist for a much longer time than was anticipated. Final convergence was obtained after 4000 time steps for the transonic cases. This knowledge has led to a new definition of the convergence criteria when using the Unsteady Navier-Stokes codes at transonic speeds and thus, a more accurate determination of the projectile aerodynamics.

IV. M825 PROJECTILE STUDY

During FY86 the capability to obtain a 3D solution for a full projectile configuration, including the base region, was developed. Due to the limited machine access of the NASA AMES Cray, only one calculation was obtained. This result was limited to using 1M words of memory. Previous calculations[13] have shown that increased grid resolution is required to adequately resolve the transonic flow field. More recently, through a cooperative effort with NASA Ames Research Center, a new capability[8] has been developed for the transonic projectile problem. This effort involved the development of a grid partitioning procedure to allow for the use of SSD. Figure 7 shows a schematic of how a typical grid can be partitioned to fit within the main memory of the Cray. The base region is partitioned as G4 is not shown here.

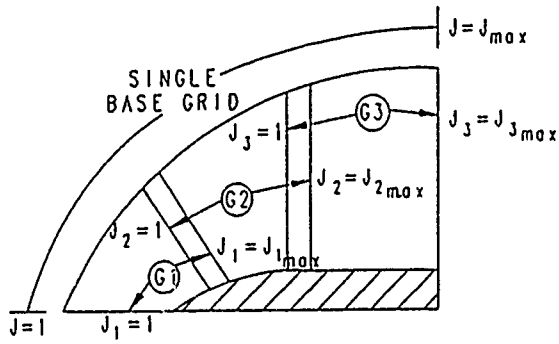


Figure 7. Grid Partitioning for Use Use of SSD

The solution procedure is to read from SSD the grid and dependent variables for section G1, advance the solution one time step, then write the solution and grid back to SSD. Next, section G2 is read from SSD. Notice that the partitioned sections are overlapped by one grid line. The computed data for $J_1 = J_{1_{MAX}}$ is the boundary condition for section G2. This is similarly done for sections G3 and G4. After one time step is completed, the solution is advanced to the next time step beginning with section G1. Once the solution has reached a steady state solution, the procedure is completed.

The external configuration of the M825, excluding the base, is similar to a M483A1 projectile. The features which have not been modeled exactly are the meplat on the fuze and the rotating band near the base. The rotating band was eliminated and the meplat was modeled as a hemisphere cap. The computational model is shown in Figure 8 and consists of a 2.84 caliber nose, a 2.7 caliber cylindrical section, and a 0.26 caliber 8° boattail. The ogive contours were matched to the original configuration as well as the undercut on the cylindrical section.

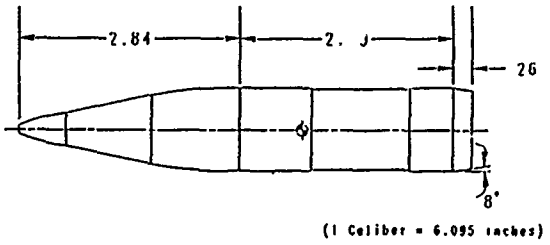


Figure 8. Computational Model, M825

The current problem of interest is the effect of the different base geometries on the overall projectile aerodynamics. Figures 9a and 9b show the standard and dome base configurations respectively. The standard base is a combination of aluminum and steel and contains a base cavity which is characterized as a flat surface. The product improved configuration is an all steel base and is characterized as a domed surface. The cavity volume is also significantly larger for the dome configuration.

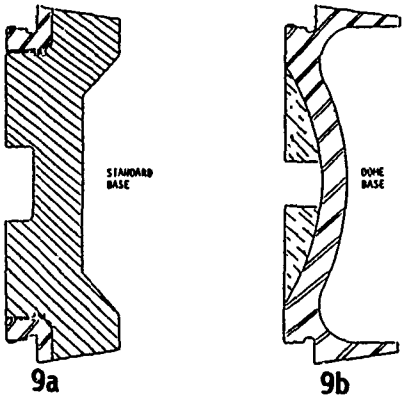


Figure 9. Base Cavity Configuration

The solution technique requires the discretization of the entire flow region of interest into a suitable computational grid. The grid outer boundary has been placed at 2.5 body lengths upstream and surrounding the projectile. The downstream boundary was placed at two body lengths. Since the calculations are in the subsonic/transonic regime, the computational boundaries must extend out well beyond the influence of the body. This ensures that the boundary conditions specified in the code are satisfied.

Figure 10 shows a grid generated for the standard configuration. The grid consists of 225 points in the streamwise direction and 50 points in the normal direction. This is broken down into two sections: a body region and a base region. The surface points for each region are selected using an interactive design program. Each grid section is then computed separately using a hyperbolic grid generation program.[12] There are 165 points along the projectile surface, including 60 points along the afterbody. The normal distribution of points in base region consists of 50 points along the base cavity. Longitudinally, the base region is matched point for point with the 60 surface points on the afterbody. An expanded

view of the base grid is shown in Figure 11a. The generally flat sections on the standard base enabled a grid to be routinely generated. However, due to the extreme concavity the grid for the dome base (Figure 11b) required an increase in the smoothing values used by the hyperbolic grid generator, as well as the addition of a grid averaging technique.

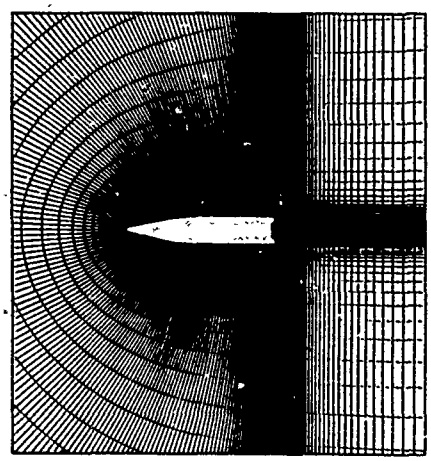


Figure 10. M825 Computational Grid

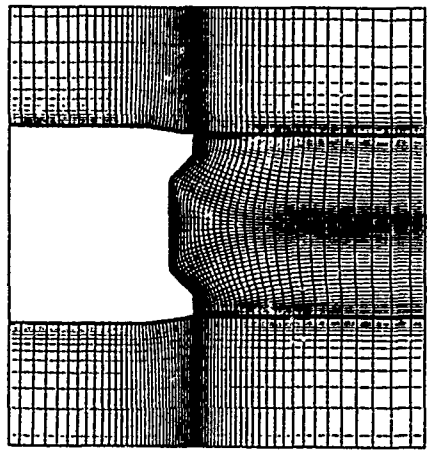


Figure 11a. Base Region Grid for the Standard Base

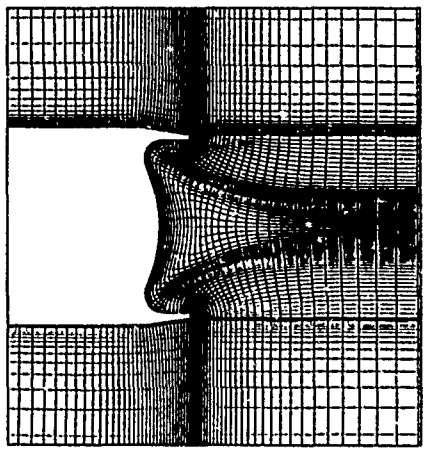


Figure 11b. Base Region for the Dome Base

Numerical computations have been made for both the standard and the dome base configurations for the range of Mach numbers from M = 0.80 to 1.5 and at α = 0° and 4°. Solutions were marched in time until the steady state results were achieved. Atmospheric flight conditions were used.

A few qualitative results are presented next. Figure 12 shows the Mach numbers contours for the dome base configuration for M = 0.98. The flow expansion at the ogive corner, a shockwave on the cylinder, and an expansion at the boattail corner can be seen. In addition, a shock system exists downstream of the base corner.

Velocity vector plots in the base region are shown in Figures 13a and 13b, respectively, for the standard and dome base configurations. As shown in Figure 13a for the standard base, the back flow, upon reaching the cavity follows the contour of the cavity and leaves the cavity pushing the flow upwards. This, in turn, creates the secondary bubbles seen near the base corner. The shear layer leaving the base corner is displaced upwards weakening the expansion at the base. Figure 13b shows the

velocity vectors for the dome configuration showing similar features such as the primary bubble and the secondary bubbles. The flow follows the contour of the cavity and upon leaving the cacity is almost parallel to the streamwise direction. This flow, thus, has less effect on the free shear layer and doesn't weaken the expansion at the base corner as much compared to the standard base. The net effect is that the size of the primary bubble for the dome base is slightly smaller than that for the standard base. The reattachment point is therefore closer to the base and should result in lower base pressure or higher base drag.

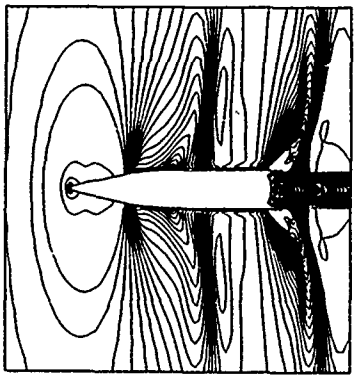


Figure 12. Mach Number Contours, $M_\infty = 0.98$, $\alpha = 0$, (Dome Base)

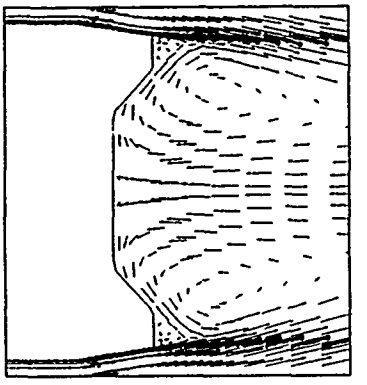


Figure 13a. Velocity Vectors in the Base Region, $M_\infty = 1.1$, $\alpha = 0$, (Standard Base)

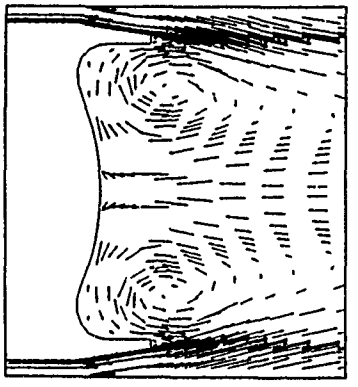


Figure 13b. Velocity Vectors in the Base Region, $M_\infty = 1.1$, $\alpha = 0$, (Dome Base)

The pressure, viscous and base drag components were determined. The pressure drag rise in the transonic speed regime was clearly predicted. The difference in pressure drag between the two projectiles (standard base and dome base) was very small at high transonic Mach numbers and gets larger ($\simeq$ 15%) at the low transonic Mach numbers ($M \simeq 0.9$). The viscous drag component is rather small and the difference in viscous drag due to the base configurations was negligible. The different base cavities are expected to impact the base drag and the results are shown in Figure 14. The difference in base drag between the standard base and the dome base is larger at the high transonic Mach numbers ($1.1 \leq M \leq 1.5$) as well as at low transonic speeds ($M < 0.94$). The difference is very small near $M \simeq 0.97$. In addition, the base drag is higher for the dome base than the standard base at high transonic Mach numbers ($0.97 \leq M \leq 1.5$) while the reverse is true at low transonic Mach numbers ($M < 0.97$). The expected rise in base drag at transonic speeds is clearly predicted.

Comparison of the total aerodynamic drag is shown in Figure 15. As shown in this figure, the difference in drag is small at near ≈ 0.97 and is somewhat larger at the high ($1.1 < M < 1.5$) and low ($M < 0.92$) transonic speeds. This plot also shows the range data for both base configurations. The overall comparison of the computed drag data with the range data is fair. As seen in the range data, the dome base has higher drag especially at higher transonic Mach numbers and this trend is seen in the computed results also.

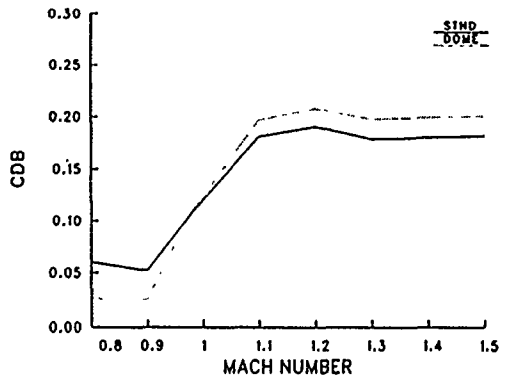


Figure 14. Variation of Base Drag with Mach Number, $\alpha = 0$

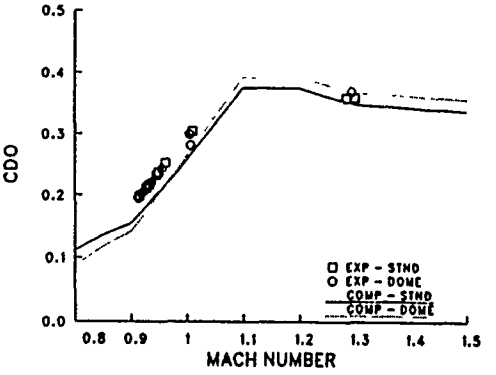


Figure 15. Variation of Total Drag with Mach Number, $\alpha = 0$

The experimental data also indicated an effect of base cavity on the aerodynamic lift and pitching moment. To determine this computationally required 3D computations. Using the method previously described, computations were performed for the M825 at 4° angle of attack. These calculations are computer intensive both in size and time and required approximately 16 million words of SSD and 20 hours on the Cray X-MP/48 for each calculation. Qualitative results of computed Mach contours and velocity vectors are shown in Figures 16 and 17 respectively for the dome base configuration.

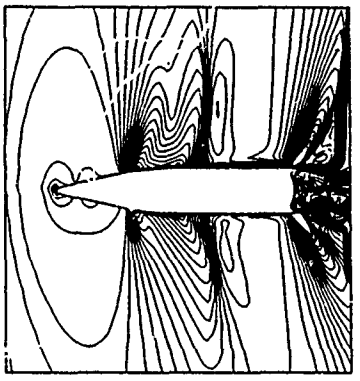


Figure 16. Mach Number Contour, $M_\infty = 0.98$, $\alpha = 4°$

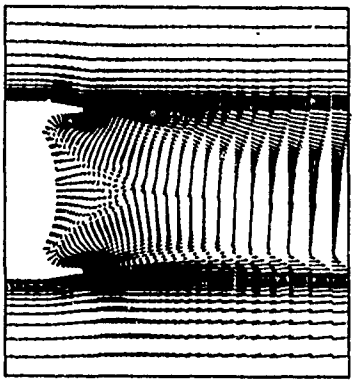


Figure 17. Velocity Vectors in Base Region, $M_\infty = 0.98$, $\alpha = 4°$

The asymmetric shock structure is clearly seen in Figure 16 with shocks existing on the cylindrical section and on the boattail. There was no apparent change in shock position between the two base configurations. Differences were evident when looking at the recirculation patterns in the base region. Figure 17 shows the velocity vectors in the base region for the domed configuration. The differences in the recirculation patterns were found to be similar to that for the axisymmetric case with the exception of the expected asymmetry. The computed data is presently being integrated to determine the aerodynamic forces and moments.

V. M864 BASE BLEED STUDY

A projectile in current development is the M864 base bleed projectile. Experimental aerodynamic coefficient data for the base bleed projectile, which has the additional complexity of a domed base, is very limited. A computational effort has begun in an effort to determine the aerodynamic coefficients and flow field characteristics associated with this configuration.

Figure 18 shows the general dimensions of the M864. The dotted lines at the base of the projectile is the actual base cavity. The gap near the centerline is the exhaust port for the base bleed. A computational grid has been developed for this shape and is shown in Figure 19a. It consists of 265 longitudinal points and 60 points in the normal direction. An expanded view of the base is shown in Figure 19b. A hyperbolic grid generation method was used to wrap the grid around the base cavity. The deep recesses in the base had caused some initial problems with grid line crossing, however the use of a cell averaging technique allowed the grids to be generated.

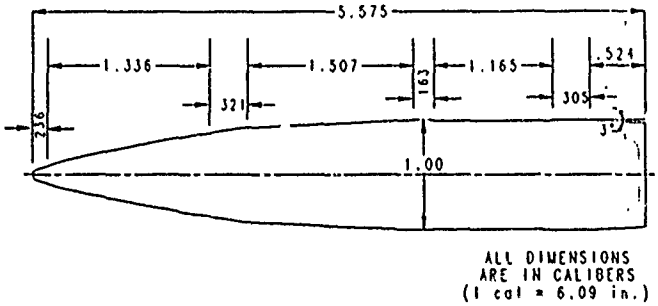


Figure 18. M864 Base Bleed Projectile

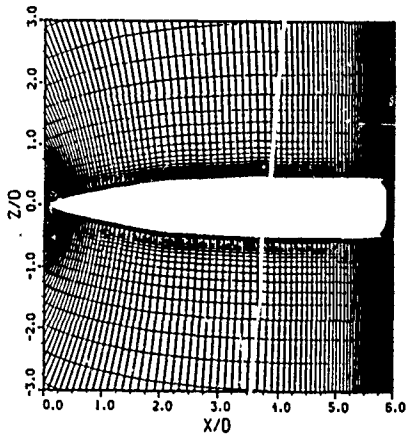


Figure 19a. M864 Computational Grid

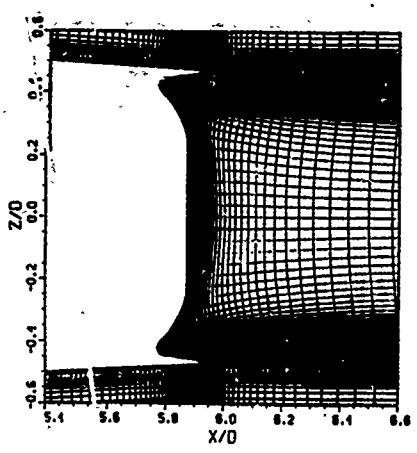


Figure 19b. Expanded View of Domed Base

A series of calculations have been run for both a flat base and the actual domed M864 with and without base bleed. The Mach number range computed was $0.7 < M < 2.5$ for $\alpha = 0°$. Figures 20a and 20b shows the computed velocity vectors in the wake region for the flat and domed base respectively. The effect of the dome base is to generate a second, low speed recirculation in the recessed cavity. This recirculation bubble is in the opposite direction of the primary recirculation pattern. Computations have also been performed for the dome base with the inclusion of base bleed. The details of the change in boundary conditions required are found in an earlier paper.[14] The mass injection parameter, I_j, defined as $I_j = \dot{m}_j/\rho_\infty u_\infty A$; was used to vary the mass flow rate. The computations to date have been for cold mass flow and with a mass injection value of 0.01. This value is the upper bound on the mass flow rate estimated for the actual M864 projectile. Although the actual mass flow rate varies over the course of the trajectory, the present calculations were run to see the effect on drag across a Mach number range. Figure 20c shows the computed velocity vectors for the dome M864 with base bleed. The addition of base bleed has increased both the size and strength of the secondary recirculation region.

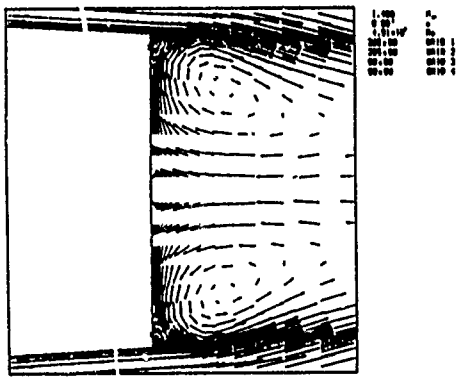


Figure 20a. Velocity Vectors for Flat Base, $M_\infty = 0.94$, $\alpha = 0.0$

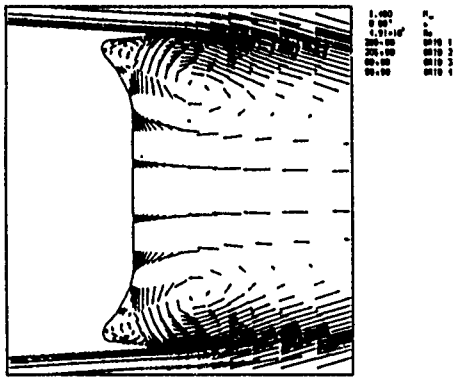


Figure 20b. Velocity Vectors for Dome Base, $M_\infty = 0.94$, $\alpha = 0.0$

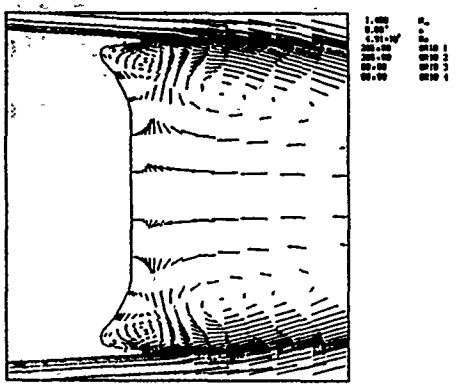


Figure 20c. Velocity Vectors for Dome Base, $M_\infty = 0.94$, $\alpha = 0.0$, $I = 0.01$

Figure 21 is a plot of the computed drag coefficient, C_{D_o}, for the flat base and dome base with and without base bleed. The computed drag coefficient is the summation of the pressure, viscous and base drag components. The results show a decrease in total drag for the dome base configuration for all Mach numbers. When the base bleed is initiated a further reduction in drag is evident except in the low subsonic region.

Recent tests performed in the BRL Transonic Aerodynamics Range have provided some limited drag data[14] for the M864. Comparison of the computed and experimental data is also shown in Figure 21. The data is for the M864 without base burn using an inert burn motor. The computed results are in general agreement at the supersonic velocities however in the transonic and subsonic regions the comparison is poor. The underprediction of the drag in this region has been noticed before and at present lacks a full explanation. Grid refinement studies, turbulence modeling and alternate solution techniques are currently in progress.

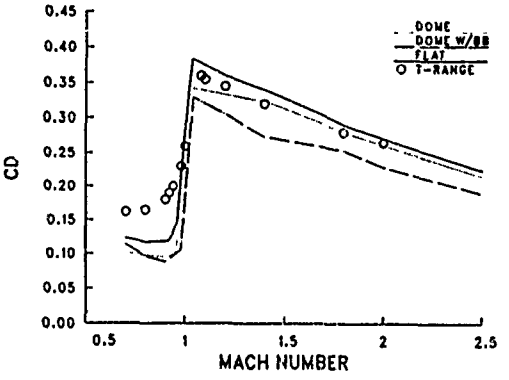


Figure 21. Total Drag Coefficient for Flat and Dome Base

VI. SUMMARY

Computational and experimental results have been shown for the roll damping coefficient of a secant-ogive-cylinder configuration. A series of cases have been computed throughout the transonic flight regime. The difference in the computed and experimental results is attributed to the modeling of the correct state of the boundary layer. Additionally, a new understanding of the solution convergence history has been obtained.

Numerical computations have been made for the M825 projectile with two base cavity configurations at transonic speeds. Computed results show the difference in the qualitative features of the base region flowfield

between the two base cavities. Additionally, the Navier-Stokes computation provide some insight as to the mechanism which account for the changes in base pressure due to different base configurations. The drag components and the total drag have been computed. Changes in the base cavity configuration have been found to affect the aerodynamic drag. Differences in drag of between 2 to 10% have been predicted. The dome base configuration produces higher drag for $1.1 \leq M \leq 1.5$, lower drag for $M < 0.97$ and qualitatively agrees with the trend of the experimental range data.

The initial computations of the domed base M864 base bleed projectile are providing insights to the effects of base cavities and base bleed on drag. The ability to run a large number of cases for various grids and conditions has provided a better understanding of the numerical modeling. Additional work is required in refining the computational mesh in the base region and the addition of hot gas.

The installation of the Cray X-MP/48 at the BRL has provided a computational tool which is essential to the development of a predictive capability for projectile aerodynamics. The use of advanced color graphic workstations is also an essential element in the analysis of the computed flow fields. The combined availability of the BRL resident Cray X-MP/48 and color graphics workstations provides researchers the required tools for current and future development in the area of computational projectile aerodynamics.

REFERENCES

1. Pulliam, T.H. and Steger, J.L., "On Implicit Finite-Difference Simulations of Three-Dimensional Flow," AIAA Paper No. 78-10, January 1978.

2. Nietubicz, C.J., Pulliam, T.H. and Steger, J.L., "Numerical Solution of the Axiumuthal-Invariant Thin-Layer Navier-Stokes Equations," ARBRL-TR-02227, US Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, March 1980. (AD A085716)

3. Steger, J.L. and Warming, R.F., "Flux Vector Splitting of the Inviscid Gasdynamic Equations with Application to Finite-Difference Methods," Journal of Computational Physics, Vol. 40, No. 2, 1981.

4. Beam, R. and Warming, R.F., "An Implicit Factored Scheme for the Compressible Navier-Stokes Equations," AIAA Paper No. 77-645, June 1977.

5. Sahu, J. and Nietubicz, C.J., "Improved Numerical Prediction of Transonic Flow," Proceedings of the 4th Army Conference on Applied Mathematics and Computing, Cornell University, May 1986.

6. Pulliam, T.H., "Artificial Dissipation Models for the Euler Equations," AIAA Paper No. 85-0438, January 1985.

7. Sahu, J., Nietubicz, C.J. and Steger, J.L., "Numerical Computation of Base Flow for a Projectile at Transonic Speeds," AIAA Paper No. 82-1358, August 1982.

8. Sahu, J. and Steger, J.L., "Numerical Prediction of Three Dimensional Transonic Flows," AIAA Paper No. 87-2293-CP.

9. Baldwin, B.S. and Lomax, H., "Thin-Layer Approximation and Algebraic Model for Separated Turbulent Flows," AIAA Paper No. 78-257, 1978.

10. Chow, W.L., "Improvement on Numerical Computations of the Thin-Layer Navier-Stokes Equation with Emphasis on the Turbulent Base Pressure of a Projectile in Transonic Flight Condition," Contract Report No. DAAG29-81-D-0100, US Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, November 1985.

11. Schmidt, L.E. and Murphy, C.H., "The Aerodynamic Properties of the 7-Caliber Army-Navy Spinner Rocket in Transonic Flight," ARBRL-MR-775, US Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, March 1954.

12. Nietubicz, C.J., Heavey, K.R. and Steger, J.L., "Grid Generation Techniques for Projectile Configurations," Proceedings of the 1982 Army Numerical Analysis and Computers Conference, ARO Report 82-3, February 1982.

13. Nietubicz, C.J., Mylin, D.C., Sahu, J. and LaForge, R., "Aerodynamic Predictions for a Projectile Configuration at Transonic Speed," AIAA Paper No. 84-0326, 1984.

14. Sahu, J., Nietubicz, C.J., and Steger, J.L., "Navier-Stokes Computations of Projectiles Base Flow With and Without Base Injection," ARBRL-TR-02532, US Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, November 1983. (AD A135738) (Also see AIAA Journal, Vol. 23, No. 9, September 1985, pp. 1348-1355.

15. Oskay, V., private communications, Free Flight Branch, Launch and Flight Division, US Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, September 1987.

Improvements to Tilt Rotor Performance Through Passive Blade Twist Control (U)

* Mark W. Nixon
U.S. Army Aerostructures Directorate
NASA Langley Research Center
Hampton, Virginia 23665-5225

Introduction

The most unique feature of a tilt rotor aircraft is its wingtip-mounted rotors which are directed to a vertical position in hover, but are tilted to a horizontal position in forward flight. Clearly, the design of a tilt rotor requires technologies which are traditional to both helicopters and fixed-wing aircraft. The overlapping of these technologies leads to design compromises because there is no solution which is optimum for both flight regimes. One such compromise exists in the twist design of a tilt-rotor blade. Typically, a tilt-rotor blade is designed with a twist distribution which produces acceptable hover performance at the cost of a lower forward flight propulsive efficiency.

The aerodynamic performance of tilt rotor aircraft would be improved if blade twist could be tailored to the requirements of both hover and forward flight. The twist distributions which are optimum for each flight condition, under the assumption of uniform inflow, are presented in reference 1. The two distributions are significantly different which suggests that there is a performance improvement to be gained through a twist change between the flight modes. However, no actual performance values are reported in reference 1. A more sophisticated twist optimization study was performed in reference 2. Here, a two-piece linear twist variation was used to approximate the non-linear twist used on a conventional metal blade of the XV-15. An optimum twist was determined for forward flight performance which improved propulsive efficiency by about 5 percent. A twist variation study for hover was also performed in reference 2. However, because the analysis assumed uniform inflow, the results showed little sensitivity to twist selection. Further investigation is required to determine an optimum twist design and its resulting performance benefits for the hover condition. This is because references 1 and 2 both use uniform inflow theory which is generally not sufficient for predicting hover performance trends.

The concept of changing blade twist between two flight modes can be realized with the use of composite rotor blades, designed to exhibit extension-twist coupling through an arrangement of off-axis ply angles and stacking sequences. In forward flight, the rotor speed of a typical tilt rotor is 20 percent less than it is in hover. Thus, there is a net change in centrifugal forces which can be used to passively obtain the desired rotor blade twist for each flight mode. The extent of the improvement in aerodynamic performance depends on how closely the actual twist developed in hover and forward flight approaches the optimum twist in each of the two modes. This in turn depends on the magnitude of twist deformation which can be produced within the material strength limits of the blade structure. As the allowable

twist deformation increases, so does the ability to obtain desirable twist distributions in both modes of flight. Although the desired twist change occurs for a 20 percent change in rotational velocity, it is the twist change caused by increasing rotational velocity from zero to its maximum value which produces the maximum blade stresses. For structural substantiation of a design, the centrifugally generated stresses must be considered simultaneously with bending stresses resulting from air and inertia loads.

In reference 3, passive twist control concepts are applied to the extension-twist-coupled design of a rotor blade for the XV-15 tilt rotor assuming a 15 percent change in operating rpm between hover and forward flight. The design was first required to match the baseline XV-15 composite blade in flapwise and lead-lag bending stiffnesses, torsional stiffness, chordwise *c.g.* location, and mass distribution per unit span. Using these requirements only about 0.5° twist change was developed. A second design study was performed which allowed deviations from the baseline XV-15 blade in the bending and torsional stiffnesses, but retained the baseline mass distribution and *c.g.* location. This resulted in a 2° twist change over the same 15 percent rpm range. The approach used in reference 3 was to alter an existing conventional design. It is likely that larger twist changes can be developed if blades are designed using both mass and stiffness to take full advantage of extension-twist-coupling benefits. An indication that larger twist changes can be obtained is found in reference 4. Here, composite tube tests resulted in twist rates of between 0.384 and 0.487 deg/in. at the material strength design-limits for structures representative of rotor blade spars. This translates to about 48 degrees of total twist for the tailorable span of the XV-15 tilt-rotor blade. However, the single cell structure tested in reference 4 was not completely representative of an actual blade cross section, and was also not subjected to simultaneous bending loads. Thus, the twist deformation which can be obtained within material strength limitations when all applicable loads are considered remains in question.

The purpose of this paper is to show that passive blade twist control through use of an extension-twist-coupled structural design is a feasible concept for tilt rotor blades. The investigation described in this report addresses the aerodynamic and structural aspects of a passive twist-control rotor blade system. First, the aerodynamically-optimum twist distributions are determined for hover and forward flight. The performance benefits gained from use of these twists over the twist employed in the conventional XV-15 blade are also ascertained. Three extension-twist-coupled designs are then developed, based on a D-shape spar and a NACA-0012 airfoil, with the intent of achieving aerodynamically-improved twist distributions. For this design study the maximum twist deformation available at 100 percent rpm, subject to material strength limitations, is determined using a coupled-beam analysis and a laminate analysis integrated with an optimization analysis. The twist deformations available to each design are used to obtain twist distributions which improve the performance associated with the conventional XV-15 blade twist in both hover and forward flight.

Nomenclature

c	chord length, in.
c.g.	center of gravity location, in.
C_T	coefficient of Thrust, $\frac{T}{\rho\pi R^2(\omega R)^2}$
E_{11}	lamina longitudinal modulus, psi
E_{22}	lamina transverse modulus, psi
G_{12}	lamina shear modulus, psi
HP_{min}	minimum horsepower required
r	radial position from center of rotation, in.
R	radius, in.

S	Lamina shear strength, ksi
T	Thrust, lb
T_{rt}	spar thickness in root section, in.
T_{sk}	airfoil skin thickness, in.
T_{tr}	spar thickness in transition section, in.
T_{tw}	spar thickness in twisting section, in.
v_0	velocity of the wake at rotor disk, ft/sec
V	free stream velocity, ft/sec
$wnsm$	weight of non-structural mass, lb/in
$wtip$	tip weight, lb
x, y, z	principal coordinate directions
X	lamina strength in the fiber direction, ksi
Y	lamina strength in the matrix direction, ksi
α	general ply angle, deg
α_{opt}	optimum angle of attack, deg
α_{sk}	laminate fiber angle in airfoil skin, deg
α_{sp}	laminate fiber angle in spar, deg
$\Delta 20$	twist deformation at tip between 80 and 100% rpm, deg
$\Delta 80$	twist deformation at tip between 0 and 80% rpm, deg
$\Delta 100$	twist deformation at tip between 0 and 100% rpm, deg
Δx_{tr}	length of transition section, in.
Δx_{tw}	length of twisting section, in.
θ	blade twist angle, deg
μ	tip-speed ratio, $\frac{V}{\omega R}$
ν	inflow velocity, ft/sec
ν_{12}	lamina primary Poisson ratio
ρ	density of air, slug/ft^3
ω	rotational velocity, rad/sec

Subscripts

n	number of repetitions of a lamina sequence
s	make lamina symmetrical

Aerodynamic Considerations

The aerodynamic part of this investigation focuses on determining the rotor blade twist distributions which are optimum for aerodynamic performance in hover and forward flight. The twist distributions which are optimum under the assumption of uniform inflow are presented first. Next, optimum twists with linear distributions are determined for each flight mode using proven analysis techniques. The performance associated with the optimum linear twists are compared to the performance associated with the twist used on the conventional XV-15 tilt rotor blade and the twists which are optimum based on uniform inflow. Finally, the process of twist distribution selection in practical extension-twist-coupled designs is explained.

Optimum Twist for Uniform Inflow

For a rotor blade operating in a free-stream with the rotational plane perpendicular to the flow direction (uniform inflow), the optimum performance is realized when the blade twist

distribution is such that the Betz condition is satisfied. The Betz condition states that the trailing vortex sheet moves aft as a rigid helical sheet, which is to say that the wake displacement velocity is radially constant (refs. 1 and 5). Using combined blade element/momentum theory to satisfy the Betz condition, reference 1 shows that the displacement velocity of the vortex wake can be approximated by

$$v_0 = \frac{\omega R}{2}\left(\sqrt{\mu^2 + 2C_T} - \mu\right) \tag{1}$$

and the optimum blade twist distribution is given by

$$\theta = \tan^{-1}\left(\frac{v_0}{\omega r} + \mu\frac{R}{r}\right) + \alpha_{opt} \tag{2}$$

In hover, μ is zero, and the wake displacement velocity v_0 is the same as the induced velocity ν so that equation 1 reduces to

$$\nu = \omega R\sqrt{C_T/2} \tag{3}$$

and equation 2 becomes

$$\theta = \tan^{-1}\left(\frac{\nu}{\omega r}\right) + \alpha_{opt} \tag{4}$$

For a radially constant chord and airfoil section, α_{opt} can be ignored because it is the same for all radial points and, therefore, does not influence the built-in twist distribution. Assuming a radially constant chord and airfoil section, the twist distributions are calculated using equations 2 and 4, and are shown in figure 1. The forward flight twist is shown to be nearly linear with about -42° of twist from root (r/R=.15) to tip. The twist distribution in hover is shown to be highly nonlinear, and has less overall twist, measured root to tip, than the forward flight twist. Conventional tilt-rotor blade twist distributions are designed by interpolating between the optimum hover and optimum forward flight twist distributions. To illustrate this, the conventional XV-15 tilt rotor twist distribution is also plotted in figure 1.

Optimum Linear Twist Distributions

For this investigation more sophisticated analysis techniques were employed to compute the aerodynamic performance associated with various linear twists. Linear twist distributions were considered because they are easily quantified, and are more commonly used in practice than nonlinear twist distributions. The Comprehensive Analytical Model of Rotorcraft Aerodynamics and Dynamics (CAMRAD, ref. 6) was used to compute performances in the forward flight condition. The hover performances were assessed using HOVT, a strip theory momentum analysis with a nonuniform inflow model, which is based on equations found in reference 7. The hover performance predictions of HOVT have been verified for three different blade designs through correlation with experiments performed at the Langley 4 x 7 meter wind tunnel (ref. 8).

The performance of several linear twists were determined for hover and forward flight using HOVT and CAMRAD, respectively. The aerodynamic blade model used in the analyses was a high-speed 350-knot blade configuration. The performance trends are illustrated in figure 2, where the performance is expressed as a percentage increase in horsepower required over the minimum horsepower required in each flight mode. The minimum horsepower required corresponds to the optimum linear twists which are shown to be -20° in hover and -42° in

forward flight. The results of figure 2 also indicate that use of a single compromised linear twist results in an increase in power-required for both modes of flight. The -42° linear twist, found to be optimum for forward flight, is geometrically equivalent to the twist distribution defined by equation 2 which is based on uniform inflow. The forward flight performance associated with the -42° linear twist and the conventional XV-15 twist are compared in figure 3. The -42° linear twist is shown to reduce horsepower required by 6.5 percent at the design velocity. The hover performance was calculated assuming nonuniform inflow for three twist distributions: the conventional XV-15 twist, the uniform-inflow twist defined by equation 4, and the -20° linear twist. The performance results are compared in figure 4 which illustrates that the hover performance associated with the twist based on uniform inflow is worse than the -20° linear twist (based on nonuniform inflow). This is expected because the aerodynamics of the hover condition are highly influenced by the inflow distribution, and in the actual case the inflow is not uniform. Figure 4 also shows that the hover performance can be improved 6.1 percent at the design gross weight by using the -20° linear twist instead of the conventional XV-15 twist. The figure shows further that the performance benefit can be traded for a 600 lb increase in payload carrying capability. Based on the results of the hover and forward flight predictions, the aerodynamic performance of tilt rotors can be significantly improved if two different linear blade twist distributions are generated, approaching -42° in forward flight and -20° in hover.

Selection of Twist Distribution

The linear twists which were shown to be optimum for hover and forward flight can not both be obtained using the extension-twist-coupled blade concept. The centrifugal forces are a nonlinear function of the rotational velocity so that the twist deformation produced between two radial positions changes nonlinearly with rotor speed. Thus, when using an extension-twist-coupled blade, only one of the tilt rotor flight modes can be specified to have a linear twist distribution. Figure 2 shows that performance in the forward flight condition is much more sensitive to variations in twist selection. Thus, it is advantageous to use the optimum linear twist for forward flight in the forward flight mode which requires use of a nonlinear twist in hover. If a linear twist of -42° is assumed in forward flight at 80 percent rpm then the twist distribution in hover at 100 percent rpm is a nonlinear variation of the desired -20° linear twist.

Assuming a -42° linear twist in forward flight, the nonlinear twist distribution obtained in hover is a function of the particular extension-twist coupled design. The shape of the twist distribution in hover is roughly the same for all practical extension-twist designs because the twist distribution is controlled by the blade weight distribution. The weight distribution cannot, in practice, be altered so much that a significant change in the shape of the twist distribution will occur over a 20 percent change in rotor speed. However, a combination of weight and laminate design can be used to significantly alter the overall magnitude of twist change measured between the blade root and the tip. The magnitude of twist between root and tip has a large influence on the hover performance, but it is not clear what magnitude is optimum. The linear twist study indicates that a twist of -20° is optimum for a linear distribution in hover, but the magnitude of twist desired for a nonlinear distribution may be different.

For the concept of twist control presented in this investigation, the -42° forward flight twist distribution is obtained by determining the twist distribution of the non-rotating, undeformed blade (sometimes referred to as the jig-shape) required for a particular extension-twist-coupled design. The twist deformation at various radial positions are determined for a change in rotational velocity between 0 and 80 percent rpm. These twist deformations are subtracted from the distribution desired in forward flight to obtain the twist distribution necessary for the

jig-shape. The twist distribution for the blade in hover is calculated by a similar procedure. In this case, the twist deformation at various radial positions are determined for a change in rotational velocity between 0 and 100 percent rpm. These deformations are then added to the twist distribution of the jig-shape to obtain the twist distribution in hover.

Structural Considerations

Geometry and Materials

The structural blade model used in this design study is based on a 20-inch chord NACA-0012 airfoil with a D-shaped spar. The basic cross-sectional dimensions and components are illustrated in figure 5. The spar and airfoil skin are composed entirely of Ciba-Geigy IM6/R6376 graphite/epoxy laminates. The properties of this material are given in Table I. The blade planform is assumed to be rectangular and to consist of three sections as illustrated in figure 6. Each section is further divided into a series of beam segments for analytical purposes; two in the root section, eight in the transition section, and ten in the twisting section.

The root section extends from the center of rotation to $0.15R$. This section is not extension-twist-coupled and consists only of a spar which is used to represent the hub and built-up root-end blade structure typical of rotor blades. The plies of the laminates in this section are all oriented at 0° to the spanwise axis. Conversely, the spar in the twisting section consists only of off-axis plies. Its laminate is assumed to be of the form $[(\alpha/\alpha+90)_s]_n$. The transition section is necessary to change laminates from 0° in the root section to an off-axis angle in the twisting section. Thus, the spar in the transition section consists of both 0° and off-axis plies. The layup used in this section is given by $[(\alpha/\alpha+90/0/90)_s]_n$. The transition section experiences twist deformation, but to a lesser degree than the twisting section because the 0° plies significantly increase the extensional stiffness of the blade. The airfoil skin surrounds the D-spar in both the transition and twisting sections, and is composed of the same type of laminate used in the spar of the twisting section, although α is not necessarily the same for both. The extension-twist-coupled laminates, $[(\alpha/\alpha+90)_s]_n$ and $[(\alpha/\alpha+90/0/90)_s]_n$, are used because elastic coupling can be obtained without thermal coupling. Thermal coupling is avoided when a symmetric laminate is composed entirely of sets of 0/90 plies even if some sets of the 0/90 plies are oriented off-axis. Thus, the laminates used in this structural model produce a twist deformation under an axial load, but do not produce a twist deformation with a change in temperature.

Design Parameters

Several design parameters are used to perform the extension-twist-coupled blade design study. The parameters, T_{tw}, T_{tr}, and T_{rt} are used to control the thicknesses of the D-spar laminates in the twisting, transition, and root sections, respectively. T_{sk} is used to control the thickness of the airfoil skin. The angles of the off-axis plies used in the D-spar of the transition and twisting sections are identical, but may differ from the angle used in the airfoil skin. Thus, α_{sp} controls the off-axis ply angles in the spar, and α_{sk} controls the off-axis ply angles in the airfoil skin. The length of the twisting section is given by the parameter Δx_{tw}. The length of the transition section, Δx_{tr}, is calculated by subtracting the root and twisting section lengths from the blade radius, R. The remaining parameters required for the design study are used to control weight. The tip weight, *wtip*, is assumed to be lumped spanwise at R. The parameter *wnsm* is nonstructural weight per unit span added to the running-weight

(weight per unit span) of the twisting section. All of the thickness design parameters are illustrated in figure 5 while the remaining parameters are illustrated in figure 6.

The calculation of blade running-weight is extremely important in this design study because the centrifugal loads result from these weights. Thus, the twist deformation produced at a given rotational velocity is directly related to the running-weight. The running-weights of each section are calculated from the existing design parameters. First, the weight per unit span of each blade section is determined from the corresponding values of T_{sk}, T_{tw}, T_{tr}, and T_{rt}. More running-weight is added to a segment by assuming that the trailing edge core is filled with honeycomb, and that the entire cross section is balanced at the quarter chord with leading-edge weights. Additional weight per unit span may be added to the twisting section by the parameter *wnsm* which is assumed to be placed at the quarter chord. This parameter is used to increase the centrifugal loads when those resulting from the tip weight and other running-weight are not sufficient to create twist deformations which approach the limits of the material strength.

Structural Substantiation

One design condition which is necessary to meet is that a positive margin of safety must be obtained everywhere in the structure when subjected to the substantiating load condition To insure conservative predictions of the blade strength, the substantiating load condition is composed of the maximum flapwise and inplane bending loads, combined with the maximum centrifugal load, even though all of these loads will not likely be maximum simultaneously. The resulting stresses are used to calculate margins of safety which, for structural substantiation, must be greater than zero everywhere on the structure.

The general procedure for determining the substantiating loads for helicopter rotor blades is outlined in reference 9. However, this procedure is not completely applicable to the tilt rot r case because the loads produced in the approach of reference 9 account for loads produced in some helicopter maneuvers which are not performed by tilt rotor aircraft. Thus, the maximum bending loads used in this design exercise were obtained from test data on conventional XV-15 rotor blades. The loads on an XV-15 tilt rotor model were determined from wind tunnel tests and are reported in reference 10. The maximum bending loads in the blade occur in the helicopter mode at the spanwise station of 0.35R. The flapwise bending moments here are about 20,000 in-lb steady and 22,000 in-lb oscillatory for a maximum load of 42,000 in-lb. The maximum chordwise bending moment is somewhat less, but the same moment of 42,000 in-lb is used to insure conservative results. These bending moments are introduced into the structural model by applying a tip load of 430 lb in the flapwise and inplane directions. This method of producing the maximum bending moment at station $0.35R$ also produces a moment distribution in the blade which is a good approximation to the moment distribution reported in the wind tunnel tests of reference 10. The centrifugal load distribution is calculated at the maximum rotational velocity and is a function of the blade weight distribution. The combination of the maximum flapwise and inplane bending moments applied simultaneously with the maximum centrifugal load is the load condition for which the blade designs in this study are substantiated.

Analyses

Structural Analyses

The rotor blade deformations are calculated with a coupled-beam analysis which is based on the theory presented in reference 11. The analysis was originally developed for a single-cell structure, but was extended to treat the case of a two-cell structure such as that used in

the present investigation. The single-cell version of the analysis was verified through comparison to an MSC/NASTRAN model in reference 12, and through experimental correlation in reference 4.

A laminate analysis, based on reference 13, is used to determine the material strength margins of safety. The margins are based on Tsai-Hill first-ply failure theory, and are calculated using material design-limit strengths. Material design-limit strengths are obtained by dividing the material ultimate strengths by a factor of 1.5. The use of material design-limit strengths insures linear elastic behavior, and a conservative prediction of the blade strength under the prescribed loads.

Optimization Approach to Blade Design

For this study, the optimization process involves coupling the mathematical optimization analysis CONMIN (ref. 14) to the analyses used to predict deformations and material strength margins of safety. The optimization process requires the definition of an objective function, a set of design variables, and a set of constraints. The optimizer attempts to minimize the objective function through perturbation of the design variables while simultaneously satisfying the prescribed constraints. The objective function and constraints are stated in terms of mathematical expressions which are recalculated for each change in design variable.

The intent of this design study is to achieve twist distributions which improve the performance of the tilt rotor in both hover and forward flight. As discussed previously, this can be accomplished by maximizing the twist deformation for an extension-twist-coupled design at 100 percent rpm. The parameter $\Delta 100$ represents the objective function, and is calculated in the coupled-beam analysis. There are nine design variables needed for the optimization which are defined as follows: three D-spar laminate thicknesses in various sections of the blade, T_{tw},T_{tr}, and T_{rt}; the laminate thickness in the airfoil skin, T_{sk}; the angle of the plies in the airfoil skin, α_{sk}, and in the spar, α_{sp}; the length of the twisting section, Δx_{tw}; the tip weight, *wtip*; and the additional weight per unit span added to the twisting section, *wnsm*. The only constraint imposed in the present study is that the material strength margin of safety is greater than zero at every point in the structure. However, the design variables affect the material strength margins at different spanwise locations to varying degrees. Thus, it is advantageous to consider the minimum margin of safety at each of the 20 blade segments as a separate constraint. With this selection of constraints, the optimizer can more easily determine how to change the design variables to correct a constraint violation. The material strength margins of safety are calculated in the laminate analysis.

The design variables are limited to realistic values through the use of side constraints. Side constraints can not be violated at anytime in the optimization process. The side constraints applied to the thickness variables are 0.0 in. on the lower bound and 0.3 in. on the upper bound. The laminate angles were bounded between 0° and 45°. The twisting section length was limited to a maximum of 105 in. so that there was a minimum of 22.5 in. of transition section. The minimum length of the twisting section was 15 in. The tip weight was limited to a maximum of 15 lb in one design, 60 lb in a second design, and was not limited in a third design. The 15 lb upper limit keeps the tip weight in the range which is typical of conventional rotor blade designs. The 60 lb limit allows the tip weight to reach an unconventional value, but keeps the weight within a value which can be realistically obtained. The absence of an upper limit on tip weight in the third design will result in the maximum twist deformation at 100 percent rpm even though the design may not be practical. Lastly, the additional running weight added to the twisting section was bounded between 0.0 and 0.3 lb/in.

A second optimization based on the final values of the first optimization is required because of the design variables and assumptions used in this study. In practice, the laminates can be

made only in certain thicknesses which are the number of plies multiplied by the thickness per ply available for the material. Also, the laminate layups must be changed in multiples of four plies to maintain the form of $[(\alpha/\alpha+90)_s]_n$, and in multiples of eight plies to maintain the form $[(\alpha/\alpha+90/0/90)_s]_n$. The valid thicknesses could have been maintained in the first optimization by applying multiple constraints on the thickness design variables. However, it is not desirable to apply multiple constraints on these design variables because large jumps in the thicknesses would result. It is likely that large jumps would impede reaching a global minimum of the objective function. A better approach is to repeat the optimization a second time without thickness design variables. The laminate thicknesses resulting from the first optimization are fixed at the nearest available thicknesses which can be processed for the material. The final design results are obtained by repeating the optimization without the use of thickness design variables. For the remaining design variables, the results of the first optimization are used as initial guesses to the second optimization.

Applications

Design Trends

Some design trends were observed through manual perturbation of the design parameters. The magnitude of twist deformation which can be achieved for an extension-twist-coupled design is highly dependent on the tip weight and blade weight distribution. The maximum twist deformation is obtained within material strength constraints when the twist rate is constant. However, this would only be possible using a tip weight in conjunction with a series of weightless blade segments. Since each segment has weight, the centrifugally generated axial load (and thus the blade twist rate) increases from tip to root. These observations indicate that it is desirable to maximize the use of tip weight and minimize the use of running-weight to obtain a twist rate as close to constant as possible. As the tip weight is increased, the total twist deformation which can be produced within the material strength limits is also increased.

Optimization Results

The optimization approach is used to develop three extension-twist-coupled designs. All three designs are based on the same geometry and assumptions discussed previously except for the upper limit allowed on tip weight as previously mentioned. Design 1 is limited to 15 lb of tip weight which is typical of conventional helicopter blade designs. Design 2 is limited to 60 lb of tip weight which is practical though unconventional, and Design 3 is not limited in tip weight which may result in an impractical design.

Results of the first optimization of each design are listed in Table II. The trend indicated by these results agrees with the trend mentioned previously of increasing twist deformation with increasing tip weight. The thickness parameters reported in Table II are redefined to the nearest processable laminate thickness before beginning the second optimization.

Results of the second optimization are listed in Table III. The trend of increasing twist deformation with increasing tip weight is still evident because the twist deformations of the second optimization are shown to be nearly the same as the twist deformations of the first optimization. Using the twist deformations shown in Table III, the twist distributions for the hover condition were calculated assuming a -42° linear twist in forward flight. The power required to hover was then determined for each design using HOVT. Using these results, the hover performance is plotted as a function of gross weight in figure 7 for the three twist designs

and the conventional XV-15 twist. This plot indicates that the differences in horsepower required between the designs are fairly constant with respect to gross weight. Each of the three designs is shown to improve performance over the conventional compromised design. However, Design 3 adds significant weight, about 1200 lb, to the total gross weight which negates any performance improvements it might produce. Design 2 adds about 200 lb to the gross weight which is not negligible, but is not enough to negate its performance improvements.

The hover performance trend with respect to the $\Delta 20$ twist deformation is illustrated in figure 8. This plot shows the change in hover horsepower associated with the the three twist designs from that associated with the conventional XV-15 twist, with a negative change in hover horsepower indicating a performance improvement. A zero value of $\Delta 20$ corresponds to a design with no extension-twist coupling so that the twist in hover is the same as the twist in forward flight. The hover performance of the conventional XV-15 twist design is achieved with a $\Delta 20$ of about 4.5°. The change in horsepower required is shown to decrease at a diminishing rate until a minimum value of about -6 percent is reached at a $\Delta 20$ of approximately 15°. This indicates that for practical designs, with nonlinear twist distributions resulting from extension-twist deformation, a $\Delta 20$ of 15° is the optimum twist change for hover which is less twist change than is desired if both twist distributions in hover and forward flight were linear. As discussed previously, the optimum linear twist distributions are -20° and -42° which would require a 22° $\Delta 20$ twist deformation. The plot also shows that the hover horsepower-required is improved over the compromised hover horsepower-required by 2.6, 4.8, and 5.8 percent for Design 1, Design 2, and Design 3, respectively.

An example of the change in twist distribution with rotational velocity is presented for Design 1 in figure 9. The twist at 80 percent rpm is shown to be a -42° linear distribution while nonlinear twist distributions are required for the blade at 0 and 100 percent rpm. The figure further shows that the built-in blade twist at 0 rpm has a large negative distribution. As the rotational velocity is increased, the twist distribution moves in a positive direction until it reaches its least negative distribution at 100 percent rpm for the hover condition. As the rotor speed is decreased to 80 percent rpm for forward flight, the twist moves back to a more negative distribution which is linear.

Finite Element Comparison

The bending and twist deformation results calculated with the coupled-beam analysis were compared with an MSC/NASTRAN built-up shell finite-element model for Design 1. The reason for the comparison is that the beam analysis had only been previously verified for single cell structures. The D-spar and airfoil skin comprise a two-cell structure. The finite element model, illustrated in figure 10, used 5640 degrees of freedom and 920 flat-plate quadrilateral elements to model the structure. The substantiating loads applied in the previous design studies were applied to the finite element model, and the resulting twist and translations are listed in Table IV along with the coupled-beam model results. The finite element and coupled-beam results correlate well with a maximum difference of about 3 percent which indicates that the coupled-beam analysis is adequate for use in this design study.

Concluding Remarks

This paper describes the aerodynamic and structural design of a passive blade-twist control concept for tilt rotor aircraft. The linear twist distributions which result in minimum horsepower required were determined for both hover and forward flight. The performance improvement resulting from use of these linear twists instead of the conventional XV-15 twist

was also determined. Three extension-twist-coupled rotor blades were designed with various amounts of tip weight. Design 1 had 15 lb of tip weight which is typical of conventional helicopter rotor blades. Design 2 had 60 lb of tip weight which is much more than that used in conventional designs, but can be realistically obtained. Design 3 was not limited in tip weight which resulted in the maximum twist deformation although the weight increase made the design impractical. All three designs were obtained by optimizing for maximum twist deformation subject to material strength constraints. The twist change available for each design was used to determine its hover twist distribution and the resulting hover performance.

The performance benefits predicted for use of the aerodynamically-optimum linear twists instead of the conventional XV-15 twist were significant, with 6.1 percent horsepower savings in hover and 6.5 percent horsepower savings in forward flight. All three extension-twist-coupled designs resulted in 6.5 percent horsepower improvements in forward flight because the optimum linear twist was obtained. In increasing the rotor speed for the hover condition, significant amounts of twist change were developed which resulted in nonlinear twist distributions. The hover twist distributions resulted in significant improvements to hover performance with savings of 2.6, 4.8, and 5.8 percent for Design 1, Design 2, and Design 3, respectively. Design 3 increased the gross weight substantially which makes that design impractical. Designs 1 and 2, however, are practical and result in significant improvements in horsepower required in both hover and forward flight. These results indicate that the passive blade-twist control concept is viable and can enhance current tilt rotor performance.

References

1. McVeigh, M.A., Rosenstein, H.J. and McHugh, F.J.: Aerodynamic Design of the XV–15 Advanced Composite Tilt Rotor Blade. 39th Annual Forum, American Helicopter Soc., 1983.

2. Johnson, W., Lau, B.H., and Bowles, J.V.: Calculated Performance, Stability, and Maneuverability of High Speed Tilting Proprotor Aircraft. *Vertica*, Vol. 11, 1987.

3. Bauchau, O.A., Loewy, R.G., and Bryan, P.S.: An Approach to Ideal Twist Distribution in Tilt Rotor VSTOL Blade Designs. AHS 39th Annual Forum Proceedings, May 1983.

4. Nixon, M.W.: Extension–Twist Coupling of Circular Tubes with Application to Tilt Rotor Blade Design. AIAA Paper 87-0772, 28th Structures, Structural Dynamics and Materials Conference, Monterey, CA, April 1987.

5. McCormick, B. W., Jr.: *Aerodynamics of V/STOL Flight.* Academic Press, c.1967.

6. Johnson, W.: A Comprehensive Analytical Model of Rotorcraft Aerodynamics and Dynamics. NASA TM 81182, 81183, 81184, 1980.

7. Gessow, A., and Myers, G.C.: *Aerodynamics of the Helicopter.* Frederick Ungar Publishing Co., c.1952.

8. Walsh, J.L., Bingham, G.J., and Riley, M.F.. Optimization Methods Applied to the Aerodynamic Design of Helicopter Rotor Blades. AIAA Paper 85-0644, 26th Structures, Structural Dynamics, and Materials Conference, Orlando, FL, April 1985.

9. *Military Specification – Structural Design Requirements, Helicopters.* MIL-S-8698, Apr. 20, 1981.

10. Bell Helicopter Company: Advancement of Proprotor Technology – Wind Tunnel Test Results. NASA CR-114363, Sept. 1971.

11. Rehfield, L. W.: Design Analysis Methodology for Composite Rotor Blades. Seventh DoD/NASA Conference on Fibrous Composites in Structural Design, Denver, CO, June 1985.

12. Hodges, R.V., Nixon, M.W., and Rehfield, L.W.: Comparison of Composite Rotor Blade Models: A Coupled-Beam Analysis and an MSC/NASTRAN Finite-Element Model, NASA TM 89024, AVSCOM TM 87-B-2, 1987.

13. Jones, R.M.:*Mechanics of Composite Materials.* McGraw-hill, 1975.

14. Vanderplaats, G. N.: CONMIN – A FORTRAN Program for Constrained Function Minimization, User's Manual. NASA TM X-62 282, August 1973.

Table I. Material Properties of IM6/R6376.

E_{11} (psi x 10^{-6})	E_{22} (psi x 10^{-6})	G_{12} (psi x 10^{-6})	ν_{12}	X (ksi)	Y (ksi)	S (ksi)
23.1	1.4	.76	.34	261.	7.1	11.9

Table II. Extension-Twist-Coupled Designs After First Optimization.

	T_{tw} ($in.$)	T_{tr} ($in.$)	T_{rt} ($in.$)	T_{sk} ($in.$)	α_{sp} ($deg.$)	α_{sk} ($deg.$)	Δx_{tw} ($in.$)	$wtip$ ($lb.$)	$wnsm$ ($lb./in.$)	$\Delta 100$ ($deg.$)
Design 1	.157	.175	.296	.034	15.6	13.8	105.	15.0	.30	19.0
Design 2	.187	.225	.296	.052	15.5	14.2	105.	60.0	.00	27.3
Design 3	.109	.269	.194	.219	13.6	13.0	105.	200.	.00	35.1

Table III. Extension-Twist-Coupled Designs After Second Optimization.

	T_{tw} ($in.$)	T_{tr} ($in.$)	T_{rt} ($in.$)	T_{sk} ($in.$)	α_{sp} ($deg.$)	α_{sk} ($deg.$)	Δx_{tw} ($in.$)	$wtip$ ($lb.$)	$wnsm$ ($lb./in.$)	$\Delta 100$ ($deg.$)	$\Delta 20$ ($deg.$)
Design 1	.154	.176	.286	.044	17.2	16.8	105.	15.0	.30	19.0	6.9
Design 2	.198	.220	.286	.044	15.0	13.3	105.	60.0	.00	27.3	9.8
Design 3	.110	.264	.198	.220	13.2	13.0	105.	201.	.00	35.0	12.8

Table IV. Finite Element and Coupled-Beam Deformations for Design 1.

	Chordwise Translation ($in.$)	Flapwise Translation ($in.$)	Twist, $\Delta 100$ ($deg.$)
Finite Element Model	.466	15.71	19.42
Coupled Beam Model	.433	15.32	19.02

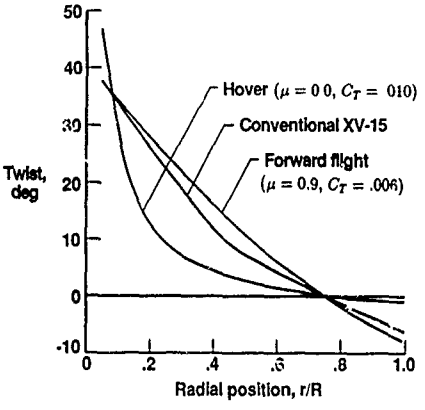


Figure 1. Twist distributions based on uniform inflow.

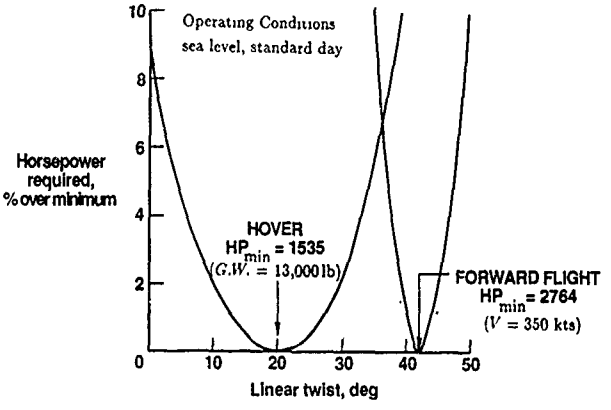


Figure 2. Optimum linear twist distributions for a typical tilt rotor in hover and forward flight.

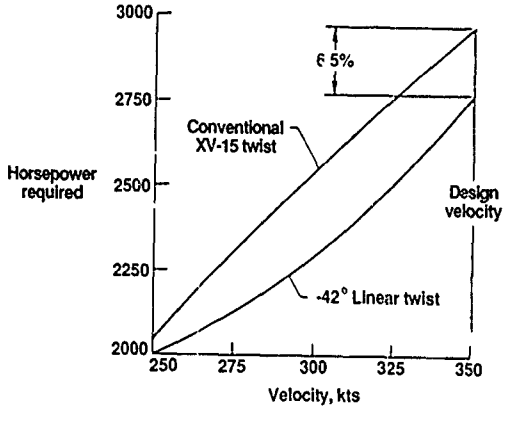


Figure 3. Comparison of forward flight performance for two blade twist distributions.

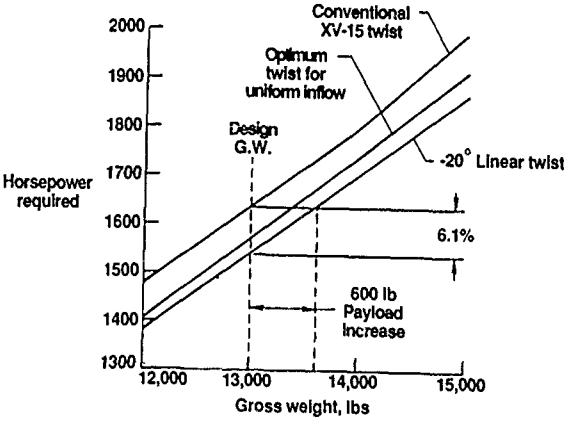


Figure 4. Comparison of hover performance for three blade twist distributions.

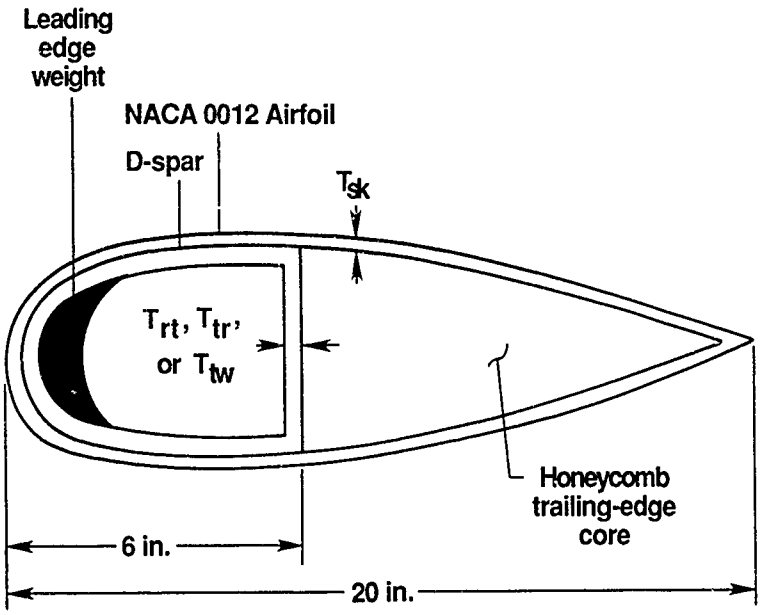


Figure 5. Cross section of the structural model.

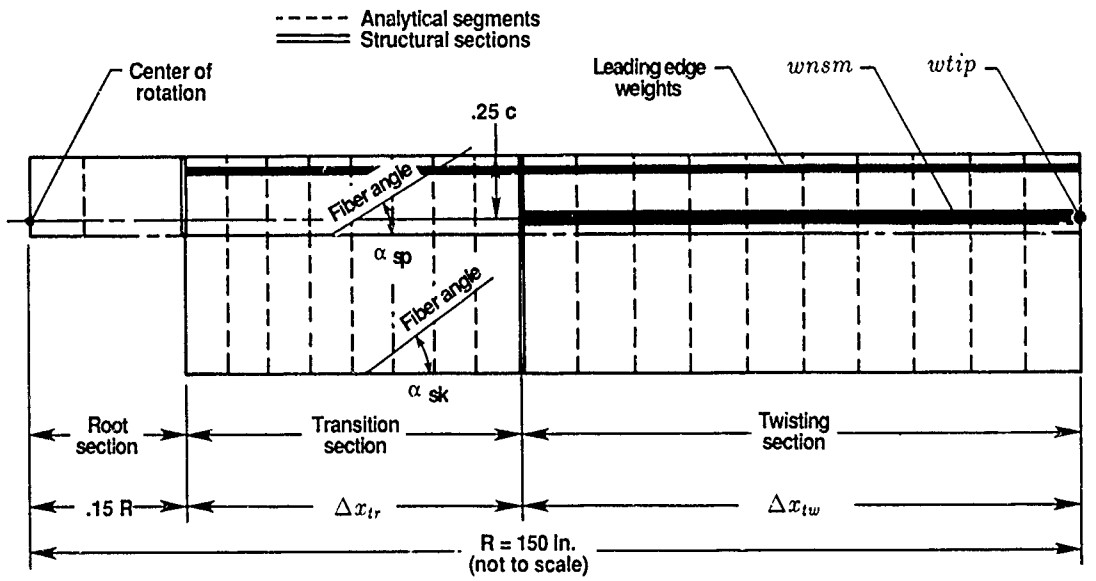


Figure 6. Planform of the structural model.

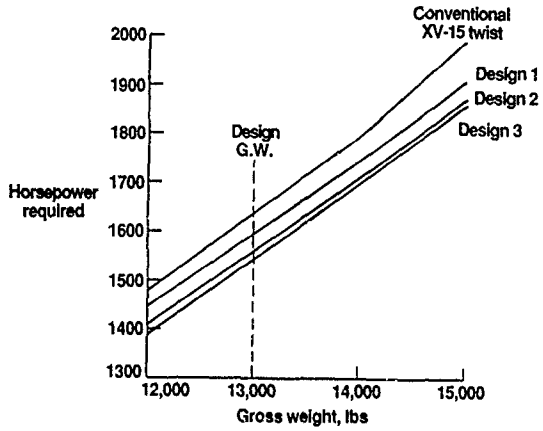


Figure 7. Horsepower required to hover as a function of gross weight.

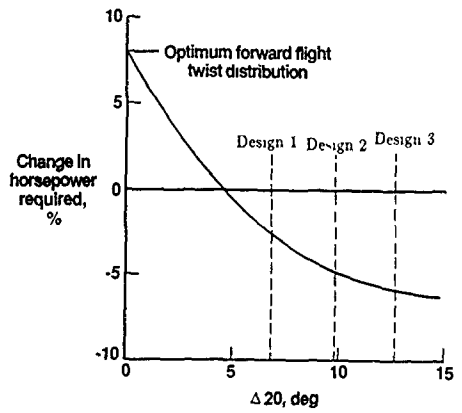


Figure 8. Change in hover horsepower required as a function of twist deformation.

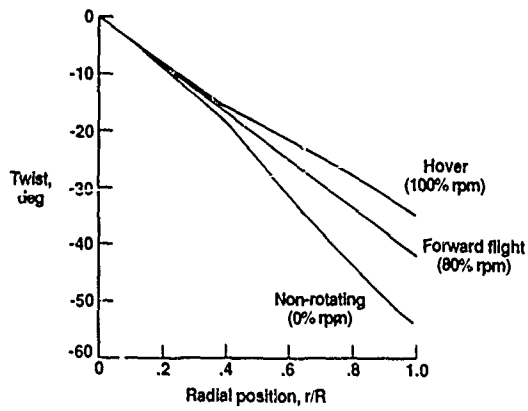


Figure 9. Blade twist distribution of Design 1 in three modes.

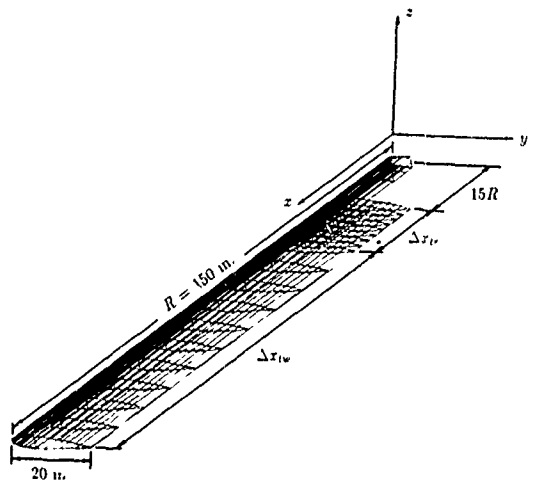


Figure 10. Finite element model of the extension-twist-coupled blade designs.

A Realistic Evaluation of Traveling Charge(U)

W.F. Oberle*, R.E. Tompkins, K.J. White, G.P. Wren, A.A. Juhasz
U.S. Army Ballistic Research Laboratory
Aberdeen Proving Ground, Maryland 21005-5066

I. INTRODUCTION

The traveling charge concept or "impulse gun", originally proposed by Langweiler,[1] is a solid propellant propulsion technique felt by ballisticians to offer the prospect of muzzle exit velocities in the 2 to 3 km/s range without the high breech pressures (700-1000 MPa)[3] required of conventional gun propulsion systems. Resulting advantages of velocities of this magnitude have been discussed by various authors[1,2] and can be summarized as improved delivery range, increased target penetration due to higher kinetic energy of the projectile, and enhanced hit probability resulting from the decreased time-of-flight.

No attempt, in this report, is made to present a theoretical analysis of the traveling charge concept or review previous experimental results. The interested reader is referred to the works of Langweiler,[1] and May et al.[2] for a discussion of the theoretical analysis and development of computer models for the traveling charge concept. Findings of previous experimental efforts can be found in the report by May et al.[2] An idealized description of the traveling charge effect has been presented in an earlier work by Smith[3] and is shown in Figure 1. The ignition process is in two stages. A conventional granular booster charge is used to rapidly pressurize the chamber and accelerate both the projectile and a very high burning rate (VHBR) propellant charge (traveling charge, TC) attached to the base of the projectile. At some point during this initial pressurization, usually past the peak pressure due to the booster charge, the traveling charge is ignited. Burning is in such a manner as to generate and eject combustion products at sufficient velocity to maintain constant thrust/pressure on the projectile base and to increase projectile velocity. Thus in a traveling charge, the average propellant gas velocity at muzzle exit is lower than for a conventional charge hence a greater portion of the chemical energy

goes into the projectile kinetic energy rather than the propellant gases. At very high velocities, the traveling charge is expected to be more efficient than conventional propelling charges, and higher projectile velocities are possible for a given maximum chamber pressure.

In summary, the traveling charge effect is characterized by:
a) The attachment to the projectile of a very high burning rate propellant which travels with the projectile down the tube.
b) Deviation from the "normal" pressure gradient which would be obtained if all the propellant, booster and traveling charge were placed in the chamber. The deviation should show lower chamber pressures and increased downbore pressures.
c) An increase in muzzle velocity over the corresponding conventional firing.

For the past several years the Ballistic Research Laboratory (BRL) has been involved in an experimental effort to demonstrate the traveling charge concept as a practical and useful gun propulsion system in a 14-mm bore diameter test fixture. The purpose of this report is to summarize the major results of this experimental effort. The areas investigated in this report are propellant combustion diagnostics, experimental gun firings, and computer modeling and predictions.

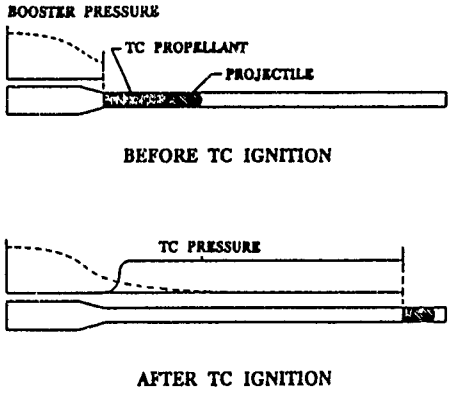


Fig. 1. Idealized Traveling Charge Concept

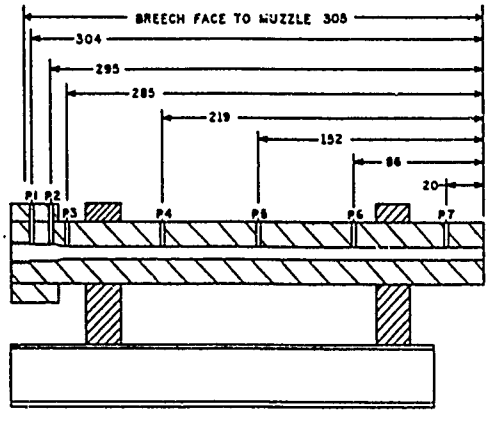


Fig. 2. 14-mm Traveling Charge Gun Test Fixture (Dimensions in cm)

II. COMBUSTION DIAGNOSTICS

As was discussed in the introduction, the TC in the gun fixture will be ignited and will burn at or after the booster has reached peak

pressure. Thus, screening of formulations to find candidates for TC were carried out in a closed chamber under conditions that were expected to be encountered in the gun fixture. A booster charge was ignited in the chamber to raise the pressure up to several hundred megapascals pressure. The traveling charge then ignited and burning times were recorded. Based on projected velocities on the order of 2 km/s, projectile base pressures of 250 MPa, and TC sample length of 25-50 mm, estimated burn times of 0.5 to 2 ms are required for a 14-mm gun fixture. Details of the test procedures and results can be found in separate reports[4,5] However, the combustion diagnostics can be summarized as follows:

1) Ignition and combustion were observed under combined shear/thermal load.
2) Shear forces alone are insufficient for ignition.
3) Side confinement and immobilization of the sample are required for rapid burn.
4) The burn time is strongly dependent on the booster pressure.
5) Although some control of ignition delay times has been achieved, a reproducible ignition delay unit has not yet been designed.

III. EXPERIMENTAL FEATURES AND COMPUTER CODE

A schematic of the gun fixture, with pressure port locations, is shown in Fig. 2. The bore diameter is 14 mm. The smooth-bore barrel length is 2900 mm. A conventional booster propellant, to be ignited in the 100 cm^3 chamber, provides the initial acceleration to the projectile. The booster propellant used for the gun firings was a non-deterred, unrolled small arms ball propellant. A non-deterred propellant was chosen in order to simplify the interior ballistic calculations. For all firings a 34-g booster charge with a 1.5 g igniter of black powder was utilized.

A sketch of the projectile used for the majority of firings is shown in Fig. 3. The overall projectile length, when using a 50-mm sample of TC propellant, is 104 mm. The nominal mass of the unloaded projectile is 14.3 grams. For most firings, a 9.6 g TC sample was epoxied into the rear cavity of the projectile (Fig. 3). From earlier calculations[6] it had been determined that the best TC performance is achieved when ignition is delayed until after the maximum booster pressure. Consequently, an ignition delay element was added to the end of the TC. The most reliable device was a layer of masking tape, with vacuum grease at the edges.

The computer code selected to model the interior ballistic event was the XNOVAKTC (XKTC) code developed by Paul Gough Associates. This code is a combination of a newer version of the NOVA[7] code together with the BRLTC[8] code. Selection of XKTC was based upon several factors. First, the code has the capability to model conventional,

traveling charge, and a combination of booster and traveling charge gun firings. Second, the code includes kinetic options which allow flexibility in investigating the traveling charge effect. The final factor in selecting XKTC was its demonstrated accuracy in predicting conventional gun performance in terms of pressure profiles, pressure oscillations, and velocity. Calibration results of XKTC for the 14-mm test fixture showed pressure discrepancies of one MPa at the breech and muzzle velocities within one m/s between conventional gun firings and simulations.

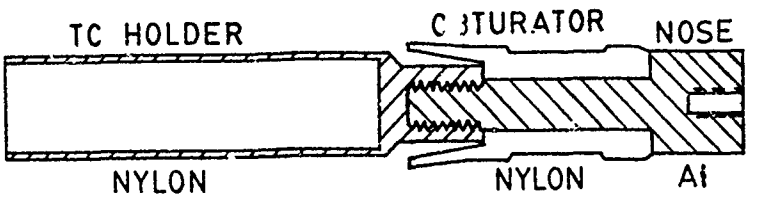


Fig. 3. Traveling Charge Projectile Schematic

IV. EXPERIMENTAL FIRINGS AND CODE SIMULATIONS

Throughout the duration of the experimental program over fifty traveling charge firings have been performed with results detailed in separate reports.[4,5] In this paper, a reduced selection of firings, which demonstrate the important features of the traveling charge, will be presented. Results for the experimental TC firings and resulting calculations are summarized in Table I, as ID 43, 44, 50 and 51. Also included in this table is ID 49. This test is identical to ID 43, 44, 50 and 51 except that the TC did not ignite. The usefulness of test ID 49 is to yield a direct comparison of the effect of the TC functioning on the interior ballistic processes.

Table I. Results of TC Firings and Simulations

REMARKS	ID	P1	P2	P3	P4	P5	P6	P7	VELOCITY (m/s)	IG DELAY# (ms)
		(pressures in MPa)								
TC firing	43	327	322	--	130	78	39	31	1846	+0.03
"	44	323	310	--	141	62	40	25	1887	+0.03
"	50	326	324	375	122	76	30	23	1820	-0.15
"	51	434	408	450	85	79	47	35	1805	-0.25
TC calculations	44	311	--	285	87	52	35	27	1840	+0.16
"	51	419	--	404	100	85	51	35	1863	-0.19
TC firing (no ignition)	49	306	295	278	133	71	38	23	1440	----
TC calculations (no ignition)	49	324	---	---	125	68	45	28	1463	----

#- TC ignition delay with respect to maximum chamber pressure.

To demonstrate the effect that the traveling charge can have on the interior ballistic process, consider the pressure and velocity histories for ID 44, a successful TC firing, and ID 49, (velocity only) where the TC did not ignite. These histories are given in Figs. 4 and 5 Ignition of the TC is shown as point D in Fig. 5. Prior to TC ignition, the pressures and velocities should be the same for both firings. Examination of the velocity histories (Fig. 5) indicates that up until the time of maximum chamber pressure the velocities are identical. After this time the TC ignites (point D), and the acceleration is substantially increased for ID 44. Thus, ignition of the traveling charge results in improved acceleration and velocity without substantial increases in chamber pressure.

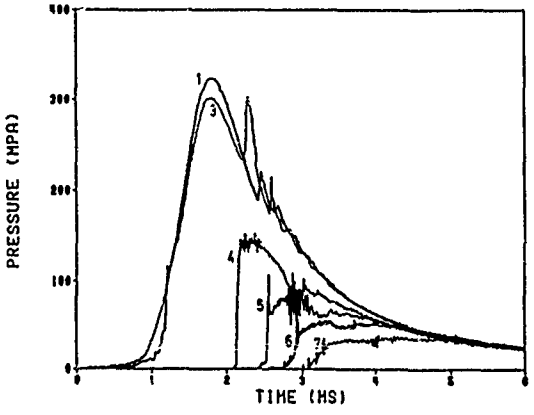


Fig. 4. ID 44, Exp. Press. Late TC Ignition

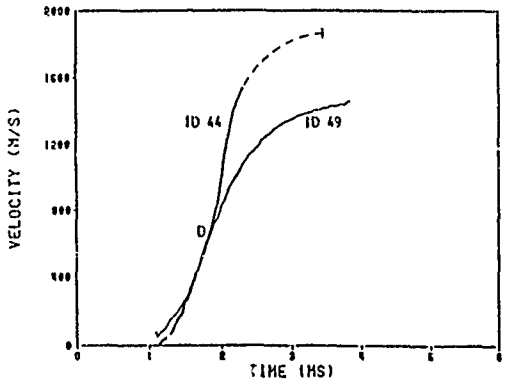


Fig. 5. Exp. Velocity, (44) Late TC Ignition, (49) No TC Ignition

Unfortunately, obtaining the increased acceleration and velocity without the large increases in chamber pressure associated with TC ignition appears to be heavily dependent on the time at which the traveling charge ignites. In ID's 43 and 44, ignition of the TC was just after attainment of the maximum pressure due to the booster propellant (see Table 1.). For ID's 50 and 51, TC ignition occurred before the maximum pressure was reached. Pressure and velocity histories for TC firing ID 51 are shown in Figs. 6 and 7. The chamber pressure is larger in magnitude by 100 MPa, although the muzzle velocity is lower by 60 m/s than for the TC firings ID 43 and 44. Additionally, close examination of the velocity history indicates that the largest acceleration for ID 51 occurs prior to maximum breech pressure whereas for 43 and 44 this occurs slightly after maximum breech pressure, corresponding to the time at which the TC ignites. The observed variability in the traveling charge firings can in all probability be attributed to the non-reproducibility of the ignition delay element.

In simulating the traveling charge firings using XKTC the same input data that was utilized in simulating the conventional firings (booster propellant burn rate, as determined from closed chamber firings, and the resistance profile, as determined by matching calculated pressure and velocity histories with experiment) was used. Only characteristics of the traveling charge were varied to try to match calculations with experiment. An estimate of the TC ignition was made from an examination of the experimental velocity histories.

In the initial set of calculations[6], a pressure-dependent burn rate law was used to describe the TC combustion. However, muzzle velocities were consistently larger than experimental velocities. Therefore, two changes, based upon experimental evidence, were made in the burning characteristics of the TC. First, since the calculated

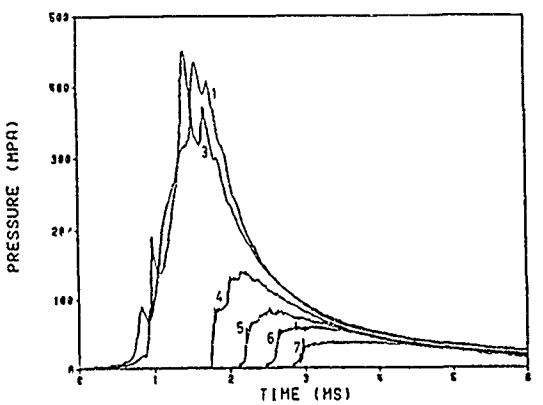


Fig. 6. ID 51, Exp Press. Early TC Ignition

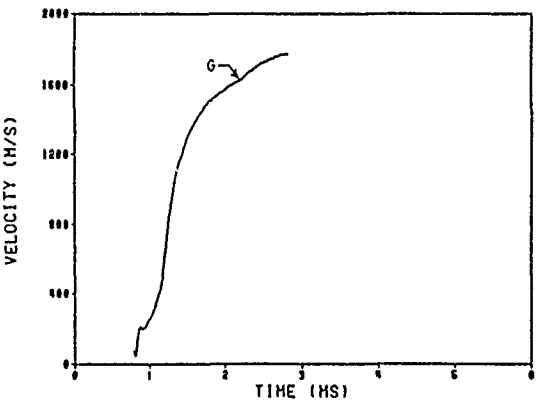


Fig. 7. ID 51, Exp. Velocity Early TC Ignition

velocities were systematically high, it was speculated that not all of the 9.6 g of TC was burning and contributing energy to accelerating the projectile. There is some experimental evidence for this hypothesis. Closed chamber firings[9] have indicated that for this type of propellant the final pressures are not always identical with those predicted from thermochemical calculations. This could be due to incomplete combustion of the formulations. Also, a witness plate in front of the muzzle for the experimental firings has indicated that a substantial amount of small particulate matter is being accelerated with the projectile. This could be interpreted as unburned TC. Thus, a series of calculations was performed in which the burnt amount of TC, out of the total of 9.6 g, was varied. Second, closed chamber firings of a number of TC formulations have indicated that cylindrical samples, similar to those used in the traveling charge, burn with some form of deconsolidation rather than in a laminar fashion. Pressure

histories indicate a decreasing mass generation rate as a function of burn time[10].

As a consequence of these two observations, the TC combustion data used in the computations were altered in the following way. First, the total amount of TC burned was reduced from 9.6 g to 5.5 g, with the difference of 4.1 g being added to the projectile mass. Second, a dual burn rate law was introduced depending an the amount of TC burned,

$$r(m/s) = 127 \quad (0 \text{ to } 4.5 \text{ g})$$

$$r(m/s) = 25.4 \quad (4.5 \text{ to } 5.5 \text{ g}).$$

Thus, the first 4.5 g burned at the rate of 127 m/s and the final 1 g burned at a rate of 25.4 m/s.

Results from the calculations are presented in Table I. Pressure and velocity histories for the simulation of ID 44 is shown in Figs. 8 and 9. The agreement with the experimental results (ID 44, Figs. 4 and 5) is reasonably good.

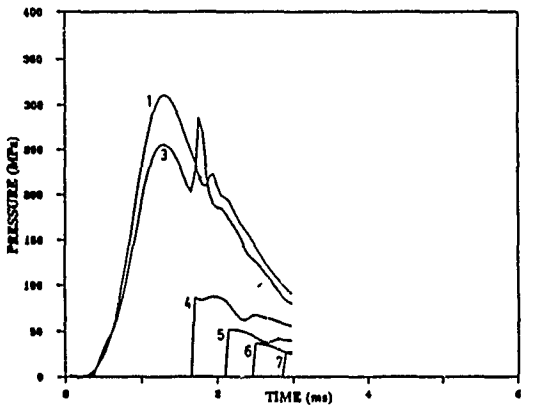


Fig. 8. Simulation of ID 44 Pressure

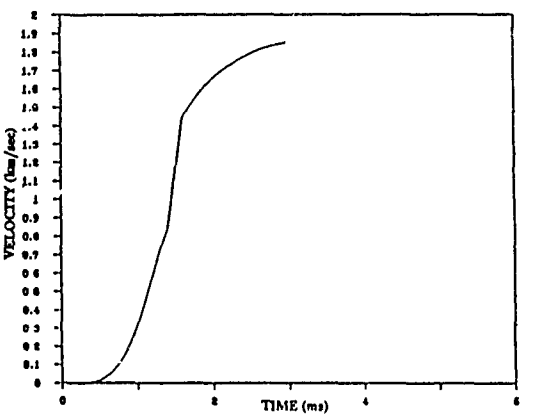


Fig. 9. Simulation of ID 44 Velocity

An attempt was made to simulate the experimental results from ID 51. In this case the ignition of the TC is taking place prior to the maximum booster pressure. The value chosen was -0.190 ms with respect to the maximum chamber pressure. The results are shown in Table I and Figs. 10 and 11. Comparing with the experimental pressure histories, Fig. 6, shows the predicted large pressure waves observed experimentally. The velocity history (Fig. 11) shows some interesting characteristics. The first element of the TC ignites at time C and burns out at time E. The second element ignites at E and burns out at F. These characteristics are also observed in the experimental data, Fig. 7. After the burnout, a decrease in acceleration is observed

until point G, at hich both the calculated and experimental velocity histories show an unusual inflection. To investigate this inflection point, a study was made of the pressure vs. distance and gas velocity vs. distance produced by the code. These results indicate that the ignition and burn-out of the TC induces large pressure pulses at the base of the projectile. At a later time in the interior ballistic cycle, this pulse dissipates after stagnating against the breech face. Slightly later, the normal pressure gradient from breech to projectile decreases. The pressure at the breech goes down but, the projectile base pressure increases. This occurs at G producing an increase in velocity through the remainder of the projectile travel. The code is showing the inflection in the velocity history. This phenomenon is somehow related to the pressure pulses generated by the traveling charge, but the exact details of the hydrodynamics are not clear at this time. The implications of these observations and its impact on the overall TC process will be a subject for future investigation.

Conclusions from the experimental gun firings and code simulations are given below.

1) Experimental data indicate a distinct difference between the character and magnitude of pressure and velocity histories of TC and conventional firings.
2) The two phase hydrodynamic interior ballistic code (XNOVAKTC) gives a good representation of the pressure and velocity histories for both conventional and TC firings. It can confidently be used to study the wave dynamics of the problem and help determine the important criteria that must be met for a successful application of the traveling charge concept.
3) The results indicate that not all of the potential energy from the traveling charge was realized in the 14-mm gun firings. The reasons for this are not clear at this time.
4) Results indicate that the VHBR formulations show degressive burning characteristics under TC conditions. This confirms what was observed in combustion diagnostic studies.
5) Successful application of the traveling charge concept will require precise control over ignition and burnout times of the TC propellant.

V. TRAVELING CHARGE COMPUTATIONAL PREDICTIONS

In the experimental section, it appeared that a critical factor in traveling charge performance was ignition time of TC relative to the maximum breech pressure due to the booster propellant. To further investigate the effect that TC ignition time could have on performance, a parametric study using XKTC was performed. Results of this study, together with the results of optimized TC performance compared to conventional performance and an investigation of situations where traveling charge is most beneficial are presented.

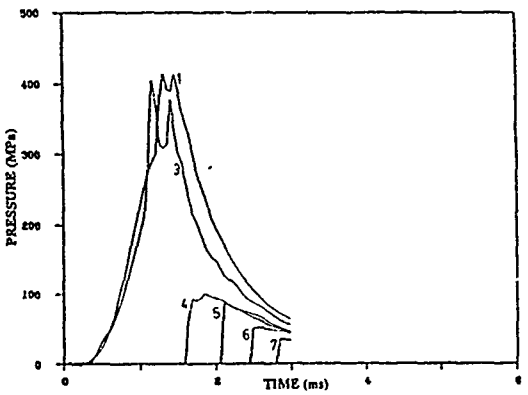


Fig. 10. Simulation of ID 51, Pressure

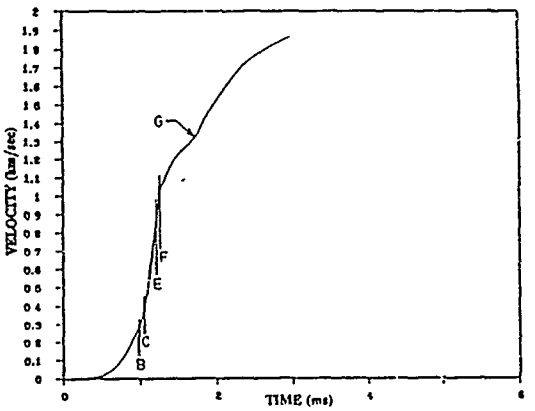


Fig. 11. Simulation of ID 51, Velocity

VI. TC IGNITION TIME/LOCATION OF TC BURNOUT

TC ignition time refers to the time at which the TC ignites relative to the start of the ballistic event. For example, an ignition time of 1.15 ms means that the TC ignites 1.15 ms after the primer is ignited. Location of TC burnout refers to the position of the projectile in its travel at which the TC has totally burnt out. The two are considered together since a change in ignition time will affect the burnout position if the burn rate of the TC is held fixed. In the study, the gun geometry utilized in the code was identical to the 14-mm experimental gun fixture. Propellant, both booster and TC, burning characteristics were taken from experimental results; however, booster mass and geometry were varied to obtain results for a variety of situations. TC mass in all cases was 8 g. In all instances, the web of the booster propellant was varied so as to maintain maximum gun pressure at 435 MPa. It should be pointed out that the maximum pressure of 435 MPa is due totally to the booster charge acting on the projectile and 8 g of traveling charge. If the burning of the TC caused pressures above 435 MPa the web of the booster was not adjusted to reduce the pressure, and the simulation was not considered in the study. The purpose of the study was to determine if there was an optimal ignition delay for the TC to provide the maximum velocity and to determine performance sensitivity to variations in the TC ignition delay.

Fig. 12 summarizes the relation of velocity to TC ignition time for a variety of different propellant geometries and charge-to-mass ratios. No direct comparison between the velocities for the different cases should be made due to the different propellant and projectile masses used in the studies. However, the shape of the curve in each case is identical. Velocity increases with delayed ignition time up to a point. This point occurs at an ignition time beyond which all

the TC does not burn out in the bore. That is, for maximum performance the TC should burnout at muzzle exit. For each curve, TC ignition times prior to those indicated on the graph result in breech pressure exceeding the 435 MPa constraint.

Also, the graphs of Fig. 12 can be used to investigate the sensitivity of performance as a function of TC ignition time relative to different charge-to-mass ratios. The slope of the graph, change in velocity divided by change in ignition time, is a measure of this sensitivity. As can be seen in Fig. 12, going from a c/m of 1 to a c/m of 2.3 for the ball propellant shows a large increase in slope. An even larger increase in slope is shown in going to a c/m of 4. Thus, the sensitivity of performance to changes in TC ignition time increases as the charge-to-mass ratio increases.

Clearly, not having TC burnout in bore should result in lowered velocities since additional energy must be expended to accelerate the parasitic mass, unburnt TC, attached to projectile. However, the reason for improved velocities with TC ignition times resulting in TC burnout near muzzle exit is not as obvious, since delaying TC ignition means that the unburnt TC must be accelerated further down the tube. One possible explanation is presented in Fig. 13. Shown are graphs of base pressure versus travel for two different traveling charge simula-

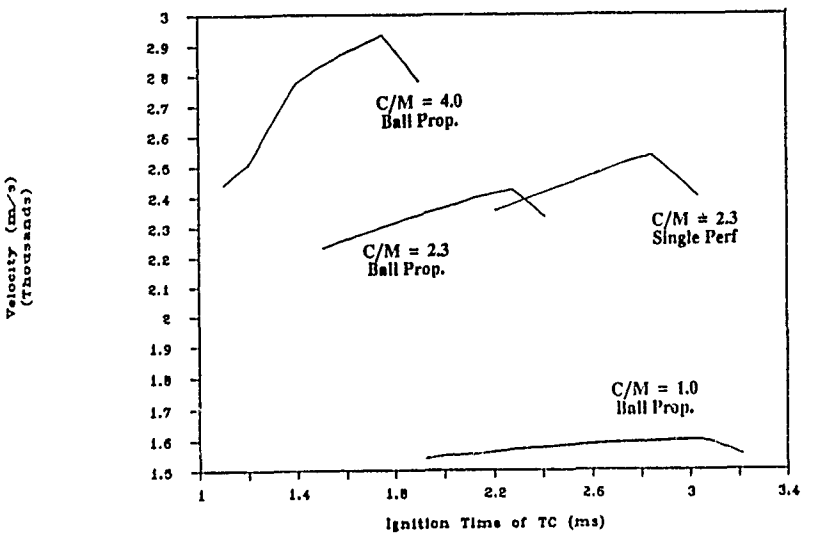


Fig. 12. Velocity Versus TC Ignition Time for a Variety of Simulations

tions with TC burnout at 135 cm (A) and 60 cm, (O) respectively. For both of the graphs there are abrupt drops in pressure, from 200 to 100 MPa for curve O and from 100 to 50 MPa for curve A. The position in the travel at which the drop begins is approximately the location at which the TC burned out. Thus, burnout of TC in-bore appears to

result in a drop in magnitude of about 50% in the base pressure. The cause for this drop may be that TC burn-out is accompanied by a loss of gas generation at the rear of the projectile and a momentary pressure drop. Since the energy imparted to the projectile is essentially the area under the base pressure versus travel curve, more energy will be delivered to the projectile by avoiding the large drop in base pressure. That is, a TC ignition time which results in TC burnout near muzzle exit may increase the energy delivered to the projectile.

VII. TRAVELING CHARGE VERSUS CONVENTIONAL SIMULATIONS

Having determined the importance of TC ignition time and burnout location, it was of interest to investigate the potential benefits which could be obtained through the use of the traveling charge concept. Thus, comparisons between optimal performance as a conventional gun versus the best traveling charge results, as predicted by XKTC, for the 14-mm gun fixture were determined (Table II). The attempt here is simply to attain the maximum velocity possible from the system given a propellant type, with its mass and web allowed to vary, under a constraint of 435 MPa maximum gun pressure ("normal chamber"). In all computations no bore resistance was included since the actual bore resistance profile for the 14-mm test fixture is unknown. The "extended chamber" refers to enlarging the chamber to include the volume that would be occupied by the TC in the traveling charge simulation. Thus, the "extended chamber" cases are conventional simulations utilizing an increased chamber volume. The "fixed total energy" case refers to fixing the mass of propellant used in the conventional simulation to be the same as the total mass used in the best TC simulation, in this instance 92.75 g (63 + 29.75), and then allowing the propellant web and chamber volume to vary to obtain optimal velocity. Finally, the "short tube" calculations refer to reducing projectile travel from 2900 mm, 200 calibers, to 1450 mm, a more realistic 100 calibers.

Several of the results from Table II are worthy of mention. First, the additional chamber volume resulting from the "extended chamber" has not resulted in an increase in velocity compared with the "normal chamber". Second, for the "normal chamber" the optimal results are obtained using a single perforated grain instead of the expected seven perf. Finally, "traveling charge simulation" predicts (2909 m/s) a velocity increase of 655 m/s, a 29% increase over the best conventional case (1-perf, "normal chamber", Table II) and a velocity increase of 829 m/s, a 40% increase, over the "fixed energy" calculation. For the "short tube" configuration the calculations predict a 27.3% increase for traveling charge over conventional charge (2408 vs. 1891).

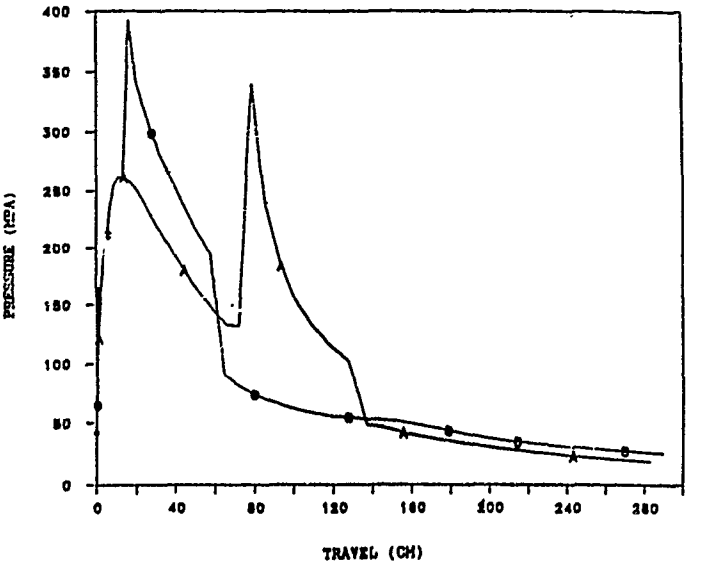


Fig. 13. Base Pressure Versus Travel for Traveling Charge Simulations

Table II. XKTC Optimization of Experimental 14-mm Gun Fixture

Booster Type (-)	Booster Mass (g)	Traveling Charge Mass (g)	Projectile Mass (g)	Velocity (m/s)
	NORMAL CHAMBER -- CONVENTIONAL FIRING			
Ball	53	-	18	2154
1-perf	58	-	18	2254
7-perf	60	-	18	2193
	EXTENDED CHAMBER -- CONVENTIONAL FIRING			
1-perf	64	-	18	2239
7-perf	64	-	18	2192
FIXED TOTAL ENERGY - CONVENTIONAL FIRING (Optimal Loading Density)				
1-perf	92.75	-	18	2080
	TRAVELING CHARGE SIMULATION			
1-perf	63	29.75	18	2909
	SHORT TUBE			
1-perf	58	-	18	1891
1-perf	63	17	18	2408

VIII. APPLICABILITY OF TRAVELING CHARGE

Finally, using XKTC, attempts were made to extend the results to other systems besides the 14 mm. In some cases, substantial increases in velocity were not obtained. The determining factor in all cases appeared to be the ratio of the total mass of both booster and TC to the mass of the projectile, charge-to-mass ratio. As this ratio increased so did the improvement due to the traveling charge. To investigate the effects of this ratio on performance a parametric study using XKTC was performed. The objective was to determine if the benefit of the traveling charge was dependent on the overall charge-to-mass ratio. A series of comparisons was made to examine the difference in velocity between conventional and traveling charge simulations as a function of c/m. In all cases a ball propellant was used for the booster charge, and the propellant diameter was varied to obtain a maximum gun pressure of 435 MPa. For the traveling charge simulations, 40 g of booster and 8 g of TC were used, and the charge mass was considered to be 48 grams for purposes of computing the charge-to-mass ratio. Again the ignition time of the TC was varied to force TC burnout just at muzzle exit. In the conventional calculations two different masses were utilized. In the first series of runs, 48 g of propellant was used so as to maintain constant energy for the comparison with TC. However, with 48 g of propellant total burnout before projectile exit did not occur, and the propellant mass was lowered to 40 g, resulting in total burnout and improved velocity. To obtain the desired charge-to-mass ratio, the mass of the projectile was adjusted. Resulting velocities for the various configurations are presented graphically in Fig. 14. As can be seen, as the c/m increases the benefits in increased velocity from using traveling charge improve.

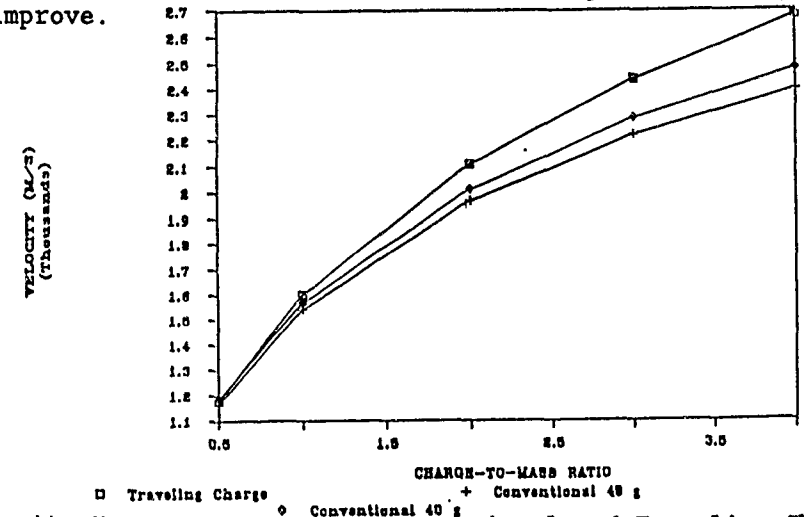


Fig. 14. Comparison between Conventional and Traveling Charge Velocity for Various Charge-to-Mass Ratios

IX. CONCLUSIONS

Combustion diagnostics have shown that several of the candidate formulations have adequate combustion properties for the traveling charge application. The diagnostics have further shown that it is possible to obtain the physical properties required to withstand base pressures during the interior ballistic cycle.

Although experimental traveling charge gun firings have not exhibited the muzzle velocities predicted by XKTC computations, down bore increases in pressure and projectile acceleration have been noted in the traveling charge firings.

Parametric studies with the computer code indicate that the traveling charge concept can offer substantially increased performance over conventional gun systems when hypervelocity performance is required. The implementation of the concept, however, requires precise timing of both the ignition and burn out of the traveling charge propellant. The critical importance of this factor cannot be over emphasized enough. Despite two years of experimental work, the most significant technological hurdle to optimum TC performance remains the development of a precise ignition delay element.

ACKNOWLEDGMENTS

The authors would like to express their appreciation for the support and help given on this paper by I. May, F. Robbins, W. Aungst, J. Newberry, and C. Lewis.

REFERENCES

1. Langweiler, H.,"A Proposal For Increasing the Performance of Weapons by the Correct Burning of Propellant," British Intelligence Objective Subcommittee, Group 2, Ft. Halstead Exploiting Center, 1247.

2. May, I.W., Baron, A.F., Gough, P.S., Baer, P.G., "The Traveling Charge Effect," BRL Memorandum Report ARBRL-MR-03034, Ballistic Research Laboratories, Aberdeen Proving Ground, MD, July 1980.

3. Smith, H.C., "An Investigation of the Use of Porous Propellants in a Traveling Charge Gun," BRL Memorandum Report No. 1554, Ballistic Research Laboratories, Aberdeen Proving Ground, MD, January 1964. (AD#441254)

4. Tompkins, R.E., White, K.J., Oberle, W.F., and Juhasz, A.A., "Traveling Charge Concept-Combustion and Interior Ballistic Studies," 23rd JANNAF Combustion Meeting, CPIA Pub 457, pp 193-213, The Johns Hopkins Applied Physics Laboratory, Laurel, MD, October 1986.

5. Tompkins, R.E., White, K.J., Oberle, W.F., and Juhasz, A.A., "Application of Very High Burning Rate Propellants To Traveling Charge," 24th JANNAF Combustion Meeting, Monterey, CA., October 1987.

6. Oberle, W.F., White, K.J., Tompkins, R.E., and Juhasz, A.A., "Traveling Charge Computations-Experimental Comparisons and Sensitivity Studies," BRL-TR-2844, Ballistic Research Laboratory, Aberdeen Proving Ground, MD, August 1987.

7. Gough, P.S., "The NOVA Code: A User's Manual. Volume 1. Description and Use," IHCR 80-8, Naval Ordnance Station, Indian Head, MD, December 1980.

8. Gough, P.S., "Extensions of BRLTC. A Code for the Digital Simulation of the Traveling Charge," Contract Report.

9. I.W. May, F.R. Lynn, A.A. Juhasz, E. Fisher, P.S. Gough, "Thrust Characterization of Very High Burning Rate Propellants," ARBRL-MR-03359, Ballistic Research Laboratory, Aberdeen Proving Ground, MD, July 1984.

10. K.J. White, D.G. McCoy, J.O. Doali, W.P. Aungst, R.E. Bowman, A.A. Juhasz, "Closed Chamber Burning Characteristics of New VHBR Formulations," BRL-MR-3471, Ballistic Research Laboratory, Aberdeen Proving Ground, MD, October 1985.

REAL-TIME FLOW CONTROL IN LARGE BLAST/THERMAL SIMULATORS

*Klaus O. Opalka
Richard J. Pearson
U.S. Army Ballistic Research Laboratory
Aberdeen Proving Ground, Md. 21005-5066

Due to the increased emphasis on tactical nuclear forces there is a growing concern for the survivability of U.S. weapon systems on a combined conventional/nuclear battlefield. The ban on above-ground nuclear testing forces the U.S. Army to look for alternative testing techniques. The construction of a test facility large enough to test full-sized military equipment like tanks or helicopters has been proposed for simulating ideal blast waves. The U.S. Army, in concert with other government agencies, is presently developing a concept of such a facility suitable to simulate both thermal and blast effects of nuclear explosions. This facility would be used to test military equipment for its survivability and vulnerability criteria and to perform research into nuclear blast phenomenology.

1. INTRODUCTION

A number of DoD instructions and Army procurement regulations specify nuclear hardness levels for over 150 weapon systems within the U.S. Army alone requiring that these weapon systems survive on a combined conventional and nuclear battlefield. There exist currently two techniques for testing the susceptibility of these weapon systems to the blast effects of nuclear weapons, one using high explosives (HE) in large quantities, and the other using special shock tubes. The simulation of nuclear blasts with high explosives requires large test areas and is limited to small yields ($\leq$20 kT). Because of the high cost, the high environmental impact and the low frequency of HE tests (one test per year) the use of specialized shock tubes has been studied with increasing interest in recent years. Specialized shock tubes, called blast wave simulators, allow frequent and repeatable testing (e.g., two tests per week) of a weapon system at a small fraction of the cost of a field test.

The U.S. Army Ballistic Research Laboratory (BRL) has taken the lead role in the design of a Large Blast/Thermal Simulator (LB/TS) on behalf of the U.S. Army Harry Diamond Laboratory and the Defense Nuclear Agency. The operating envelope for the proposed LB/TS is shown in Figure 1. Each blast wave in this envelope is defined by a shock overpressure (p_{so}) and a positive phase duration (PPD). The shock overpressures range from 13.8 to 241.3 kPa and the PPDs from 0.15 to 4.30 seconds. The yields

represented by this set of shock overpressures and PPDs range from 1 to 600 kT for a tactical nuclear air burst with a scaled height of burst of $61m \times W^{1/3}$, i.e., the height is equal to 61 metres times the weapon yield in kilotons to the 1/3 power. Blast waves will be simulated by releasing compressed gas from several high-pressure steel driver tubes into a large expansion tube constructed of prestressed concrete. The thermal simulation will be effected through aluminum/oxygen combustion near the target.

Initially, the U.S. design studies were based on the Large Blast Simulator at the Centre d'Etudes de Gramat (CEG), France. However, a much larger facility was needed in order to accommodate the full range of anticipated targets. A broader range of shock overpressures and weapon yields was also required to cover the test conditions specified in the operating envelope. Parametric studies were initiated to answer questions about the necessary size and the expected performance of such a facility simulating the required range of blast waves. An existing implicit, factored, finite-difference scheme was used to develop the quasi-one-dimensional BRL-Q1D code[1] for simulating the flow in blast wave simulators. This numerical approach was chosen because it could be realized almost immediately and was more suitable to the quick production of a large number of data points.[2-4] An experimental effort was begun also, but required much more preparation time for building a model facility. The results of the experimental work have been used to validate the computational and analytic work.[5,6]

2. THE COMPUTATIONAL SIMULATION

The BRL-Q1D code[1] is a computer program written in the FORTRAN 77 language. The code makes the simplifying assumptions that the compressible and time-dependent flow in the LB/TS is quasi-one-dimensional, adiabatic, and inviscid resulting in the onedimensional Euler equations. The code incorporates an implicit finite-difference technique developed by Beam and Warming[7] which is applied to these quasi-one-dimensional Euler equations. The governing equations are derived for arbitrary geometries in their conservative form and transformed into a uniform computational grid. Linear and nonlinear gridding options are available. Central spatial differencing casts the difference equations into a block-tridiagonal structure which is solved for the increments in the dependent variables at each successive time step. The theory, code structure, input requirements and output options of the BRL-Q1D code are described in detail in Reference 1.

The purpose of the proposed facility is to simulate decaying blast waves similar to those which are generated by nuclear explosions.[8] A simulated blast wave is said to approximate a given free field blast wave if the peak overpressure, the positive phase duration and the impulse correspond to the free field parameters associated with the given yield and shock overpressure. The characteristics of an idealized blast wave at a given distance from the center of the explosion are illustrated in Figure 2. A pressure gauge mounted at this distance would record an instantaneous pressure rise to a peak value p_{so}, unique for this distance, at arrival time t_a. The pressure then decays to ambient in the time period T_+, drops to a partial vacuum thereafter, and finally returns to the ambient

References are listed at the end of the paper.

atmospheric pressure. The impulse of the blast wave is given by the area under the positive overpressure curve and is proportional to the weapon yield.

The shock front of the blast wave passing through the air not only increases the pressure, but also the density ρ and the temperature Θ, and accelerates the air in the direction of its travel. An observer at the gauge location would experience a strong wind blowing away from the center of the explosion after the passage of the shock front. The dynamic pressure wave is characterized by the peak dynamic pressure q_s, the dynamic pressure positive phase duration T_{+q} and the dynamic pressure impulse I_q.

The static overpressure $p(t)$ and the dynamic pressure $q(t)$ decay exponentially with time and their "history" can be described by a number of mathematical equations[8] among which the Modified Friedlander equation is the favored form. It describes the pressure history through the use of three parameters, i.e., the primary shock overpressure p_{so}, the positive phase duration T_+ and the pressure decay constant β for the static overpressure history and the peak dynamic pressure q_s, the dynamic pressure positive phase duration T_{+q} and the pressure decay constant β_q. The constants can be determined from the appropriate impulses.

A tabulation of a 40 kT free field blast wave for a height of burst (HoB) of 208.5 m (i.e., 61 m/kT$^{1/3}$ HoB) is used as the reference data base for simulating blast waves of various overpressures and yields. In the table, the static and dynamic pressures, the time of shock arrival and positive phase durations, and the impulses of the static and dynamic pressures are listed as functions of the distance from ground zero (i.e., the ground point directly under the center of explosion). Sach's scaling laws[8] are employed to adapt the blast wave parameters involving distance and time to a desired weapon yield.

3. THE U.S. LB/TS CONCEPT

Large blast wave simulators are basically shock tubes whose cross-sectional areas vary along their lengths. Straight shock tubes do not produce the flow durations which are observed in decaying blast waves. To obtain long flow durations comparable to those of decaying blast waves the outflow of the driver gas has to be retarded. This is effected by converging the flow area at the exit of the drivers, so that the driver takes the shape of a bottle. Furthermore, a single driver is technically impractical because of the size of the facility and the required supply pressures in the driver. A number of smaller drivers has to be employed to accomodate the necessary initial test conditions.

3.1 THE LAYOUT OF THE FACILITY

A schematic layout of the proposed U.S. LB/TS facility is shown in Figure 3. The shock wave in the expansion tube is driven by an array of nine steel driver tubes [1] which are filled before the test with heated nitrogen. Each driver tube measures 1.83 metres ID (72 inches) and is 60 metres long; the maximum driver volume required by the test conditions is 1450 cubic metres. A specially developed insulation consisting of E-fiberglass in a high temperature polymeric resin matrix will cover the inside of the

driver walls. The driver gas is stored in tanks as liquid nitrogen [15]. When the drivers are to be filled, it is evaporated and superheated in nine pebble bed heaters [17] in which heat is exchanged between hot steel pebbles and the cold nitrogen. The desired gas temperature in the drivers is gained by mixing the superheated nitrogen with bypassed cold nitrogen in a secondary, small pebble bed mixer before it enters the drivers. A large concrete reaction pier [2] is used to anchor the driver tubes.

Convergent nozzles of 22° half-angle [3] connect the driver tubes with the 0.914 metres ID (36 inch) throat sections (Throat/Test-Section Area Ratio = 1:28). The throat valves and diaphragms [4] are mounted in the throat at the exit of the convergent nozzle (the bottle neck). Fast opening and closing valves are the preferred flow control devices over diaphragms because they eliminate the need to adjust the length of the driver tubes for simulating the desired yield. The diaphragms are spent in each test and have to be replaced by new ones before the next test, necessitating a great amount of labor. The rupturing of the diaphragms is controlled by exploding cutting charges which cannot be exposed to heat for safety reasons. Therefore, a double diaphragm system would have to be employed, increasing the labor intensity of the system. On the other hand, diaphragms may be needed because valves that open fast enough to let the escaping gas form a shock front are not state-of-the-art and it is presently not known if they can be constructed.

The expansion tube [5] is 190 m long, of semicircular cross-section and formed of prestressed concrete. A steel test section [6] with a cross-sectional area of 165 m^2 is located in the expansion tube, 100 m downstream from the nozzle exit. The walls of the test section will be equipped with a large number of ports for cameras, lighting and instrumentation [10]. A soil tank [12] for testing shallow buried structures will be located in the floor of the test section. The thermal loading that is associated with blast waves resulting from nuclear explosions will be simulated by thermal radiation sources (TRS) [7], mounted in the floor of the test section just ahead of the target area. Air curtains fed through nozzles from a plenum in the floor [9] will be placed to both sides of the TRS nozzles to contain the flames and the combustion products. Eighty jet pump ejectors [8] would be mounted in the top of the test section to remove the TRS combustion products from the simulator. A rarefaction wave eliminator [13], controlling the flow of gas exiting the expansion tube will be located at the end of the expansion tube.

3.2 THE LB/TS OPERATING SEQUENCE

The operation of the U.S. LB/TS includes several phases, such as the instrumentation of the target and its placement in the test section, the calibrating of the recording instrumentation in the data acquisition center, the programming of the throat valves and of the RWE, the installation of diaphragms and cutting charges, the preparation of the aluminum-oxygen torches, the heating of the driver gas, and the pressurization of the drivers. After final safety checks are performed, the firing sequence is initiated.

In a combined thermal/blast simulation, aluminum and oxygen will first be mixed in the TRS nozzles and ignited. The hot aluminum oxide produced by the combustion irradiates the target, producing the thermal simulation. The aluminum oxide is confined and entrained by the air curtains. Natural convection and air curtain entrainment move the

aluminum oxide upwards toward the ceiling of the test section, where it is sucked out by the ejector pumps.

After the ejectors have removed the aluminum oxide, the shock wave is initiated by either very rapidly opening the throat valves, or by breaking the diaphragms to start the flow through valves which are already open. The shock wave then forms and moving down the expansion tube, begins the blast loading of the target. After the shock is initiated, the preprogrammed throat valves would be automatically closed to meter the flow, sending forward rarefaction waves which cause the pressure behind the shock to decay in the same manner as in a free field blast wave.

At the end of the expansion tube the flow interacts with the RWE. The RWE partially reflects the primary shock and generates a rarefaction wave which interacts with the reflected shock to cancel each other. The RWE then closes such that its exit pressure matches the ambient pressure. When the RWE is properly set, neither a rarefaction nor a compression wave will move upstream into the test section. Finally, rarefaction waves generated in the drivers move through the test section lowering the pressure and decelerating the flow. The simulation ends when the the ambient conditions are restored.

3.3 FLOW PHENOMENA IN AN LB/TS

In order to simulate the U.S. LB/TS in a quasi-one-dimensional context, the total cross-sectional area at any lengthwise location has to be determined by lumping the individual driver and nozzle areas together into one. For computational reasons, a divergent nozzle has to be added to help the expansion of the computational flow along. A diaphragm is assumed in the center of the throat section. The RWE is not modelled here. The computational model of the U.S. LB/TS shown in Figure 4 is used as a reference for the following discussion. It is noted here that our experimental evidence suggests that the one-dimensional representation of the flow overstates these wave phenomena.

The flow patterns encountered in an LB/TS are much more complex than those encountered in a straight shock tube and there are a number of problems encountered in simulating a blast wave. Figure 5 shows a schematic comparison of the flow patterns in a straight shock tube and in an LB/TS. The initial flow pattern in a straight shock tube is made up of a primary shock(2) moving into ambient air(1), followed by a contact surface(3) which separates the warm gas processed by the shock from the cold gas initially in the driver(4). A rearward facing rarefaction wave in the driver(5) accelerates and cools the driver gas. For low shock overpressures (e.g., < 28 kPa for the proposed LB/TS), the flow in an LB/TS is subsonic everywhere, similar to that in a straight shock tube, with a steady expansion in the convergent nozzle and a steady compression in the divergent nozzle.

Generally, the flow in an LB/TS is distinguished from the flow in a conventional shock tube by flow choking in the throat of the nozzle. As the driver empties, the subsonic flow expands isentropically in the convergent nozzle(6) such that it becomes sonic in the throat(7). The flow then becomes supersonic as it continues to expand in the divergent nozzle(8). But because the flow behind the primary shock(2) is subsonic, a

recompression shock(10) must form to match the static pressure behind the primary shock and across the contact surface(3). For moderate shock overpressures (28-70 kPa for the proposed LB/TS), the flow forms a standing shock part way through the divergent nozzle. The subsonic flow behind it goes through an unsteady compression in the remaining part of the nozzle.

The main problem with unheated driver gas in these cases is caused by the temperature difference across the contact surface. The air in the expansion tube is compressed and heated by the shock. The driver gas is cooled below ambient temperature by rarefaction waves and the expansion in the nozzle. Since the pressure is the same on either side of the contact surface, the density of the cool driver gas behind it must be higher than the density of the heated air ahead of it. This density discontinuity causes an undesirable increase in dynamic pressure as the contact surface passes disturbs the blast wave simulation at the test section.

For high shock overpressures (>70 kPa for the proposed LB/TS), the BRL-Q1D code computations indicate that the recompression shock is swept out of the nozzle and down the expansion tube. It is followed by a region of supersonic flow at extremely low pressure. In extreme cases, the recompression shock may be swept past the test section, destroying the blast simulation because the low static, and the high dynamic pressure following the recompression shock do not properly simulate a blast wave. At later times, the recompression shock returns to the nozzle exit where it is partially reflected(16) and partially transmitted(13) moving upstream into the drivers.

Late-time wave patterns are illustrated in Figure 5, also. The decay of static and dynamic pressure necessary for the simulation of a blast wave is produced by rarefaction waves which are reflected from the closed ends of the drivers(14). Moving forward, the rarefactions interact with the convergent nozzle and are partially transmitted and partially reflected. The parts of the rarefactions which are reflected from the convergent nozzles move back into the drivers, lowering the driver pressure again and accelerating the flow into the expansion tube. This passing back and forth of rarefactions in the drivers continues until the pressure has returned to ambient.

The transmitted parts of the rarefactions(15) overtake the recompression shock, the contact surface and the primary shock. This overtaking causes a decrease in forward velocity of the contact surface and recompression shock and limits the distance they move down the expansion tube. The rarefactions overtaking the primary shock decrease its strength. For high shock overpressures and short driver tubes the rarefactions can overtake the primary shock well before it reaches the test section causing an unacceptable loss of shock strength.

Rarefactions are also generated at the open end of the expansion tube(16) and proceed upstream after the primary shock(2) has exited and the accelerated gas continues to flow out of the tube. After the contact surface has exited the tube, a back-facing compression (17) forms at the open end and moves back into the tube to bring the overexpanded driver gas back to ambient pressure. Behind it, ambient air(18) moves back into the expansion tube. Thus series of rarefaction and compression waves move up

and down the expansion tube, restoring the ambient conditions. Again, it is noted that our experimental evidence suggests that the quasi-one-dimensional representation of the flow in the LB/TS overstates these wave phenomena.

4. WAVE SHAPING TECHNIQUES

In the foregoing discussion of flow phenomena in an LB/TS it was assumed that the simulation is initiated by bursting diaphragms. The early-time pressure decay in the flow then is primarily caused by the driver end walls. The end walls of the drivers reflect the rarefaction waves arriving from the diaphragm location and send them downstream through the nozzles into the expansion tube where they overtake and weaken the primary shock. This gives the pressure wave a stepped shape. By giving the drivers various lengths, the rarefaction waves can be made to arrive at different times, giving the blast wave a refined, but still stepped shape. But the driver lengths have to be changed in order to vary the equivalent weapon yield at any given blast wave overpressure. Another method for breaking up the steps in the pressure curve would be to make use of baffles in the driver. However, the positions of these baffles would have to be adjusted for each overpressure and weapon yield simulated in the LB/TS. Both methods are rudimentary and appear therefore not feasible. The decay of the static and dynamic pressure necessary for the simulation of a blast wave may be controlled in other ways.

4.1 DRIVER GAS HEATING

In a parametric study of driver pressures and flow durations[2] it was quickly discovered that an LB/TS with cold driver gas would not be able to cover the required operating envelope. Heating of the driver gas is imperative and accomplishes three purposes. First, the proper amount of heating eliminates the density difference across the contact surface between the gas originally in the driver and the shocked gas originally in the expansion tube. By eliminating the density discontinuity across the contact surface, the dynamic pressure history of the simulated blast wave is properly presented.

Secondly, driver gas heating requires weaker recompression shocks because the gas contains more energy for expanding before a recompression shock becomes necessary. Thirdly, heating reduces the required driver pressure for any given, simulated shock overpressure and thereby reducing the cost. Driver heating does, however, accelerate the decay of the primary-shock strength at low-yield simulations because the rarefaction waves catch up faster with the primary shock. This problem can be overcome by the use of throat valves instead of diaphragms.

4.2 THROAT VALVES

The advanced design concept of the U.S. LB/TS employs computer controlled, fast-acting throat valves to meter the gas flow out of the drivers. Fast opening and closing throat valves offer the most elegant method of blast wave shaping because they make changes in the driver geometry to adjust the yield simulation unneccessary, are reusable and eliminate the need for changing diaphragms. If sufficiently fast opening valves can be constructed, they will replace the diaphragms of the present-day design. Currently, the

primary function of the valves is to shape the pressure decay of the blast wave in the expansion tube after the flow is initiated by bursting a diaphragm.

The shape of the blast wave will be controlled by closing the valve as a function of time in a predetermined manner after the shock has been initiated. In order for the primary shock to be considered properly formed, its rise time (i.e., the time it takes the blast wave pressure to go from 5% to 95% of the maximum pressure) must be under 0.5 ms. The closing functions for the two extreme blast wave conditions in the operating envelope, i.e., 241 kPa/1 kT and 14 kPa/600 kT, were determined computationally and the resulting blast waves are shown in Figure 6. At the upper limit of the operating envelope, where the shortest valve closing time is expected, the computation predicted a value of 82 milliseconds for the 241 kPa/1 kT blast wave. At the lower limit of the operating envelope, where the longest valve closing time is expected, the computation predicted a value of 4.32 seconds for a 13.8 kPa/600 kT blast wave. A delay of 44 ms between the shock initiation and the start of the valve closing prevents the rarefactions generated by the valve closing from overtaking the primary shock before it reaches the test section and from causing a premature decay of its strength.

While shaping the decay of the blast wave is the principal function of the throat valve, the ability to initiate the primary shock by rapidly opening the valve is highly desirable. Because throat valves are reusable, they eliminate the need for changing diaphragms which requires a great deal of labor. Also, diaphragms would be very expensive and difficult to handle because of the high driver pressures and temperatures encountered in the blast wave simulator. If the valves can be opened fast enough to initiate the shock at high overpressures, this method would produce a significant cost savings in the operation of the LB/TS. The time in which the valve has to open in order to produce a shock rise time of 0.5 ms at the test station was determined computationally[9] for the full range of shock overpressures. The opening time required to allow proper shock formation at the highest primary shock overpressure of 241 kPa was found to be 64 ms, and at the lowest shock overpressure of 14 kPa it was 26 ms. Such short opening times may be difficult to achieve and therefore, the throat valves will be combined with a diaphragm to initiate the shock at low shock overpressures. Most likely, the diaphragm would be located downstream of the valve and opened by cutting charges since heating is not required at the low-pressure end of the operating envelope.

4.3 RAREFACTION WAVE ELIMINATOR (RWE)

An RWE is a device at the open end of a shock tube that controls the flow into and out of the tube. An RWE is needed because flow disturbances from the open end will travel upstream and alter the flow pattern in the test section during the test period if they are not prevented from doing so. The simple method of extending the duct behind the test section far enough so that the disturbances arrive only after the test period is completed is too costly on the large scale of the proposed LB/TS. Another method for eliminating, or at least minimizing, the effects of rarefaction waves on the blast wave simulation uses an RWE.

An RWE is basically a converging nozzle used to adjust the blast wave pressure exiting the tube to match the ambient pressure. The nozzle effect is created by partially obstructing the flow exit area of the tube. The nozzle area is generally referred to as open area ratio, normalized by the total, unobstructed exit area of the expansion tube. A passive RWE has a fixed open area ratio preset before, and held fixed during the test period corresponding to the simulated blast wave conditions. An active RWE has a changing open area ratio which is continuously adjusted during the test period in accordance with the flow characteristics in the simulated blast wave. By changing the open (nozzle) area during the test in a predetermined manner, the rarefaction waves can be eliminated. Figure 7 gives an example of the influence of an RWE on the flow pattern in the test section.

Initially, three active RWE design concepts were considered,[10] (1) a rotating louver design (basically an enlarged version of the RWE at the Centre d'Etudes de Gramat, France), (2) a rotary fan blade design, and (3) a folding door concept. Among these the rotating louver design evolved as the prime candidate for further investigations because it has the smallest moments of inertia and therefore the least power requirements which are nevertheless considerable. The proposed RWE is composed of nearly 100 aerodynamically shaped louvers which must all rotate through 180° at a nonlinear angular acceleration rate during the positive and negative phases of the passing blast wave. The louvers and the supporting structures are relatively massive to maintain a necessary level of structural rigidity. Due to their large inertia, a complicated, high-powered, hydraulic system is required for actuating the louvers.

The position of the louvers will be controlled by a computer that evaluates an input data function and sends appropriate signals to the hydraulic actuators. This positioning function is a critical design element. Three methods are being considered for generating this function. The first one is a real time flow control cycle. Sensors would be placed in the flow field upstream of the RWE which measure the flow parameters, send signals to the control unit which analyses them in real time and instructs the hydraulic system how to move the louvers. With the second method, the control unit would be preprogramed with an empirical positioning function based on theoretical calculations and/or experimental data. It would drive the louvers without referring to real time flow measurements. The third method would be to combine the first two methods in a complimentary manner. The real time flow measurements would be used to correct a preprogramed positioning function.

5. CONCLUSION

Our computational studies with the BRL-Q1D code have shown that a full-scale blast/thermal simulation facility can be designed. The studies have also shown that driver heating, an active RWE, and computer-controlled throat valves should be used. Small-scale experiments with heated driver gas have been completed, to support the computational results. Small-scale RWE and throat-valve models are being developed for experiments. Also, a 1/6-scale test bed using a throat valve and an active RWE is in the design stage and, when built, will be used for proving the concepts used in the full-scale U.S. LB/TS design.

REFERENCES

[1] Opalka, K. O. and Mark, A., "The BRL-Q1D Code: A Tool for the Numerical Simulation of Flows in Shock Tubes with Variable Cross-Sectional Areas," BRL-TR-2763, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, October 1986.

[2] Opalka, K. O., "Large Blast-Wave Simulators (LBS) with Cold-Gas Drivers: Computational Design Studies," BRL-TR-2786, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, March 1987.

[3] Pearson, R. J., Opalka, K. O. and Hisley D. M., "Design Studies of Drivers for the U.S. Large Blast/Thermal Simulator," Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, U.S.A; in Proceedings of the 9th International MABS Symposium, Southend-on-Sea, Essex, England SS3 9XE, September 1985

[4] Mark, A., Opalka, K. O., et al., "Simulation of Nuclear Blasts with Large-Scale Shock Tubes," Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, U.S.A; in Proceedings of the 8th International MABS Symposium, Spiez, Switzerland, 20-24 June 1983.

[5] Hisley, D. M., Gion, E. J. and Bertrand, B. P., "Performance and Predictions for a Large Blast Simulator Model," BRL-TR-2647, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, April 1985

[6] Kingery, C. N. and Coulter, G. A., "Rarefaction Wave Eliminator Concepts for a Large Blast/Thermal Simulator," BRL-TR-2634, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland, February 1985

[7] Beam, R. M., and Warming, R. F., "An Implicit Factored Scheme for the Compressible Navier-Stokes Equations II: The Numerical ODE Connection," Paper No. 79-1446, AIAA 4th Computational Fluid Dynamics Conference, Williamsburgh, VA, 23-24 July 1979

[8] Baker, W. E., ed. "Explosions in Air: Engineering Design Handbook, Part One," AMC Pamphlet No. 706-181, Headquarters, U.S. Army Material Command, Alexandria, VA, July 1974

[9] Barry, R. A. and Lassahn G. D., "A Computational Study of the Effects of Throat Valve Opening Times on Shock Formation in the LB/TS", EG&G Idaho, Inc. Idaho National Engineering Laboratory, Idaho Falls, Idaho, January 1988

[10] Guice, R. L. and Gottlieb, J. J., "Passive and Active Reflection Eliminators," in Proceedings of the 10th International MABS Symposium, Oberjettenberg, Germany, September 1987

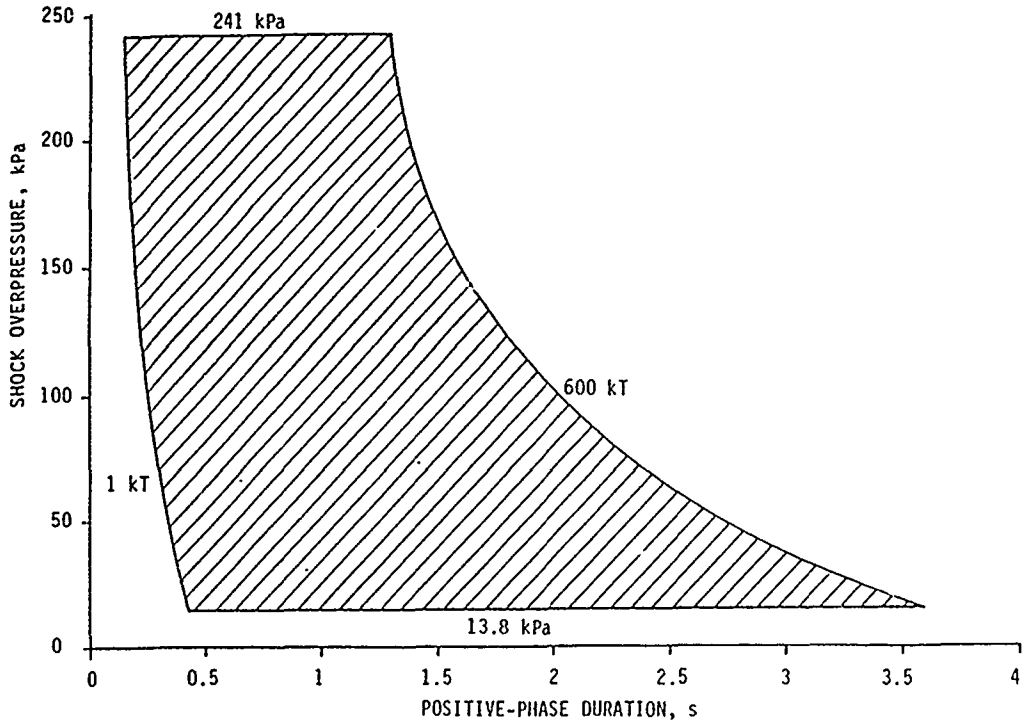


Figure 1: Operational Envelope of the U.S. LB/TS

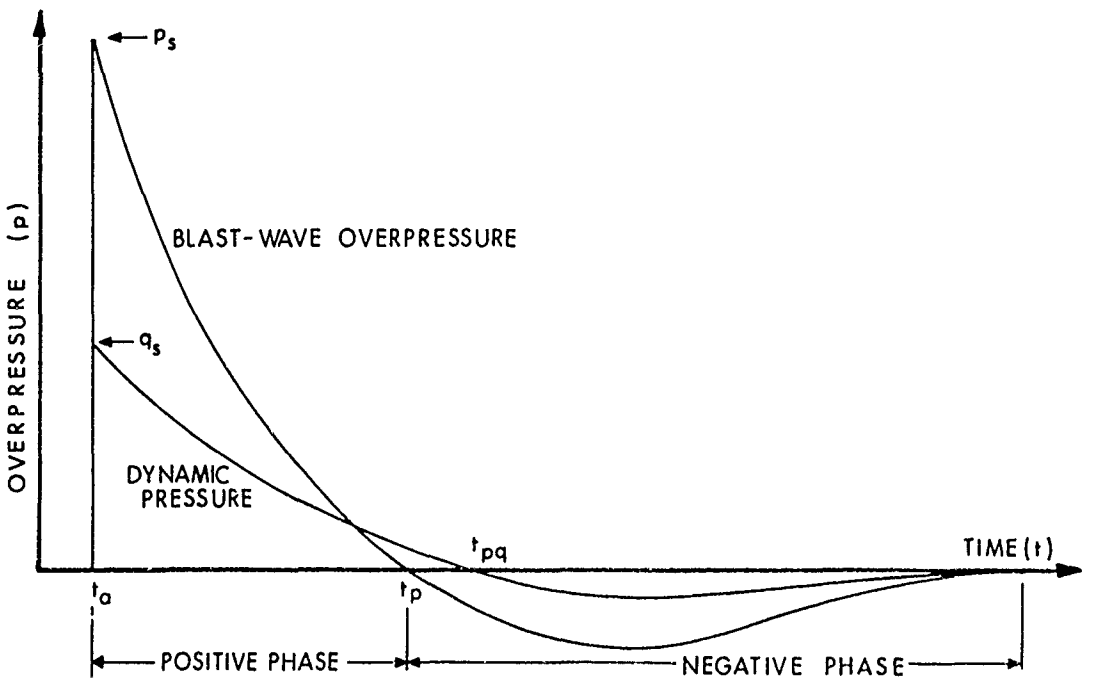


Figure 2: Pressure History of an Ideal Blast Wave

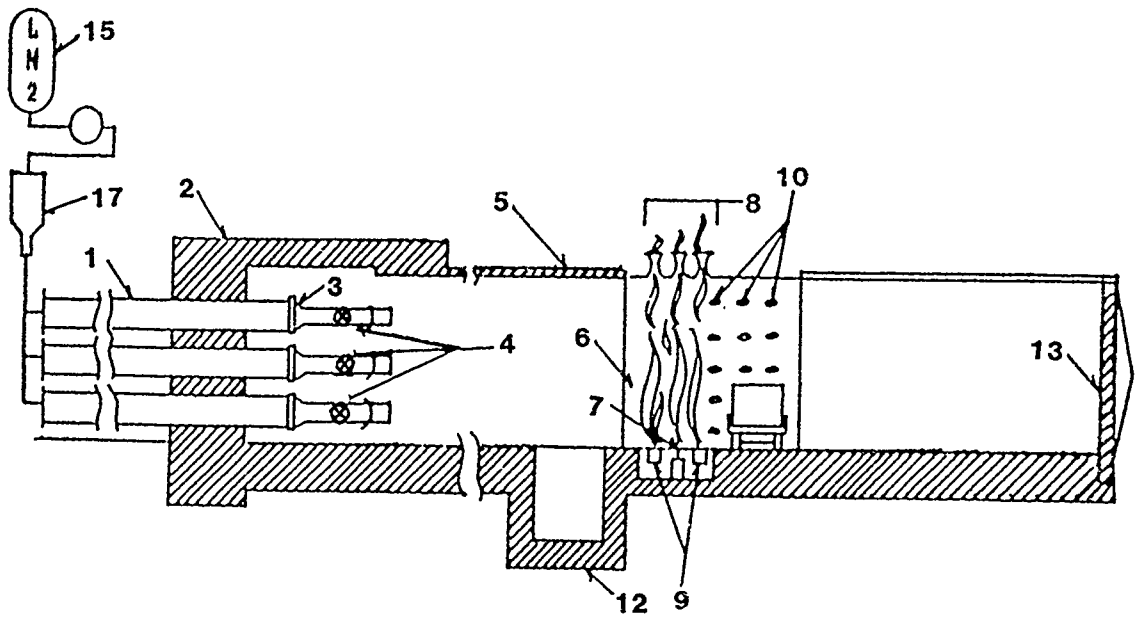


Figure 3: Schematic Layout of the US-LB/TS Design

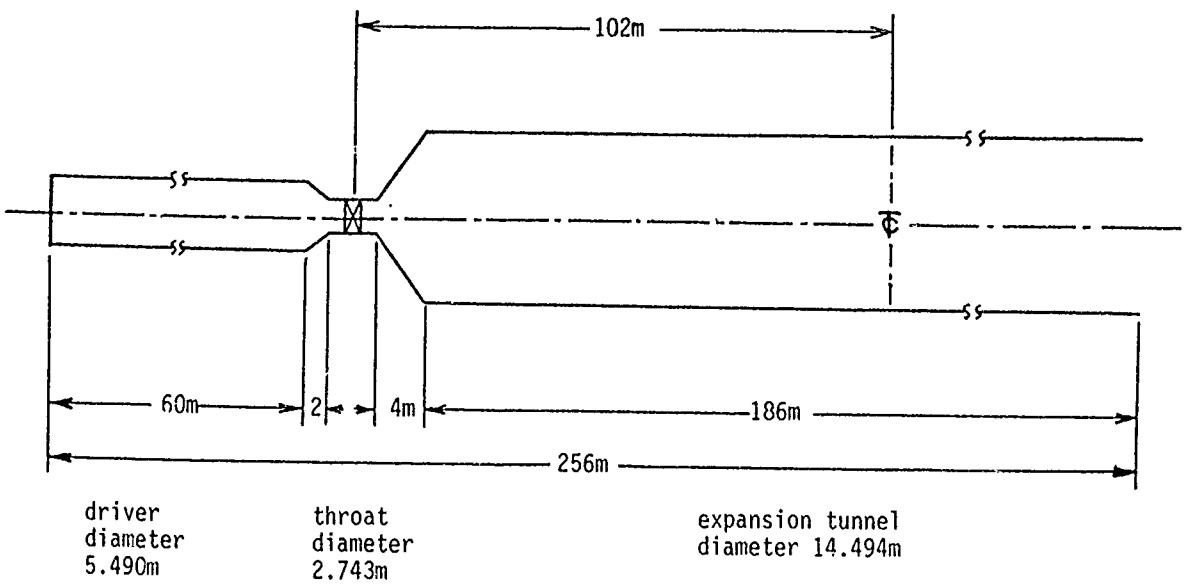
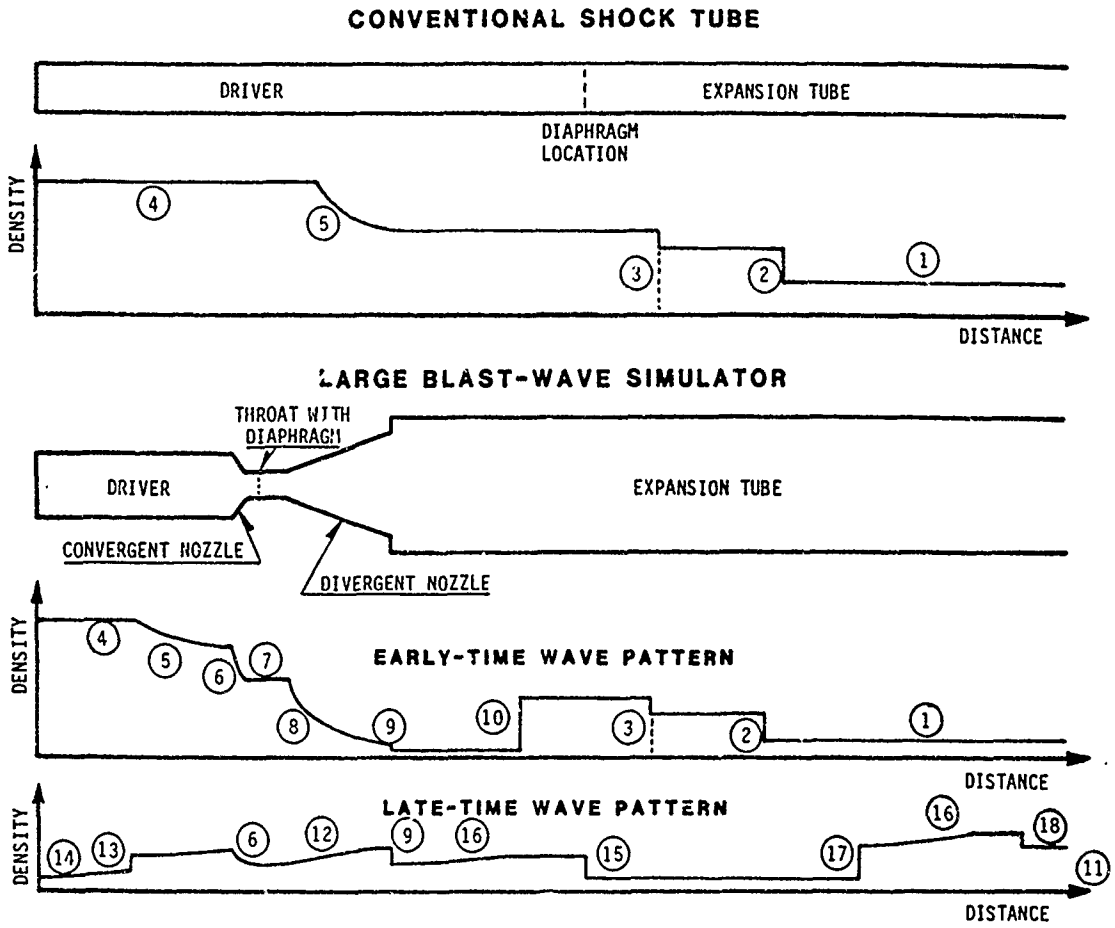


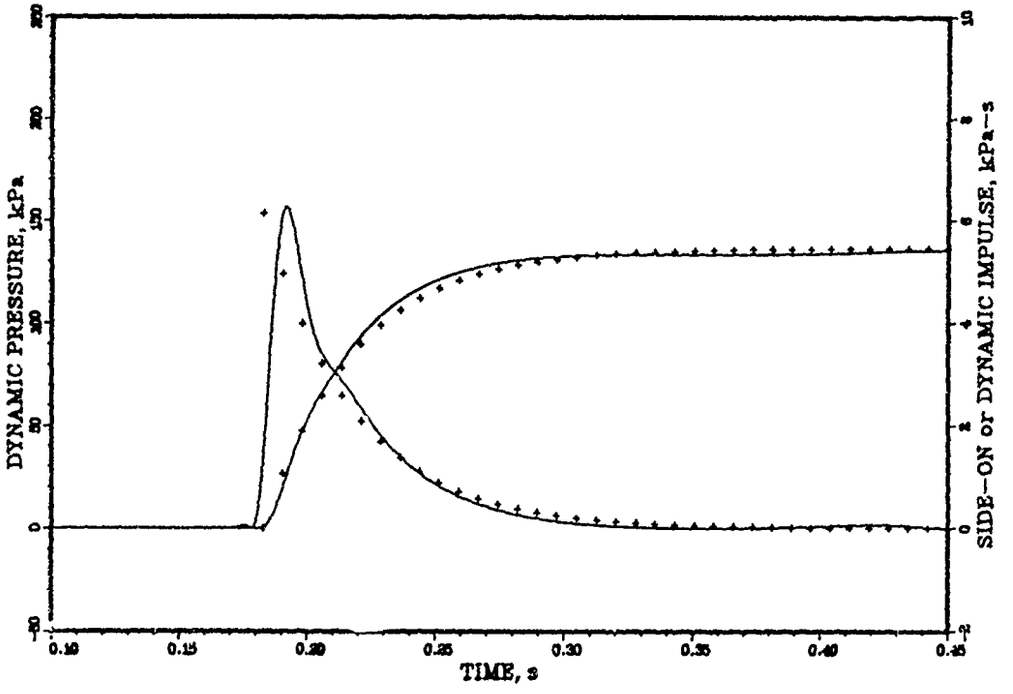
Figure 4: The Computational LB/TS Model



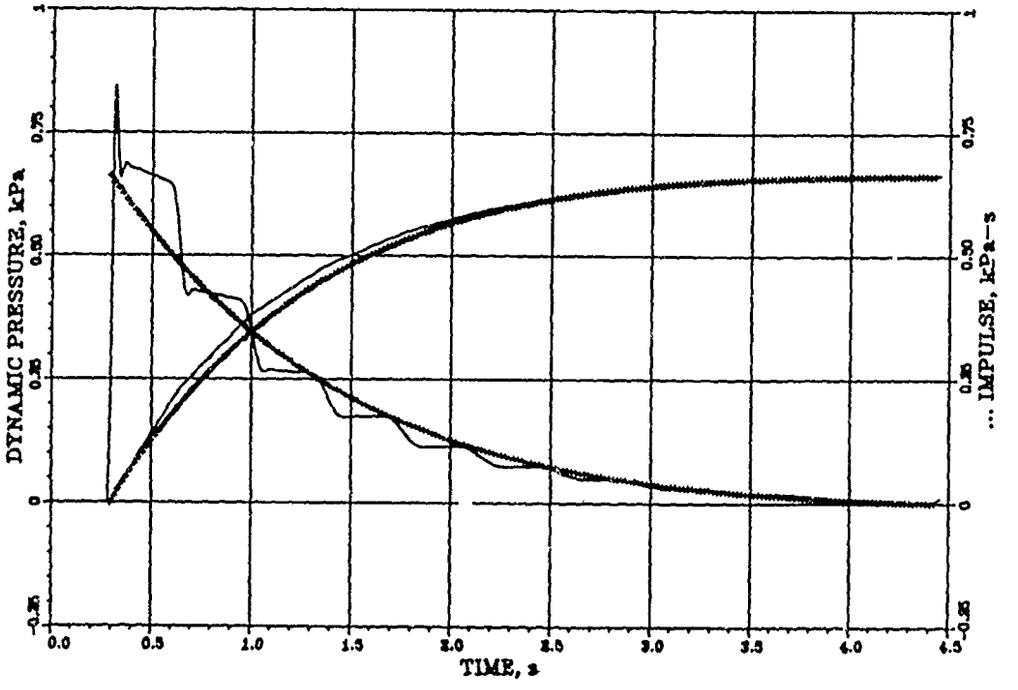
LEGEND

EARLY-TIME FLOW PATTERN		LATE-TIME FLOW PATTERN	
1	Ambient Conditions	11	Inflow at Ambient Pressure
2	Principal Shock	12	Isentropic Compression of Subsonic Flow
3	Contact Surface	13	Recompression Shock Partially Transmitted into Drivers
4	Initial Driver Conditions	14	Rarefaction Wave Reflected from Driver End
5	Rarefaction Wave	15	Recompression Shock Partially Reflected from Nozzle Exit
6	Isentropic Expansion of Subsonic Flow	16	Rarefaction Wave Transmitted through Nozzle
7	Choked Flow in Throat	17	Back-Facing Shock from Open End of Expansion Tube
8	Isentropic Expansion of Supersonic Flow	18	Contact Surface from Open End of Expansion Tube
9	Density Discontinuity at Nozzle Exit		
10	Recompression Shock		

Figure 5: Illustration of Physical Flow Phenomena.



(a) 241 kPa & 1 kT Blast Wave



(b) 13.8 kPa & 600 kT Blast Wave

Figure 6: Computationally Simulated, and Ideal, Blast Waves.

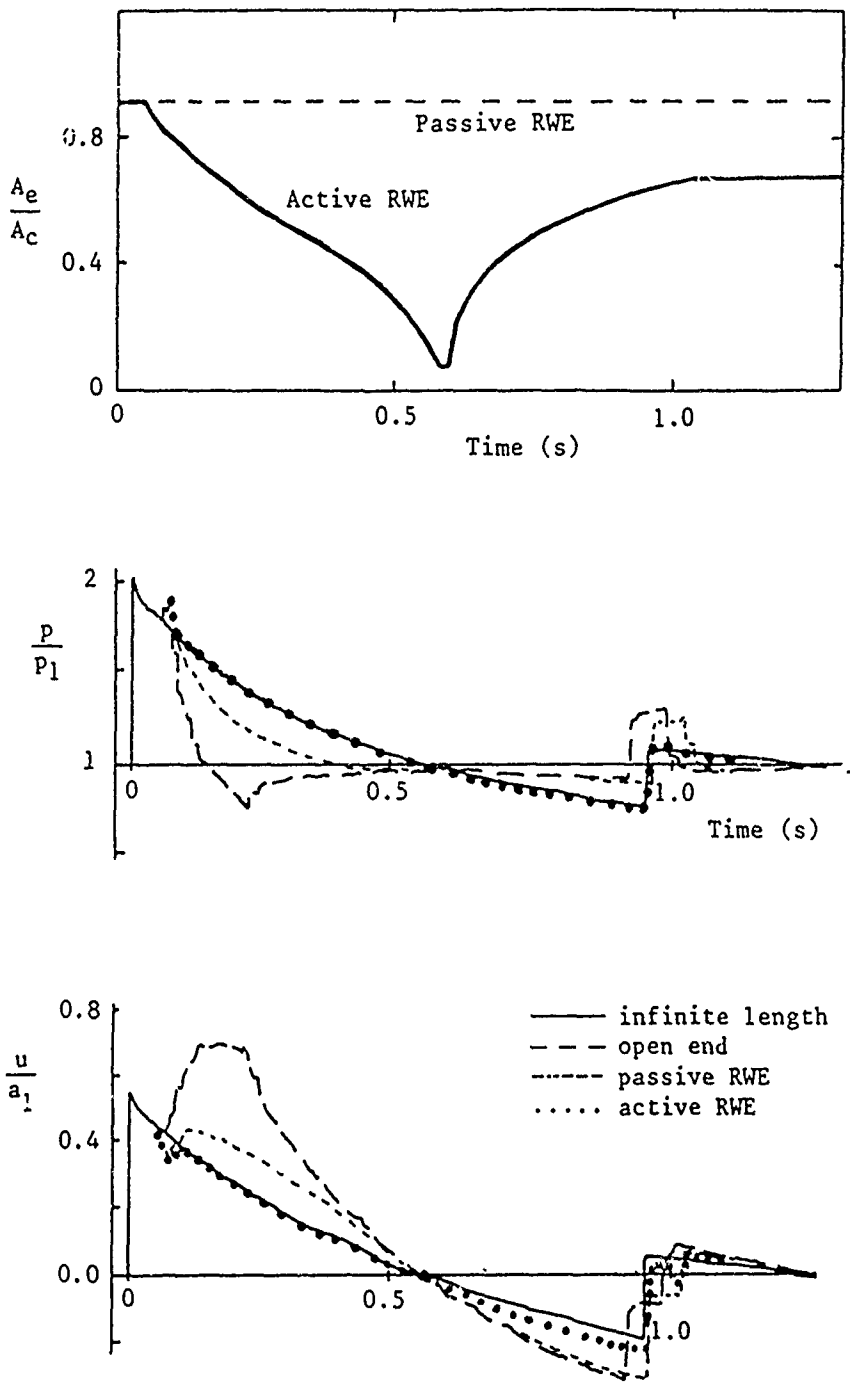


Figure 7: The Effect of an RWE on the Blast-Wave Simulation[10]

Stochastic High-Resolution Modeling Support for Live-Fire Test Programs (U)

Aivars Ozolins
Ballistic Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

I. BACKGROUND

SQUASH (Stochastic Quantitative Analysis of System Hierarchies) is a vulnerability simulation code that gives probabilistic predictions of physical damage and loss of combat effectiveness for impacts of large caliber munitions against mobile ground targets. It was developed in response to increased attention being focused on full-scale destructive tests and the related need to have a model whose damage predictions could be realistically compared with test results. Other models are not as suitable. They lack the fidelity and resolution needed to make detailed single-shot predictions. They are, instead, more useful for first moment predictions of function damage which are computed by averaging over many possible hit locations. So far, SQUASH has been used for preshot predictions of M1/M1A1 target vehicles (Abrams tank). Analysis and post-shot predictions will be made after completion of all test shots; predictions for other targets will be started in the near future.

The current emphasis on destructive test programs is due to Congressional legislation (National Defense Authorization Act for FY 87)[1] which mandates "live-fire" testing of all major weapon systems before entering full-scale production. The objectives are to estimate the vulnerability of "combat-configured" systems to all threat munitions likely to be encountered in combat, estimate the lethality of munitions against their principal foreign targets, and provide early insights into design modifications that may be needed to improve performance. The willingness to budget costly test programs acknowledges that there are great risks in depending solely on computer generated vulnerability estimates, and that some questions of vulnerability such as crew casualties or fire initiation are so vital or so complex that they are best answered by direct testing.

Officially, live-fire testing is not for model development. There is, however, close coordination between the testing community and the modeling community to insure that test results can be used to improve model capabilities. There are two reasons why this cooperation is vital. The first is that all vulnerability codes are empirical and, until now, the paucity of high quality vulnerability data has hindered their development and limited their application - especially with regard to many modern weapon systems. The other reason is that many activities depend on comprehensive vulnerability profiles that can only be generated using computer models that evaluate all likely attack directions and all possible hit locations. Examples of such activities are cost and operational effectiveness analyses (COEA's), spare parts replacement and repair analyses (SPARC), combat

simulations (i.e. war games), analyses of some foreign vehicles which may not be available for testing, and analyses of concept vehicles which may only exist on paper.

The underlying framework of the SQUASH code (developed by the author) represents a radical departure from other existing vulnerability codes. It is based on the conviction that the most reliable way to handle model-test comparisons is to use stochastic processing to model naturally unpredictable (i.e., random) outcomes and to structure vulnerability computations as individual simulations of real (observable or measurable) physical events. Thus, the initial munition impact and all subsequent target-threat interaction by-products and damage mechanisms would be simulated separately and with built-in uncertainties. The advantages are that statistical confidence intervals can be set up to determine if apparent discrepancies are serious, significant discrepancies can be more easily resolved, and calibration (or validation) of the code can be directly related to observable or measurable quantities. The obvious disadvantage of this approach is the additional computational burden.

II. SQUASH OVERVIEW

This section deals with computational procedures, capabilities, and basic elements of the SQUASH code. The sections that follow touch on the problems of validation, the role of SQUASH to live-fire tests, and the role of SQUASH to other models. The objective here is to provide a broad review. Space limitations do not permit discussing any single topic in great detail.

A. Random Variables

That stochastic variability is an important characteristic of combat damage, even for similar attack conditions, can be inferred from numerous studies involving terminal ballistics and vulnerability. Reports of armor penetration, projectile breakup, behind-armor debris characteristics, destructive testing of vehicles and vehicle components, etc., all support the idea that many observed effects are to some degree unpredictable and due to random or stochastic elements components.

Technically, it does not matter if the source of variability is natural or artificial; the consequences are the same. Trajectories of behind armor debris fragments, for example, are largely unpredictable and might be said to have a natural randomness. The variable penetration of shaped-charge jets, on the other hand, is partly caused by manufacturing tolerances and misalignment of parts.[2] In the first case, we cannot predict if a fragment will sever a small, but functionally critical, wire. In the second case we cannot predict if a jet will damage a component near its penetration limit.

Stochastic processing was implemented in SQUASH by designating appropriate model parameters as random variables. A summary of the more useful ones is given in Table 1. Each variable can be independently controlled (turned on or off), and new variables can be easily added or removed from the list. Such flexibility is useful not only for modeling stochastic behavior, but also for making sensitivity studies and for examining the effects of uncertainties in data inputs and numeric modeling parameters.

Table 1. Random Variables	
Item	Details
Impact Location	Uniform PDF (0,N) to select a shotline from N-1 shotlines clustered about designated impact point
CE Penetration	Gaussian PDF μ: from penetration submodel calibrated to test data σ: from variability of test data
Initial KE Penetration	Gaussian PDF μ: from penetration submodel calibrated to test data σ: from variability of test data
Residual KE Penetration	Residual(s) may be up to 200 broken pieces inside a Gaussian cone (i.e. spatial density declines with angular separation from cone axis) 1 σ cone widths from x-ray data Symmetry axis computed from obliquity and armor overmatch. No. of pieces from Poisson PDF (μ: from x-ray data & miscellaneous sources) Fragment properties (masses, speeds, & shapes) stochastically selected, but conserving total calculated residual mass and energy.
Lethal Spall Fragments	Fragment distributions are modeled using Gaussian shaped cones. 1 σ cone widths, and directions of symmetry axes from "witness plate" test data Total number of fragments - from Poisson PDF (μ: from calibrated submodels, or empirical tables)
Component Damage	No. of spall fragment impacts - from Poisson PDF (μ : from integrating spall densities over presented area) Kill assessments - from uniform PDF [0,1] (component is killed if sampled value < average PK)

Abbreviations: μ = distribution mean KE = kinetic energy round
σ = distribution 1 sigma CE = chemical energy round (shaped-charge)
PDF = probability density function

B. Other Methodology Enhancements

Besides stochastic processing, several other methodology improvements were incorporated into SQUASH. They are identified and described in Table 2 . Some were added as better solutions to old problems. Others were added to deal with important characteristics of new target vehicles, or because traditional approaches were incompatible with the fidelity and resolution requirements of making single-shot analyses or were incompatible with stochastic processing.

With respect to the last two reasons, consider the case of deflected shotlines. That KE rounds can deflect and sometimes shatter has been observed for years. Although exact trajectories are unpredictable, the general direction can be expressed as a function of armor obliquity and overmatch.[3] This makes residual penetrators ideal candidates for stochastic processing. It would be nonsense, however, to statistically process damage at the component level if deflection was not taken into account and predicted damage was to components along or near the original (undeflected) shotline while actual damage was to components far removed.

Currently, SQUASH handles KE penetrators, shaped-charge warheads, and self-forging fragments. Future work will focus on adding other munitions like artillery shells and mines, and other damage mechanisms like secondary spall, shock, blast, and ricochet.

Table 2. Methodology Enhancements

Item	Comments
KE Penetration	Grabarek's penetration equations[4] are commonly misused by directly applying them to several target elements in sequence. A new iteration procedure for multiple target elements prevents accumulation of extrapolation errors.
Direct Lethality Component PKs (ballistic impacts)	PKs are customarily functions of material, mass, speed, and shape of threat. A better method (now being implemented) relates PKs to target damage parameters of hole size and penetration depth. Perforation (mass removal) is both a frequent cause of component failure and a test observable - while masses and speeds often have to be inferred. When detailed engineering and test data is unavailable, the fallback position is to establish minimum PKs for small threats like spall and maximum PKs for primary penetrators. Intermediate PKs for medium sized threats, like broken penetrators, are obtained by interpolation using hole size and penetration depth.
Elliptical and Multiple Spall Distributions	The impact of some munitions against modern armors can produce irregular spall patterns. These are approximated by superposition of multiple distributions with elliptical density contours.[5] Customarily, spall is modeled using single cylindrically symmetric distributions.
Multiple Threats	SQUASH has bookkeeping to evaluate damage from multiple warheads, salvo fired weapons, and independent consecutive hits. Other codes cannot be readily converted to handle such cases.
KE deflection and KE breakup	SQUASH is currently the only operational vulnerability code that does not portray KE residuals as always being single threats with undeflected trajectories. See discussion above and table 1.
On-line Ray Tracing	Interrogation of target geometry (ray tracing) is normally done off-line. In SQUASH, ray tracing can, also, be done on-line (during runtime). This is for future methodology improvements involving new damage mechanisms like secondary spall or ricochet, and for generating inputs to other models.

C. High Resolution Geometry

To simulate vulnerability, one must realistically portray, not only the physics but, also, the geometry to which the physics applies. The ability to make 3-dimensional computer models of entire vehicles has existed for several years. Such models represent objects as true solids that can be constructed to whatever level of detail is appropriate for a specific application - provided that suitable design information is available.

Figures 1 and 2 are typical illustrations of target geometry that exhibit levels of detail suitable for SQUASH. The first is a generic view of a target's exterior. The "real image" quality is the product of a lighting model. The other is a 'see-through' line drawing of a target's interior which reveals traces of 9 closely spaced main penetrator shotlines and those critical components that were intersected by the shotlines and all potentially lethal behind-armor debris fragments. By definition, a component is critical if loss of that component diminishes any of the target's assigned combat capabilities such as mobility or firepower.

The software system that was used for Figures 1 and 2 was developed at the BRL. It is one of the more elegant systems currently available. A complete CAD (Computer Aided Design) package which includes modeling, rendering, and ray-tracing is publicly available.[6] At the BRL, it is being used to build computer targets for such applications as vulnerability and lethality analyses, signature analyses (radar, infrared, optical), and structural analyses (heat flow and shock).

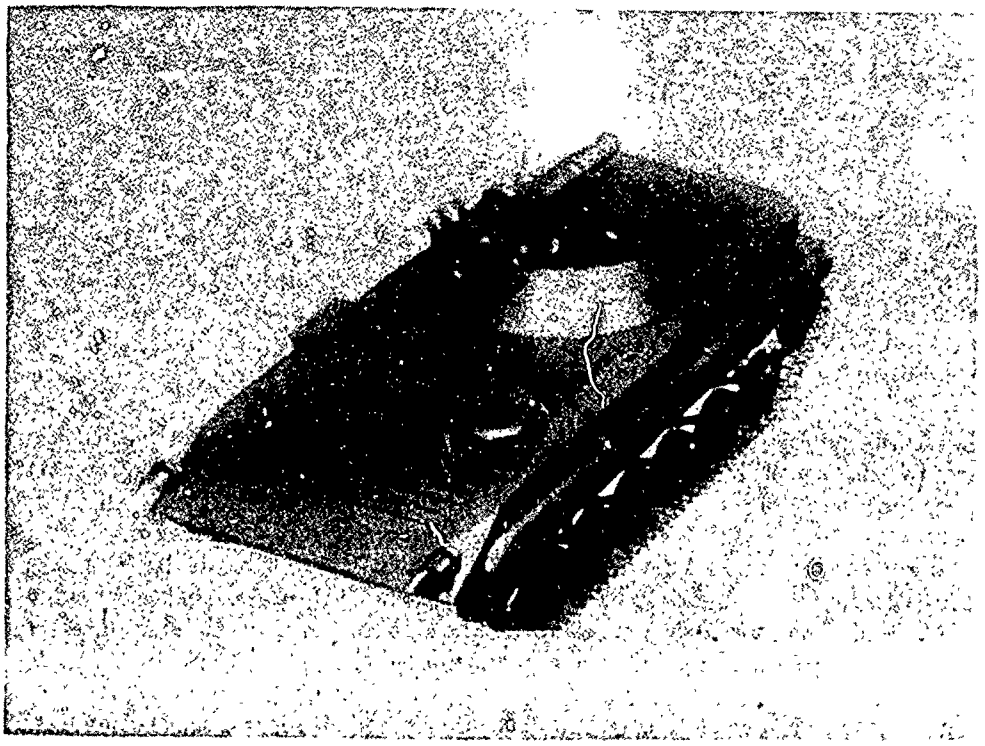


Figure 1. Target Description - Typical View of Exterior

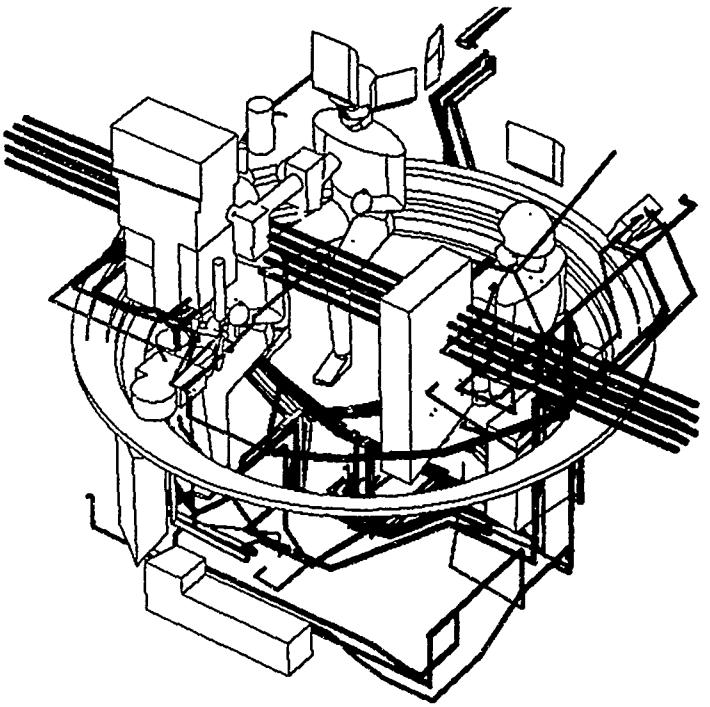


Figure 2. Target Description - Partial View of Interior

D. Computational Procedure

The most distinguishing feature of SQUASH is the use of repeated or Monte Carlo simulations of individual shots to produce predictions of possible damage states. Associated with each state is a computed probability which represents the likelihood of observing that state as an outcome of a corresponding live-fire test. Figure 3 is a schematic diagram of how damage states are generated.

Step (A) in Figure 3 shows one of many possible critical components in a vehicle which can be damaged by an impacting munition. Possible damage mechanisms are illustrated using three arrows entering step (A). These may be such entities as residual penetrators, behind armor spall, and shock. The effect of each damage mechanism is separately evaluated against each critical and vulnerable component The results are separate component kill probabilities (PKs) which are drawn as 3 arrows leaving (A). It should be recalled that each Monte Carlo trial will produce different PKs. This is due to stochastic processing that is applied to physical target-threat interaction variables such as penetration capability and number of lethal spall fragments.

In step (B), the PKs for all damage mechanisms are combined into a single composite PK which is the overall probability of functionally destroying a component. In principle, one has to acknowledge that, at least for some components, there may be important synergistic effects. At

present, no empirical data is available that deals with this problem, so independence is assumed. However, the SQUASH code does recognize individual components, and individual damage mechanisms, so there would be little problem in applying synergism if or when appropriate data becomes available.

In the next step, composite PKs are compared against random numbers to determine which components are killed (K) and which survive (S) a given Monte Carlo trial. The operational states of all critical components that are both vulnerable and "visible" to any damage mechanism is represented by step (D). Shown is a specific combination of K and S states for only 7 components. In practice 30 to 50 would be more typical. The illustrated mixture represents a single distinguishable target damage state. That this is one of many possible states that can be sampled by repeated shot simulations is illustrated in step (E). Two target damage states can be distinguished from each other when the operational state of at least one component is different.

As indicated in (E), the maximum number of damage states for N components is 2 raised to the Nth power. If N was 20 the number would be 1,048,576. To sample that many states would require a minimum of 1,048,576 Monte Carlo trials. To do so is neither practical nor necessary. Depending on the variability and magnitude of individual component PKs, some damage states are more probable and may be sampled many times. Consider a component that can be directly hit by the main penetrator on each trial and is, at the same time, very vulnerable to that penetrator. Excluding consideration of other components, those states that list it as being killed are more likely then those that list it as having survived. Simple comparison tests were made using different Monte Carlo sequences and different shot conditions to establish 1000 trials as a useful limit beyond which only small differences in outcome probabilities can be expected.

Step (F) maps damage states into loss of combat functions. The mapping operation uses a previously agreed upon "Damage Assessment List" to spell out the consequences of losing specific vehicle operating capabilities. For example, loss of manual traverse (a backup system) may mean no loss of function, loss of power traverse may mean 10% loss of firepower, but loss of both may mean 95% loss of firepower. Obviously, not all combinations involving all operating capabilities can be specified; those that are not, appear as separate table entries. Thus, loss of bore evacuator might be separately specified as 5% loss of firepower.

The unanswered question at this point is how to deal with the many undefined combinations. The universal practice is to make the assumption that remaining capabilities are multiplicative. Therefore, for loss of both bore evacuator and power traverse, the remaining firepower capability would be (1 - .05) * (1 -.1) or .855. The implication here is that damage does not affect the relative value of remaining capability. For example, if loss of commander produces 50% loss of firepower to an intact vehicle, then loss of commander would degrade by 50% whatever firepower was available to an already damaged vehicle.

The last step (G) gives a typical histogram of mapped damage states for a full set of Monte Carlo trials. Such histograms can be regarded as high resolution "fingerprints" that characterize a target's vulnerability to a specific munition striking a specific area from a specific direction. The ones shown are for relative loss of mobility (M-kill) and firepower (F-Kill). Other common "kill" types are catastrophic (K-kill), and a combined M-or-F kill. Sometimes, surveillance and communication kills may be of interest as well.

The discrete nature and sometimes large variance of function loss has obvious and important implications for live-fire testing. First moment estimates of vulnerability from non-stochastic models or the distribution mean from SQUASH cannot be sensibly compared to single live-fire test results. First moment values may, simply, not coincide with any possible loss of function state, and discrepancies between single shots and predictions may be large - even for an, otherwise, perfect model. Take, for example, a shot that has 10% probability of detonating live ammunition. What can one conclude about how good or bad the model is if the model accurately predicts a K-kill of .1, and a test shot produces spectacular and total destruction of the vehicle (assessed as K-kill = 1)?

This problem would not surface if value judgements were based on a combined analysis of all test data. However, until the SQUASH program was developed to reveal the variety of possible outcomes, the tendency was to place undue emphasis on individual predictions. After all, if each test shot cost hundreds of thousands of dollars, and predictions are single numbers that can be paired with corresponding shots, it would be difficult to show that a model might, in fact, be good when individual comparisons showed large apparent differences.

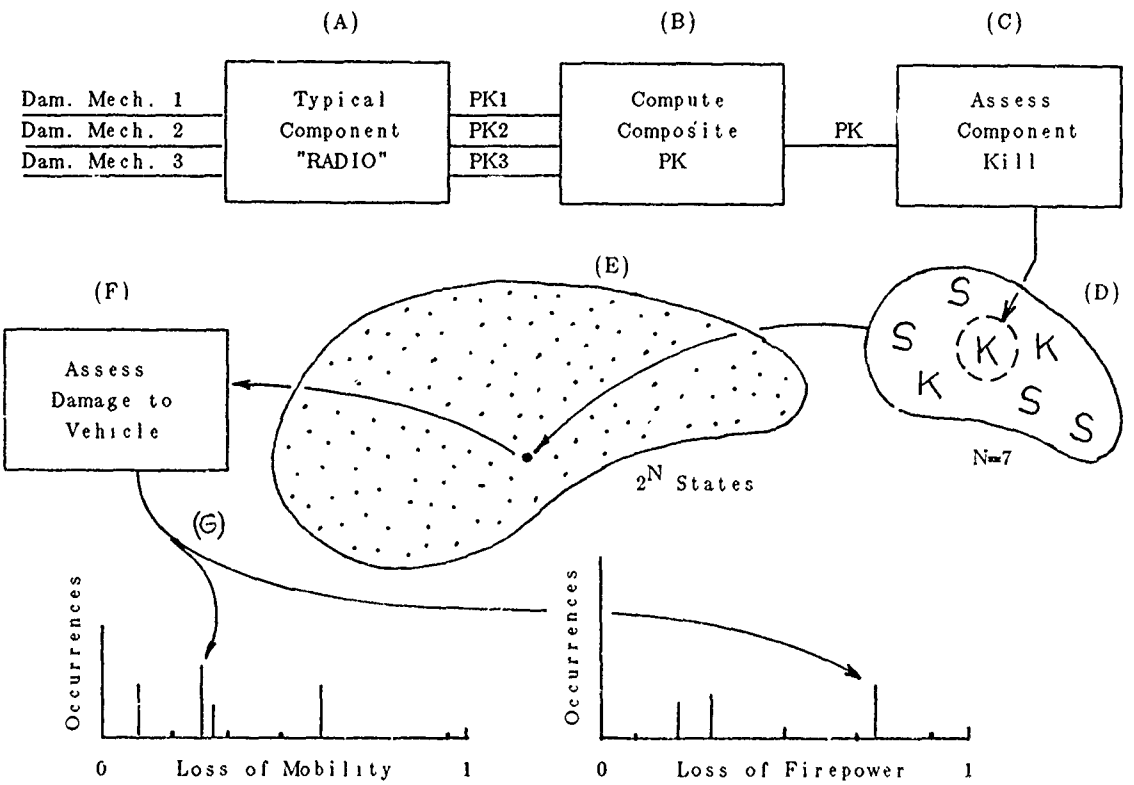


Figure 3. Procedural Schematic

E. Program Inputs

To get realistic outcomes, SQUASH tries to physically simulate all defined target-threat interactions. This does not imply, however, that SQUASH is not an empirical model. It is, and like all other vulnerability codes, the accuracy of its outputs is affected by the quality of its data inputs - the only difference then being the specific character of the inputs. An input summary table is presented in Table 3.

Given that a target description is already available, the most difficult inputs to prepare are component vulnerabilities. To get these, one must first perform a "criticality analysis" of all functionally important target systems. This is to identify those components that are vital to a vehicle's effectiveness in combat. Once identified, an engineering analysis must then be made of each component to quantitatively determine how readily it can be disabled by different damage mechanisms - an arduous task that involves finding out how it is constructed, how it functions, and what it is made of.

A critically analysis also provides vital information about the interdependence of target subsystems - i.e., do they function independently, or are they functionally interconnected. For series arrangements, killing any member will kill system function; for parallel arrangements, all members must be killed to kill system function.

Table 3. Principal SQUASH Inputs

Category	Subcategory
Armor Properties	Ballistic Resistance (Stopping Power) Spall Characteristics
Threat Properties (ballistic)	Penetration Capability Degradation Characteristics
Component Attributes	Material Critical Component - Critical System Associations Ballistic Resistance Vulnerability Characteristics (for each damage mechanism)
Target Geometry (ray-trace data)	Component Identifiers Line-of-Sight Thicknesses Surface Obliquities Component Separation Distances Air Space Codes (external, crew compartment, ammo box, etc.) Burst Point Coordinates (for fragment cones)
Damage Assessment	Criticality Data (system interdependence) Loss of Function Mapping Rules
Stochastic Control	Variability Data Selection Flags

F. Nature of Outputs

The outputs of SQUASH deal with predicting the influence of different damage mechanisms, and predicting the probabilities and patterns of damage that may occur during live-fire tests. For illustration purposes, a partial set of outputs is presented in Figures 4 to 7. These are typical of the outputs needed to make single shot predictions.

The first of the figures highlights items of interest for which variability information may be desired. The first numeric column gives the number of Monte Carlo trials for which that line item was relevant. For example, 81% for the number of spall fragments implies that the main penetrator defeated armor and produced spall 81% of the time. The remaining columns give computed means and standard deviations.

If closer examination of any item is needed, any of the underlying probability distributions can be output. Figure 4 also shows a typical distribution of residual penetrations. Of interest here is the presence of several peaks. These occur when main penetrator trajectories encounter different armors - a distinct possibility, even when trajectories are tightly clustered about an impact point. For that reason and because precise impacts cannot be guaranteed for all test shots, hit location can be designated a random variable.

Figure 5 is a composite of two outputs which summarize predicted damage to individual components. The components listed are those whose probability of kill was greater than 0 for at least one Monte Carlo simulation. The first 3 numeric columns give the minimum, maximum, and mean kill probabilities that were recorded. The remaining columns breakdown the mean or "composite" PKs by damage mechanism.

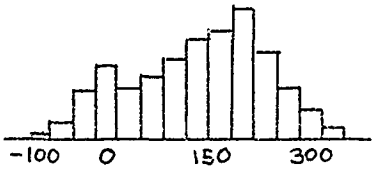
Killing some components such as fuel tanks and ammunition can produce a K-kill or catastrophic loss of vehicle. A summary table of K-kill information is shown in Figure 6. Here, there was only a small K-kill probability from fuel fires.

Probabilities for killing a group of components in combination is illustrated in Figure 7. Any number of groups can be defined to examine specific problems which are important to a limited number or type of component. To provide flexibility, groups do not have to be mutually exclusive. Thus, an engine oil cooler could be grouped with engine components, with components that affect only mobility, with components that contain fluids (leakers), or with spall sensitive components.

Two types of information are provided. The first gives ranked probabilities for each predicted combination (next to last column) along with a cumulative probability (last column). The other (along the bottom) gives probabilities of killing a specific number of group members. For some component groups, knowing how many can be killed is useful for some types of analyses. Losie[7], for example, found it more difficult to predict which rounds of stowed ammunition would detonate than how many.

Not shown here as part of SQUASH outputs are loss of function histograms, two samples were included earlier in Figure 3. Such histograms are important because they present the final payoff numbers that deal with the question - given a hit, how much combat capability can be lost?

SUMMARY TABLE (1000 TRIALS)				SELECTED HISTOGRAM
ITEM	% OF TRIALS	SAMPLE MEAN	SAMPLE (STD. DEV.)	HOR. = RESIDUAL PENET'N VERT = RELATIVE OCCURRANCES
^ M-Kill	100	0.225	0.249	
^ F-Kill	100	0.221	0.308	
^ MF-Kill	100	0.287	0.287	
$ K-Kill	100	0.028	0.170	
Residual Penet'n (mm)	100	115	113.9	
Available Penet'n (mm)	100	789	21.0	
Exit Hole Diam. (mm)	81	9	5.0	
No. of Spall Fragment	81	187	73.7	



Degradation (0 to 1) $ Binomial (0:failure 1:success)

Figure 4. Typical Case: Variability of Outputs

COMP. ID	NAME	MIN.	MAX.	COMP-OSITE	JET + BLAST	RESID'L JET	FRAG-MENTS
1100	Commander	0.000	0.938	0.013	0 000	0.013	0.000
1200	Gunner	0.000	0.984	0.032	0.000	0.000	0.032
1300	Loader	0.000	1.000	0.301	0.000	0.300	0 005
1400	Driver	0.000	1.000	0.257	0.000	0.000	0.257
2030	Cable / Elec #4 / Turret	0.000	1.000	0.063	0.000	0.000	0.063
2170	Cable / Fire Cntrl#6 / Hull	0.000	1.000	0.021	0.000	0.000	0.021
2370	Cable / Elec #1 / Hull	0.000	1.000	0.200	0.000	0.000	0.200
2385	Cable / Elec #3 / Hull	0.000	1.000	0.143	0.000	0.000	0.143
2402	Cable / Elec #5 / Hull	0.000	1.000	0.114	0.000	0.000	0.114
2480	Cable / Elec#15 / Hull	0.000	1.000	0.063	0.000	0.000	0.063
2630	Hull Distrib. Box	0.000	0.945	0.215	0.000	0.000	0.215
2640	Hull Networks Box	0.000	0.892	0.135	0.000	0.000	0 135
2650	Turret Networks Box	0.000	1.000	0.252	0.000	0.251	0.002
3030	Hyd Line: to Aux Pump	0.000	0.840	0 018	0.000	0.000	0.018
3040	Hyd Line: Aux Pump > Reserv	0.000	0.840	0.012	0.000	0.000	0.012
3095	Hyd Line: Pump Drain	0.000	1.000	0.015	0.000	0.015	0.000
3110	Hyd Line: TDM to Accum.	0.000	0.936	0.041	0.000	0.000	0 041
3120	Accumulator	0.000	1.000	0.573	0.000	0 356	0.371
3150	Turr Distrib. Manifold (TDM)	0.000	0.672	0.061	0.000	0.000	0 061
3225	Hyd Line: TDM > Az. Servo	0.000	0.875	0 043	0.000	0 000	0.043
3250	Azimuth (Az.) Gearbox	0 000	0 750	0 024	0.000	0.000	0.024
3285	Hyd Line: Man El Pump>Servo	0.000	0.875	0 045	0.000	0 000	0.045
3290	Elevation Servomechanism	0.000	0.622	0.110	0.000	0.000	0.110
3365	Race Ring	0 000	0.948	0.352	0.000	0.000	0 352
5015	Track: Left Side	0 072	0.469	0.334	0.309	0.025	0.000
5981	Support Roller: Left Side	0.000	1.000	0.220	0 000	0 220	0.000
6010	Fuel Tank: Left Bow	0 000	1.000	0.317	0 000	0.317	0 000
6300	Fuel Line. Left Bow > Manif	0.000	0.840	0.007	0.000	0 000	0 007
7055	Hyd Line: Park Brake > Manif	0.000	0.936	0.036	0.000	0.000	0 036
7320	Engine: Electronic Cntrl Unit	0.000	0.972	0.234	0.000	0.234	0.000

Figure 5. Typical Case: Individual Component Kill Probabilities (PKs)

Fuel Fires		0.028
All Ammo Combined		0.000
Ready Rounds	0.000	
Ammo (Hull Compartment)	0.000	
Ammo (Bustle)	0.000	
All Ammo and Fuel Combined		0.028

Figure 6. Typical Case: Contributions to Catastrophic Kill

COLUMN 1	2480	Cable / Elec#15 / Hull	AVG PK = 0.063
COLUMN 2	5015	Track: Left Side	AVG PK = 0.334
COLUMN 3	5981	Support Roller: Left Side	AVG PK = 0.220
COLUMN 4	6010	Fuel Tank: Left Bow	AVG PK = 0.317
COLUMN 5	6300	Fuel Line: Left Bow > Manif	AVG PK = 0.007
COLUMN 6	7055	Hyd Line: Park Brake > Manif	AVG PK = 0.036
COLUMN 7	7320	Engine: Electronic Cntrl Unit	AVG PK = 0.234

	COLUMNS 123456789 123456789 123456789 123456789	EVENT PK	CUM. PK	
1	sssssss	0.309	0.309	1
2	ssKssss	0.119	0.428	2
3	sssKssK	0.118	0.546	3
4	sKsssss	0.098	0.644	4
5	sKKssss	0.097	0.741	5
6	sssKsss	0.063	0.804	6
7	sKsKssK	0.050	0.854	7
8	sKsKsss	0.028	0.882	8
9	ssssssK	0.012	0.894	9
10	KKsKssK	0.012	0.906	10
11	KssKssK	0.011	0.917	11
12	Kssssss	0.010	0.927	12
13	sssssKs	0.010	0.937	13
14	sssKsKK	0.009	0.946	14
15	sKsKsKK	0.008	0.954	15
16	KKsssss	0.008	0.962	16
17	KssKsss	0.006	0.968	17
18	KKsKsss	0.004	0.972	18
19	ssssKss	0.003	0.975	19
20	sKKsssK	0.003	0.978	20
21	sssKsKs	0.003	0.981	21
22	sKssssK	0.002	0.983	22
23	KssssKs	0.002	0.985	23
24	KssssKK	0.002	0.987	24
25	sssssKK	0.001	0.988	25
	(12 Lines Intentionally Deleted)			
37	ssKsssK	0.001	1.000	37

K = kill
s = survive

Number Killed	0	1	2	3	4	5	6	7
Probability	.309	.315	.270	.081	.023	.002	0	0

Figure 7. Typical Case: Combination Kill Probabilities for Propulsion Group Components

III. MODEL VALIDATION

Stochastic selection of model parameter values produces variation of predicted outcomes. For many Monte Carlo trials, the outcomes form distributions which reveal the possible variability of test results. To determine the credibility of model predictions, one would, ideally, like to have test shots replicated many times. Standard statistical tests could then be applied which, at some confidence level, would indicate whether predicted distributions can be distinguished from the true, but unknown, distributions from which real test samples are drawn.

Because of cost and the need to evaluate many different shot conditions, live-fire shots are not usually replicated. This complicates validation. The problems of stochastic model validation have been examined by Baker and Taylor.[8] They have developed and recommend non-parametric ranking techniques in conjunction with Mann-Whitney hypothesis tests for equality of distributions. Such tests may be applicable to SQUASH. In addition, several specific tests have been suggested by the Probability and Statistics Branch of SECAD at the BRL.[9] These would make assessments for component damage, for armor penetration, and for loss of combat functions. Shnidman[10] has also proposed a test for component damage that may be useful.

Due to lack of shot data, no statistical comparisons have yet been made; M1/M1A1 live-fire tests are still in progress. When shot results have been released for analysis, and the data has been given a preliminary review, other and more discriminating tests may suggest themselves.

As already mentioned, SQUASH was structured as a true simulation model. The author believes that the most important long term benefit of this is that validation can and should involve directly observable or measurable physical quantities - as opposed to focusing attention primarily on final loss of function values. If validation is limited to loss of function, one cannot safely extrapolate the use of the code to target-threat combinations which were not part of the test program. If, on the other hand, validation is applied directly to each important physical submodel, then the code can be used with greater credibility as a shot simulator for untested conditions, provided that the validated submodels are still applicable.

IV. APPLICATIONS AND FUTURE ROLE

The primary role of SQUASH is to provide computational support for live-fire test programs, a major objective being the vulnerability evaluation of test vehicles to attack by designated munitions. As indicated earlier, testing is the best way to obtain direct evidence about a target's vulnerability and the most desirable way to answer specific questions of vital importance about its combat effectiveness. However, due to cost, such testing is, necessarily, of limited scope and cannot be used alone to generate comprehensive vulnerability profiles. Test data must be supplemented with synthetic computations of vulnerability which examine thousands of possible hit locations against all threats which are likely to be encountered in combat.

To use synthetic calculations one should adopt procedures that will make effective use of costly test data. It seems reasonable that, as a minimum, these procedures should establish initial model credibility, provide for effective calibration to test data, and verify the effectiveness of subsequent modifications. In line with such objectives, the current plan for using SQUASH is to

(1) generate preshot vulnerability predictions,
(2) make detailed model-test comparisons,
(3) update empirical inputs and make necessary code adjustments,
(4) generate post-shot predictions to verify the effectiveness of any changes,
(5) statistically test model outputs as part of an overall methodology assessment, and
(6) generate comprehensive vulnerability estimates

Once calibrated, SQUASH will be used as a shot simulator to build data inputs for less detailed vulnerability codes. At present, the most attractive and feasible option is to generate new damage correlation tables for VAMP which is BRL's compartment level production code. There are several reasons why this approach is being pursued. The primary reason is that SQUASH is not computationally fast enough to do routine production work. Also, only compartment level models can process target descriptions with incomplete interiors - i.e., concept vehicles and some foreign vehicles. Finally, it promises to be a good way to introduce test results into other codes and, at the same time, enforce consistency. A long standing problem has been the proliferation and independent evolution of vulnerability codes. Tests of consistency have, occasionally, been attempted, but general consistency has never been satisfactorily demonstrated.

The use of SQUASH as a shot simulator has, also, other applications not directly related to live-fire testing. The author believes that it will prove to be a useful diagnostic and research tool for general methodology development and for investigating special problems that require good discrimination levels, but do not need full production capabilities. Typical problems that fit these conditions are determining the effectiveness of specific target or munition design changes, testing the usefulness of new compartment model correlation parameters, and evaluating the need for or the best way to implement new damage mechanisms.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the efforts of co-workers Cynthia Dively and Scott Henry who used SQUASH to make the first set of live-fire predictions and who made many valuable recommendations for improving program operation and program outputs. He, also, wishes to thank Helen Polak who helped write several output routines, and Lawrence D. Losie whose expertise in probability theory and statistical procedures helped solve many implementation problems. Mr. Losie also provided helpful recommendations for handling self-forging fragments and built the new "direct lethality" routines for evaluating component vulnerabilities.

REFERENCES

1. O'Bryon, J.F., "Live Fire Testing: The Legislation and its Impact", Army Reasearch, Development & Acquisition Bulletin, p (1 - 3), May - June 1987.

2. Segletes, S.B., "Improved Drift Velocity Computations for Shaped-Charge Jets", Ballistic Research Laboratory, Aberdeen Proving Ground, MD, Report No. BRL-TR-2823, June 1987.

3. Backman, M.E., 'Terminal Ballistics", Naval Weapons Center, China Lake, CA, Report No. NWC TP 5780, Feb. 1976 (ADA 02182

4. Grabarek, C., "Penetration of Armor by Steel and High Density Penetrators", Ballistic Research Laboratory, Aberdeen Proving Ground, MD, Memorandum Report No. 2134, Oct. 1971.

5. Ploskonka, J.J., "Required Modifications to the SLAVE Vulnerability Model for Evaluating Modern Armored Tanks", Ballistic Research Laboratory, Aberdeen Proving Ground, MD, (Unpublished Report).

6. SECAD/VLD Computing Consortium, "Ballistic Research Laboratory CAD Package: A Solid Modeling System and Ray-Tracing Benchmark Distribution Package - Release 1.21 (2 June 1987)", Ballistic Research Laboratory, Aberdeen Proving Ground, MD, (POC: Dr. Paul H. Deitz, Michael J. Muuss, or Douglas Gwyn).

7. Losie, L.D., "Comparison of BRL's EFP Vulnerability Model Calculations With Test Results from Joint Chicken Little Controlled Vulnerability Tests", Ballistic Research Laboratory, Aberdeen Proving Ground, MD., (In Preparation).

8. Baker, W.E., Taylor, M.S., "A Nonparametric Statistical Approach to the Validation of Computer Simulation Models", Ballistic Research Laboratory, Aberdeen Proving Ground, MD, Report No. BRL-TR-2696, Nov. 1985.

9. Thomas, J., Bodt, B.A., Webb, D.W., "SQUASH: Recommendations for Three Statistal Measures of Comparison With Test Data", Probability and Statistics Branch / SECAD, Ballistic Research Laboratory, Aberdeen Proving Ground, MD, (Private Communications), March 1987.

10. Shnidman, R., "New Concept for Statistically Testing Component Kill Predictions", Vulnerability Methodology Branch / VLD, Ballistic Research Laboratory, Aberdeen Proving Ground, MD, (Private Communications), Dec. 1987.

Mathematical Investigation of The Reaction-Diffusion Equations For Heterogeneous Catalytic Systems (U)

*Louis J. Piscitelle, Mr.
Department of the Army
US Army Natick Research, Development and Engineering Center
Natick, MA 01760-5017

INTRODUCTION

A basic problem in chemical protection is the prevention of a given substance from penetrating into a clean area, e.g., a tent or the inside of a soldier's uniform. This may be accomplished by using some barrier material that is impermeable to the substance. One problem with this method is that the exterior surface of the material is still contaminated. Another problem is that if the barrier is impermeable to the substance it is also impermeable to air and water vapor and this can result in a heat stress problem for the soldier. The use of a reactive material that is capable of breaking down the substance into harmless products appears to be an attractive solution to this basic problem.

A catalyst is a chemical which accelerates a given chemical reaction but is not consumed during the reaction. For the case considered in this paper, the reaction will be the breakdown of a substance, called the substrate, into products. By combining with the substrate, the catalyst forms an intermediate, called a complex, which then breaks down to form products and to regenerate the catalyst. In so-called homogeneous catalysis, all of the catalyst is readily available to the substrate and the reaction is a time-dependent process only. However, if the catalyst is attached to some solid matrix, the support, the substrate must diffuse through this material or its pores to reach the catalyst. This type of system is called heterogeneous. In such systems the interpretation of experimental results is complicated by the interaction of the reaction and diffusion processes. Also, the mathematical model becomes more complex because the reaction-diffusion process is dependent upon both spatial position and time.

Heterogeneous catalytic systems in which the contribution of the external surface reaction is significant are not well understood. Most of the work in the area of heterogeneous catalysis has examined problems where the internal surface area of the support for the catalyst is much greater than the external surface area. Also, because of the difficulties involved in solving the reaction diffusion equations for systems of this type, many researchers have assumed steady state conditions exist and have chosen a simplified kinetic model. The body of research dealing with heterogeneous catalysis and the reaction diffusion equations which describe the process is extensive. Hence, no attempt will be made here to provide a complete bibliography. An excel-

lent reference and source of classical results and papers is to be found in Aris [1]. The survey paper by Fife [2] on the asymptotic theory of reaction diffusion equations provides a list of papers of a more modern flavor.

This paper will examine the reaction diffusion equations which describe a heterogeneous catalytic system with significant surface reaction. A governing set of nonlinear partial differential equations is derived from the chemical reaction equations using the law of conservation of mass. The system is parabolic and the nonlinearities are polynomial. The equations describe the time evolution of the concentrations of the substrate, catalyst, complex and product as well as their spatial distribution.

The study will be both analytical and numerical. Qualitative methods from dynamical system theory will be used to discuss the mathematical structure of the system of reaction diffusion equations. The numerical solutions are obtained by using the method of lines. This technique consists of discretizing the space dimension using an appropriate finite difference scheme and than numerically integrating the resulting set of ordinary differential equations see Hyman [3]. A FORTRAN code for this method of solution has been written by J.M. Hyman and is quite robust.

PHYSICAL MODEL

The film is modelled as a slab which is either very thin or has perfectly sealed edges so that the problem is expressible in terms of one space dimension. The exterior conditions are assumed to be known and the flux of substrate from the fluid surrounding the film to the film's surface is modelled by an effective mass transfer coefficient. The chemical reaction is assumed to be catalytic and the catalyst is taken to be fixed within the supporting matrix. This assumption has interesting consequences in that while the substrate and product are free to diffuse within the film, the catalyst and the catalyst-substrate complex formed are immobile. Hence, there are three processes of importance in this system. They are:

1) The ability of the substrate to diffuse from the fluid to the surface of the film. A related problem is that of the adsorption of substrate onto the surface which can described by various isotherms (e.g., Langmuir, Freundlich, etc.).

2) The intrinsic reaction rate of the catalysis. This is what one would observe if the reaction was occurring in solution, i.e., homogeneous catalysis, or if the diffusion rates were much greater than the reaction rates.

3) The ability of the substrate to diffuse into the film and reach the catalyst which is not on the surface.

GOVERNING EQUATIONS

The governing equations for the reaction are assumed to be of the form

$$S+E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C \tag{1}$$

$$C \xrightarrow[k_2]{} P+E \tag{2}$$

where S, E, C and P represent the concentrations of the substrate, catalyst, complex and product respectively. Also, k_1, k_{-1} are the forward and backward reaction rates for the formation of complex, respectively and k_2 is the reaction rate for the breakdown of the complex into product and catalyst. At this stage, the effects of product poisoning of the catalyst, multiple products or other intermediates have not been included. The on-going work in this area will address these effects in the future.

Using equations 1 and 2 and assuming one space dimension x, we obtain the following set of reaction diffusion equations:

$$S_{,t} = D_S\, S_{,xx} - k_1 SE + k_{-1}C \tag{3}$$

$$E_{,t} = D_E\, E_{,xx} - k_1 SE + k_{-1}C + k_2 C \tag{4}$$

$$C_{,t} = D_C\, C_{,xx} + k_1 SE - k_{-1}C - k_2 C \tag{5}$$

$$P_{,t} = D_P\, P_{,xx} + k_2 C \tag{6}$$

where the subscripts (,t) and (,xx) imply $\partial/\partial t$ and $\partial^2/\partial x^2$ respectively and D_S is the diffusivity of the substrate, etc. Note, that the assumption that the catalyst and complex are immobile has not as yet been used.

Equations 3 through 6, represent a set of nonlinear reaction-diffusion equations where the nonlinear terms are polynomial. We next examine the equations for the boundary conditions. All of the rate constants given thus far have been per unit volume. In order to describe boundary conditions which include reaction, these rates are converted to rates per unit surface area. Let the catalytic body have a surface area to volume ratio of A/V . Then a boundary condition for the right hand surface has the general form:

$$h(S_{fluid} - \underline{S}) - V/A\,(k_1 \underline{S}\underline{E} + k_{-1}\underline{C}) - D_S \underline{S}_{,x} = 0 \tag{7}$$

where S_{fluid} is the concentration of substrate in the fluid and all other concentrations are evaluated at the film surface.

Equation 7 contains the effects of the substrate's transfer from the surrounding fluid to the film surface, its reaction on the surface and its diffusion into the film. Equation 7 is a consequence of a mass balance at the surface and states that the rate at which substrate enters the film through the fluid boundary layer at the surface minus the rate at which the substrate is destroyed at the surface gives the net rate at which substrate enters the film.

At this point it is advantageous to put the governing equations in dimensionless form. Let

$$u = S/S_{fluid}, \quad v = E/E_o, \quad w = C/E_o \quad \text{and} \quad q = P/S_{fluid} \qquad (8a,b,c,d)$$

where the subscript (o) indicates initial values of the variable and

$$\lambda = \frac{k_2}{k_1 S_{fluid}}, \quad \kappa = \frac{k_{-1} + k_2}{k_1 S_{fluid}}, \quad \varepsilon = \frac{E_o}{S_{fluid}} \qquad (9a,b,c)$$

and

$$\tau = k_1 E_o t, \quad \eta = \frac{A}{V} x \qquad (9a,b)$$

also define

$$\delta_C = \frac{D_C}{D_S}, \quad \delta_E = \frac{D_E}{D_S}, \quad \delta_P = \frac{D_P}{D_S} \qquad (\text{form } 10a,b,c)$$

and finally

$$\phi^2 = \left(\frac{V}{A}\right)^2 \frac{k_1 E_o}{D_S}, \quad \alpha = \frac{V}{A}\frac{h}{D_S} \qquad (11a,b)$$

This choice of dimensionless groups is rationalized as follows. The time scale $1/k_1 E_o$ is essentially the time for the formation of complex. The diffusivities are all scaled relative to the diffusivity of the substrate in the film. The concentrations of substrate and product are scaled by the concentration of substrate in the surrounding fluid. The concentrations of catalyst and complex are scaled by the initial concentration of the catalyst, since the concentrations of each must be bounded by this value. The parameters λ, κ, and ε are standard dimensionless parameters which occur in the homogeneous catalysis problem. In general, ε is a small parameter.

In terms of the dimensionless variables and parameters defined in equations 8 through 11, the reaction diffusion equations (equations 3 through 6) become:

$$u_{,\tau} = \frac{1}{\phi^2} u_{,\eta\eta} - uv + (\kappa - \lambda) w \qquad (12)$$

$$\varepsilon v_{,\tau} = \frac{\varepsilon \delta_E}{\phi^2} v_{,\eta\eta} - uv + \kappa w \qquad (13)$$

$$\varepsilon w_{,\tau} = \frac{\varepsilon\delta_C}{\phi^2} w_{,\eta\eta} + uv - \kappa w \tag{14}$$

$$q_{,\tau} = \frac{\delta_P}{\phi^2} q_{,\eta\eta} + \lambda w \tag{15}$$

The assumption that the catalyst and complex are both immobile will now be examined. Let $\delta_E = \delta_C = 0$, then equations 13 and 14 may be rewritten as:

$$\varepsilon v_{,\tau} = -uv + \kappa w \tag{13'}$$

$$\varepsilon w_{,\tau} = uv - \kappa w \tag{14'}$$

Now add equations 13' and 14' to obtain the result

$$(v + w)_{,\tau} = 0 \tag{17}$$

which implies that $v + w = g(x)$. For the calculations performed in this paper the function g is taken to be a constant equal to 1. Hence

$$v = 1 - w \tag{18}$$

The function $g(x)$ is not a function of time. Since initially none of the catalyst has been used, no complex can exist. Hence $g(x)$ is just the initial spatial distribution of the catalyst within the film. This has been taken to be uniform, but an interesting and important consideration would be the effect of a different distribution upon the ability of the film to breakdown a substance and hence prevent penetration. Based upon the above simplifications, the set of reaction diffusion equations which will be examined are:

$$u_{,\tau} = \frac{1}{\phi^2} u_{,\eta\eta} - u(1 - w) + (\kappa - \lambda)w \tag{19}$$

$$\varepsilon w_{,\tau} = u(1 - w) - \kappa w \tag{20}$$

$$q_{,\tau} = \frac{\delta_P}{\phi^2} q_{,\eta\eta} + \lambda w \tag{21}$$

The boundary conditions and the initial conditions for the problem will be described in the section on applications. There are however, some interesting features of the set of partial differential equations given by equations 19 through 21. If the equations are examined under the assumption that the diffusion coefficients are large, then the system behaves as if $u_{,\eta\eta} = 0$ i.e., reaches a linear distribution through the film thickness. Then at each point the system is essentially the following set of ordinary differential equations:

$$\dot{u} = -u(1 - w) + (\kappa - \lambda)w \tag{22}$$

$$\varepsilon\dot{w} = u(1 - w) - \kappa w \tag{23}$$

$$\dot{q} = \lambda w \tag{24}$$

These equations (22 - 24) are the classical enzyme equations examined using singular perturbation techniques by Heineken [4]. They have also been investigated by Carr [5] using invariant manifold theory and it was demonstrated that the system possessed a center manifold on which the flow agrees, to lowest order, with the differential equation originally derived by making the pseudo-steady state hypothesis of Michaelis-Menten kinetics (this theory said that a reasonable approximation to the full solution could be obtained if one assumed the concentration of the complex to be changing slowly in time and to depend upon the concentration of the substrate via an algebraic equation rather than a differential equation). Thus the full system of partial differential equations (equations 19-21) has embedded within it a system which has rich and important structural behavior.

NUMERICAL CONSIDERATIONS

The program, MOL1D, used to numerically integrate the system of partial differential equations is a research code written in FORTRAN by Hyman [3]. It allows the use of both a variety of spatial discretization schemes (the user may also write the particular finite difference scheme to be called) and a number of good integrators for the time variable. The method of solution has been described in this paper's introduction. The structure of the reaction diffusion equations is such that the difficulties in obtaining a numerical solution do not lie in the spatial approximation. These systems are usually stiff and it is the time integration of the resulting set of ordinary differential equations which require care. For this reason, a spatial (x) mesh of 30 points was chosen and a sixth order symmetric finite difference scheme was used to approximate the spatial derivatives. Gear's backward differentiation formula was used to integrate the time variable. The computations were performed on a problem with constant diffusion coefficients. Reaction-diffusion equations with variable diffusion coefficients can also be handled by the MOL1D code.

The problem was examined for sensitivity to the number of grid points used in the spatial discretization, the order of the finite difference scheme used and the particular time integrator chosen. The final choice was based upon obtaining acceptable accuracy with minimal time penalty. Even so, a typical run is on the order of 2 hours.

APPLICATIONS

The practical problem outlined in the introduction is now considered. A material which is reactive (catalytic) is attached to a permeable support matrix. This paper determines the ability of the resulting film to provide protection against the penetration of some substance.

The film is modelled as a slab of thickness T and it is assumed that a fluid is in contact with both the left and right faces of the slab, see figure 1.

Employing the data used by Heineken et al. [4], it is found that the dimensionless constants, equations 9a-9c, for the kinetics portion of the system are:

$$\lambda = .375, \quad \kappa = 1.0, \quad \varepsilon = 0.1 \tag{25a,b,c}$$

It is also assumed that the diffusivity of the product is on the same order of magnitude as that of the substrate and that these are both an order of magnitude smaller that the mass transfer coefficient for the fluid. The assumption that the product behaves as the substrate is meant only to be a starting point for the full parametric study which can be performed. Based upon these relative magnitudes, the dimensionless parameters, equations 10a-c and 11a,b, become:

$$\delta_C = 0, \quad \delta_E = 0, \quad \delta_P = 1.0 \tag{26a,b,c}$$

$$\phi^2 = 10., \quad \alpha = 10.0 \tag{27a,b}$$

respectively.

The concentration of substrate in the fluid in contact with the right surface of the film is a constant nondimensionalized to unity. The concentration of the substrate in the fluid in contact with the left hand surface of the film is held at zero. Initially, the concentrations of the substrate, complex and product in the film are zero. The concentration of the catalyst is as stated earlier uniformly distributed throughout the film and has the nondimensionalized value 1. It is also assumed that the product is unable to leave the film and enter the fluid. The substrate, of course, does not have this restriction. The initial conditions for this problem may then be written as:

$$u(\eta,0) = 0., \quad w(\eta,0) = 0., \quad p(\eta,0) = 0. \text{ for } 0 \le \eta \le 1. \tag{28a,b,c}$$

and the boundary conditions for the right surface of the film are:

$$\alpha(1. - u) - \phi^2 u(1. - w) + (\kappa - \lambda)w = u_{,\eta}, \quad at\ \eta = 1. \tag{29a}$$

$$w_{,\eta} = 0., \quad at\ \eta = 1. \tag{29b}$$

$$\delta_P q_{,\eta} = \lambda\phi^2 w, \quad at\ \eta = 1. \tag{29c}$$

Since the boundary conditions for the left hand surface are similar, they will not be written here.

DISCUSSION OF RESULTS

This section describes the results of integrating the reaction diffusion equations given by equations 19-21, subject to the initial and boundary conditions given in equations 28a-c and 29a-c and the parameter values given in equations 25a-c, 26a-c and 27a,b. Figures 2., 3. and 4. show the substrate, complex and product concentrations, respectively, as a function of position in the film at dimensionless time $\tau = 2$. The left and right hand surfaces of the film are at $\eta = 0$ and $\eta = 1$, respectively. The concentration of substrate in the film is (as expected)

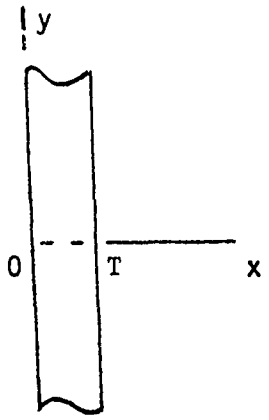


FIGURE 1. CROSS SECTION OF FILM

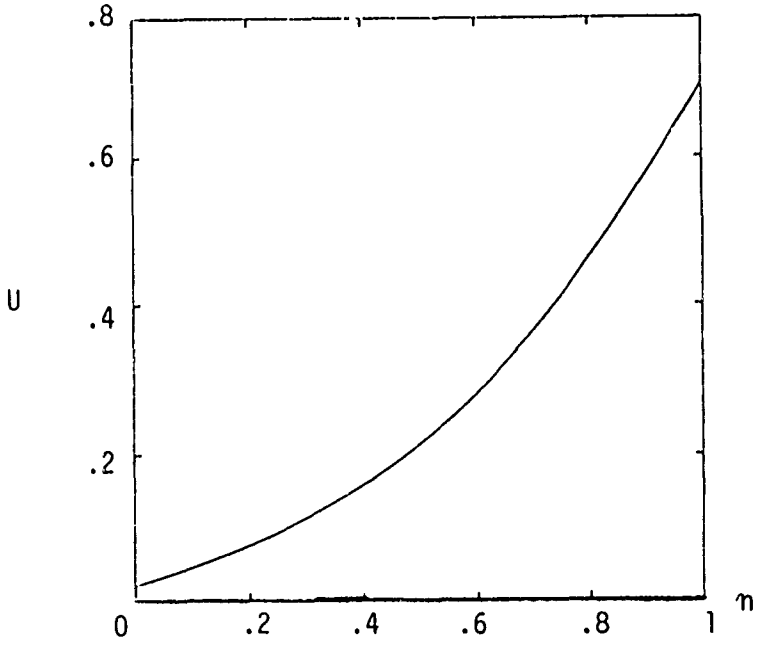


FIGURE 2. SUBSTRATE CONCENTRATION vs POSITION

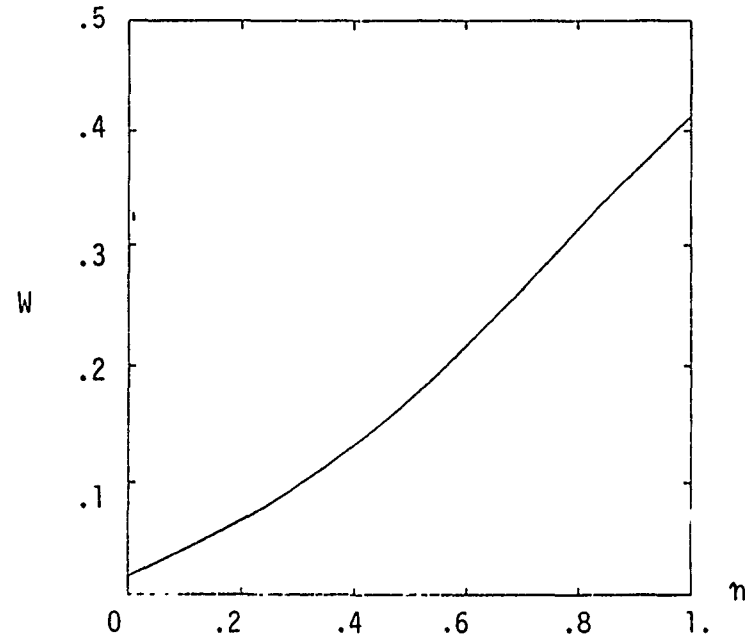


FIGURE 3. COMPLEX CONCENTRATION vs POSITION

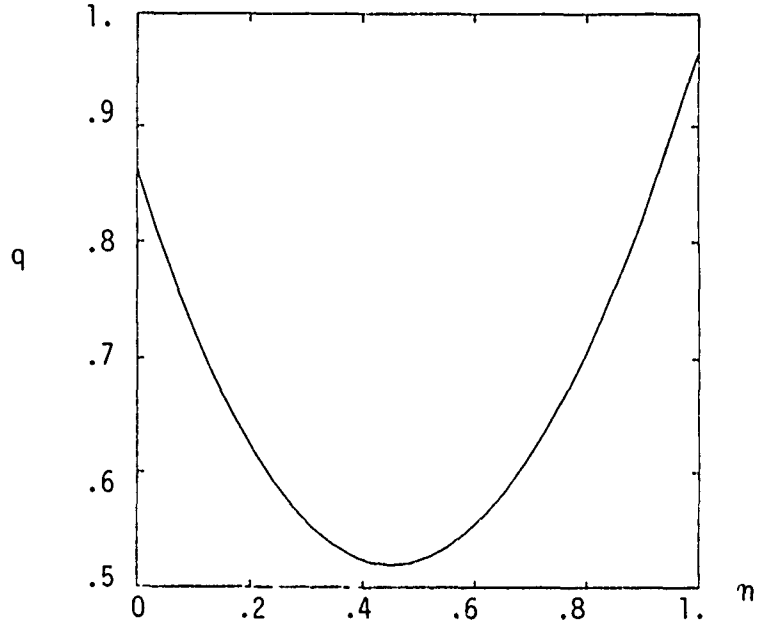


FIGURE 4. PRODUCT CONCENTRATION vs POSITION

highest at the right hand surface. The concentration of product is seen to have a parabolic shape. This is due to the boundary condition that allows product to be created at each surface but does not allow product to leave the film by diffusing into the surrounding fluid. Hence, the interior is served by both surfaces. The surface concentration is higher than the concentration at the center because of the order of magnitude difference between the rate of product creation and the rate of diffusion.

Figure 5. shows the concentration of the complex at the right hand surface of the film as a function of time. The behavior is quite consistent with catalytic reactions which follow Michaelis-Menten kinetics. While this problem is one of heterogeneous catalysis, the structure of the partial differential equations is such that the behavior should be similar. The system given by equations 19-21, has a small parameter, ε , multiplying the time derivative of the complex concentration. This is suggestive of a singular perturbation problem for which one would expect the behavior displayed in figure 5., namely a rapid increase in the complex concentration to it's nearly constant value. Figure 6. is a very clear illustration of this behavior. Figure 6 is a graph of the complex concentration against the substrate concentration from $\tau = 0.$ to $\tau = 2.$. The initial portion of this curve from the origin to about (.5 , .35) occurs in about .3 dimensionless time. After this time, both the substrate concentration and the complex concentration vary slowly, but the complex concentration varies three times as slowly. The existence of a center manifold for this system of partial differential equations with the boundary conditions given here has not yet been proven. This is a problem which will be examined.

Along these lines, an interesting calculation can be performed to determine if the quantitative results of the program are reasonable. In equation 19 set $\varepsilon = 0$. This is a reasonable approximation because the variable is small. Now the resulting algebraic equation is:

$$w = \frac{u}{u + \kappa} \tag{30}$$

The complex concentrations are approximated by equation 30 quite well. If a comparison is made at time $\tau = 2.$, the maximum error in the approximation at any point in the film is about six percent. Table 1 below gives an overview of the error made at various positions within the film by using the approximation in equation 30.

TABLE 1.

Errors in Approximation of W by Eq. 30

Axial position η	0.0	0.31	0.69	1.0
U	.018	.117	.355	.706
W	.0168	.1011	.2599	.4134
W (by eq. 30)	.0178	.1047	.2622	.4137
Per cent error (in W)	6.0	3.5	0.9	0.07

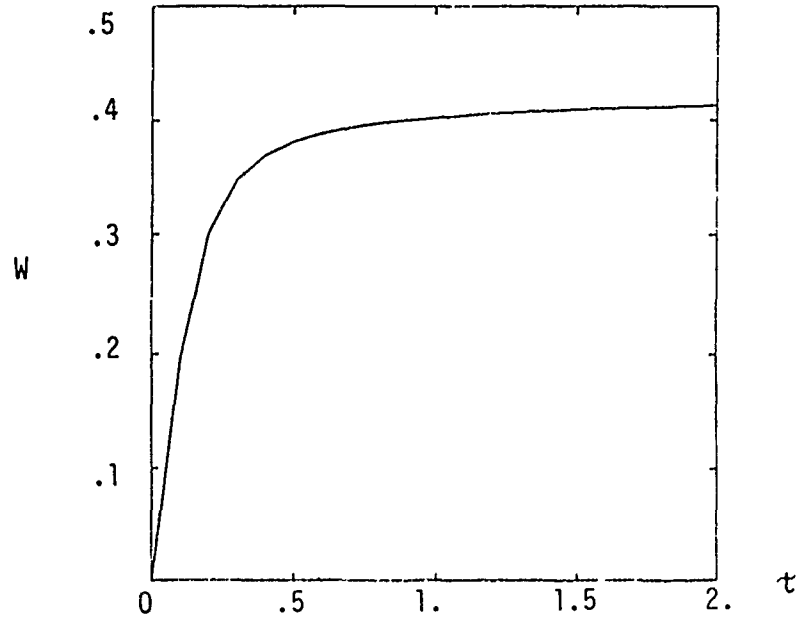


FIGURE 5. COMPLEX CONCENTRATION vs TIME
at RIGHT SURFACE

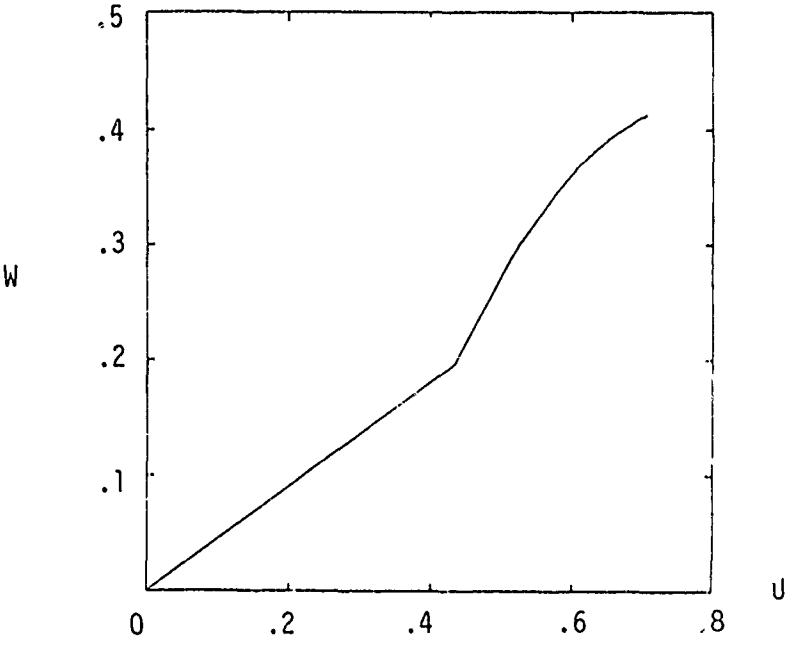


FIGURE 6. COMPLEX CONCENTRATION vs
SUBSTRATE CONCENTRATION

At this point in the study, the author considered the question of what the concentration of substrate would be in the film if there had been no reaction. This was considered in order to determine what amount of substrate would be prevented from penetrating the film after a given period of time as a direct result of the reactivity of the film. The linear diffusion problem (which is created by eliminating the production terms in the right hand sides of equations 19-21) is classical and hence obtaining its solution posed no difficulty. Figure 7 shows a plot of the concentration of substrate at the left hand surface of the film as a function of time, for the cases of reaction and no reaction in the film. Figure 7 shows approximately a 50 per cent reduction in the concentration at time $\tau = 2$. It also shows that the concentration with reaction is significantly reduced at all times. Figure 8 shows a plot of the substrate flux as a function of position in the film at time $\tau = 2.$ for the cases of reaction and no reaction. The flux of substrate out of the film at the left hand surface $\eta = 0.$ is shown to be reduced by about 60 per cent for the case where the film is reactive. The fact that the flux into the film is higher at the right hand surface for the reactive case is easily understood by noticing that since the film is breaking down substrate, the concentration of substrate at the surface will be lower and hence the driving potential from the difference in fluid-surface substrate concentration also be greater, resulting in the higher flux.

CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK

The method of lines is an accurate and reasonably quick method for solving problems resulting in systems of reaction diffusion equations. The study performed in this paper indicates that the ability of a film to reduce the amount of substrate penetrating is directly related to the reactivity of the material. The results obtained were physically reasonable and the potential now exists to carry out further work in this area. This effort is important because of the difficulty and expense involved in the testing of reactive materials to determine their protection potential. It is also important because it will provide a technical basis for making design decisions.

There are a number of open problems in this field. The boundary conditions considered in this paper have been time independent. The author has performed some calculations for time dependent boundary conditions, which appear to be yielding very interesting results but require further study. The addition of more intermediate reactions and of poisoning of the catalyst will be of both theoretical and practical value. Finally, the question of determining the optimal distribution of catalyst throughout the film is a challenging and important practical problem.

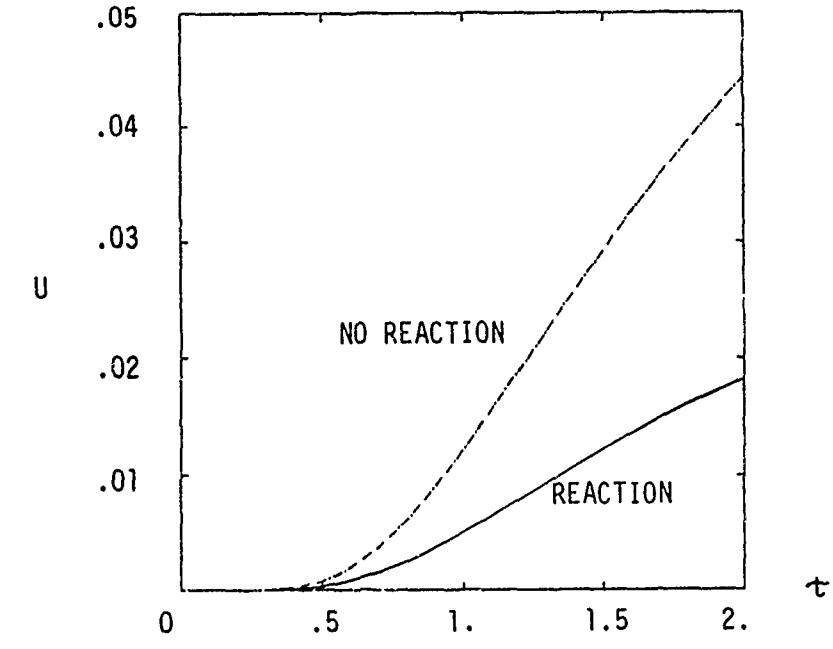


FIGURE 7. SUBSTRATE CONCENTRATION vs TIME at LEFT SURFACE

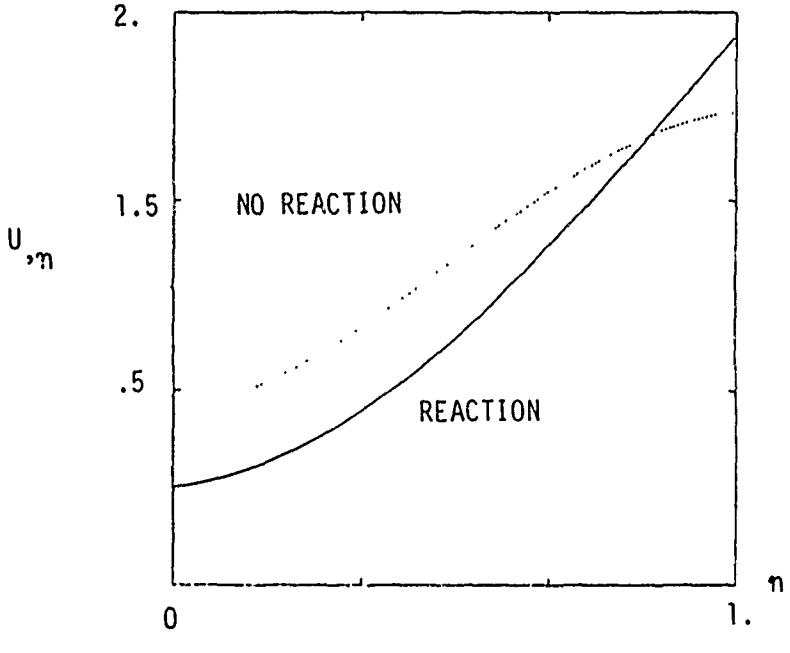


FIGURE 8. $U_{,\eta}$ at TIME 2.0 vs POSITION

REFERENCES

1. Aris, R., THE MATHEMATICAL THEORY OF DIFFUSION AND REACTION IN PERMEABLE CATALYSTS, Clarendon Press, Oxford, 1975.

2. Fife, P. C., "Results and Open Questions in the Asymptotic Theory of Reaction-Diffusion Equations," NONLINEAR EVOLUTION EQUATIONS, Academic Press, New York, 1978.

3. Hyman, J. M., "The Method of Lines Solution of Partial Differential Equations," Courant Institute of Mathematical Sciences, Mathematics and Computing Laboratory report COO-3077-139, 1976.

4. Heineken, F. G., Tsuchiya, H. M., Aris, R., "On the Mathematical Status of the Pseudo-steady State Hypothesis of Biochemical Kinetics," Mathematical Biosciences 1, 95-113, 1967.

5. Carr, J., APPLICATIONS OF CENTRE MANIFOLD THEORY, Springer-Verlag, New York, 1981.

STAGES OF SKILL ACQUISITION AND TRAINING DEVICE/SIMULATOR FIDELITY

CAROL A. POLLACK-NELSON, MS
*RAY S. PEREZ, PhD
OK-CHOON PARK, PhD
U.S. ARMY RESEARCH INSTITUTE
ALEXANDRIA, VA 22333-5600

I. INTRODUCTION

The introduction of complicated and expensive equipment into the working environment makes training on such equipment impractical. To provide novices with the opportunity to master necessary skills and procedures without overwhelming them with information or incurring inordinate costs and safety risks (to either the operator himself or the equipment), simulators are, with ever-increasing frequency, becoming the choice mode of training presentation.

The variety of tasks for which simulators have been used ranges from training simple procedural tasks to those requiring complex trouble-shooting skills. The question which has been posed by numerous resear-chers concerns the level of physical and functional fidelity a simulator should provide. As a result, the bulk of simulator research has involved comparisons of devices with different fidelity levels presented during training, and the resultant transfer to the actual equipment.

While many researchers have found evidence supporting the hypothesis that higher fidelity is needed for training (Brictson & Burger, 1976; Gray, 1979; Browning, Ryan, Scott & Smode, 1977), others have found no significant differences between devices of lower and higher fidelity and have, therefore, recommended using the former to save costs (Cox, Wood, Boren & Thorne, 1965; Grimsley, 1969; Prophet & Boyd, 1970). In an attempt to resolve this conflict in findings, some experimenters have begun to hypothesize that transfer is not solely a function of training device fidelity. Rather, there are several other variables that must be examined in conjunction with fidelity so that optimal transfer can be realized. Both task type (Hays, 1980) and the phase of learning the student is in (Fitts, 1964; Fink & Shriver, 1978; Kinkade & Wheaton, 1972) are two such variables which need to be considered when deciding the level of fidelity which is optimal to train a task.

TASK TYPE

Predictably, Many of the studies which have demonstrated success with high fidelity simulators were used in training tasks that required difficult discriminations and/or the responses to be learned were critical to the system's operation. Kinkade & Wheaton (1972), however, stated that unless the training objective is the acquisition of high level skills (e.g., decision-making), the device probably does not need to be of high fidelity. That is, for simple procedural tasks, lower fidelity levels may be sufficient to produce acceptable transfer. In terms of specific tasks, Fink & Shriver (1978, cited by Hays, 1980), suggested that the level of fidelity called for by maintenance training simulators differs from those devices used to train operational procedures because the former must accommodate the decision-making processes that occur during maintenance.

An example of a task which requires decision-making processes is a troubleshooting maintenance task. The literature defines a troubleshooting task as one which requires the individual to identify a faulty system component and possibly to repair that component as well. Successful completion of the task requires almost exclusive reliance on cognitive processing skills (Allen, Buffardi & Hays, 1985). According to Morris & Rouse (1985), the technician's abilities to: (a) repair or replace faulty components; (b) to perform tests; and (c) to employ a strategy, determine how well s/he will fare on the troubleshooting assignment.

STAGE THEORY OF LEARNING

In addition to the task type and difficulty, researchers must also consider the phase of learning the trainee is in, when prescribing the fidelity level a task requires. Many researchers have delineated different stages of learning they believe the trainee passes through when acquiring new skills (Fink & Shriver, 1978; Fitts, 1964); Kinkade & Wheaton). Fink & Shriver (1978) identified four stages, each having a different general training objective and therefore requiring different types of training devices. In the first stage, the training objective is to "acquire enabling skills and knowledges". To accomplish this, trainees should be provided with simple demonstrators such as wall charts, films, TV, along with the nomenclature and identification of parts location in order that a basic understanding of the equipment is achieved.

In the second stage of learning, the trainee begins to "acquire uncoordinated skills and unapplied knowledges" and should be provided with part-task trainers and/or procedures trainers. The coordination of these skills and the ability to apply knowledge marks the third stage. At this point, troubleshooting logic trainers, skills and/or job segment trainers

are needed. In the fourth stage, job proficiency in the actual job setting is the general training objective and the operational equipment and actual equipment trainers (AET) would be most useful. It should be re-emphasized that according to the Fink & Shriver model, the actual equipment trainers or operational equipment need not be brought into the training scheme until the final stage of learning.

Fitts's (1964) theory of basic learning processes is a stage theory of learning. Fitts purports that there are three stages one passes through when learning a new skill. First, is the "cognitive stage", at which time the initial encoding of the new skill occurs. At this point, the trainee is able to generate a "crude approximation" of the desired behavior. Additionally, the student may engage in what Fitts refers to as "verbal mediation", when the student goes through the task verbally to assist in their acquisition of a new skill. During the second, "associative stage", the learner's performance of the skill becomes more natural and errors made during the initial phase of learning begin to disappear. So does any verbal mediation which may have occurred during the previous stage. The third stage of skill acquisition, the "autonomous stage", is a never-ending process which is characterized by gradual improvements in performance as the skill becomes more "second nature" with time.

IMPLICATIONS OF THE STAGE THEORY AND THE EFFECTS OF TASK TYPE

Viewing the acquisition of new skills as a developmental or evolutional process, rather than an all-or-none occurance, helps explain the contradictory findings in the literature regarding the level of fidelity that is needed in a training device. Fink & Shriver suggest that, part-task and procedures trainers are sufficient for training in the earlier stages of learning. They speculate that the use of high fidelity devices during the initial stages of training may account for the finding of negative transfer on some studies. D. Andrews (1983) points out, placing a new trainee in a high fidelity simulator often leaves him "bewildered and frustrated by cues with which he cannot cope, even under training conditions" (p.71). This point was demonstrated in a study by Martin & Waag (1978) who found a high fidelity flight simulator detracted from the learning acquisition of Air Force novices, presumably because too much information was provided too soon. Not only does overstimulation have a negative effect on learning, but it may also place too much reliance upon inappropriate environmental cues for the acquisition of new skills.

The identification of different learning stages forces one to consider the possible negative impact training may have on a novice. First, in early learning stages, the training developer must be careful not to overload the trainee with information. Further, it suggests that the training developer must not only consider how much information but the type of

information that is appropriate to provide the trainee. Classic studies in cognitive psychology state that 7 bits of, non-meaningful, information (plus or minus 2) is the most one can understand and incorporate into memory. Excess information beyond that point poses the problem of information overload. Only after some initial period of acquisition, is the trainee able to (chunk) organize the new information into a meaningful whole with existing material. At this point, s/he is ready to take on new information, or utilize that which s/he already knows.

Second, it could be expected that the length of time one remains in a particular stage is likely to be related to the type and/or difficulty of the task. The more difficult the task (e.g., troubleshooting, decision-making), the longer it could be expected to take a trainee to pass through the first two stages before reaching autonomy. Because the trainee remains in each stage longer during a difficult task, the importance of providing him/her with training devices which are suitable to a particular stage of learning, is greater than it would be for a simple, procedural task where relatively little time is spent in the first two stages (i.e., before autonomy is achieved). Thus, when training a task with a high cognitive load (e.g., troubleshooting), presentation of appropriate training materials is critical. Rather than have trainees work on only one device, as has typically been done in the past, they should have the opportunity to work with equipment that is specially suited to match the different stages of skill acquisition. The model of the acquisition of the cognitive skills in a troubleshooting task are:(a) the Cognitive Stage, involves the initial encoding of cognitive information (e.g.,facts and concepts) needed to perform the task;(b) the Associative Stage,involves the linkage between the cognitive information and skill performance (e.g.,troubleshooting strategies); and (c) the Autonomous stage,includes the gradual improvement of skill performance with repeated practice.

PRESENT RESEARCH

The literature seems to indicate that rather than compare the exclusive use of high or low fidelity simulators for training a troubleshooting task, the joint application of these two types of trainers might lead to optimal transfer. The question this study seeks to answer is whether or not a combination is better than the sole use of a high or low fidelity simulator throughout training. According to the Fitts paradigm, the presentation of less complicated devices at the start of training should prevent negative learning that might occur if the novice were overwhelmed with too much information.. Followed by training on a high fidelity simulator would allow for some practice with a device which is similar physcially and functionally, to the actual equipment and would prevent negative transfer that might occur if the novice were never provided with the opportunity to become familiar with a replication of the actual equipment.

This study tests the hypothesis that optimal transfer will occur with the progressive presentation of varying levels of low to high levels of fidelity during the stages of skill acquisition and would be superior to either low or high fidelity alone.

II.METHODS

A. Subjects

24 male and 30 female undergraduate students from the George Mason University participated in the study. All were paid volunteers and were recruited either in their psychology classes or through advertising on a bulletin boards located in the Psychology department. All races and religions were represented. Demographic information about the sample was obtained for the purpose of investigating individual differences. Information on any differences can be found in the Results section of this paper.

B. Apparatus

1. Subject Equipment: Simulator Training and Test

a. Actual Equipment Trainer

Two types of training devices were used in this study. The actual equipment trainer (AET) was employed at different points in the training (as specified below in the Procedures section of this paper) and during all the testing phases, at which time transfer was assessed. The AET consisted of a large circuit board containing 28 electromechanical relays and five solid-state pull up panels. Each of the 28 relays on the circuit board could be identified by a number and an associated "relay checkpoint" which was seen as a red circle next to the relay number. The five pullup panels were situated to the left of the circuit board, in an aluminum frame. Each wooden panel could be lifted upwards, one at a time, to reveal a small green light. If the panel was functioning properly, this light would be on. The panels, labeled "Panel A" through "Panel E" corresponded to relays 16, 26, 20, 5, and 1 (from A through E, respectively). Amber lights on top of the aluminum frame reminded the subject that a tray was down (correct position) or up (test position).

The relays and pullup panels were interconnected to 8 output devices: a fan, a water pump, a solenoid valve assembly, three lights, a TV monitor, a sound generator and a speaker. If there were no faults in the system, the 8 output devices would operate properly. However, when the experimenter caused a fault to occur in the system, the associated relay(s) or pullup panel(s) and one or more of the output devices would fail

to function. Pictures of this device can be found in the Appendix. This device is of both high physical and functional fidelity.

b. 2-D Mockup

The 2-D mock-up of this system, which was used in some training conditions (as specified in the Procedures section below) simply consisted of a "pen-and-ink" symbolic representation of the AET. Relays, pullup panels and output devices were represented by rectangles. Lines between the rectangles represented the relay interconnections. While this mock-up did not bear a strong physical resemblance to the AET (low physical fidelity), it was of high functional fidelity.

c. Hand-Held Tester In order to test the relays/pullup panels, in the AET, to see if they were working or not, the trainee needed to operate a hand-held testing device. The dials on front of this hand-size box could be arranged so that they could display the numbers 1-28 in the window; each number corresponded to a particular relay. Extending from the hand-held tester was a probe which had to be touched to the relay checkpoint in order for a test to be carried out. If the relay was functioning properly, a green light would light up on the face of the tester. If it was not working properly, the green light would not come on. This hand-held tester was attached to a MINC-11 computer in the experimenter's room.

The hand-held tester was also used when the trainee wished to signal final solutions to the computer. When the subject felt that s/he had finally discovered the faulty relay/pullup panel, s/he dialed in the number of the relay that was suspect. Then, s/he would depress the "solution" button and then, the "test" button. If his/her suspicions were correct, a red light would show up on the face of the hand-held tester and a high-pitched tone would be emitted. In the event that wrong "solution" was selected, neither the light nor the tone would be produced.

d. Tape Recorded Message

All the instructions for how the task was to be performed were administered via a tape recorded message.

e. Cards Indicating Defective Outputs (Mock-Up Only)

Since the low fidelity simulator was unable to indicate, upon testing, which of the output devices were "down", subjects were presented with an index card for each problem performed on the mock-up, indicating the output devices that were not working.

f. Informed Consent

A page explaining who is conducting the experiment and that no harm will come to any participant. At the bottom of this page is a place for the subject's signature. Also on the page is the date and the signature of the experimenter.

g. Background Questionnaire

This questionnaire assesses the subject's prior experience with technical tasks, asks them how confident they feel about performing such tasks, and also collects some basic demographic information.

h. Bennetts Mechanical Comprehension Test

This test assesses one's mechanical ability. The test is composed of 68 multiple choice questions, where each question is followed by three answers, one of which is correct.

i. Vocational Interest Inventory

This questionnaire, developed by the Department of Labor asks the reader to check off as many or as few of the listed "interest areas" that s/he would like to do. The items are grouped according to type of activity. Scores tell what area the test taker would be best at, assuming that people do best in areas in which they are interested.

2. Experimental Apparatus:

a. Reference System (AET) Faulter Panel

This device, located in the experimenter's room, faulted the relays and pullup panels on the large circuit board with which the subjects worked. This panel consisted of 28 lights and corresponding switches. Each of the 28 lights/switches represented and controlled the functioning of a relay and/or pullup panel on the circuit board. To fault a component, the experimenter simply threw the switch that corresponded to the relay s/he wished to disrupt. This action caused the component to stop working and thus, all interconnected relays, pullup panels and output devices to also cease functioning. The lights on the faulter panel made it easy for the experimenter to identify which switch has been thrown.

b. 2-D Simulator Faulter Panel

This device, used by the experimenters to fault relays and output devices on the low fidelity simulator, or 2-D mockup of the AET, was similar to the faulter panel used with the AET reference system. By throwing a switch associated with a particular relay, the experimenter was able to

fault that relay and therefore, cause one or more of the output devices to fail.

c. MINC-11 Computer

The MINC-11 computer was used to record the subject's test and solution attempts via the hand-held tester. A simulator software program, specially devised by Jon Allen and his associates at the George Mason University (1984) was used to acquire, evaluate, respond and record a trainee's performance. This program is provided in the Appendix.

C. Procedure Subjects were randomly assigned to one of four groups: (1) Low fidelity, (2) Low + High Fidelity #1, (3) Low + High Fidelity #2, (4) High fidelity. The difference between the groups was the type of training they received. The chart below illustrates these four conditions:

	Training Phase = Six Problems:			Transfer Phase:
Condition	Problems 1 & 2	Problems 3 & 4	Problems 5 & 6	Problems 7-9
Low Fid.	2-D	2-D	2-D	3-D
Low+Hi #1	2-D	2-D	3-D	3-D
Low+Hi #2	2-D	3-D	3-D	3-D
High Fid.	3-D	3-D	3-D	3-D

Depending upon the experimental condition to which they were assigned, subjects were ushered into either the room housing the low fidelity simulator (Conditions #1, 2 & 3) or the high fidelity room, housing the AET (Condition #4). Following a verbal welcome by the experimenter, they were asked to take their time in reading a consent form and to sign their name and date.

Upon re-entering the experimental room, subjects were told that they would be asked to troubleshoot six problems on either a simulator or an actual relay panel. All of the instructions explaining how the relay panel functions and the troubleshooting task were provided via tape recorder. The tape was stopped periodically to ensure that subjects understood what was being said and how they were to proceed with the task. When the tape was finished, the experimenter briefly recapitulated the instructions and told the participant that there would be no time limit for each problem. Each problem was initiated by the sound of a buzzer.

At this point, the experimenter left the subject room and entered the

experimenter's room housing the computers that controlled the simulator fault panels. For each problem, a single relay in the panel was knocked out. As a result, one or more of the output devices ceased to function. Additionally, any other relays that were in the path of the faulted relay (on its way to the defunct output device) would also fail to function. The subject's task was to identify the single relay that was causing one or more of the output devices and possibly some other relays to malfunction.

Subjects in all four conditions solved the same six training problems in exactly the same order. The problems were of slightly increasing difficulty. Difficulty was determined prior to the study, and was based on the number of relays and output devices that would be affected by faulting the specific relay. The first two problems knocked out a total of three relays and/or outputs. The next two problems knocked out a total of five relays and/or outputs. The last two problems knocked out a total of seven relays and/or outputs.

Participants in the first condition ("Low Fidelity") completed all six training problems on the 2-D mock-up. Participants in the fourth condition ("High Fidelity") completed all six problems on the 3-D AET. Subjects in the second and third conditions started out on the low fidelity simulator and at some point were moved to train on the high fidelity simulator. This transition was handled by asking them to go into the next room which housed the AET. They were then presented with brief instructions which explained that the device they would work on for the remaining (two or four) problems was similar to the simulator on which they were just working. The tape recorded instructions briefly explained how the AET worked. It was stopped periodically to ensure that subjects understood how the equipment operated. When it was clear that they were ready to proceed, the remainder of the training session was completed.

After the six training problems were completed, subjects were asked to fill out a biographical information sheet. During this time, the experimenter would prepare the three final test problems. Upon completion of the Biographical Questionnaire, subjects were informed that there were three troubleshooting problems to be completed. Since subjects in the last three conditions were already familiar with the AET, they simply began to work without additional instructions. Those assigned to the first condition had never seen the AET before and were provided with the same instructions that the second and third conditions heard when they were transferred into the high fidelity room during training. All subjects received the same three test problems to solve on the AET. These three problems were more difficult than the training problems. The first problem knocked out 10 relays and/or output devices. The second problem knocked out 13 relays and/or outputs and the final test problem knocked out 16 relays and/or outputs.

III. Hypotheses

The hypothesis examined in this study was that subjects receiving a sequential combination of low and high fidelity training would perform better on the three troubleshooting test problems than subjects trained on low or high fidelity simulators exclusively. Performance was operationally defined as: (1) time to solve each problem and, (2) the number of erroneous fault identifications made (errors). In addition to examining the main effects of the treatment conditions, four individual difference variables were studied for their relationship with the dependent variables. These variables are listed below:

1. Performance on the Bennett's Mechanical Interest Inventory
2. Preference for Mechanical tasks, as assessed by the Vocational Interest Inventory
3. Level of Confidence as assessed on Biographical Information Sheet
4. Amount of Prior Experience as assessed on the Biographical Information Sheet

It was expected that those with the higher scores on the Bennett's Test would have better troubleshooting ability and therefore, positively influence their performance on the two dependent variables. A vocational preference for mechanical tasks was also expected to exert a positive influence on the dependent variables, as was their level of confidence and amount of prior experience with related task .

III. RESULTS

To determine the effect of treatments on the dependent variables, time and errors, a multiple analysis of variance was run. From the MANOVA, we observed that there were no significant differences between the four treatment groups in terms of the time it took them to solve the transfer problems and the number of errors made.

To analyze the role of the four covariate variables: confidence, mechanical aptitude, technical experience, and mechanical task interest, correlations were run with the dependent variables, time and errors. Below, the results of the correlational analyses are delineated.

1. Confidence- The confidence measure was rated on a 7-pt scale ranging from "Not too confident" (score of 1) to "Highly confident" (score of 7). The mean confidence rating was 5.012. Men expressed that they were more confident ($x = 5.403$) with technical tasks than women ($x = 4.621$) and this difference was statistically significant ($t = 3.421$, $p < .0012$). Differences between men and women were controlled for in advance by balancing the four treatment groups for gender.

Correlations run between the subject's confidence level and their performance during transfer revealed no significant relationship between confidence and the amount of time required to reach the correct solution during transfer. A significant relationship was found, however, between confidence and the number of errors made ($r = -0.33$, $p < 0.02$). That is, people who are more confident make fewer errors.

2. Mechanical Aptitude- The Bennett Mechanical Aptitude Test (BMAT) measures one's understanding of physical relationships and consists of 68 multiple choice questions. The average score on the test was 46.25. Men scored significantly higher on the BMAT with a mean of 50.33 compared to women, whose mean was 42.17 ($t = 3.70$, $p < 0.0005$).

Correlational analyses examining the relationship between mechanical ability and transfer performance, revealed that mechanical aptitude is related to the dependent variables in a similar fashion to confidence. While no significant relationship was found between aptitude and time to solve the transfer problems, there was a highly significant relationship between one's mechanical aptitude and the number of errors they made ($r = -0.55$, $p < 0.0001$). That is subjects possessing higher mechanical aptitude made fewer errors during transfer troubleshooting problems.

3. Technical Experience- The experience variable was rated on a 7 -point scale ranging from "No experience" (score of 0) to "Expert" (score of 6). Very little variance was obtained on this measure with only 6 of the 52 subjects reporting a score higher than 0. Additionally, none of these six subjects exceeded a score of 4. No further analyses were conducted with this variable.

4. Vocational Interest- The Interest Check List was developed by the Occupational Analysis and Industrial Services Division of the U.S. Department of Labor. The Check List included the following dimensions of interest: "Engineering: Technical Work", "Mechanical Work", "Electrical Work", and "Machinery and Machine Operating". Correlations run between scores on these variables and troubleshooting performance revealed a significant correlation for the interest category "Machinery and Machine Operating"($r = -0.301$, $p < 0.028$). This indicates that those subjects having an interest in machinery tended to make the fewest number of errors during transfer troubleshooting.

In light of the significant correlations between three of the covariates (Confidence, Mechanical Aptitude, and Preference for Machine Operating Tasks), covariate analyses were run to determine the effects of treatment on the number of errors made when covariate variables are controlled for. Results indicated a significant difference between treatment groups in the number of errors made when scores on the BMAT ($F = 6.08$, $p <$

0.0005) were held constant. Thus, had subjects' mechanical aptitude been controlled for in advance, it is very likely that significant differences would have resulted between the treatment groups in terms of the number of errors that were made.

V. DISCUSSION AND CONCLUSIONS

This study revealed several interesting findings. First, although no main effects were found, several significant correlations between covariate variables and the dependent variable, number of errors, resulted. In addition, once differences in mechanical ability were controlled for, significant differences between the treatment groups resulted. Based on the means obtained in the analysis of main effects, one could conclude that, had mechanical aptitude been controlled for, subjects in the conditions receiving a combination of low and high fidelity training would have made statistically signficantly fewer errors. Means also demonstrated that subjects in the high fidelity training group performed the troubleshooting task with more errors than those receiving both low and high fidelity training. However, these subjects (trained on high fidelity equipment exclusively) fared better than those trained exclusively on low fidelity equipment.

These findings have implications regarding the amount of time required by the different treatment groups to solve the three troubleshooting test problems. The examination of fault identification procedures used by the subjects in this study could also be used to explain the effects of the varying levels of fidelity. Since,in the task they performed for this study, subjects were simply asked to IDENTIFY the faulted relay for each problem; they were not asked to FIX the faulted relay. Feedback was provided instantaneously to the subjects regarding the accuracy of the fault identifications they made. Upon correctly identifying the relay the experimenter knocked out for a particular problem, visual and auditory reinforcement were provided indicating a correct solution. This, in turn, led to the start of the next problem. Incorrect identification of a working relay as the one believed to be faulty did not terminate the problem. Instead, subjects were provided with feedback that indicated that their solution was incorrect and as the taped instructions stated earlier, they were to continue with their course of troubleshooting.

In an actual troubleshooting situation, incorrect fault identification would require time consuming testing. In an actual troubleshooting task, workers would be required to conduct some sort of an electronic test, or in some cases, fix the part they suspect to be faulted in order to determine if in fact, they had located the source of the problem. Either of these actions (e.g., to fix the part or conduct an electronic test) would incur some period of time. As a result, those workers making

the most erroneous fault identifications would also be likely to take more time to solve a troubleshooting problem. In the present experiment, however, there was no time penalty associated with each erroneous fault identification. If there had been such a penalty, it is likely that an increased number of errors would have been associated with an increase in transfer troubleshooting time. It is difficult to statistically impose a time penalty on to each error since the degree of an ad hoc time penalty would be arbitrary.

Preliminary, analysis of the troubleshooting strategies used by the subjects in each group indicate that, although no difference were found for the dependent measures differences do exist among the subjects in the study with regard to their use of a strategy for troubleshooting.

In conclusion, the authors believe that the sequential combination of low and high fidelity training devices merits further investigation both in the lab and the field. Further, it is believed that for individuals with roughly equivalent mechanical aptitudes, a sequential combination of low and high fidelity simulation training will result in fewer errors and, consequently could take less time on a transfer task than training programs which offer low or high fidelity training exclusively.

APPENDIX

Instructions:

Welcome to the training room.

Today, you will be trained to troubleshoot problems which occur in a piece of electromechanical equipment. As you may know, troubleshooting involves trying to find out why something is not working properly. For example, when your phone or television set breaks, the maintenance person who works on it tries to find the faulty part (e.g., to troubleshoot the problem). While you will not repair the equipment in today's experiment, your job will be to discover which of its components is malfunctioning, or causing the equipment to fail. In doing the task, you will develop a strategy that best helps you solve the problems with which you will be faced. Please do not be concerned if you have no previous experience with either troubleshooting or with electrical devices; none is necessary for the purpose of this experiment.

This experiment will be conducted in two phases. In the first phase, the training stage, you will work on either the actual, electromechanical equipment, or a 2-D replication/simulation of the equipment, or you will work on both the actual device and the 2-D simulator. In the second phase, following training, you will be asked to perform troubleshooting tasks on the actual equipment. I will now begin to explain the troubleshooting instructions. It is important that you understand both how the equipment operates and the task to be completed. As I go through the instructions with you, please feel free to ask any questions that come to mind. Also, you are welcome to examine the equipment and begin to acquaint yourself with it.

The electromechanical device you will be working on consists of a circuitry display which is connected to several output devices including a TV, a water pump, an air valve, a speaker, and several lights. Additionally, there is a set of five relay panels, each of which functions as a circuit and, like the large circuitry display, controls the output devices. If all the circuits were functioning properly, all of the output devices that you see here, would work when turned on. That is, the TV would display different bars of color, the water pump would circulate the water in the tank, the air valve would pass air through it, the speaker would make noise and the lights would go on. If, however, any of the relays in a particular circuit (either on the large display or in one of the five pull-up panels) is defective, it will result in at least one of the output devices failure to function properly. That is, the TV would display a random pattern (known as "noise") instead of the colorful bars, and/or the water pump would not circulate the water, and/or the air

valve, speaker, or lights would not go on. You should be aware that when a circuit is dead, it not only causes one or more of the output devices to fail to operate, but may also cause other relays to malfunction as well.

In order to test a particular output device, to see if it is working, you must depress the checker switch associated with that output device. All switches are clearly labeled as to which output device they will check. Having determined which of the output devices are not operating (at least one will fail to operate), your job will be to determine which one of the relays is causing the system's failure to operate properly. To do this, you will need to examine the interconnection between the various circuitry components and eventually try to isolate the one particular relay which is at fault. Please note that while more than one output device may fail to function, only one of the relays will be at fault (for each problem).

Each output device is at the end of a particular relay or circuit. You want to zero in on the malfunctioning relay that is preventing the output device or devices from working properly. To test the relays and see which one isn't working, use the hand-held testing device. Set the numeric thumbwheel to the number of the relay you want to test. Then, touch the attached probe to the relay terminal on the large circuitry display. This relay terminal is marked as a red circle on the relay you are testing. When the probe is touching the relay terminal, depress the "test" button located on the hand-held unit. If your suspicions about the relay you just tested (e.g., that it is malfunctioning) are correct, then no light will come onto the hand-held unit. If you were wrong, and the relay you tested is working as it is supposed to, then a green light will come on to the hand-held unit.

While most of the relays can be tested directly on the large circuitry display board using the hand-held unit, five of them cannot be tested directly. The five relays that cannot be tested directly are marked "do not test" on the circuit board. To test these relays, you simply raise the pull-up panel which corresponds to the suspected faulty relay using the attached handle. Beneath the uplifted panel there is a green indicator light. If that light is on, then the relay which corresponds to the panel is actually working. If, however, your suspicion that the relay was malfunctioning, was correct, then the green light will not be lit.

Once you feel that you have found the relay which is causing the output device or devices to fail, indicate your solution by setting the numeric thumbwheel on the hand-held device to the number of the relay you believe is broken. Then, press the "solution" button, release it and press the test button. If you have reached the correct solution, the red light will come on the test unit, signalling a defective relay. If,

References

1. Allen, J.A., Buffardi, L., Mandes, E., & Hays, R.T. (1984). The relationship of simulator fidelity to task and performance variables. (George Mason University Annual Report, Contract MDA 903-82-K-0464). Fairfax, VA.: George Mason University.

2. Andrews, D.H. (1983). The relationship between simulators, training devices and learning: A behavioral view. In International Conference on Simulators, September, pp. 70-75. New York: Institution of Electrical Engineers.

3. Brictson, C.A. & Burger, W.J. (1976). Transfer of training effectiveness: A7E Night Carrier Landing Training (NCLT) Device 2F103, NAVTRAEQUIPCEN Technical Report 74-C-0079-1. Orlando, Fla.: Naval Training Equipment Center.

4. Browning, R.F., Ryan, L.E., Scott, P.G., & Smode, A.F. (1977). Training effectiveness evaluation of device 2F87F, P-3C Operational Flight Trainer. Training Analysis and Evaluation Group Report No. 42, Orlando Training Analysis and Evaluation Group, U. S. Navy.

5. Cox, J.A., Wood, R.O., Boren, L.M., & Thorne, H.W. (1965). Functional and appearance fidelity of training devices for fixed-procedures tasks. Technical Report: 65-4, Alexandria, Va.: Human Resources Research Organization.

6. Fink, C.D. & Shriver, E.L. (1978). Simulators for maintenance training; some issues, problems and areas for future research. Technical Report: AFHRL-TR-78-27, Air Force Human Resources Laboratory, Technical Training Division, Lowry Air Force Base, Colorado.

7. Fitts, P.M. (1964). Perceptual-Motor Skill Learning. In Categories of Human Learning, (Ed.) A.W. Melton, London: Academic Press.

8. Gray, T.H. (1979). Boom Operator Part-Task Trainer: Test and evaluation of the transfer of training. AFHRL-TR-79-37, Williams AFB, AZ.: Flying Training Division, AFHRL.

9. Grimsley, D.L. (1969). Acquisition, Training, and Retraining: Training Category IV Personnel with Low Fidelity Dvices. HUMMRRO Technical Report 69-12, Monterey: Human Resources Office, Division 3.

10. Hays, R.T. (1980). Simulator Fidelity: A Concept Paper. Technical Report 490, Alexandria, VA.: U. S. Army Research Institute.

11. Kinkade, R.G. & Wheaton, G.R. (1980).Training device design. In H.P. Van Cort & R.G. Kinkade (Eds.), Human Engineering Guide to Equipment Design, U.S. Government Printing Office, Washington, D.C. 20402.

12. Martin, E.L. & Waag, W.L. (1978). Contributions of platform motion to simulator effectiveness: Study 1-Basic contact. Interim Report: AFHRL-TR-78-15, Brroks Air Force Base, Tex.: Air Force Human Resources Laboratory.

13. Morris, N.M. & Rouse, W.B. (1985). Review and evaluation of empirical research in troubleshooting. Norcross, GA.: Search Technology, Inc.

14. Prophet, W.W. & Boyd, H.A. (1970). Device task fidelity and transfer of training: Aircraft cockpit procedures training. Technical Report, 70-10, Alexandria, VA.: Human Resources Research Organization.

Residual Stress Measurements in Armament-System Components by Means of Neutron Diffraction(U)

*Henry J. Prask, Dr.+
Chang S. Choi, Dr.+
Energetics & Warheads Div., AED, ARDEC
Picatinny Arsenal, NJ 07806-5000

INTRODUCTION

Macro-residual stresses are the distribution of long-range stresses retained within a body after external stimuli have been removed. They arise, for example, from nonuniform heat treatment or plastic deformation and vary continuously in the body over regions which are very large compared to inter-atomic distances. With regard to performance or durability of hardware, the net stress on a body in service is the superposition of external forces and the residual stress system within the body. This is directly relevant to fatigue failure and fracture and, therefore, to the choice of fabrication methods by which the residual stresses are produced. Additionally, certain metals (including uranium and aluminum) are subject to stress-corrosion cracking (SCC) which is surface cracking produced by the interaction of a corrosive environment and tensile residual stresses at the surface. For most metals, the effect of detrimental residual stresses can be minimized by stress-relieving heat treatments. However, where desirable compressive stresses to reduce SCC are present at the surface this procedure is not feasible.

Several techniques are available for the nondestructive determination of residual stress: x-ray diffraction, magnetic methods, eddy current, ultrasonics. However of these, the best established for quantitative characterization is x-ray diffraction which is, except in very special cases, a surface probe. Although surface stresses are of prime concern in many armament applications, certain alloys present serious difficulties for the x-ray technique because of grain size, surface contamination, or texture (i.e. preferred grain orientation); among these are uranium alloys and in some cases aluminum alloys[1,2]

+Guest scientist at Reactor Radiation Division, National Bureau of Standards, Gaithersburg, MD 20899-0001.

Neutron diffraction closely parallels x-ray diffraction in methodology and analytical formalism. However, because neutrons interact with nuclei and x-rays with electrons, neutrons are typically about a thousand times more penetrating than x-rays in the wavelength range for diffraction ($0.7\text{Å} \leq \lambda \leq 4\text{Å}$). In addition, different elements exhibit significantly different relative scattering powers for neutrons and x-rays. A utilization of the unique aspects of conventional neutron diffraction for residual stress analysis was suggested and partially demonstrated by us previously[3]. However, the energy-dispersive neutron diffraction (EDND) technique, described in this paper and in references 4 and 5, minimizes texture and absorption effects - which limit conventional measurements - and enhances instrumental resolution.

In this work we describe the application of EDND to the characterization of sub-surface tri-axial stress distributions in depleted uranium - 0.75 wt% titanium samples ("U-0.75Ti") with different thermo-mechanical histories, and to two types of 7075-T6 aluminum-alloy ogives for the M483A1 155mm artillery shell.

NEUTRON METHOD

1. General Considerations

In both x-ray and neutron diffraction determination of residual stress, what is measured is strain which is manifested by changes in the distances between atomic planes, or d-spacings, in the sample. A unique advantage of neutron diffraction arises from the different relative scattering cross-sections and penetration relative to x-rays. This is illustrated in Table I in which $t_{1/2}$, the thickness at which half the beam intensity is lost through scattering and absorption processes, is listed for some metals of technological importance. The values are based on cross-sections from standard references and the difference in wavelengths used for neutrons and x-rays is not significant.

Table I. X-Ray/Neutron Comparison

Element (Atomic No.)	$t_{1/2}$(1.54Å X-rays)	$t_{1/2}$(1.08Å Neutrons)
Al (13)	0.0530 mm	71.0 mm
Ti (22)	0.0076	15.9
Fe (26)	0.0027	6.1
Cd (48)	0.0035	0.057
W (74)	0.0021	6.5
U^{238} (92)	0.0015	13.6

It should be mentioned that the $t_{1/2}$ values in Table I do not represent the depth of penetration in a residual stress measurement. This

is dependent on a number of factors such as source intensity, coherent scattering cross-section, and beam-spot size. However, the $t_{1/2}$ values clearly show that neutrons in the normal diffraction wavelength range are several orders-of-magnitude more penetrating than x-rays. Also, the penetration does not decrease monotonically with atomic number as with x-rays but is, essentially, a random function of atomic number. It is clear from the Table that both uranium and aluminum are especially good materials for neutron examination.

The properties of neutrons suggested the possibility of measuring sub-surface residual stresses in metallurgical samples employing tight collimation and the scattering geometry shown schematically in Figure 1. A scattering angle, Ω, of 90° is used to minimize the examined differential volume, ΔV. The perspective view in Figure 1 shows ΔV defined by two rectangular apertures in an absorbing material, e.g. cadmium; however, the apertures could be any shape including circular. The plan view in Figure 1 indicates how the sample can be translated in the beam so that ΔV can be examined as a function of depth.

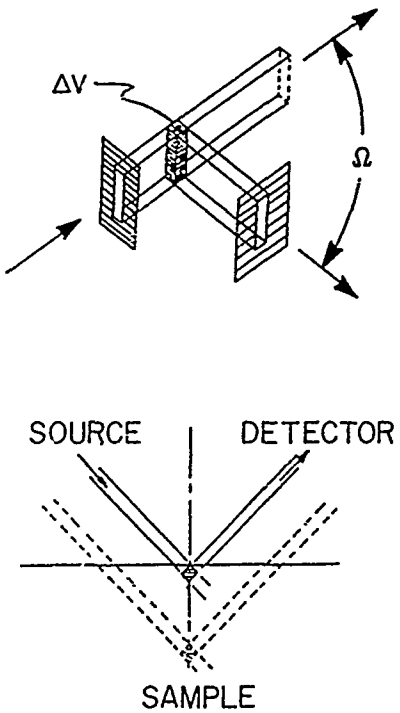


Figure 1. Schematic of how absorbing masks are used with Bragg's Law to define the examined volume, ΔV (upper); and plan view of ΔV translation through the sample (lower) in the EDND measurements.

The strains from which residual stresses are inferred are obtained from measured d-spacings through Bragg's Law: $\lambda = 2d(hkl)\sin(\Omega/2)$ where λ is wavelength, d(hkl) is the separation of atomic planes with Miller indices hkl, and Ω is stepped - usually with the sample orientation stepped by $\Omega/2$ - and sharp resonances in scattered intensity are observed at scattering angles where the Bragg condition is fulfilled. Precise

determination of $\Omega(hkl)$, the peak position, yields d(hkl) directly. Although success has been achieved with this mode of measurement, as reviewed in reference 4, serious problems have been encountered with highly attenuating or highly textured samples. Here, texture means the existence of preferred crystallographic orientation of grains or crystallites relative to a coordinate system fixed in the sample.

With reference to the plan view in Figure 1, it is clear that as Ω is varied, path length to and from ΔV changes and intensity as a function of Ω is possibly distorted leading to a false shift in d(hkl). Similarly, gradients in preferred grain orientation in the sample over the changing beam-in/beam-out paths can produce intensity variations which shift the apparent $\Omega(hkl)$. Since the strains, $(d-d_0)/d_0$ - where d_0 is the d-spacing at zero stress, are on the order of 0.0001, a small anomalous shift nullifies the stress measurement.

2. Energy-dispersive Neutron Diffraction

To avoid the difficulties described in the previous sub-section, we have developed and tested a mode of diffraction measurement which is "unconventional" for residual stress determination. In our measurements, we have made use of the fact that Bragg-condition resonances can also be observed at fixed Ω with varying wavelength. With the scattering angle fixed, changing attenuation and texture gradients are less distortive. In addition, the examined volume ΔV, remains exactly the same throughout each scan.

The instrument used for energy-dispersive neutron diffraction (EDND) is called a triple-axis spectrometer. Crystals of known d-spacing are placed before (monochromator) and after (analyzer) the sample; the Bragg relation is then used to select and step the wavelength incident on the sample. In principal, the analyzer crystal - which we step at the identical wavelength as the monochromator - is not needed. However, utilization of the analyzer significantly enhances instrumental resolution and produces peak profiles which are Gaussian in shape and straight-forward to analyze with least-squares curve-fitting techniques.

In our system, pyrolytic graphite crystals were used for monochromator (002 plane) and analyzer (004) with a pyrolytic graphite filter between monochromator and sample to reduce higher order wavelength contributions and background. The collimation employed was 50'-20'-27'-80' from source to analyzer, with a resultant resolution $\Delta\lambda/\lambda = 0.0073$ at $\lambda = 2.692$Å. The Cd absorbers were cut with 4mm x 4mm square apertures.

3. Stress-Strain Relations

The relation between stress and strain applicable to diffraction measurements has been presented, for example, by Evenschor and Hauk[6].

With reference to Figure 2, r, θ, and z are specimen-fixed axes, and the strain $\epsilon'_{\phi\psi}$ is measured along L'_3 ; then

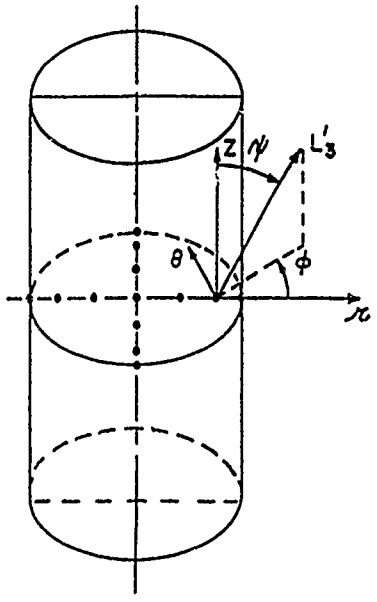


Figure 2. Coordinate system and measurement mesh for the 2.5 cm diameter by 10 cm long DU samples. The solid circles represent the centers of the 4x4x4mm^3 ΔVs examined.

$$\epsilon'_{\phi\psi} = (d_{\phi\psi} - d_0)/d_0$$
$$= (\epsilon_{11}\cos^2\phi + \epsilon_{12}\sin 2\phi + \epsilon_{22}\sin^2\phi)\sin^2\psi \qquad (1)$$
$$+ (\epsilon_{13}\cos\phi + \epsilon_{23}\sin\phi)\sin 2\psi + \epsilon_{33}\cos^2\psi$$

where $d_{\phi\psi}$ is the lattice spacing along $\vec{L}'_3$ and d_0 is the unstressed lattice spacing. The stresses are related to the measured strains through

$$\epsilon'_{\phi\psi} = 1/2\ S_2(hkl)[\sigma_{11}\cos^2\phi\sin^2\psi + \sigma_{22}\sin^2\phi\sin^2\psi$$
$$+ \sigma_{33}\cos^2\psi + \sigma_{12}\sin 2\phi\sin^2\psi + \sigma_{13}\cos\phi\sin 2\psi \qquad (2)$$
$$+ \sigma_{23}\sin\phi\sin 2\psi] + S_1(hkl)[\sigma_{11} + \sigma_{22} + \sigma_{33}].$$

The $S_i(hkl)$ are diffraction elastic constants ("XEC") for the (hkl) reflection which, in general, depend on the material and the reflection examined. For an elastically isotropic solid the XEC are given by

$$1/2S_2(hkl) = (1 + \mu)/E \text{ and } S_1(hkl) = -\mu/E \qquad (3)$$

where μ, E are Poisson's ratio and Young's modulus, respectively.

EXPERIMENTAL AND RESULTS

1. Uranium alloy

A. Samples: U-0.75Ti is an alloy of considerable importance in K.E. penetrator applications. In order to decrease production costs, fabrication-process changes have been proposed which can affect desired storage lifetime by increasing susceptibility to SCC. Fabrication processes - such as swaging - can improve mechanical properties through induced texture anisotropy. In both cases anisotropy in sub-surface stresses and the introduction of near-surface tensile stresses are of considerable concern. To date, the only (partial) determination of residual stress in uranium has been by the Sach's boring-out method[7].

The basic fabrication method includes gamma-phase solutionizing, quenching, aging, and cold work. For convenience the residual stress determinations were made on 10 cm long pieces cut from the mid-point of 46 cm long, 3.3 cm (starting) diameter rods. However, with slight modification of the instrument uncut rods could also be examined. One group was γ-phase solutionized in an AVS vacuum furnace, water-bath quenched, and rotary straightened ("mild" cold work). Of these, some were machined to the final 2.5 cm diameter, one was moderately cold worked and machined to the final 2.5 cm diameter, and one was severely cold worked to the final 2.5 cm diameter. In addition, one sample was solutionized in an induction furnance and water-spray quenched, aged, rotary straightened and machined to the final diameter. In all cases aging after quenching was 1-2 hours at 400°C or less. The sample types examined are listed in Table II.

Table II. U-0.75Ti Residual Stress Specimens

Sample	Stock	Heat Treat.	Cold Work
1	Rolled	Vac., BQ[a]	RS[b]
2	Rolled	Ind., SQ	RS
3	Extruded	Vac., BQ	RS
4	Rolled	Vac., BQ	RS, 7% CW
5	Rolled	Vac., BQ	RS, 31% CW

[a]All specimens solutionized at 850°C, aged at ≈400°C; BQ=water-bath quench, SQ=spray quench, Vac.=vacuum furnace, Ind.=induction furnace.
[b]RS=rotary straightened; CW=cold work.

B. Residual Stress: Since the determination of residual stress in technological samples by means of eqn. 2 depends directly on measurement

of strain values, a precise value for the unstressed d-spacing, d_0, is essential. We have been successful in determining stress-free lattice parameters from conventional diffraction measurements in some cases; however, cold working produces texture which makes lattice parameter determinations too uncertain relative to the precision required for residual stress chacterization. In addition, U-0.75Ti presents special problems because heat treatment to produce a stress free reference sample changes the composition. Therefore, we utilize the overall equilibrium conditions required by elasticity theory to determine d_0. That is, since the body is static with no external force applied, the residual stresses on any plane must balance so that in cylindrical geometry:

$$\int \sigma_{zz} r dr d\theta = 0 \qquad (4a)$$

$$\int \sigma_{\theta\theta} dr dz = 0 \qquad (4b)$$

and at any surface the stress orthogonal to that surface must vanish. Stresses inferred from measured strains can be adjusted, by adjusting d_0, to fulfill the equilibrium conditions.

One critical element in the above procedure is the determination of near-surface strains since stress gradients, especially for σ_{zz}, can be large near the cylindrical surface. In this work we have carefully measured d-spacings in a strain-free copper powder in cylindrical geometry identical to the U-0.75Ti cylinders. We find that measured strains, even for ΔV only half inside the surface (see Figure 2), agree within two standard deviations with the mean d-value. From this we conclude that no systematic instrumental errors limit the determination of d_0 using the overall equilibrium approach.

Utilizing the (112) reflection we have determined strains in the mid-point r-θ plane of the 10 cm long U-0.75Ti pieces. The points indicated in Figure 2 correspond to the centers of the $4x4x4mm^3$ beam spots examined. Determination of stresses was made using eqns. 2, 3 and 4a with a composition-weighted average for Poisson's Ratio (0.22) and the elastic modulus (170000 MPa) for the α and α' phases present[8]. We utilize σ_{zz} balance (eqn. 4a) to determine d_0 for strain evaluation because it is appropriate for the measurement geometry examined. An independent d_0 from $\sigma_{\theta\theta}$ equilibrium (eqn. 4b) is used as a check but is less precise because of edge effects at the ends of the cylinders. We find that this approach works very well for all the mildly cold-worked samples, 1-3, and the moderately cold-worked sample, 4, yielding self-consistent d_0 values, reasonable $\sigma_{\theta\theta}$ balance, and $\sigma_{rr} \approx 0$ at the surface. Application of this approach to the severely cold-worked, highly-textured sample 5 does not fulfill these requirements.

In the case of sample 5 we make use of a variation of the model of Noyan[9], in which a directional dependence of the diffraction elastic constants, $1/2S_2$ and S_1, is proposed for severely plastically deformed

metals. Texture measurements show a much higher degree of preferred orientation for sample 5 than for any of the others. Furthermore, the texture is fiber-like and monotonically increasing from the axial to the radial direction. This directionality is taken into account in the stress calculation by: 1) utilizing the d_0 obtained for samples 1-4, 2) adding an increment to all d(hkl) measured in the axial direction, and 3) adjusting this increment until σ_{zz} balance is obtained. The validity of this semi-empirical plastic deformation correction is confirmed by the fact that reasonable $\sigma_{\theta\theta}$ balance is achieved and $\sigma_{rr}\approx 0$ at the surface as with the other samples.

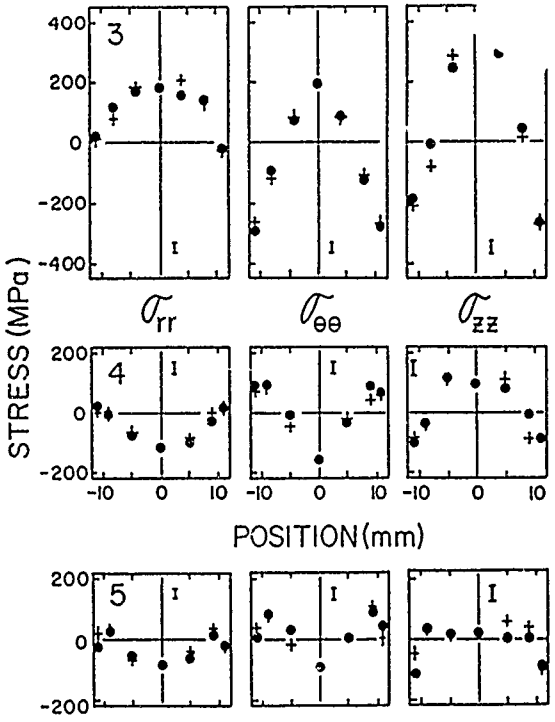


Figure 3. EDND determined residual stress distributions for mildly cold-worked U-0.75Ti specimen(s), 3, the moderately cold-worked specimen, 4, and the severely cold-worked specimen, 5. The crosses and solid circles correspond to measurements along the two orthogonal diameters shown in Figure 2. Representative standard deviations are indicated near the borders of each figure.

The EDND measurements, with the analysis method described above, provide the first determination of tri-axial residual stress distributions in U-0.75Ti material for penetrators. Results are shown in Figure 3, where the values for sample 3 are representative of the stresses determined for all three of the mildly cold-worked specimens, 1-3. Based on these results the following conclusions are drawn:

i. All of the specimens which underwent mild cold work (i.e. only rotary straightening) retain the residual stress distribution expected for a quenched-rod[7], with tensile stresses at the center and compressive hoop and axial stresses at the surface;

ii. Starting material and method of heat treatment play a very minor role in affecting the final residual stress distributions;

iii. Moderate cold work significantly alters the residual stress distribution of quenched rods; at the center the change in stress distribution corresponds, approximately, to the superposition of a 400MPa compressive stress on each stress component; at the surface, the stress distribution changes to maintain overall equilibrium; additional cold work does not produce further changes proportional to the degree of cold work;

iv. Moderate cold work can cause tensile residual stresses at the surface (cf. $\sigma_{\theta\theta}$, Fig. 3, sample 4); however, additional cold work appears to cause a possible recovery of compressive hoop stress at the surface.

2. 7075-T6 Ogives

A. General: In the past few years several ogive failures have occurred during ballistic acceptance testing of the 155mm M483A1 projectile. The purpose of this projectile is to deliver submunitions.

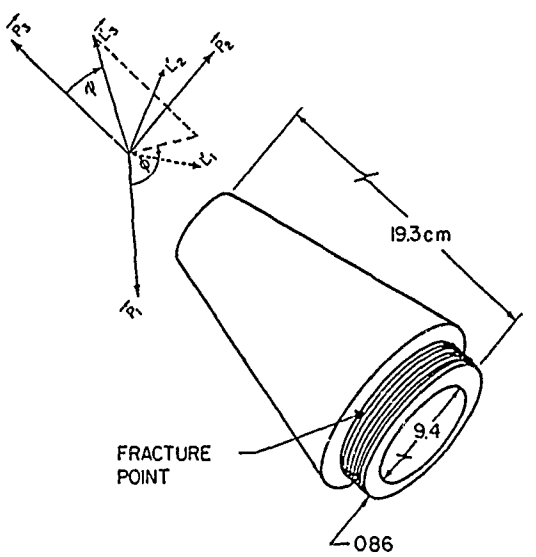


Figure 4. M483A1 ogive with forward end at upper left. The region of interest is under the most forward thread, as indicated.

It is constructed with an ogive at the forward end - that contains a small explosive charge - which is threaded to the projectile body and sealed from it by a pusher plate. The aft end is sealed by a shear plate. At detonation, the pressure developed in the ogive cavity creates an aft-directed force that acts on the pusher plate transmitting it through the submunitions to the aft-end shear plate. For effective operation, the force must be sufficient to fracture the shear plate and accelerate the

submunitions, thus scattering the submunitions behind the forward-traveling projectile body. The malfunctions occurred when the ogive body failed before the aft-end pusher plate.

The ogive is manufactured from 7075-T6 aluminum and is pictured in Figure 4. Production ogives are manufactured by two different suppliers who use somewhat different manufacturing methods. Manufacturer B produces the ogives by cold forging the cavity to finished dimensions, heat treating, and machining the outside dimensions. Manufacturer A produces a preform by forging at 332°-382°C, machines the cavity, heat treats, then finish machines the outside dimensions. Both suppliers use the same starting material. In full-scale tests the primary failure mode, exclusive to the B-type ogive, is a circumferential fracture at the first loaded thread. A single failure of an A-type ogive, by longitudinal fracture in the conical region, has also been reported.

Several material characterization studies and simulation tests have been conducted (summarized in reference 10). The materials studies showed some difference in the microstructure of A versus B ogives, but no substantial material property differences. The results of a simulation study, in which ogives were sealed and a charge exploded within, showed substantial differences in behavior. The B ogives failed at the thread at containment pressures substantially below that which is required to fail A ogives. Furthermore, these tests showed that the A ogives did not fail at the thread, but rather by ductile rupture with the crack running in the longitudinal direction. It was concluded that residual stress differences might be the cause of the performance difference between A and B ogives.

Prior to the present work three studies of residual stress in the ogives have been reported[10-12]. Two of these employed hole-drilling and one was an x-ray diffraction study utilizing layer removal, all destructive techniques. In the hole-drilling work of Kapp et al.[10], it was found that at the shoulder near the threads of interest A-type ogives exhibited compressive longitudinal stresses while B-type ogives exhibited tensile longitudinal stresses. In the x-ray diffraction study[11], three A- and three B-type ogives were examined. Considering the results through the thickness of the ogives under the threads, the report concludes that there did not appear to be any highly distinguishable differences in the results for the two types. The most recent study[12] examined four B-type ogives by the hole-drilling technique at positions on the shoulder near the threads, as in reference 10. In contrast to the work of Kapp et al., they found that the longitudinal stresses were in one case (of fourteen positions examined) slightly tensile and were otherwise compressive, ~7 to ~70 MPa.

B. Experimental: One ogive of type B and one ogive of type A were studied by EDND. In Figure 5, a view of the three 4x4x4mm^3 beam spot positions in the plane of the potential failure site is shown. An

identical set of measurements was made at 180° to the arbitrarily chosen 0° position. Partial measurements performed on a second type B ogive gave strain values very similar to the first type B ogive.

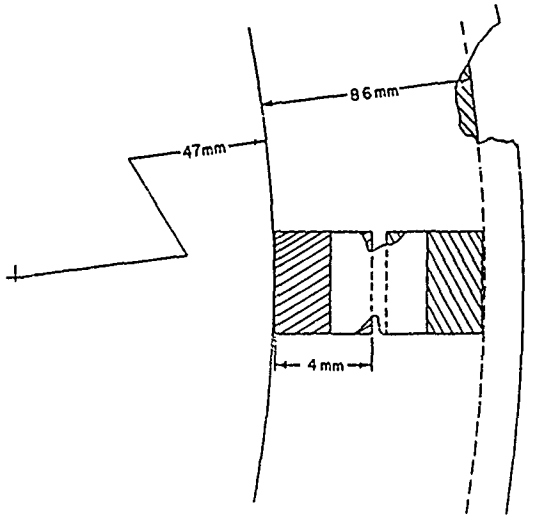


Figure 5. Partial, sectioned schematic of ogive at the failure plane as viewed parallel to the cylinder axis. Cross-sections of the 4x4x4mm³ beam spots at the measurement positions are shown.

As with U-0.75Ti, overall equilibrium (Eq. 4a) was used to determine $d_0(200)$. It was assumed that the average of the 0° and 180° data was representative of the full circumference. The axial direction equilibrium was then determined through the thickness (8.6 mm) by two averaging procedures: one in which the three measurement positions were given equal weight (except for the circumferential area element subtended), and a second in which the center-point contribution was subtracted out of the inner and outer points. The d_0s arrived at to balance σ_{zz} stresses were in excellent agreement for both types of averaging and for both types of ogives. The diffraction elastic constants used in the stress/strain calculations were the theoretical values of Bollenrath et al.[13]

C. Residual Stress: In Figure 6, final residual stress values are shown for both ogives for 0° and 180° positions at the first thread position. The nondestructive EDND stress determination indicates very significant differences in the two ogive types. In the A-type ogive it is found that just below the thread root ("OD") the σ_{zz} stresses are about -30 MPa (compressive) whereas the B-type shows +20 to +60 MPa (tensile) stresses. This is completely consistent with the principal failure mode differences that have been observed. In contrast, $\sigma_{\theta\theta}$ at the ID of A-type ogives are somewhat compresssive, while in the B-type $\sigma_{\theta\theta}$ stresses are highly compressive at this position. Qualitatively, this is consistent with longitudinal ductile rupture being the primary failure mode for A-type but not B-type ogives.

It is also of interest that the measured stresses are very nearly symmetric [$\sigma(0°)\approx\sigma(180°)$] in A-type ogives, and noticeably asymmetric in the B-type. In terms of failure it would seem that material strength in

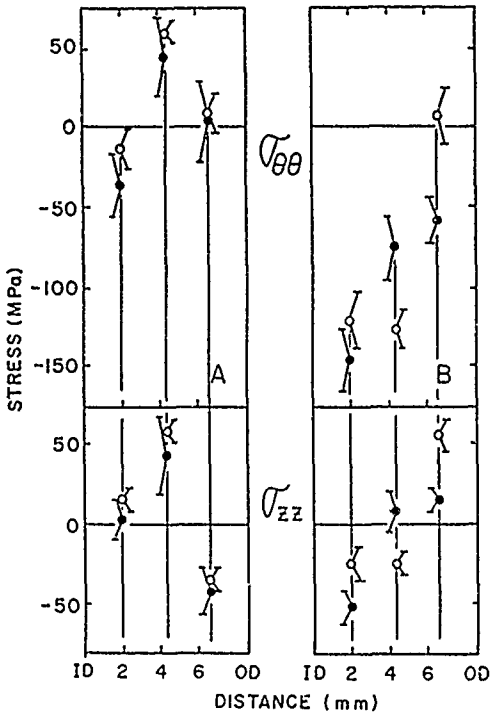


Figure 6. $\sigma_{\theta\theta}$ and σ_{zz} residual stress distributions of the A and B type ogives as determined by EDND. "OD" corresponds to the thread-root position. The open and closed circles correspond to measurements made 180° apart.

the axial direction would depend more on the maximum tensile stress at any point around the OD rather than the average σ_{zz}. If so, the σ_{zz} stress differential for failure is at least 90 MPa (13 ksi) for the two ogives examined. Work is continuing on additional finished ogives of each type, and on samples obtained at intermediate steps in the fabrication process.

CONCLUSIONS

In the present work we have shown that neutron diffraction techniques, in particular energy dispersive neutron diffraction, can provide information essential to failure analysis and improvement of performance of armament systems. We have developed methods by which zero-stress d-spacings can be obtained from macroscopic equilibrium conditions, including - in some cases - highly textured samples. Application of these methods has yielded the first determination of tri-axial residual stress distributions in aluminum-alloy and U-0.75Ti armament-system components. These measurements have shown a dramatic

effect due to finish-machining and cold-work on the near- and sub-surface residual stresses. Finally, because of the nondestructive nature of EDND, characterized items can be further processed and re-characterized to develop a complete understanding of fabrication steps and induced residual stresses, which can be followed by destructive, performance testing.

ACKNOWLEDGEMENTS

The authors express their appreciation to Mssrs. W. Sharpe, D. Robertson and M. Carlini of ARDEC, and G. Dudder of Battelle Pacific Northwest Labs for helpful discussions.

REFERENCES

1. C. P. Gazzara and R. E. Hinxman, III, "Measurement of X-ray Residual Stress in Depleted Uranium," Final Rept., AMMRC, 1984.

2. F. Wells, "Acoustoelastic Birefringence Measurements in Rolled Aluminum Plate Products," Proc. 34th Def. Conf. on NDT, Oct. 1985, pp. 222-236.

3. C. S. Choi, H. J. Prask, S. F. Trevino, H. A. Alperin, and C. Bechtold, "Application of Neutron Diffraction to Nondestructive Testing Problems," NBS Tech. Note 995, 1979, pp. 34-39.

4. H. J. Prask and C. S. Choi, "NDE of Residual Stress in Uranium by Means of Neutron Diffraction," J. Nucl. Matls., 126, 1984, 124-131.

5. H. J. Prask and C. S. Choi, "Residual Stress Measurements in Armament-System Components," in Residual Stress in Design, Process and Materials Selection (ed. W. B. Young), ASM International 1987, pp. 21-26.

6. P. D. Evenshor and V. Hauk, Z. Metallkde. 66, 1975, 167-8.

7. F. A. Simonen et al., "Residual Stresses in Depleted Uranium Bars," Battelle PNW Lab. Rept. AM&S No. 5, Sept. 1982.

8. G. Dudder, private communication; G. H. Llewellyn et al., "Computer Simulation of Immersion Quenching of Uranium-0.75wt% Titanium Alloy Cylinders," Oak Ridge Y-12 Plant Rept. Y-2355, 1986.

9. I. C. Noyan, "Determination of the Elastic Constants of Inhomogeneous Materials with X-ray Diffraction," Matls. Sci. Eng., 75, 1985, 95-103.

10. J. A. Kapp, R. J. Fujczak, and R. T. Abbott, "An Evaluation of the Service Failure of Aluminum Nose Cones Using Four Test Techniques", ARDEC Tech. Rept. ARCCB-TR-87006, 1987 (based on work completed in 1985).

11. R. N. Bolin and P. S. Prevey, "X-Ray Diffraction Determination of the Principal Residual Stress Distributions in Six M483 AA 7075-T6 Ogives P/N 10542916," Lambda Res. Inc. Rept. 117-004, Dec. 1984.

12. L. Natalini and W. Rider, "Residual Stress Measurements of M483A1 Ogives by the Blind Hole Drilling Technique," ARDEC Tech. Rept., in press.

13. F. Bollenrath, V. Hauk, and E. Muller, "Zur Berechnung der vielkristallinen Elastizitatskonstanten aus den Werten der Einkristalle," Z. Metallkde. 58, 76-82 (1967).

Weapons Design and the Inner Ear: Critical Insights from Mathematical and Physiological Models (U)

*G. Richard Price, Dr.
Joel T. Kalb, Dr.
U.S. Army Human Engineering Laboratory
Aberdeen Proving Ground, Maryland 21005-5001

THE IMPULSE NOISE PROBLEM

In the quest for superior weapons performance, acoustic impulses have reached such high levels that weapons design and operational use are now constrained by human tolerance for impulse noise [1]. Furthermore, the noise problem is not limited to a single weapon or system. With continuing emphasis on light weight, increased muzzle velocity and greater range, it is virtually certain that the operator will be exposed to higher sound pressures and impulse noise will continue to be a critical issue for the Army in the years ahead.

The impulse noise problem is not new, its roots probably dating back to the invention of gunpowder. From a systems engineering standpoint, the most desirable alternative would be to design the noise problem out of the system. There are a number of damage-risk or design criteria in use in the world that address this issue (China, France, Germany, Great Britain, the Netherlands and the United States all have their own criteria [2-8]); however, there are serious problems with them. First, the latest data on hazard indicate that they all seriously overrate the hazard from low frequency impulses [9]. Secondly, none of the rating systems is theoretically based, which means that they have limited utility in suggesting design alternatives. The situation is critical because we know that the existing methods are inadequate and we need a new approach to rating impulse noise hazard.

Given the physics of rapid combustion and the explosive propulsion of supersonic projectiles, it follows that major reductions in noise at the source are neither possible nor feasible. The problem therefore becomes one of accommodating the characteristics of the human ear, which in turn depends on discovering the mechanisms by which intense impulses damage the inner ear. The payoff from this exercise includes both

reduction in liability for hearing disability but even more importantly, maintenance of operational effectiveness in the field [10].

In support of these goals, a basic research program was begun at the Human Engineering Laboratory about 10 years ago aimed at identifying the mechanisms by which the ear is damaged. It led to the hypothesis that damage due to intense sounds is a function of mechanical stress within the structures of the inner ear [11]. Furthermore, the acoustic properties of the external and middle ears were seen as playing a crucial role by shaping and limiting the energy that is transmitted to the inner ear. Impulses with their peak energies in the mid-range (about 3.0 kHz) were predicted to be much worse for the ear than impulses with their peak energies at lower or higher frequencies [12]. The same logic also argued that because of the ear's tuning, losses should be seen in the mid-range, regardless of where the spectral peak had been in the impulse in air. To test these contentions, experiments with real ears and exposures to primer, rifle and howitzer impulses were conducted at HEL [13-15]. The data were consistent with both the predicted greater susceptibility to impulses with their peak energies in the mid-range and with the prediction of greatest loss in the mid-range, regardless of the spectral peak in air.

Conflict with existing DRCs

The critical importance of these experiments for the Army becomes apparent when the data are compared to the predictions of the damage-risk criteria (DRC) in use. The situation is portrayed graphically in Fig. 1,

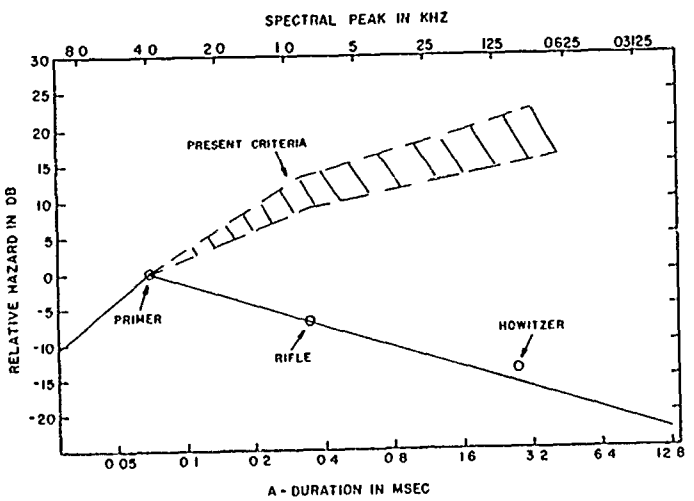


Fig. 1. Relative hazard of primer, rifle and howitzer impulses at different spectral locations as determined experimentally (data points) and as predicted by current damage-risk criteria.

where we can see that the relative hazard, shown by the solid line, declines as the spectral peak moves to lower frequencies, i.e. the impulses are from larger and larger sources (primer, rifle, howitzer). Moreover, the criteria in use in the world differ in the specific way they rate hazard; but they rank the relative hazard of these same impulses in the shaded zone of the figure. The divergence of the curves indicates that all the damage risk criteria are in fundamental conflict with the experimental data. Additional experiments from other laboratories have tended to support the contention that low frequency impulses, even though they may have more energy in them, are much less hazardous than traditionally expected[1,16]. Furthermore, no acoustic measures commonly used in other settings reproduce the appropriate ranking of hazard [17].

MATHEMATICAL MODELING

If mechanical stress in the inner ear is indeed the fundamental mechanism responsible for hearing loss at high intensities, as hypothesized [18-19]; then that is where inquiry should be focused. However, measuring mechanical stresses within the inner ear is a technical feat that is not presently possible. The living structures to be observed are micron-sized, have low visual contrast, are difficult to get to, and are extremely susceptible to any experimental manipulation that disturbs the physiological milieu. On the other hand, a mathematical model of the ear could serve as a heuristic tool to provide physical insight, suggest critical experiments and lead to the application of the findings to solving or ameliorating the impulse noise problem.

Approaches and Goals

Because of space limitations we will try to present only enough technical detail to make the implications of the model intelligible and, for the intellectually curious, note that we intend to publish a fuller account in an appropriate publication. Several general considerations guided the modeling. First, we wanted to be able to relate the various elements of a model to specific physiological entities and to give them physically realistic values, to allow for maximum physical insight. Second, to insure maximum utility, the model should be calculable with modest computer resources. Third, the main focus of the model was on the behavior of the ear at high sound intensities where loss was thought to be a function of mechanical stress. Lastly, the model had the goal of beginning with a pressure history in the free field and passing the energy through the outer, middle and inner ears to provide an estimate of stress within the inner ear. These last goals required that models of

the head, outer ear, middle ear, and inner ear all be integrated and allowances made for non-linear elements in them.

The Models

The details of the models presented will be appropriate for the cat ear. This choice was made because the physiological and acoustic values were best known for the cat and also because the hearing loss data to be explained were produced with the cat ear. The ears of other mammals are highly similar in structure and the principles used in the modeling should transfer with only modest adaptations to fit their anatomic details. A schematic diagram of the human ear with the anatomical elements labeled is provided in Fig. 2 as an aid in following the development of the models.

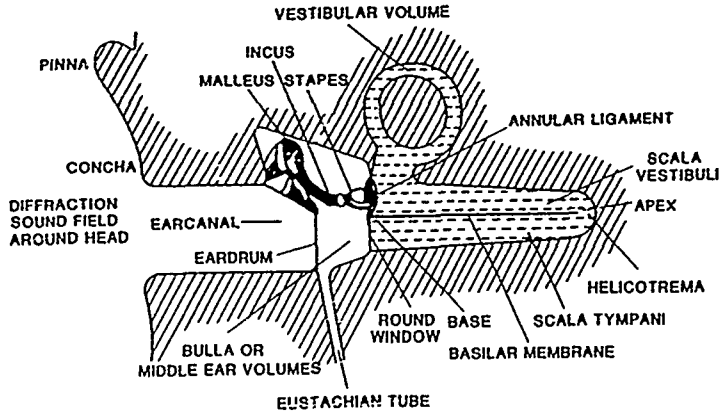


Fig. 2. Schematic diagram of the ear.

The ear has been modeled with electro-acoustic elements. Mathematically, it has been developed in two alternate forms, one in the time domain and one in the frequency domain [20]. The time domain model can be characterized as equations of motion formulated as a system of coupled ordinary differential equations that allow both linear and non-linear calculations in the time domain. The frequency domain version can of course only be applied to linear processes. Where the non-linearity of the stapes was treated (discussed later), we performed a non-linear transformation of the data and then proceeded with the remaining elements of the model. Both models share the same set of variables and values so that their performance is comparable. The two models are complimentary in that they each are structured to allow efficient computation of particular parameters. The circuit diagram of the external and middle ear portion of the models appears in Fig. 3.

From The Free-Field to the Cochlea

The head is modeled as a spherical baffle around which the sound wave diffracts and in which the ear canal is located. The acoustic field up to the ear canal entrance is approximated by a simple network proposed by Bauer [21] and Wiener, Pfeiffer and Backus' [22] two tube model of the ear canal follows it. The middle ear elements are based on a model by Zwislocki [23] with additional impedance elements given by Lynch,

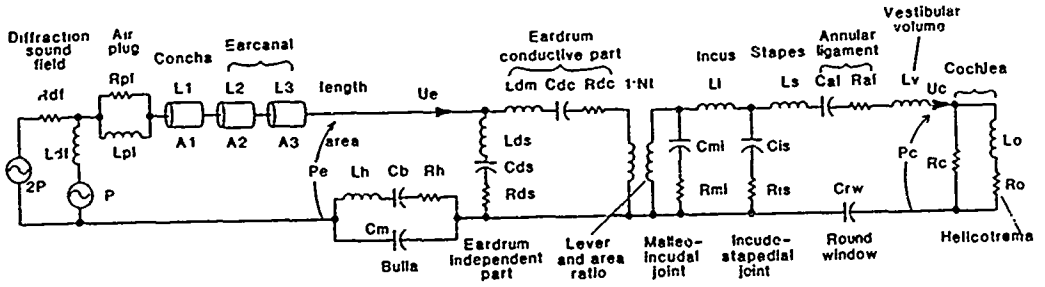


Fig. 3. Circuit diagram of the model.

Nedzelnitzky and Peake [24]. These elements account for the fact that the ear drum does not move as a unit, for losses in the ligaments joining the ossicles, for the areal ratios of the ear drum and stapes and the lever ratios of the ossicles, and for the resonances of the volumes included in the middle ear.

We can gain confidence in the performance of the integrated model by comparing values calculated from the model with acoustical quantities that have in fact been measured. If for the moment we allow the cochlea and its loads to be represented as linear mass and resistance elements, then this comparison can be seen in the four panels of Fig. 4. The agreement between the models and the data is good except for frequencies above 10.0 kHz in Panel A and above 5.0 kHz in Panel C. We believe the lack of agreement in these areas is due primarily to measurement difficulties inherent at high frequencies (noted by the original authors) and that in any event there is nothing critical happening at these frequencies where hearing hazard is concerned. In general, we believe that the integrated model does a creditable job of reproducing the quantities that can in fact be measured.

The Cochlea

The cochlea is modeled as a two-chambered, fluid-filled box with rigid side walls. The partition between the chambers is rigid, except for a tapered basilar membrane (BM) which becomes progressively more compliant toward the apex (mechanical stiffness, mass and viscous losses

changing exponentially). In addition, the cross sectional area of the cochlea decreases exponentially from base to apex. In the differential equation model, the cochlea is discretized into 512 coupled resonant elements whose characteristics change with location. The values chosen for the mechanical properties produce the envelopes of traveling wave displacements that show a changing Q and an increase in amplitude with increasing frequency as seen by von Bekesy [28]. We believe that these envelopes are characteristic of the BM when driven at high amplitudes

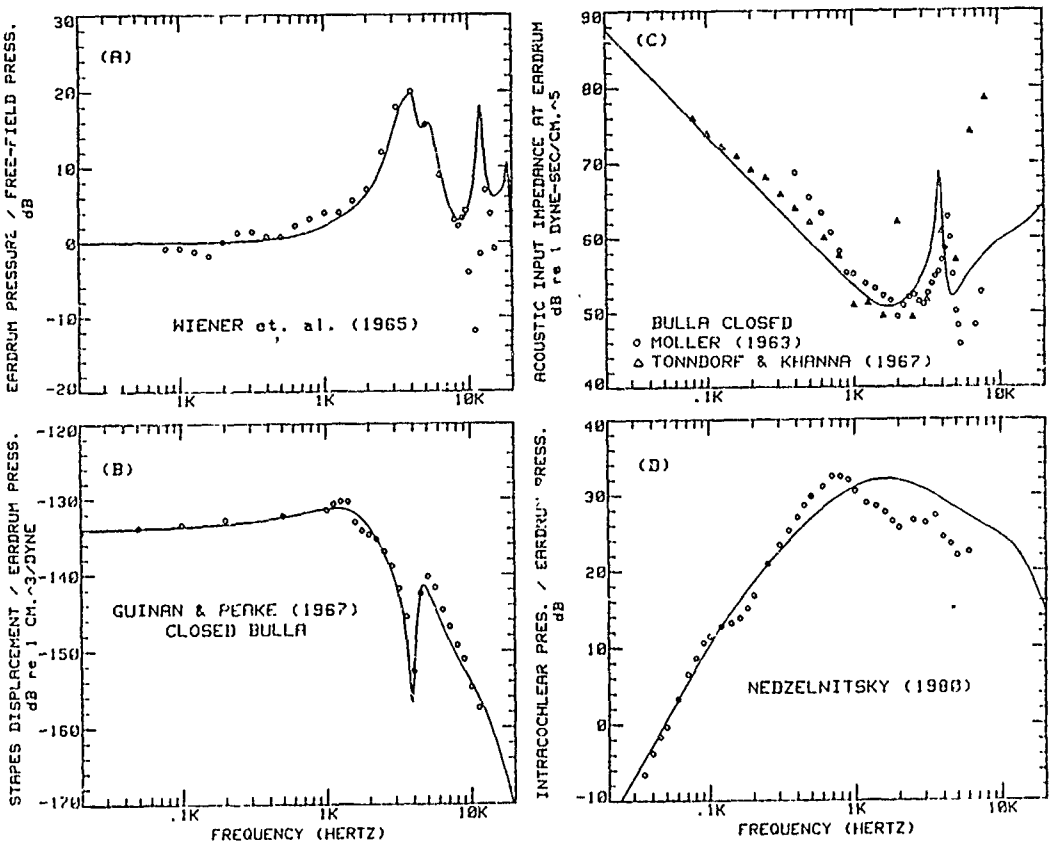


Fig. 4. Comparison of calculated values (solid lines) with measurements of four acoustical quantities (data points). Data in Panel A from Wiener, Pfeiffer and Backus (1965), in Panel B from Guinan and Peake (1967), in Panel C from Tonndorf and Khanna (1967) and Moller (1963) and in Panel D from Nedzelnitzky (1980).

where the sharp tuning produced by active processes that operate at low intensities cannot have an appreciable effect and are the same as those

shapes actually seen in studies that have measured intracochlear motion [29]

The mathematical models can be used to calculate almost any quantity - pressures, displacements, accelerations, stresses, etc. For the purpose of understanding the mechanisms which produce loss, we have focused first on displacements of the stapes and BM.

The non-linear stapes

The stapes is commonly modeled as a linear element. However, at very high intensities the middle ear becomes observably nonlinear [30-31] We hypothesize that the primary reason for the non-linearity is the annular ligament, which holds the stapes in place at the entrance to the cochlea and because of its mechanical structure can exert a controlling influence on cochlear input. Furthermore, this non-linearity has major implications for impulse noise hazard because it exerts a frequency dependent limiting action which controls the energy actually transmitted to the inner ear where it can do damage [32-33]

The annular ligament is modeled (in the time domain) as a visco-elastic element which becomes increasingly 'lossy' as the stapes displacement increases. In the frequency domain model, a non-linear transformation is performed. In Fig. 5 the model's output is compared with actual measurements of middle ear displacement made by Guinan and Peake [31] in the anesthetized cat. It can be seen in this figure that the

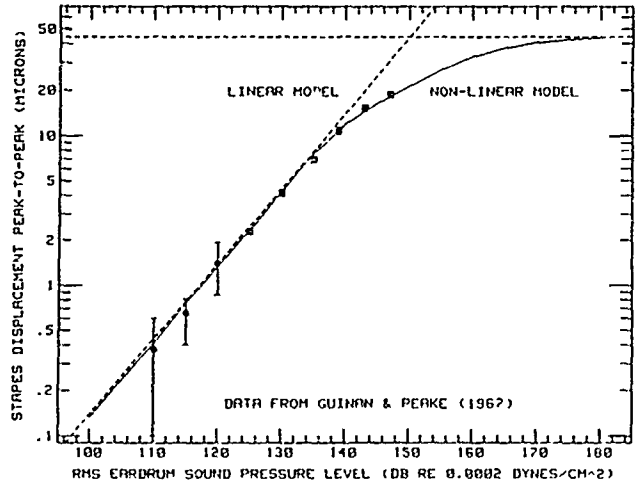


Fig. 5. Comparison of non-linear model output for stapes displacement with measurements by Guinan and Peake (1967).

model fits the available data and limits maximum displacement to about 20 microns in either direction. The time domain model suggests that the amount of loss in the ligament has a major effect on the calculated displacements when the ear is driven by very intense stimuli. With no loss, the middle ear 'rings' and the insertion of loss tends to suppress the ringing. The amount of loss used in the model is just enough to suppress the ringing. This observation suggests that critical data might be obtained from experiments measuring stapes displacements and/or intracochlear pressures at high sound pressures.

APPLICATION OF THE MODEL

The influence of the conductive properties of the middle ear on the impulse is immense. The effect of the peak-limiting property can be appreciated by considering the displacement that would have taken place, had the middle ear been linear. In response to impulses in the crew areas of large caliber weapons, calculations indicate that the stapes would try to move in the range of 1000 to perhaps 6000 microns. With peak limiting present (as is the case in the real ear) the displacements are limited to less than 50 microns, a reduction of one to two orders of magnitude!

The effect of the middle and external ears on the energy that enters the cochlea is portrayed in Fig. 6. In this somewhat complicated figure the energy in 1/3 octave bands entering the cochlea has been calculated for Friedlander waveforms with A-durations characteristic of a rifle and a howitzer at three intensities each. These waveforms are close

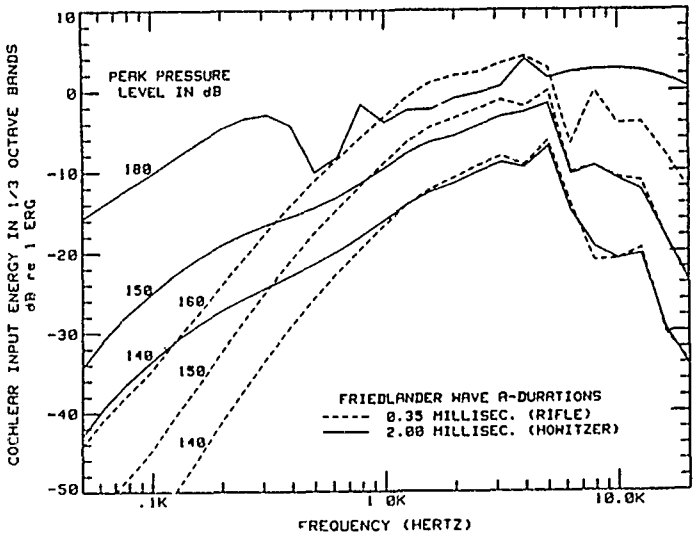


Fig. 6. Calculated 1/3 octave energy entering the cochlea for Friedlander waveforms simulating rifle and howitzer impulses at 3 intensities.

simulations of weapons impulses; but do not have the 'hash' on them associated with actual measures. Two major points can be made from this calculation. First, even though the spectral peaks of these impulses in air would have been at 140 and 800 Hz respectively, at the input to the cochlea they all have their greatest energy at about 3000 Hz. The ear acts like a fairly sharply tuned band-pass filter; therefore all impulses naturally tend to have their greatest effects in the ear's mid-range. Secondly, note that with increases in intensity, the energy increases non-linearly for both impulses. However, because of the peak limiting, the howitzer type impulse at 180 dB peak pressure actually has less energy in the mid-range than the 160 dB rifle type impulse. This process explains much of why the low frequency impulses are less hazardous than previously thought.

The external and middle ears are themselves relatively resistant to damage, the primary site for hearing loss being structures on the BM. Through use of the time domain model we can calculate the displacements of all the points along the BM when the ear is exposed to a stimulus. In the next figures we will show the envelope of the displacements in the upward direction. Such displacements correlate best with the damage[20] and have the additional advantage of making physiological sense as possible damage producers. Briefly, the latest conception of cochlear function associates upward displacement with forces that stretch critical parts of the organ of Corti. Given that tissue tends to fail in tension, upward displacements could well be especially hazardous. Conversely, compressive movements (downward displacements), would be expected to be less damaging.

How does the model's performance match data from real ears exposed to actual weapons impulses? This comparison can be seen in Figs. 7 through 10. In Fig. 7 we see the response of the ear to the 160 dB howitzer impulse. Panel A is the pressure history of the impulse, Panel B is the calculated stapes displacement, and Panel C is the calculated BM displacement. Fig. 8 is the same type of data, except for the 155 dB rifle. Notice that the pressure waveform and the stapes displacements for the howitzer are both much larger than they are for the rifle, as one might expect. However, the BM displacements show the counter-intuitive result that the 155 dB rifle produces <u>more</u> displacement than the 160 dB howitzer. The hearing loss data match the model's assessment. The experiments showed that the howitzer and rifle are equally hazardous when the rifle is about 9 dB less intense than the howitzer [13].

In the Fig. 9 we see the equivalent plots for the 166 dB howitzer, with an additional twist. In this presentation we are able to see which

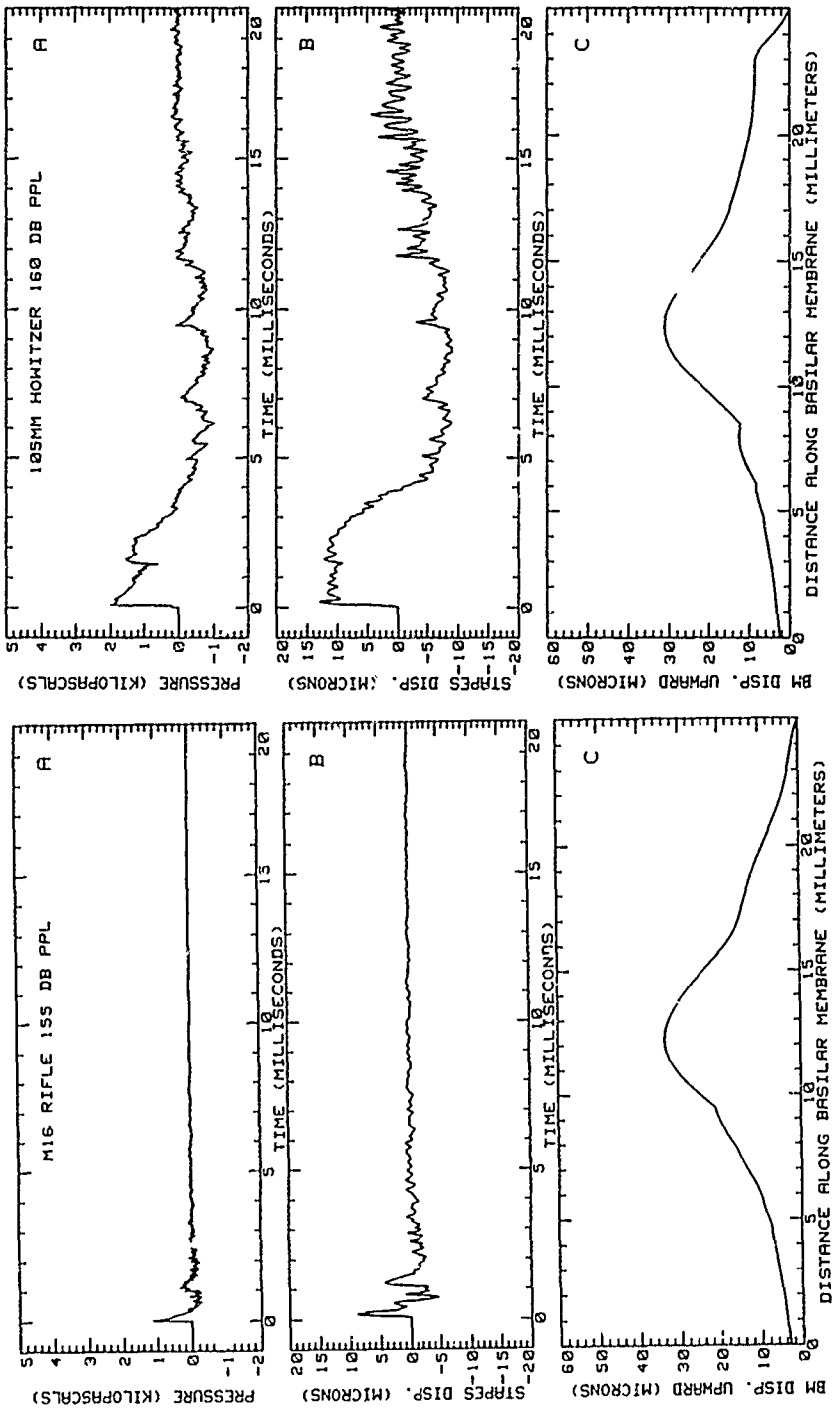


Fig. 7. Calculated stapes and basilar membrane displacements for the 160 dB howitzer impulse.

Fig. 8. Calculated stapes and basilar membrane displacements for the 155 dB rifle impulse.

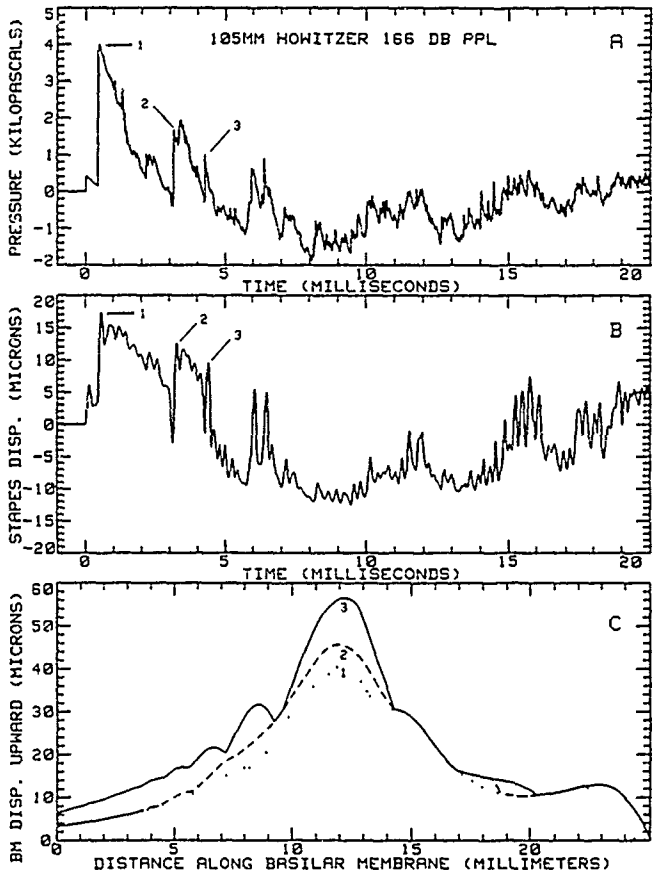


Fig. 9. Calculated stapes and basilar membrane displacements for the 166 dB howitzer impulse.

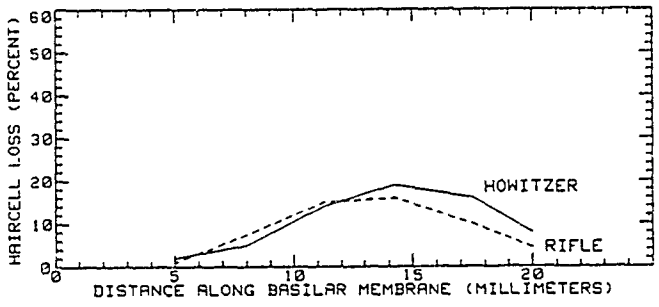


Fig. 10. Hair cell loss as a function of cochlear location following exposure to howitzer and rifle impulses.

part of a complex wave form is really the most damaging, with a surprising result. Panels A and B are the same as before; but panel C contains three curves. Each numbered curve in Panel C is the displacement pattern produced by the correspondingly numbered part of the wave forms in the upper two panels. The primary peak (1), which has attracted all the attention from noise measurement experts in the past, produced the lowest envelope of BM displacement. The ground reflection, peak number 2, produced the intermediate envelope. Surprisingly, the apparently insignificant peak numbered 3 produced the largest envelope of displacements, and presumably such peaks were responsible for the large losses in hearing sensitivity that this exposure produced [17]

Based on this type of analysis several generalizations can be made. First, high pressures with extended durations (characteristic of low frequencies) tend to push the stapes into clipping and result in greatly attenuated transmission during peak pressures (positive or negative). Second, the greatest driving of the ear is possible when pressures cross ambient. If the stapes starts from a rest position, the largest displacement it can manage is about 20 microns; however, if the stapes has been driven to an extreme position, it can move twice as far. Therefore, the real hazard to the ear often comes not with the initial peak but later in the waveform where the multiple zero crossings are occurring. Third, the rate of movement of the stapes is also critical because it corresponds to the frequency content of the movement. Lastly, the ear appears to be much more susceptible to movements in a particular direction - upward movement of the BM, which generally corresponds to outward movement of the stapes and negative pressures in air. In years gone by, we have almost exclusively focused on peak positive pressures and now it would appear that we've been facing in the wrong direction.

The specific shapes of the displacement envelopes in Fig. 7-9 are particularly interesting. No matter which weapon produced the impulse, the peak of the displacement envelope is in the middle of the cochlea, which is consistent with the prediction of maximum hazard to hearing in the mid-range (Fig. 6). The verification of this prediction can be seen in the data in Fig. 10. The data points are the percent hair cell loss (permanent inner ear damage) at different locations within the cochlea resulting from exposure to rifle and howitzer impulses [34]. The parallelism between the predicted displacements and the actual cellular losses is obvious. The model calculation even reproduces the slight rightward shift of peak cell loss for the howitzer exposure.

The foregoing discussion has shown briefly that the mathematical models are reproducing the essential elements of hearing hazard and are doing it in a theoretically consistent manner. At this time, no other method of rating of rating hazard comes close to being accurate. We are

very much encouraged that these models are on the right track and will prove to be sound basis for a new design criterion for impulse producing weapons.

Furthermore, the models suggest new avenues for reducing hazard from weapons impulses without performance penalties. Calculations with the models indicate that great benefit could be obtained if we could attenuate energy in the mid-range. This could be done with small barriers that would produce a sound shadow (worn by the soldier or mounted on the weapon) or perhaps with modest sized devices on the muzzle. Also, with the model, for the first time, we have a way for calculating the effectiveness of hearing protectors with impulsive sounds. It would appear that attenuation in the low frequency region is not particularly critical and that protectors need to be good in the mid-range. It may even be possible to simultaneously maintain good speech communication and adequate hearing protection.

In conclusion, we are very much encouraged to see the models accurately predicting the ranking of hearing hazard as well as its location and doing it in a theoretically consistent manner. There is presently no other method for doing this. Furthermore, the models are formulated so that they have heuristic value, suggesting design alternatives, hearing protector designs, etc. We are now working to refine the predictive capacity to include a greater range of impulses and conditions and we plan to continue to validate the predictions with experiments with real ears. Barring the unexpected, we believe that these models have the potential to serve as the basis for a both a weapons design standard and a new damage-risk criterion for impulse noise.

REFERENCES

1. Patterson, J. H., Mozo, B. T., Marrow, R. H., McConnel, R. W., Lomba-Gautier, I., Curd, D. L., Phillips, Y., and Henderson, R. (1985). "Direct Determination of the Adequacy of Hearing Protective Devices for Use with the M198, 155mm Towed Howitzer", USAARL Report No. 85-14, U.S. Army Aeromedical Research Laboratory, Fort Rucker, AL.
2. CHABA (1968). "Proposed Damage-Risk Criterion for Impulse Noise (Gunfire)," Report of Working Group 57, NAS-NRC committee on Hearing, Bioacoustics and Biomechanics, Washington, D.C.
3. Cheng, M., Liang, Z., Meng, Z, and Li, X. (1987). "Investigation of Military Standard for Impulse Noise", Proc. Inter-Noise 87, 2, 913-916.
4. MIL-STD 1474(B) (1979). "Noise Limits for Army Materiel," U.S. Army Missile Command, Redstone Arsenal, AL.

5. Ministry of Defense (1982). "Acceptable Limits for Exposure to Impulse Noise from Military Weapons, Explosives and Pyrotechnics," Interim Def Stan 00-27/1, Ministry of Defense, Directorate of Standardization, First Avenue House, London, WC1V6HE, England.
6. Ministry of Defense (1982). "Recommendation on Evaluating the Possible Harmful Effects of Noise on Hearing," Technical Coordination Group "Human Factors and Ergonomics", Direction Technique des Armaments Terrestres, 9211 Saint-Cloud Cedex, France.
7. Pfander, F. (1975). Das Knalltrauma. New York: Springer-Verlag.
8. Smoorenburg, G. F. (1980). "Damage Risk Criteria for Impulse Noise," Rept. 1980-26, Institute for Perception TNO, Soesterburg, The Netherlands.
9. Price, G. R. (1987). "The need for a new DRC for impulse noise," In: Proceedings of the Third Conference on Weapon Launch Noise Blast Overpressure, U.S.Army Ballistics Research Lab, Aberdeen Proving Ground, MD 21005-5001.
10. Price, G. R., Kalb, J. T., and Garinther, G. R. (1988). "Toward a Measure of Auditory Handicap in the Army", Ann. Otol. Rhinol. Laryngol. In press.
11. Price, G. R. (1977). "Toward a Theoretically Based DRC for Impulse Noise," J. Acoust. Soc. Am. 62, S95.
12. Price, G. R. (1981). "Implications of a Critical Level in the Ear for Assessment of Noise Hazard at High Intensities," J. Acoust. Soc. Am. 69, 171-177.
13. Price, G. R. (1986). "Hazard from Intense, Low Spectral Frequency Impulses," J. Acoust. Soc. Am. 80, 1076-1086.
14. Price, G. R. (1983). "Relative Hazard of Weapons Impulses," J. Acoust. Soc. Am. 73, 556-566.
15. Price, G. R. and Wansack, S. (1985). "A Test of Predicted Maximum Susceptibility to Impulse Noise," J. Acoust. Soc. Am. 77, S82.
16. Dancer, A., Lenoir, M., Buck, K, and Vassout, P. (1983). "Etude de l'influence du niveau de crete et la duree de bruits impulsionnels, du type bruit d'arme produit en champ libre, sur l'audition du cobaye," Institut Franco-Allemand de Recherches de Saint-Louis, FR, Raport 127/83, 70 pp.
17. Price, G. R. (1986). "Impulse Noise Hazard as a Function of Level and Spectral Distribution," In Basic and Applied Aspects of Noise-Induced Hearing Loss, R.J. Salvi, D. Henderson, R. P. Hamernik, and V. Coletti, Eds., pp 379-392, Plenum, N.Y.
18. Price, G. R. (1983). "Mechanisms of Loss for Intense Sound Exposures," In Perspectives in Modern Auditory Research: A Conference in Honor of Dr. E. G. Wever, edited by R.R. Fay, pp 335-346 (Amphora, Groton, Conn.).
19. Price, G. R. (1979). "Loss of Auditory Sensitivity Following Exposure to Spectrally Narrow Impulses", J. Acoust. Soc. Am. 66, 456-465.

20. Kalb, J. T. and Price, G. R. (1987). "Mathematical model of the ear's response to weapons impulses," In: Proceedings of the Third Conference on Weapon Launch Noise Blast Overpressure. U.S.Army Ballistics Research Lab, Aberdeen Proving Ground, MD 21005-5001.
21. Bauer, B. B. (1967). "On the Equivalent Circuit of a Plane Wave Confronting an Acoustical Device", J. Acoust. Soc. Am. 42, 1095-1097.
22. Wiener, F. M., Pfeiffer, R. R. and Backus, A. S. N. (1965). "On the Sound Pressure Transformation by the Head and Auditory Meatus of the Cat", Acta Otolaryngol. 61, 255-269.
23. Zwislocki, J. J. (1962). "Analysis of Middle-Ear Function. Part I: Input Impedance", J. Acoust. Soc. Am. 34, 1514-1523.
24. Lynch, T. J., Nedzelnitzky, V. and Peake, W. T. (1982). "Input Impedance of the Cochlea in Cat", J. Acoust. Soc. Am. 72, 108-130.
25. Moller, A. (1963). "Transfer Function of the Middle Ear", J. Acoust. Soc. Am. 35, 1526-1534.
26. Tonndorf, J. and Khanna, S. M. (1967). "Some Properties of Sound Transmission in the Middle and Outer Ears of Cats", J. Acoust. Soc. Am. 41, 513-521.
27. Nedzelnitzky, V. (1980). "Sound Pressure in the Basal Turn of the Cat Cochlea", J. Acoust. Soc. Am. 68, 1676-1689.
28. von Bekesy, G. (1949). "On the Resonance Curve and the Decay Period at Various Places on the Cochlear Partition", J. Acoust. Soc. Am. 21, 245-254.
29. Rhode, W. S. (1980). "Cochlear Partition Vibration - Recent Views", J. Acoust. Soc. Am. 67, 1696-1703.
30. von Bekesy, G. (1936). "Zur Physik des Mittelohres und Uber das Horen Bei Fehlerhaftem Trommelfell", Akust. Zeits. 1, 13-23. Cited in Bekesy, E. G. (1960). Experiments in Hearing, Ed. G. G. Wever,(McGraw-Hill, NY) pp 104-115.
31. Guinan, J. J. and Peake, W. T. (1967). "Middle-Ear Characteristics of Anesthetized Cats", J. Acoust. Soc. Am. 41, 1237-1261.
32. Price, G. R. (1974). "Upper Limit to Stapes Displacement: Implications for Hearing Loss", J. Acoust. Soc. Am. 56, 195-197.
33. Price, G. R. and Kalb, J. T. (1986). "Mathematical Model of the Effect of Limited Stapes Displacement on Hazard from Intense Sounds", J. Acoust. Soc. Am. 80, S122.
34. Price, G. R. and Lim, D. J. (1983). "Susceptibility to Intense Impulses", J. Acoust. Soc. Am. 74, S8.

Rapid Backscatter Simulation of Complex Targets (U)

Karl D. Reinig, Mr.
Harry Diamond Laboratories, U.S. Army LABCOM
2800 Powder Mill Rd, Adelphi, MD 20783-1197

1 Introduction

The expanding technology base for millimeter and near-millimeter electromagnetic waves has spawned countless target-sensing research and development projects. High-frequency radars, with their ability to see through many battlefield obscurants, are being developed for applications such as target detection and identification, guidance to targets, and fuzing. When the target is expected to be couched in clutter, these objectives become major components of the smart (or perhaps competent) munitions problem. The developers of these systems are required to design sophisticated data processors to make trajectory and detonation decisions based on the waveform of the backscatter signature. The need for validated modeling techniques to rapidly simulate the backscatter waveform for a wide variety of targets, trajectories, and radar frequencies is obvious. Such simulations are an indispensable tool in the design of processor algorithms. They also provide a discerning test bed for competing strategies, which is orders of magnitude more economical than collecting trajectory data by the thousands or, alternatively, conducting live fire tests, even if test trajectories were held to a handful.

This paper reports on the development of algorithms to rapidly simulate the radar backscatter from highly detailed targets. Practically no restriction is placed on the simulation trajectories. The flexibility in trajectories makes the technique suitable for simulations ranging from synthetic aperture radar (SAR) images for target detection or identification to centroid-seeking terminal guidance. A geometric optics approach is used to estimate the radar return from a complex target surface described by B-spline or B-spline-like functions.

The B-spline surfaces have position, slope, and curvature continuity wherever desired; can be constructed directly from a knowledge of the true target surface; and have no limit to the detail which they may contain. Implementation of the geometric optics approach requires calculation of the positions on the surface (the specular points) from which a ray leaving the source would be reflected back to the receiver. The specular points, along with their respective local principal radii of curvature, are used to calculate discrete radar-cross-sections which are then coherently summed to give the total expected return from the target.

Previous studies have demonstrated the usefulness of the geometric optics approach for computing the expected return from a composite of simple analytic shapes [1]. However, as the targets of interest become more complex or the desire to match their surfaces more accurately increases, the use of simple analytical shapes to describe the target surface often becomes impractical. Three-dimensional faceted models exist for most targets of interest, including tanks, helicopters, and jet aircraft. These faceted models generally use several thousand facets to describe the target surface and contain a great deal of detail. Unfortunately, the faceted models do not directly give useful geometric optics information. For example, both principal radii of curvature of any point on a flat facet model are unbounded everywhere except at facet edges, where they are undefined. However, if the target surface is modeled by second-order continuous functions, the principal radii of curvature will be continuous as well and, the model can be used to give meaningful geometric optics information. In addition, as demonstrated by this study, the second-order continuity of the model may be exploited to reduce the problem of finding all the specular points associated with each new trajectory position to that of tracking the motion of existing points. The current focus of the study is on recursive subdivision techniques for isolating specular points, as well as the locations along the trajectory at which discontinuities of specular paths occur. For a complex target, the reduction to a tracking problem results in multiple orders of magnitude in computational savings. The technique can be extended to the multiple-bounce case (in which those rays that bounce from one point on the target surface to another before returning to the source/receiver are included in the expected return). Because of the combinations of surface patches which must be searched for each multiple-bounce specular point, the reduction of the problem to a multiple-bounce tracking problem promises even greater savings. While the

multiple-bounce return will often be small when compared to the single-bounce return, the addition does extend the modelling technique to include the effects of polarization. Finally, the use of B-spline-like surfaces over arbitrary topological meshes is demonstrated as a method for extending the simulation technique to work directly from existing faceted models.

2 Single-Bounce Return Problem

Figure 1 shows the geometry of the single-bounce specular return problem from an isolated patch of a target surface. $\mathbf{R}_t(\Delta)$ is the current position of the projectile along a linear trajectory. $\mathbf{R}_s(u,v)$ describes the target surface as a function of the two parameters u and v. And $\mathbf{G}(u,v,\Delta)$ is the difference between the two vectors $\mathbf{R}_t(\Delta)$ and $\mathbf{R}_s(u,v)$. For the present it will be assumed that the projectile has an unobstructed view of the surface being considered. A necessary and sufficient condition for a point on the surface to be a specular point relative to the position $\mathbf{R}_t(\Delta)$ is that the l_2 norm of $\mathbf{G}(u,v,\Delta)$ be either a local maximum or minimum with respect to the two target surface parameters u and v. Finding all the specular points of a given surface (for a given trajectory position) is therefore the same as finding all u and v which satisfy the two

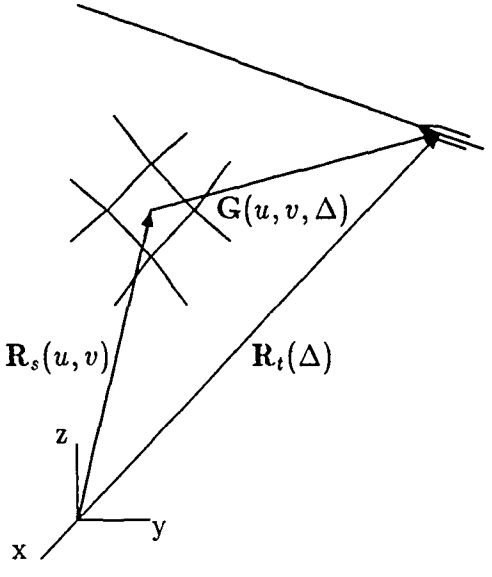


Figure 1: Specular Return Geometry

nonlinear equations

$$F_1(u,v,\Delta) = \frac{\partial \|\mathbf{G}(u,v,\Delta)\|^2}{\partial u} = 0 \tag{1}$$

and

$$F_2(u,v,\Delta) = \frac{\partial \|\mathbf{G}(u,v,\Delta)\|^2}{\partial v} = 0. \tag{2}$$

If $\mathbf{R}(u,v)$ is given by a tensor product of cubic B-splines over a uniform knot set, both $F_1(u,v,\Delta)$ and $F_2(u,v,\Delta)$ may be written explicitly in terms of u and v, for each value of Δ. In general though, solving for the u and v (call them u^* and v^*) which satisfy equations (1) and (2) requires a numerical technique. Simple application of Newton's method for nonlinear equations will find solutions to (1) and (2) provided the search is begun "close enough" to (u^*,v^*). The question of how close is close enough is a complex one, which ultimately depends on the variation of the surface being considered. Whatever method is used to find the specular points associated with a position along the trajectory, if the procedure is to be repeated for each location on the trajectory (each sample of the backscatter waveform), the computation time will prove to be generally unacceptable. However, the computation time could be greatly reduced if the simulation used knowledge of the locations of specular points for one trajectory position to find the specular points of the next.

3 Specular Paths

Suppose that at the trajectory coordinate corresponding to $\Delta = \Delta_s$, u_s and v_s satisfy equations (1) and (2); i.e., (Δ_s, u_s, v_s) are the coordinates of a specular point. We wish to observe the motion of the specular point as Δ is varied. The question is whether a unique solution exists for the values of $u(\Delta)$ and $v(\Delta)$ satisfying equations (1) and (2) as Δ is varied about Δ_s. Define the Jacobian $\mathbf{J}(u,v,\Delta)$ by

$$\mathbf{J}(\mathrm{u},\mathrm{v},\Delta) = \begin{pmatrix} \frac{\partial^2\|G\|^2}{\partial u^2} & \frac{\partial^2\|G\|^2}{\partial u \partial v} \\ \frac{\partial^2\|G\|^2}{\partial u \partial v} & \frac{\partial^2\|G\|^2}{\partial v^2} \end{pmatrix} \tag{3}$$

For a second-order continuous target model, each of the terms of $\mathbf{J}(u,v,\Delta)$ exist and are continuous. The implicit function theorem for simultaneous equations then guarantees the existence of a finite interval about Δ such that $u(\Delta)$ and

$v(\Delta)$ satisfy equations (1) and (2), provided $\mathbf{J}(u_s, v_s, \Delta_s)$ is not singular. The locations (u, v, Δ) at which the Jacobian is singular are known as twinkles, a term introduced by Longuet-Higgins to describe the locations of annihilations or creations of points of reflected or refracted light from a surface in motion [2]. In fact, the phenomena of reflection of light from a time-varying surface and the current problem are quite similar, and this author has gained much insight to the problem from Longuet-Higgins' elegant treatment of the former. The existence of twinkles is a double-edged sword. Because they exist, we cannot simply track the motion of specular points as the trajectory location varies since new specular paths may be created at some other location on the target and would not be found. On the other hand, if for any given trajectory determining all the points (u, v, Δ) for which $\mathbf{J}(u, v, \Delta)$ is singular were possible, searching the entire surface for specular points at different positions along the trajectory would no longer be necessary. It would only be necessary to find all the specular points corresponding to the initial trajectory position (call these initial specular points) and then track their motion, picking up or losing specular paths only at twinkles.

4 Local Search Techniques

Local gradient techniques can also be used to search for twinkles. Let

$$F_1 = \frac{\partial^2 \|\mathbf{G}\|^2}{\partial u}$$

$$F_2 = \frac{\partial^2 \|\mathbf{G}\|^2}{\partial v}$$

and

$$F_3 = \frac{\partial^2 \|\mathbf{G}\|^2}{\partial u^2} \frac{\partial^2 \|\mathbf{G}\|^2}{\partial v^2} - \left(\frac{\partial^2 \|\mathbf{G}\|^2}{\partial u \partial v} \right)^2$$

Then the Newton step (S_u, S_v, S_Δ) toward the parametric coordinates of a twinkle satisfies

$$\begin{pmatrix} \frac{\partial F_1}{\partial u} & \frac{\partial F_1}{\partial v} & \frac{\partial F_1}{\partial \Delta} \\ \frac{\partial F_2}{\partial u} & \frac{\partial F_2}{\partial v} & \frac{\partial F_2}{\partial \Delta} \\ \frac{\partial F_3}{\partial u} & \frac{\partial F_3}{\partial v} & \frac{\partial F_3}{\partial \Delta} \end{pmatrix} \begin{bmatrix} S_u \\ S_v \\ S_\Delta \end{bmatrix} = \begin{bmatrix} -F_1 \\ -F_2 \\ -F_3 \end{bmatrix}$$

Thus, local search techniques exist for finding all potential locations on the target surface, as a function of the trajectory position, for which the specular paths are discontinuous.

A simplistic application of Newton's method has been used in the past to find twinkles and initial specular points on a complex target to demonstrate the specular path tracking method [3]. The algorithm for finding twinkles used a Newton's search starting from the center of each patch and the center of the trajectory to find solutions to $F_1 = F_2 = F_3 = 0$. Similarly a Newton's search starting from the center of each patch with Δ fixed at zero was used to search for initial specular points. If either search wandered outside the extent of the patch (or trajectory in the case of the twinkle search), the search was discontinued for that patch. If either search led to a solution, it was considered to be the only one for that patch. Thus, at most, one initial specular point and one twinkle could be found per patch (there certainly can be more). Moreover, even if only one twinkle or initial specular point existed on a patch, there was no guarantee that it would be found. Note that when a twinkle is found, a simple rotation of coordinates and a Taylor series expansion about the twinkle can be used to effectively predict the motion of specular points in the vicinity of the twinkle, including whether the twinkle represents the beginning or ending of a specular path [3,4]. Also interesting to note, if a twinkle or initial specular path is missed, the conditions predicted at a subsequent twinkle will generally be violated, and therefore the effect may be detected and corrected for [4]. However, the current study is concentrating on global methods which are in some sense assured to find all the twinkles and initial specular points for a given trajectory.

5 Recursive Subdivision

Among the interesting features of the B-spline surface is the relative ease with which it may be recursively subdivided; i.e., the four by four set of points describing a single bi-cubic patch can be transformed into a five by five set of points giving four patches which together describe the original patch [5,6]. For any given patch, it is a simple matter to calculate the normal at the center of the patch. If a lower bound on the dot product of the center normal with all other normals on the patch were known, it would be a simple matter to calculate a "normal cone" emanating from the patch (do not forget to include the extent of the patch) which contains all possible trajectory locations corresponding to

specular points on the patch. By simply checking to see if the initial trajectory is located within the cone, we can label the patch either as a candidate patch for an initial specular point or as one to be discarded from the search. When a patch is subdivided, each of the four new cones describing the space of potential specular trajectory locations will be smaller. In general, some of the new patches will remain candidates for containing an initial specular point and some will not. In the limit, the cones become lines and the area of each patch approaches zero. Thus the simple test and subdivide scheme can be used to reduce the area which may contain initial specular points to a set of patches with arbitrarily (within the accuracy of the computer) small extent.

Now consider the twinkle search problem. The condition that the Jacobian of equation (3) be singular for a given trajectory location requires that one of the two principal radii of curvature of the surface be the same as the distance to that location [4]. As an immediate consequence, note that patches which are convex cap – the two principal curvature vectors point towards the inside of the target – need never be searched for twinkles. More generally, by determining bounds over an entire patch on the maximum and minimum of both principal radii of curvature, one can construct four concentric spheres about the patch. If the trajectory does not intersect the interspace between the bounding spheres for the maximum principal radii of curvature or the interspace of the bounding spheres for the minimum principal radii of curvature, the patch does not contain a twinkle. As the patches are subdivided, the bounds on the two principal radii of curvature become tighter. In the limit, two concentric spheres exist about the patch, one of which the trajectory must pass through in order for that patch to contain a twinkle. Recall that a twinkle must also satisfy the conditions of a specular point. Thus the trajectory must also lie within the normal cone described above for the patch to remain a candidate. By checking to see if the trajectory passes through the intersection of the normal cone and the interspace of the appropriate curvature bounding spheres, one may conduct a recursive subdivision search for twinkles as well. Bounds on both the normals and principal curvatures of a patch may be found from the fact that the zero-, first-, and second-order derivatives of the B-spline surface vector components may be written as convex combinations of appropriately chosen functions of the control mesh points. The potential exists to greatly reduce the time required to search for such bounds by precomputing (independent of any trajectory) locations at which either principal curvature is stationary with

respect to the two surface parameters. Progress has also been made in the area of stopping criteria for the subdivision techniques. The author hopes to report on rapid methods for determining when initial multiple-bounce specular points or multiple-bounce twinkles have been isolated in the near future.

6 Single-Bounce Example

Together, figures 2 and 3 show an example of the use of twinkles for tracking single-bounce specular paths on a B-spline surface. The example surface is defined by just under 2600 patches arranged in a polar format. Figure 2 shows the B-spline surface control mesh and desired trajectory, as well as the locations along the trajectory where specular path discontinuities are expected (based on a search for twinkles). Both the twinkle search and the search for initial specular points were carried out by applying the recursive subdivision techniques described above for a set number of times and then simply performing a Newton search over the remaining candidate patches. Each twinkle location along the trajectory is labeled B or D, depending on whether the twinkle represents the birth or death of a pair of specular paths, and lines have been drawn connecting them with their associated locations on the target surface. In addition, the predicted paths of the specular points in the area near each twinkle are shown. Figure 3 shows the results after tracking the specular paths for 1000 locations along the trajectory. A comparison of figures 2 and 3 shows that the specular paths did in fact remain continuous everywhere except at the twinkles and moved as predicted in the regions near each twinkle. The search for the initial specular points took approximately 35 seconds on a Convex C1 minicomputer (in scalar mode). Therefore, without the use of twinkles, the entire simulation would have taken over 9 hours. The twinkle search took about 2 minutes. However, because completely searching the entire target for specular points at each trajectory location was no longer necessary, the entire simulation took only about 5 minutes.

7 Nth-Order Specular Points

The geometric optics assumptions provide for multiple-bounce return in which it is assumed that the wave is reflected about one or more locations on the surface before returning to the source/receiver. Space constraints do not allow

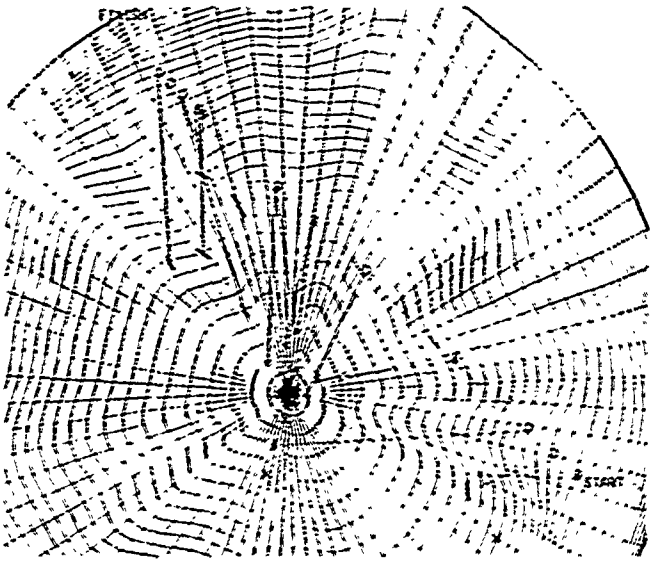


Figure 2: Twinkles and Initial Specular Points

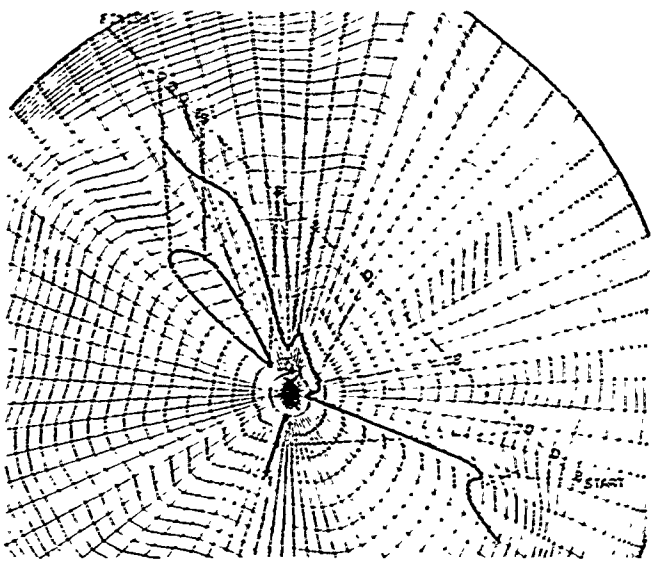


Figure 3: Specular Point Paths

for a detailed discussion of the subject in this paper. However, an important development of this study has been the generalization of the twinkle analysis to include the multiple-bounce case; i.e., it has been shown that for second-order continuous surfaces, multiple-bounce specular points move in paths which are discontinuous only at multiple-bounce twinkles. The analysis leading to the conclusion of multiple-bounce twinkles, as well as an example of the motion of double-bounce specular points, may be found in other works [3,4]. The reduction of the multiple-bounce problem to a tracking problem is particularly important since the brute-force n-bounce search must be conducted over all combinations of the surface patches taken n at a time, while tracking multiple-bounce specular points is completely independent of the number of patches.

8 B-Spline-Like Surfaces From Facet Models

It is important to note that the use of twinkles and initial specular points to track specular paths across a target surface requires only that the surface be generally second-order continuous. B-splines are just one method of surface representation that can be used to produce such target models. A close cousin to the B-spline is the recursively generated B-spline surface on arbitrary topological meshes introduced by Catmull and Clark [6]. The method takes arbitrary topological meshes like those of a typical faceted model and gives rules for recursively chopping off the corners, resulting in a "smoother" overall surface. Patches which are defined over a rectangular control mesh are subdivided the same as a standard B-spline surface. Other patches acquire a rectangular mesh as subdivision continues, except at a fixed number of locations known as extraordinary points. Doo and Sabin [7], and later Ball and Storry [8], showed that the limiting surface is continuous in both position and slope everywhere and in curvature everywhere except at the extraordinary points. In a world without numerical roundoff, the existence of a fixed number of isolated points of curvature discontinuity would have no effect on the continuity of specular paths since specular paths would never run exactly through the extraordinary point. In practice however, specular paths may approach close enough to the offending point that the numerical algorithms may be adversely affected. Ignoring for now the possible effects of the extraordinary points, the resulting limiting surface appears ideal for application of the recursive subdivision techniques discussed above. The biggest advantage of the more generalized B-spline

surface is that the large number of highly detailed faceted models which already exist could be used directly by the backscatter simulation algorithms.

As part of the current study, algorithms were recently written to demonstrate the usefulness of the subdivision technique for arbitrary topologies. One program takes as input an arbitrary faceted model in Movie.BYU* format. The subdivision rules of Catmull and Clark are then applied to the original facet vertices to form new vertices along with a new connectivity array. The output of the program is then a new subdivided faceted model in Movie.BYU format. Figure 4 shows a faceted model of a typical tank turret often used, along with a shading model to give a visual representation of the turret. Figure 5 shows the results after a single subdivision on the turret. Note that all faces on the subdivided turret are four sided (this always occurs after the first subdivision), and most patches lie on a rectangular mesh. Notice as well that the resulting mesh gives a "visually pleasing" representation of the original facet detail. Further subdivisions do little to change the shape of the control mesh. If the user is satisfied that the desired detail has been captured in the resulting mesh, then a search for the specular points and their local principal radii of curvature for a given trajectory should give all the needed information for a simulation of the associated radar backscatter using the geometric optics approach. If there are portions of the model which the user would like to change, they can be easily incorporated into the original facet model using existing graphics editing techniques.

A modification of the program allows patches to be discarded without affecting the connectivity array of neighboring patches. By using the neighboring faces to determine bounds on the normal cone of each face, the program tests each patch to determine if it is a candidate for containing specular points. Patches which are no longer candidates are discarded before further subdivision. Once all the remaining subpatches are on regular rectangular meshes, the program proceeds the same as in section 6 for finding initial specular points. The program was repeated for one hundred locations along the trajectory; figure 5 shows the results as well. Figure 6 gives a closer view of the specular points. The specular paths are all continuous except at locations where pairs are either annihilated or created (a fact made much clearer when color is used to record

*A general-purpose computer graphics software system distributed by the Brigham Young University.

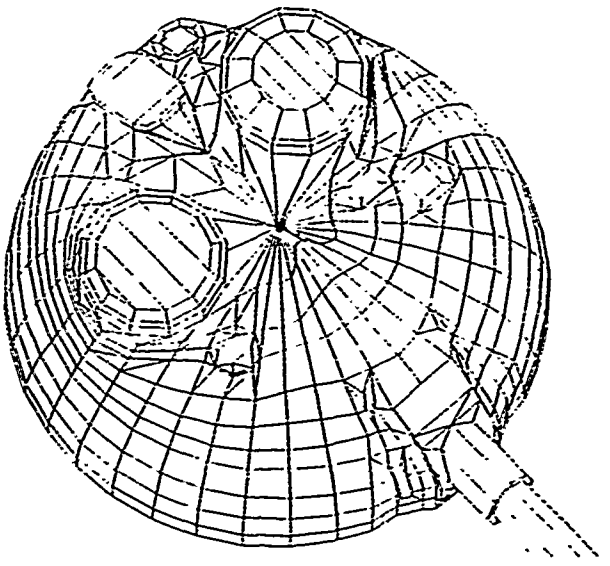


Figure 4: Faceted Tank Turret

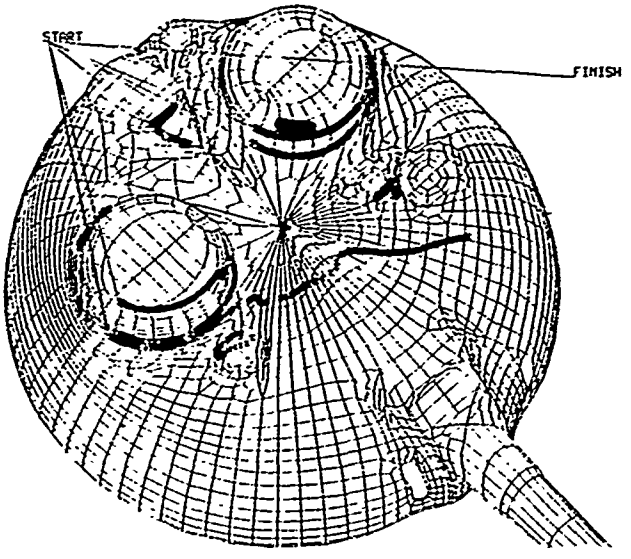


Figure 5: Faceted Tank Turret After One Subdivision

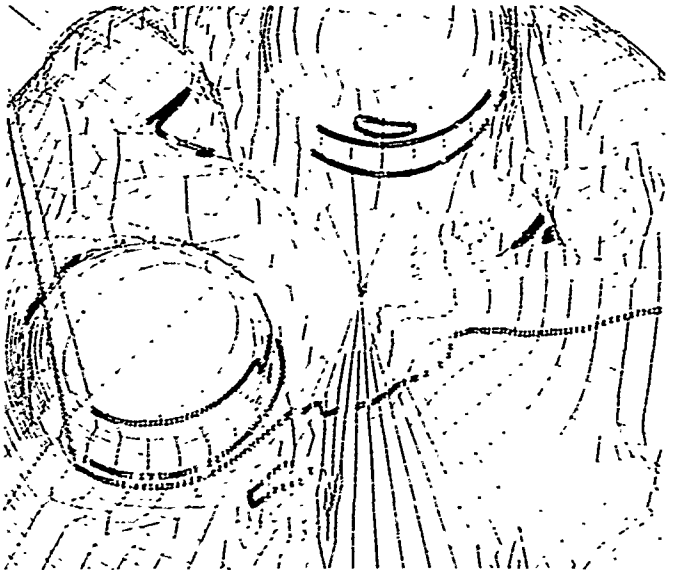


Figure 6: Closeup of Specular Paths

the chronology of the paths). Note that if tests are developed for bounding the principal radii of curvature of these more generalized B-spline patches, the subdivision technique for finding twinkles will follow directly.

9 Summary and Conclusions

Using the well-known geometric optics assumptions, we can reduce the problem of calculating the backscatter from a known surface relative to some position in space to that of finding the corresponding specular points along with their associated principal curvatures. B-splines may be used to create surfaces which capture the detail of the desired target while maintaining second-order continuity wherever desired. Ordinarily, the search for specular points on any detailed surface must be repeated for every location along the desired trajectory. Unfortunately if the return waveform is to be sampled on the order of a thousand times per meter, the computation time of the simulation becomes generally too large. The problem can be overcome by first searching for all the locations along the trajectory for which specular paths may be discontinuous (twinkles). Between twinkles the problem is reduced to that of tracking existing specular points. Near twinkles, the behavior of the specular paths can be predicted, and

the appropriate paths added or discarded as required. The same strategy can be used for tracking multiple-bounce specular paths.

The techniques for finding and tracking specular points may be extended to work with limit surfaces derived directly from existing faceted models. The extension will make immediately available hundreds of faceted models of targets of interest which may be easily modified using generally available graphics editing software.

References

1. John F. Dammann, Jr., *Air Target Models for Fuzing Simulations,* Harry Diamond Laboratories, HDL-TR-1960 (September 1982).

2. M. S. Longuet-Higgins, *Reflection and Refraction at a Random Moving Surface. I. Pattern and Paths of Specular Points,* J. Opt. Soc. Am. **50** (1960), 838.

   Karl Reinig, *A Rapid, Backscatter Simulation Technique for Complex B-Spline Target Models,* Proc. Fifth Army Conf. on Appl. Math and Comp., (June 1987).

3. Karl Reinig, *A Backscatter Simulation Approach for Detailed B-Spline Target Models,* Harry Diamond Laboratories, HDL-TR-2139 (in preparation).

   Elaine Cohen, Tom Lyche, and Richard Riesenfeld, *Discrete B-Splines and Subdivision Techniques in Computer-Aided Geometric Design and Computer Graphics,* Computer Graphics and Image Processing **14** (1980) 87-111.

   E. Catmull and J. Clark, *Recursively Generated B-Spline Surfaces on Arbitrary Topological Meshes,* Comput.-Aided Des., Vol **10**, No. 6 (November 1978), 350-355.

   D. Doo and M. Sabin, *Analysis of the Behavior of Recursive Division Surfaces,* Comput.-Aided Des., Vol **10**, No. 6 (November 1978), 356-360.

4. A. A. Ball and D. J. Storry, *A Matrix Approach to the Analysis of Recursively Generated B-Spline Surfaces,* Comput.-Aided Des., Vol **18**, No. 8 (October 1986), 437-442.

Numerical Modeling of Contaminant Dispersion in Air by Bouyancy-Driven Flows Within Fabric Structures (U)

*Struan R. Robertson, PhD
US Army Natick Research, Development and Engineering Center
Natick, MA 01760-5017

INTRODUCTION

The diffusion of contaminant through fabric walls and the subsequent mixing with the interior air is not well understood. Modern computational techniques permit the modeling of contaminant dispersal within enclosures by diffusion and by thermally induced flows. Further, ventilating strategies can be studied in order to learn how to control and minimize effects of unwanted contaminants. One method that has been found to be robust and effective for a wide variety of fluid flow problems is the control volume method as applied by Spaulding and Patankar [1]. This method has been called the "Semi-Implicit Method for Pressure Linked Equations - Revised" or SIMPLER for short. A FORTRAN code for this algorithm was provided by Charmchi and McKelliget [2] and subsequently modified by the author for the types of problems studied here.

Using the code requires a fairly intimate understanding of the governing equations and their approximation as embodied in the SIMPLER algorithm. This work describes the application of the method to several two dimensional problems in square and peaked pentagonal enclosures. The purpose is to demonstrate both how to model such problems and the effectiveness of the approach, not to determine the best strategies for contaminant minimization.

The governing equations for the problem are the Navier-Stokes equations for the velocity field, the energy equation for the temperature field and the transport equation for the contaminant. These equations are nonlinear and coupled. In the approximation process the continuity equation is treated independently in order to form a pressure equation and a pressure correction equation [1]. Thus, the problems studied here will involve six coupled finite difference equations. The velocity, temperature and concentration equations can all be put into the same generic form. A staggered grid is used. Grid points centered within a control volume are used for all scalar variables such as temperature, pressure and concentration and all properties such as conductivity, viscosity, density etc. A second set of grid points lying on the x-faces of the control volumes is used for the x-velocity components. A third set of grid points lying on the y-faces of the control volumes is used for the y-velocity components.

GOVERNING EQUATIONS

The governing equations can be cast in the generic form,

$$\partial(\rho\phi)/\partial t+(\rho v_i^o\phi)_{,i} = (\Gamma\phi_{,i})_{,i}+S \tag{1}$$

where a repeated index implies summation over the range of the index and a comma preceding an index (,i) implies $\partial/\partial x_i$.

For two dimensional problems in Cartesian coordinates the Navier-Stokes (momentum) equations are

$$\partial(\rho u)/\partial t+(\rho u^o u)_{,x}+(\rho v^o u)_{,y} = (\mu u_{,x})_{,x}+(\mu u_{,y})_{,y}-p_{,x}+F_x \tag{2}$$

and

$$\partial(\rho v)/\partial t+(\rho u^o v)_{,x}+(\rho v^o v)_{,y} = (\mu v_{,x})_{,x}+(\mu v_{,y})_{,y}-p_{,y}+F_y \tag{3}$$

where the superior o on u and v is used to indicate that these are values from the previous step in the numerical solution procedure. This approximation serves two purposes: it linearizes the momentum equations and it puts them in the form of the generic equation (1).

The energy equation for the temperature field is

$$\partial(\rho cT)/\partial t+(\rho cu^o T)_{,x}+(\rho cv^o T)_{,y} = (kT_{,x})_{,x}+(kT_{,y})_{,y} \tag{4}$$

and the equation for agent concentration is

$$\partial C/\partial t+(u^o C)_{,x}+(v^o C)_{,y} = (DC_{,x})_{,x}+(DC_{,y})_{,y}. \tag{5}$$

The continuity equation,

$$\partial\rho/\partial t+(\rho u)_{,x}+(\rho v)_{,y} = 0 \tag{6}$$

when substituted into equations (2)-(5) yields the usual equations associated with these problems. However, since the pressure field is not known a priori, it is used in the finite difference formulation to form a pressure equation and a pressure correction equation [1].

For the problems of interest here, the following simplifying assumptions are made:

1. Temperature variations are small so μ, k and D are assumed to be constant.

2. Flow is assumed to be steady so all derivatives with respect to time, t, vanish.

3. Temperature variations of ρ give rise only to bouyancy forces, the flow being incompressible.

4. Although the density is assumed constant, for purposes of determining the bouyancy force only, the following relation between density and temperature is used [3].

$$(\rho_o - \rho) = \rho\beta(T - T_o). \tag{7}$$

The next step is to determine the bouyancy force for use in the momentum equations. To this end, the pressure is broken down into the static pressure and the velocity pressure, thus

$$p = p_v + p_s \tag{8}$$

where, for an enclosure of height H with the y-direction positive upward, the static pressure is

$$p_s = (H - y)\rho_o g. \tag{9}$$

In this case, $p_{,y} = p_{v,y} - \rho_o g$. In the second momentum equation (3) with $F_y = -\rho g$ the term

$$F_y - p_{,y} = (\rho_o - \rho)g - p_{v,y}$$

upon using eq. (7) this becomes

$$F_y - p_{,y} = \rho g\beta(T - T_o) - p_{v,y} \tag{10}$$

where the first term to the right of the equal sign is the bouyancy force. In the first momentum equation (2)

$$F_x = 0.0 \quad \& \quad p_{,x} = p_{v,x}. \tag{11}$$

Eqs. (10) and (11) are substituted into eqs. (2) and (3) then all the equations are nondimensionalized as follows. Define the following variables with T_h the high temperature, T_c the low temperature and $T_o = (T_h - T_c)/2$ and similarly for C,

$$X = \frac{x}{H}, \; Y = \frac{y}{H}, \; U = \frac{\rho H u}{\mu}, \; V = \frac{\rho H v}{\mu}, \; \Theta = \frac{(T - T_o)}{(T_h - T_c)},$$

$$P = \frac{\rho H^2 p_v}{\mu^2}, \; \Psi = \frac{(C - C_o)}{(C_h - C_c)}.$$

Introduce the following dimensionless groups

$$Pr = \frac{\mu c}{k}, \quad Ra = \frac{g\beta(T_h - T_c)H^3\rho^2 c}{\mu k}, \quad \eta = \frac{\rho D}{\mu},$$

where Pr is the Prandtl number, Ra is the Rayleigh number, μ the dynamic viscosity, ρ the density and β the coefficient of thermal expansion. The governing equations (2) - (6) become, respectively,

$$\partial(U^{\circ}U)/\partial X + \partial(V^{\circ}U)/\partial Y = \nabla^2 U - \partial P/\partial X\ , \tag{12}$$

$$\partial(U^{\circ}V)/\partial X + \partial(V^{\circ}V)/\partial Y = \nabla^2 V - \partial P/\partial Y + Ra\,\Theta/Pr\ , \tag{13}$$

$$\partial(U^{\circ}\Theta)/\partial X + \partial(V^{\circ}\Theta)/\partial Y = (\nabla^2\Theta)/Pr, \tag{14}$$

$$\partial(U^{\circ}\Psi)/\partial X + \partial(V^{\circ}\Psi)/\partial Y = \eta\nabla^2\Psi\ , \tag{15}$$

$$\partial U/\partial X + \partial V/\partial Y = 0\ . \tag{16}$$

Noting that equations (12) - (15) are of the same form as the generic equation (1), the following table of analogies can be written. Note that the pressure terms in eqs. (12) and (13) are part of the source term. The code, however, treats these internally so they do not appear in the table.

TABLE 1.
Parameter Analogies

Generic Eq.	Φ	ρ	Γ	S
X-Momentum Eq.	U	1	1	0
Y-Momentum Eq.	V	1	1	$Ra\,\Theta/Pr$
Temperature	Θ	1	1/Pr	0
Concentration	Ψ	1	η	0

The code is written for Cartesian, axisymmetric and plane polar coordinate systems only. On the boundaries U and V will be either normal or tangential to the boundaries. For the momentum equations, therefore,

U or V = 0 normal to the boundary

or

U or V specified normal to the boundary in case of inflow,

or

U and V not specified in the case of outflow.

For the energy equation,

Θ or $\partial\Theta/\partial n$ is specified on the boundary.

For the mass diffusion equation,

Ψ or $\partial\Psi/\partial n$ is specified on the boundary.

For the momentum equations (12) & (13) the terms $-\partial P/\partial X$ and $-\partial P/\partial Y$ are handled by the code.

MODELING CONSIDERATIONS

The SIMPLER algorithm has certain peculiarities that need clarification. The most confusing aspect of the code is the mesh. In fact there is not one mesh but three. Scalar variables such as temperature, pressure and concentration; and material properties such as density, conductivity and viscosity are evaluated at "grid points" internal to the control volumes. Velocity components are evaluated at the mid-points of the control volume faces to which they are normal. The user defines locations where the velocity components are calculated. For example, in Cartesian coordinates the X locations for U and the Y locations for V are specified. The code generates the grid points automatically. There is one more grid point in each direction than there are velocity locations.

Consider Fig. 1 for spacings in the X-direction. In the figure the XU_i are the U locations and the X_i are the grid locations for scalars. Note that $XU_1 = XU_2$, $X_1 = XU_2$ and $X_L = XU_L$. For internal points $X_i = (XU_{i+1}+XU_i)/2$, $i=2,3,...,L-1$. The same holds true for the Y-direction. This leads to a 2D mesh like that in Fig. 2.

The control volumes for the grid points are formed by using the velocity locations in the X and Y directions. The velocities are evaluated where the grid lines intersect the control volume lines. Referring to Fig. 3, U is evaluated where the horizontal grid lines intersect vertical control lines at XU_i. As can be seen in Fig. 4, The control volumes for the U and V equations are different from each other. Also, U_{ij} and V_{ij} are not at the same location: U_{ij} is at XU_i, YV_j and V_{ij} is at X_i, YV_j.

Boundary fluxes can be treated by one of two methods. In the first, the boundary flux is replaced by an equivalent internal source term for those cells adjacent to the boundary. In the second, a balance is written between the applied flux and the flux from the boundary to the adjacent interior point. The boundary value is then expressed as a function of the applied flux and the value of the adjacent interior point. The second method has been found to be the more effective of the two. The second method is, therefore, applied to thermal boundary conditions. Its application to other variables is identical. Thus, letting q be the applied flux per unit area,

$$q = -k\,\partial T/\partial x. \tag{37}$$

Referring to Fig. 5,

$$q = -k\,(T_P - T_B)/\delta.$$

Three cases are considered;

a) constant flux

In this case the appropriate boundary condition is

$$T_B = T_P + q\delta/k \tag{19}$$

b) convective flux

In this case with T_f being the bulk fluid temperature,

$$q = h(T_f - T_B) \tag{20}$$

so that the appropriate boundary condition is

$$T_B = ((\delta h/k)T_f + T_P)/(1 + \delta h/k) \tag{21}$$

c) radiative flux

In this case with T_f being the source temperature,

$$q = \alpha(T_f^4 - T_B^4). \tag{22}$$

Let,

$$h_R = \alpha(T_f^2 - T_B^2)(T_f + T_B) \tag{23}$$

so that

$$q = h_R(T_f - T_B) \tag{24}$$

which is the same form as the convective flux, where h_R is evaluated using T_B from the previous step. Alpha is a combination of the Stefan-Boltzmann constant, view factor, emissivity of source and absorptivity of the sink.

For the problems discussed here the boundary conditions are either symmetric, specified field variable or convective like flux. The convective boundary condition for thermal problems, in terms of the nondimensional temperature, is

$$\Theta_B = (\Theta_P + Nu\Delta\Theta_f)/(1 + Nu\Delta), \tag{25}$$

where $Nu = Hh/k$ is the Nusselt number and $\Delta = \delta/H$ is the distance from the boundary node to the adjacent interior node.

For a symmetry (adiabatic) boundary the appropriate boundary condition is

$$\Theta_B = \Theta_P. \tag{26}$$

For each type of boundary condition, there are certain auxilliary conditions involving the source terms and the diffusion constant Γ. Boundary conditions for the contaminant Ψ are treated in exactly the same way as those for Θ. These are summarized in Table 2.

TABLE 2.

Auxilliary Conditions

Boundary Condition	Variables	Boundary			
		Left	Right	Bottom	Top
Symmetry	U, V, Θ,Ψ	$V_{1,j} = V_{2,j}$ $\Gamma_{1,j} = 0$	$V_{L,j} = V_{L-1,j}$ $\Gamma_{L,j} = 0$	$U_{i,1} = U_{i,2}$ $\Gamma_{i,1} = 0$	$U_{i,M} = U_{i,M-1}$ $\Gamma_{i,M} = 0$
Inflow	U, V	$\Gamma_{1,j} = 0$	$\Gamma_{L,j} = 0$	$\Gamma_{i,1} = 0$	$\Gamma_{i,M} = 0$
Outflow	U, V	$U_{2,j} = U_{3,j}$	$U_{L,j} = U_{L-1,j}$	$V_{i,2} = V_{i,3}$	$V_{i,M} = V_{i,M-1}$

APPLICATIONS

Two regions are modeled, a square and a peaked pentagonal. The square presents no particular difficulties with respect to setting up a model. The peaked pentagonal, however, does because the code permits only rectangular regions to be modeled. The peaked pentagonal is modeled by defining a square computational domain as shown in Fig. 7. The properties in the slashed region are chosen 1. to make it "infinitely" conductive with respect to both thermal and contaminant diffusion and 2. to ensure that the velocities vanish. The appropriate values that must be enforced in the slashed region are given in Table 3.

TABLE 3

Properties for a blocked off region

Velocity Equations	U = 0, V = 0 Γ = 1.0E20. S = 0
Temperature and Contaminant Equations	Θ & Ψ not specified Γ = 1.0E20 S = 0

Note that Γ is large in both cases but recall that it essentially represents viscosity for the Navier-Stokes equations and the diffusivity for the thermal and concentration equations.

Square Enclosure

Fig. 8 shows the boundary conditions on the square enclosure, the values for *Pr* and η and the range of values for *Ra*. The top and bottom are thermally insulated and impervious to contaminant. The left side has fixed minimum values of -0.5 for both Θ & Ψ while the right side has fixed maximum values of 0.5 for both. The velocity is zero on all four walls. The mesh is equispaced with 92 by 92 grid points. Fig. 9 shows the evolution of the streamlines with increasing Rayleigh number which is analogous to increasing the height of the enclosure, other parameters remaining equal. The flow becomes more complex and starts to breakup into cells. Fig. 10 shows the temperature contours. If the heat transfer were purely conductive, the contours would be equally spaced vertical lines. The fluid flow caused by bouyancy forces "turns" the contours on their side, the effect becoming more marked with increasing Rayleigh number. The same can be said of Fig. 11 for the contamination contours.

Fig. 12a shows the variation of temperature along the horizontal midline. Note that only half the curve is shown because the other half is negatively symmetric about the centerline. This is true of the other plots. In concurrence with the contour plots the gradient is steepest near the wall and gets steeper with increasing Rayleigh number. Because of the nondimensional form of the solution the value of Θ = 0.0 is the average of the maximum and minimum temperatures. The same can be said of the plots for contaminant in Fig. 12b. Fig. 12c shows the temperature variation on the vertical midline. The transition from the maximum at the top surface to the Θ = 0.0 at the centerline is more gradual but with a noticeable change in the curvature with the highest Rayleigh number. Similarly for the variation of contaminant on the vertical midline in Fig. 12c. The results for the case with Ra = 1.0e6 agree with earlier results of Zimmerman[3]. Higher Rayleigh numbers were not considered in that report. These problems took from 1000 iterations for the lowest Rayleigh number to 1500 for the highest for the solution to converge.

Pentagonal Enclosure

Fig. 8 shows the pentagonal enclosure with the boundary conditions and the values for the other parameters. The roof is heated convectively by a source (see eq. 25) at 0.5, the walls by a source at 0.0, both with Nu= 85. The floor is fixed at -0.5. Fig. 13 shows the evolution of the streamlines as Ra increases from 1e7 to 0.5e9. Note how the flow starts with 6 cells and evolves to a flow with essentially 2 cells when Ra = 0.5e9. These cases converged in 1500 or less iterations.

Fig. 14 shows the evolution of the temperature field with increasing Rayleigh number.

The contaminant is fixed on the left side of the roof and left wall at 0.5 on on the right side of the roof and right wall at -0.5. The floor is impervious. Fig. 15 shows the evolution of the contaminant contours with increasing Rayleigh number. The shape of the contours can be closely correlated to the streamlines in Fig. 13. It is clear that the contaminant is dispersed primarily by

the thermally induced bouyancy flow and not by simple diffusion.

DISCUSSION

It is often assumed in the study of the effects of contaminants that they become uniformly distributed within an enclosure. This assumption is intuitively flawed since the dispersion of contaminant obeys the same mathematical equation as the temperature field and it would not be expected that the temperature would be uniform. The simple cases studied here support the intuitive inference. They also show that the dispersion is controlled primarily by convective currents and not diffusion. Further, as the size of the enclosure increases the flows become more complex.

The present study has shown the feasibility of modeling contaminant dispersal. In the future, more realistic three dimensional problems that can be experimentally verified should be studied. This should be done by first modeling possible choices for three dimensional enclosures with a variety of realizable boundary conditions. From among the various cases one suitable for an experiment would then be selected. The computer model and the experiment would then be improved interactively, resulting in a better understanding of the process.

REFERENCES

1. Patankar, S. V., NUMERICAL HEAT TRANSFER AND FLUID FLOW, Hemisphere, Washington, D. C., 1980.

2. McKelliget, J. and Charmchi, M., private communication, U. Lowell, Lowell, MA., 1986.

3. Gebhart, B., HEAT TRANSFER, McGraw-Hill, N.Y., 1971.

4. Zimmerman, E. B., "Natural Convection in Square Partitioned Enclosures", AD-A153-707, 1985.

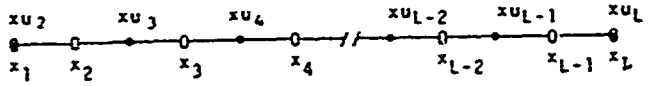


Figure 1. X - locations of velocity and grid points.

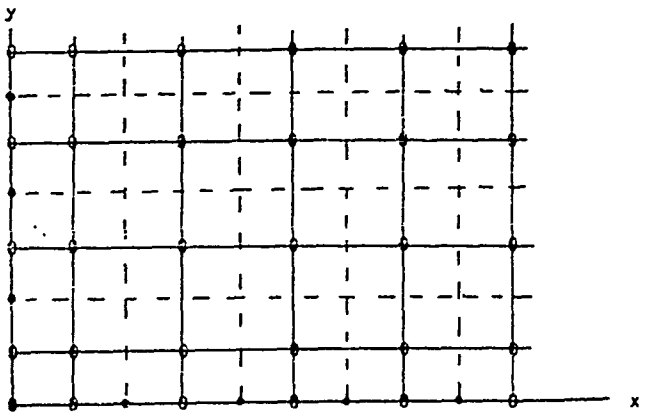


Figure 2. The locations of grid points and control volume interfaces for scalars.

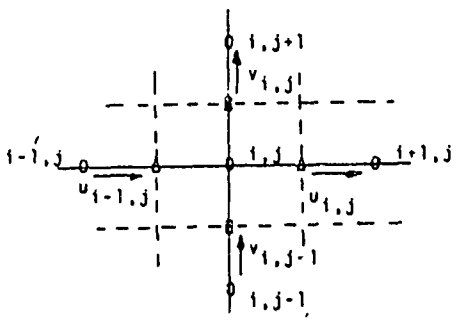


Figure 3. Points where u and v are evaluated.

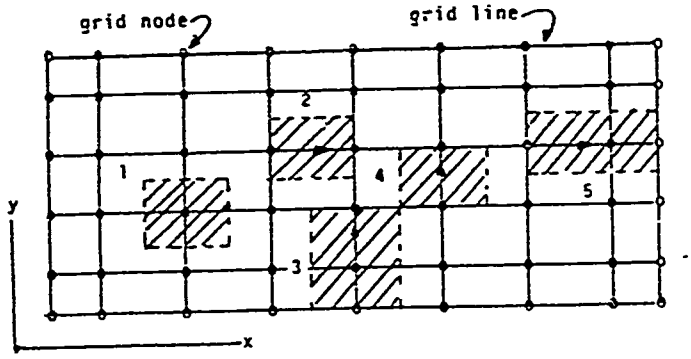


Figure 4. Control volumes: 1. scalars, 2. u, 3. v boundary, 4. v, 5. u boundary.

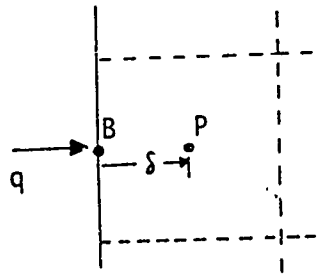


Figure 5. Boundary flux.

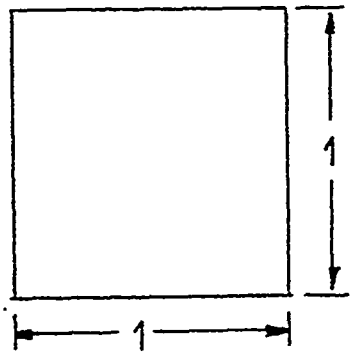


Figure 6. Square enclosure.

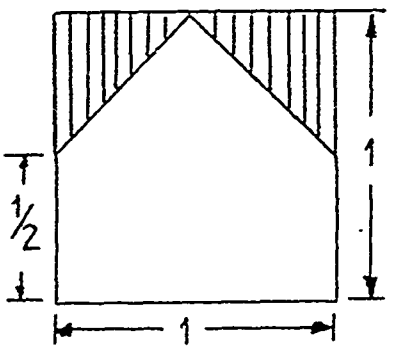
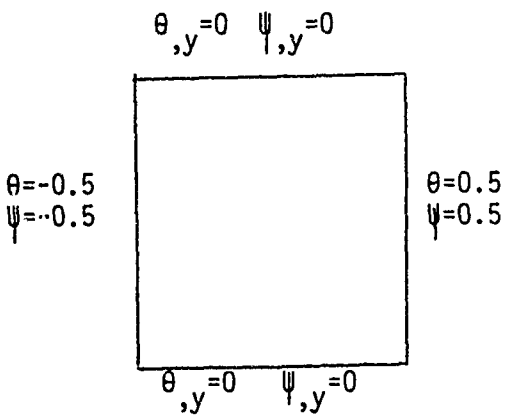


Figure 7. Pentagonal enclosure.



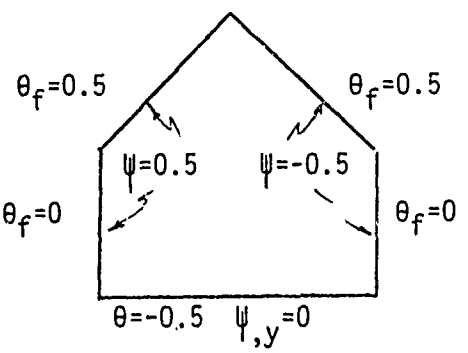
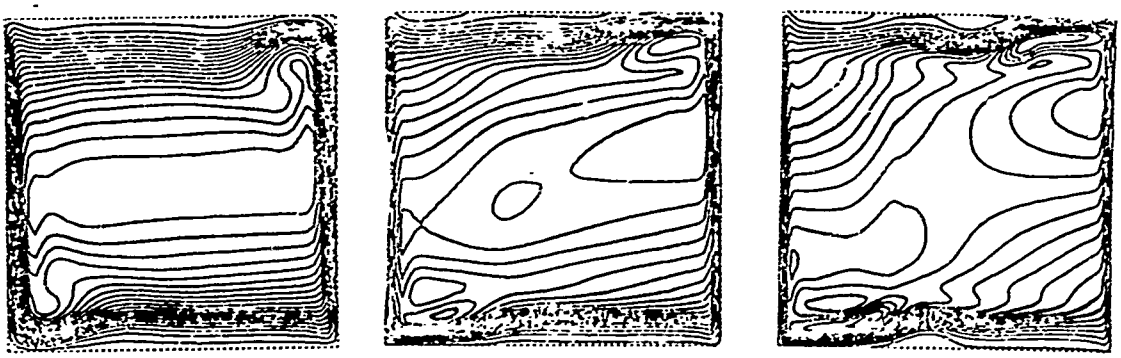


Figure 8. Boundary conditions. In both cases Ra is from 1e6 to 0.5e9, Pr = 0.71 and η = 0.5. For the pentagonal enclosure Nu = 85 for convective heating of the walls and roof.



Ra=1e7 Ra=1e8 Ra=0.5e9

Figure 9. Streamlines

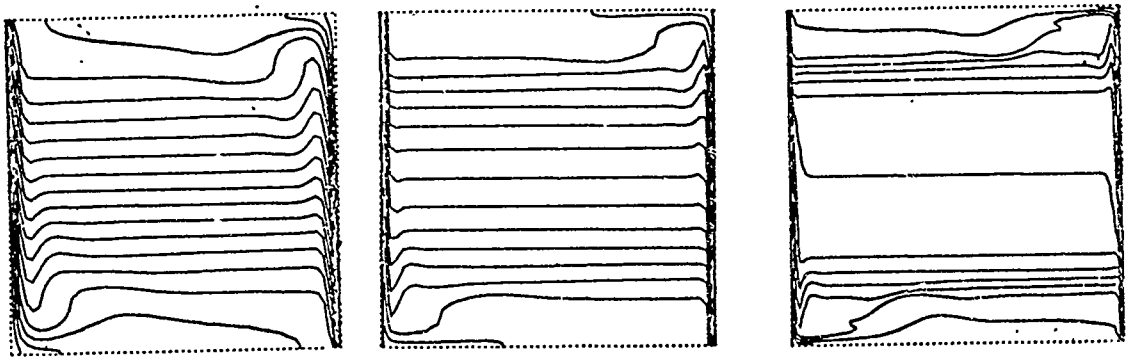


Figure 10. Temperature

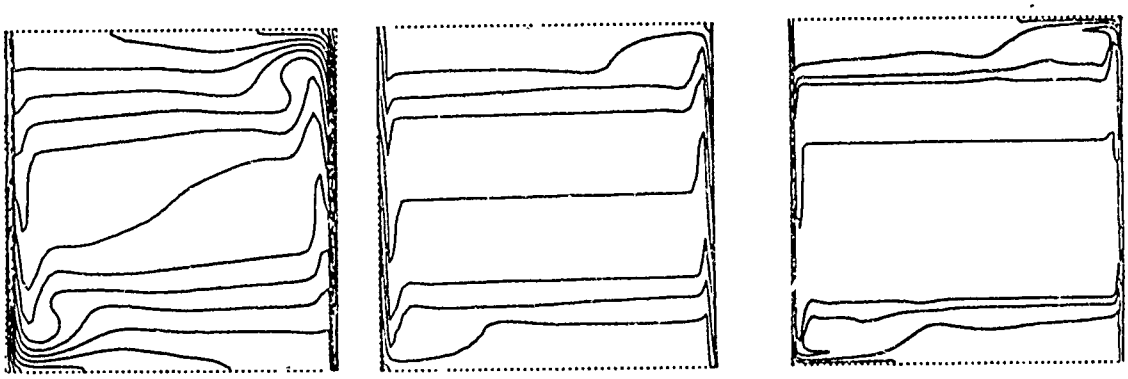


Figure 11. Contaminant

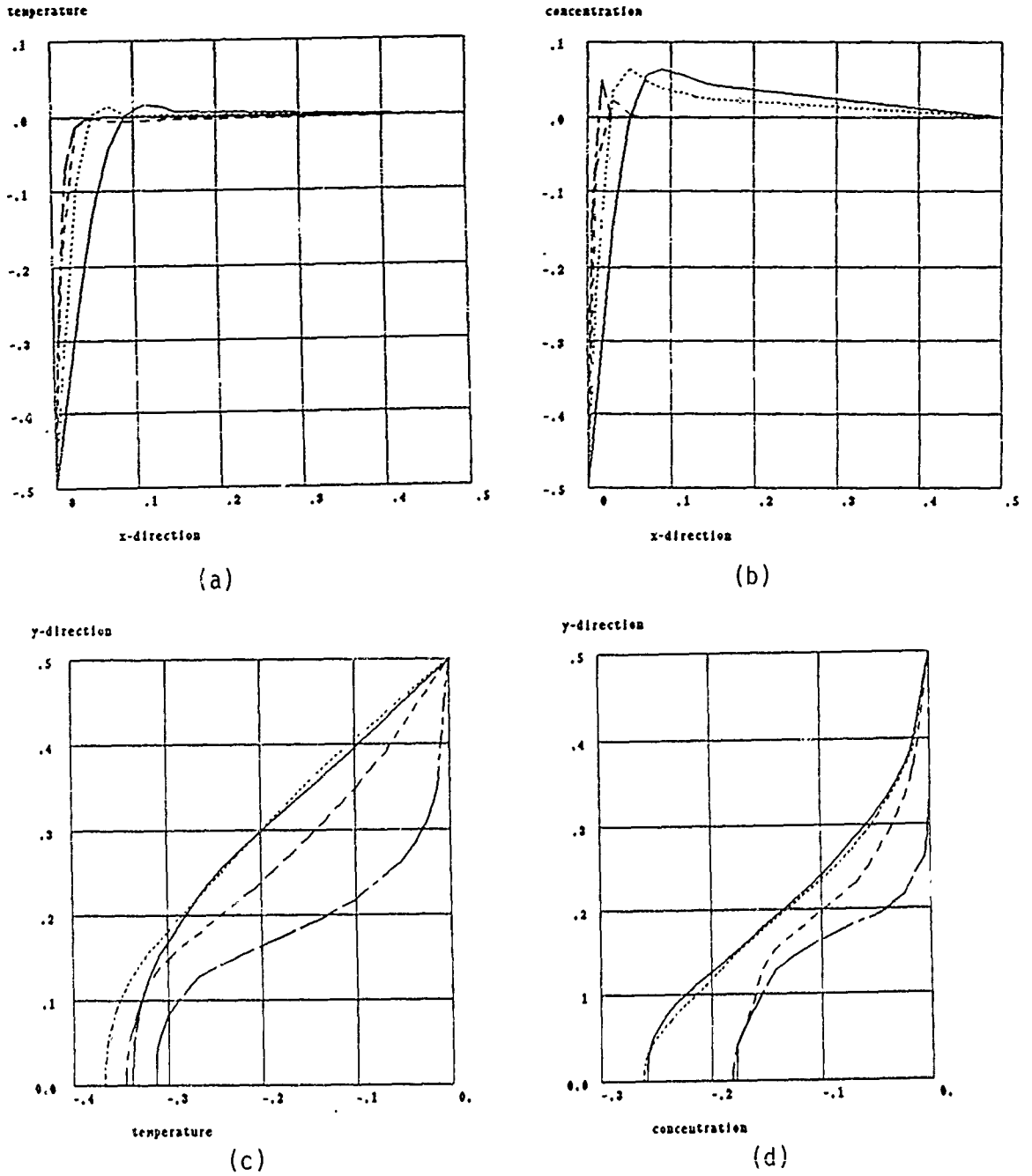
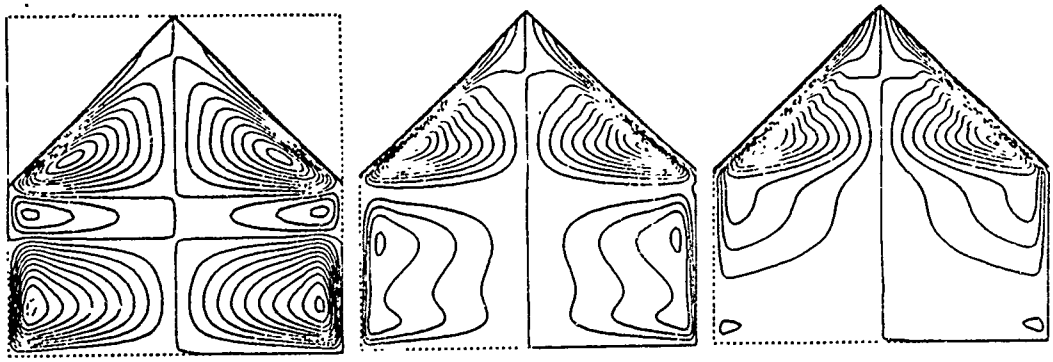


Figure 12. a) Θ on the horizontal midline, b) Ψ on the horizontal midline, c) Θ on the vertical midline, d) Ψ on the vertical midline, Ra = 1e6 _____, 1e7 ....., 1e8 _ _ _ _ , 0.5e9 _ _ _ _



Ra=1e7 Ra=1e8 Ra=0.5e9

Figure 13. Streamlines

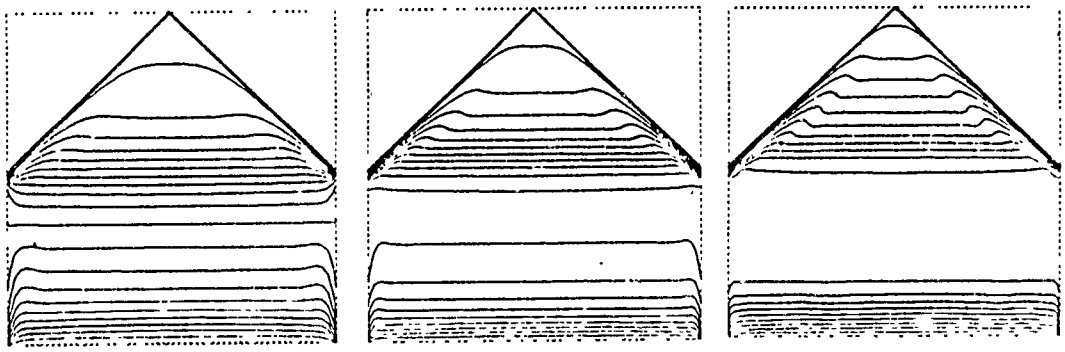


Figure 14. Temperature

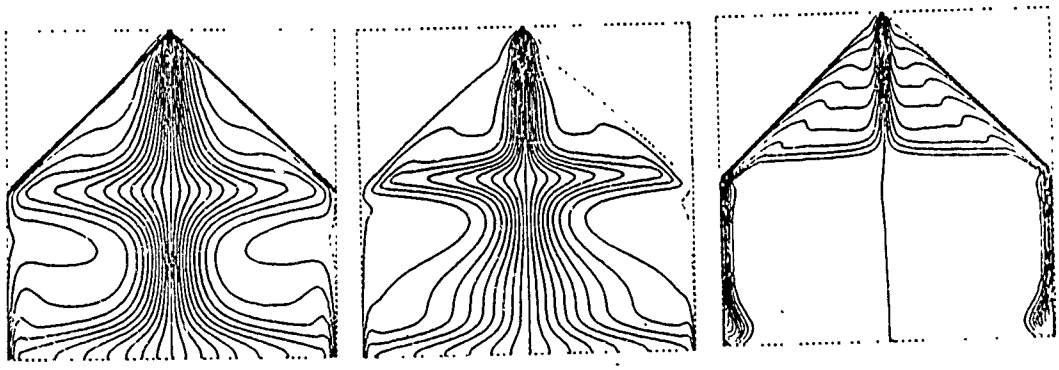


Figure 15. Contaminant

Vibration-Immune Quartz Crystal Oscillators

Vincent J. Rosati, Mr., and Raymond L. Filler, Dr.

US Army Electronics Technology and Devices Laboratory
Fort Monmouth, New Jersey 07703-5000

Introduction

The phase noise performance of quartz crystal oscillators under vibration has become recognized as one of the limiting factors in the performance of certain radar and navigation systems. For example, assuming a 10 MHz oscillator that uses a resonator with an acceleration sensitivity of 5×10^{-10} per g, a 1g vibration at 100 Hz produces a bright line on the oscillator spectrum that is only 92dB below the carrier, a level that is about 60dB worse than required.[1,2] Random vibration, the type usually encountered on vehicles and aircraft, fills the spectrum with "white" noise.It must be recalled that all atomic frequency standards (rubidium, cesium) derive their output from quartz crystal oscillators: they suffer the same phase noise degradation when vibrated.

Figure 1 shows the single-sideband phase noise of a good commercial oscillator at rest, and the phase noise of the same oscillator while undergoing only 0.01 g^2/Hz random vibration from 10 Hz to 1 kHz . The degradation of phase noise below 100 Hz is more than 50dB. The impact of this vibration-induced noise on coherent radar systems, such as JOINT-STARS, is shown in Figure 2. This curve relates the phase noise of a 10 MHz oscillator to the probability of detection of a slow-moving target. (The 10 MHz oscillator is frequency-multiplied to 10 GHz to supply the receiver local oscillator signal.) Note that unless the vibration-induced phase noise is suppressed, the probability of detection is zero!

Clearly, methods for reducing vibration effects must be developed, with repeatability, producibility, and cost in mind.

Ideally, a resonator with zero sensitivity to acceleration should be used. Such resonators do not exist, although some progress is being made toward producing them.[3] Next in order of preference is a passive method of suppression that uses no electronic or mechanical controls. Likewise, some progress is being made but presently no compact, lightweight method exists. A possible passive method is described later in this paper.

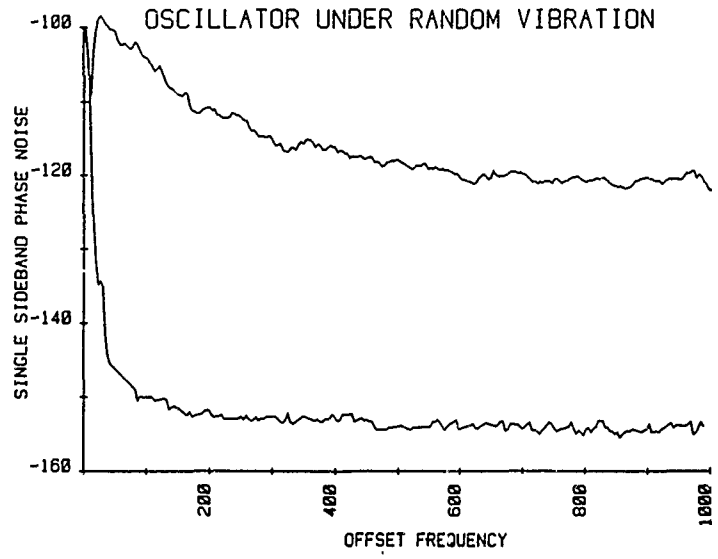


Figure 1. Single-sideband phase noise of a quality oscillator: 1. at rest (lower trace), and 2. subjected to random vibration (upper trace).

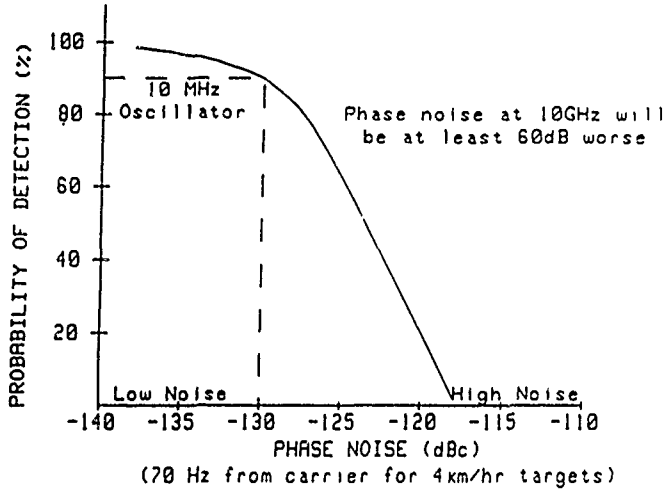


Figure 2. To achieve a 90 percent probability of detection of a slow-moving target, the phase noise must be lower than -130dB with respect to the carrier. If the noise is worse than -120dB, the probability approaches zero.

Direct-to-Crystal Vibration Suppression

The active method to be described is based on a simple principle. Suppose that an acceleration causes an upward shift of resonator frequency. One can counter that shift by applying a properly phased voltage to a tuning element in the oscillator to oppose the upward shift, thereby cancelling the acceleration-induced frequency shift.[4] We have chosen the resonator itself as the tuning element. The method is called polarization-effect tuning or direct-to-crystal vibration suppression.

The polarization effect, sometimes referred to as the voltage-frequency effect, is the change of the resonant frequency of piezoelectric resonators due to an impressed electric field.[5,6] The frequency change is a highly linear function of the applied voltage.

The direct-to-crystal tuning method consists of sensing the vibration with an accelerometer, amplifying and phase-inverting the sensing signal, and feeding it directly to the electrodes of the SC-cut resonator in the oscillator, through isolation resistors. The SC-cut resonator, via the polarization effect, is tuned to oppose the frequency change caused by the acceleration.

For SC-cut crystals, vibration amplitudes of 10g can be suppressed by the application of less than 20 volts, a value that is achievable in practice. AT-cut crystals theoretically have zero voltage sensitivity but values up to five parts per hundred billion have been observed. However, this low value renders the direct-to-crystal vibration suppression method impractical for AT-cut crystals, because the tuning voltage would have to be hundreds of volts.

Figure 3 is a pictorial description of the direct-to-crystal approach. The uppermost set of drawings illustrates the linear nature of the SC-cut voltage-frequency effect. Next, if the stimulus is changed from DC to AC, observation of the oscillator output on a spectrum analyzer will show sidebands occurring above and below the carrier frequency (F_o), separated from it by the excitation frequency (F_v). This result is explained by FM theory for low-index modulation. The next lower set of drawings shows a spectrum analyzer display of a crystal oscillator being vibrated at a frequency F_v. Notice that the two spectra are identical. The bottom set of drawings shows the result of applying an electrical replica of the vibration frequency directly to the crystal, with amplitude and phase adjusted to cause suppression of the vibration-induced sidebands.

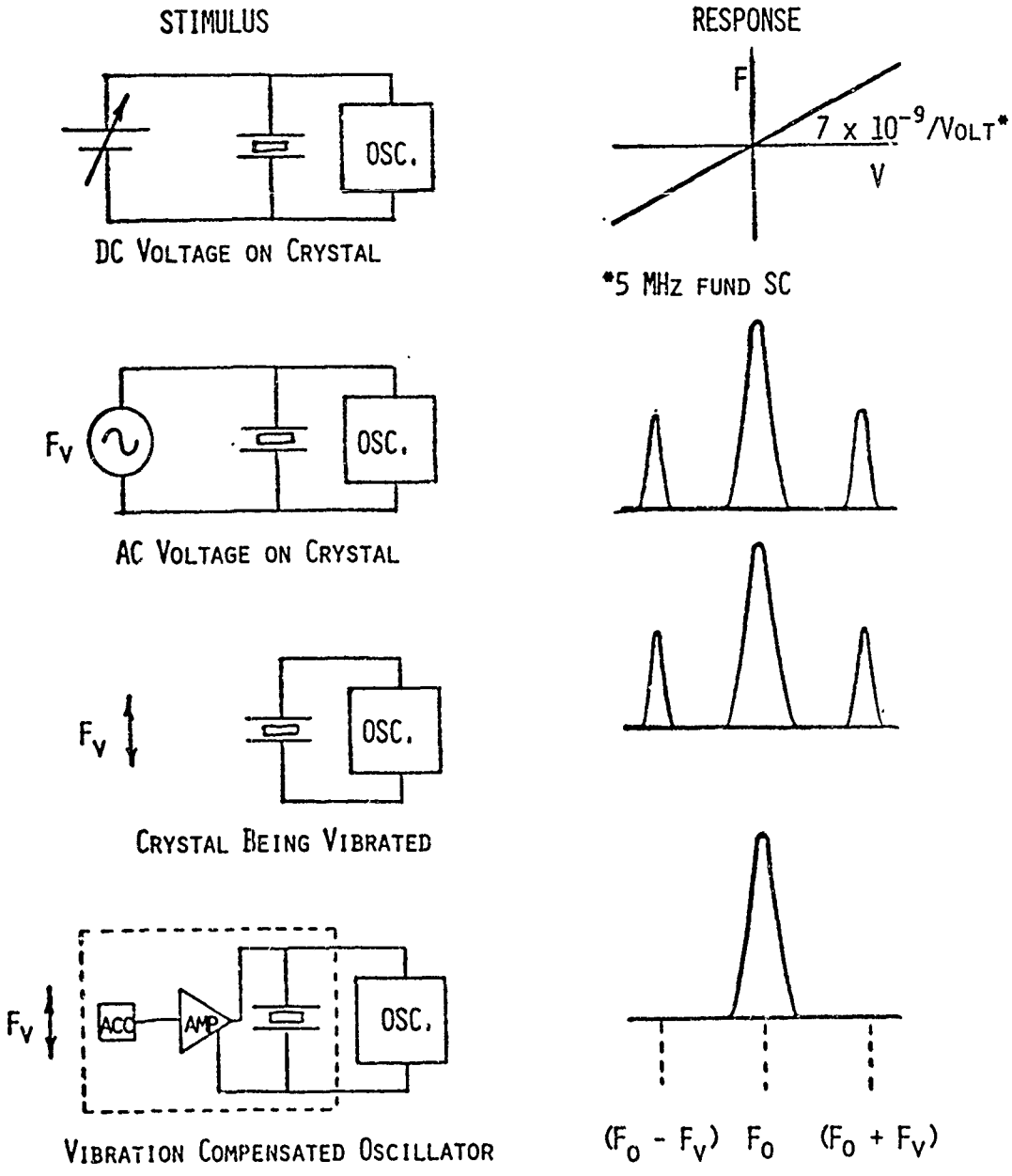


Figure 3. Pictorial diagram of the development of the direct-to-crystal approach for suppressing vibration-induced noise sidebands. The polarization effect of SC-cut quartz crystal resonators is exploited. Properly tuning the crystal in opposition to vibration-induced sidebands results in their virtual elimination.

An analysis was performed to determine the sensitivity of the method to phase and/or amplitude errors in the accelerometer feedback path. The analysis model, which should apply to all methods of suppression (except vibration isolation), is the cancellation of one sine wave with another, as shown in Figure 4.

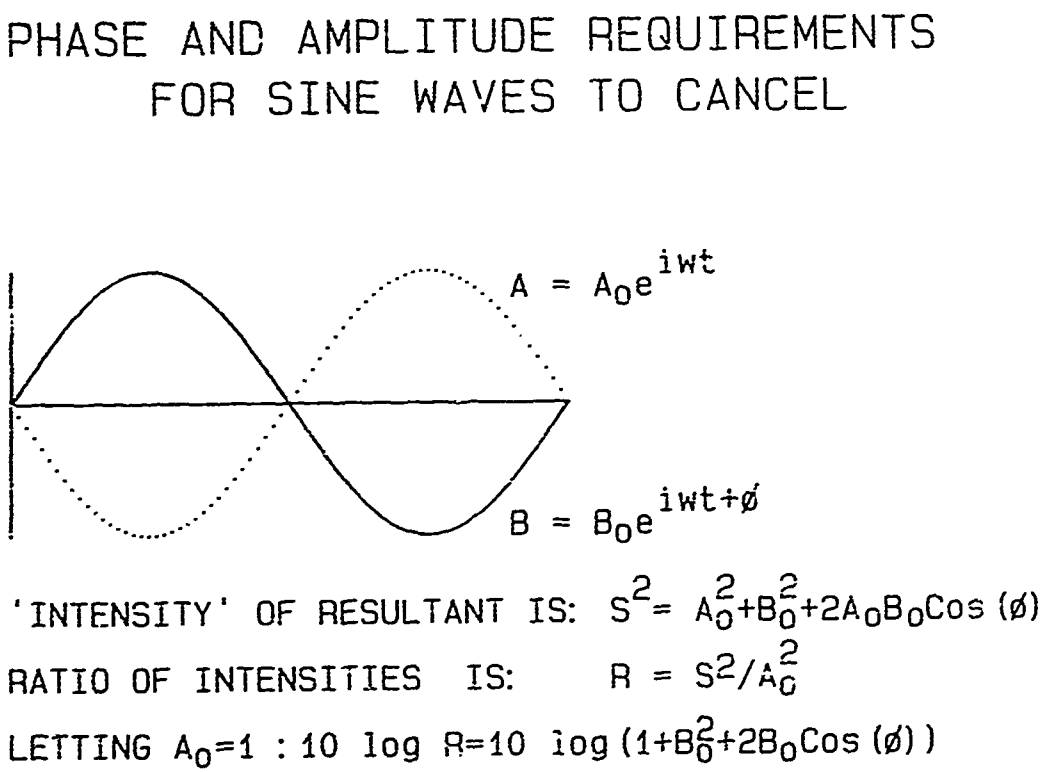


Figure 4. The constraints on the amplitude and phase accuracy for cancellation of sine waves are derived from this model.

Figures 5 and 6 show the depth of suppression obtainable as the amplitude or phase is varied. Note that to achieve 50 dB of suppression, the amplitude must be held constant to better than 0.3 percent, while phase must be held to within better than 0.2 degrees. Figure 7 is a qualitative depiction of the two-parameter sensitivity "surface".

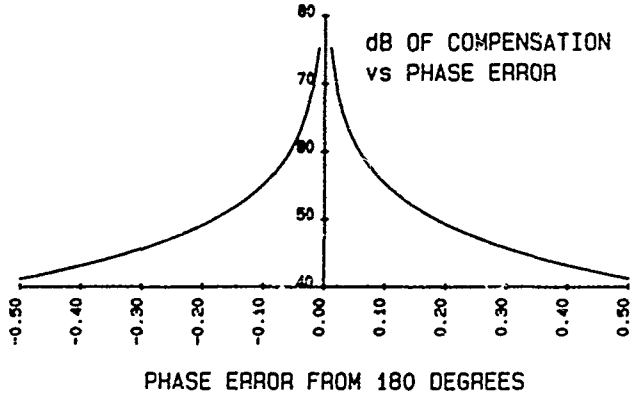


Figure 5. The variation of compensation vs. phase error is shown.

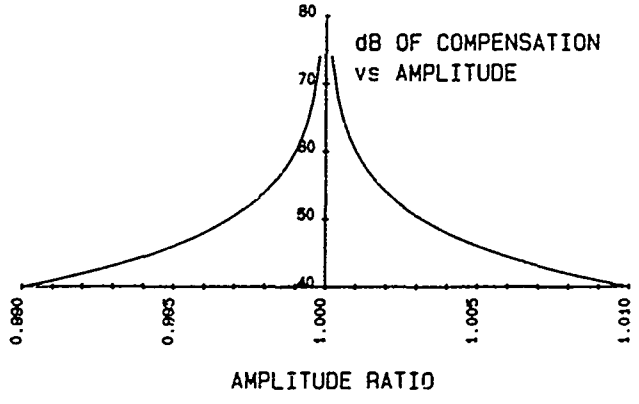


Figure 6. The variation of compensation vs. amplitude ratio is shown.

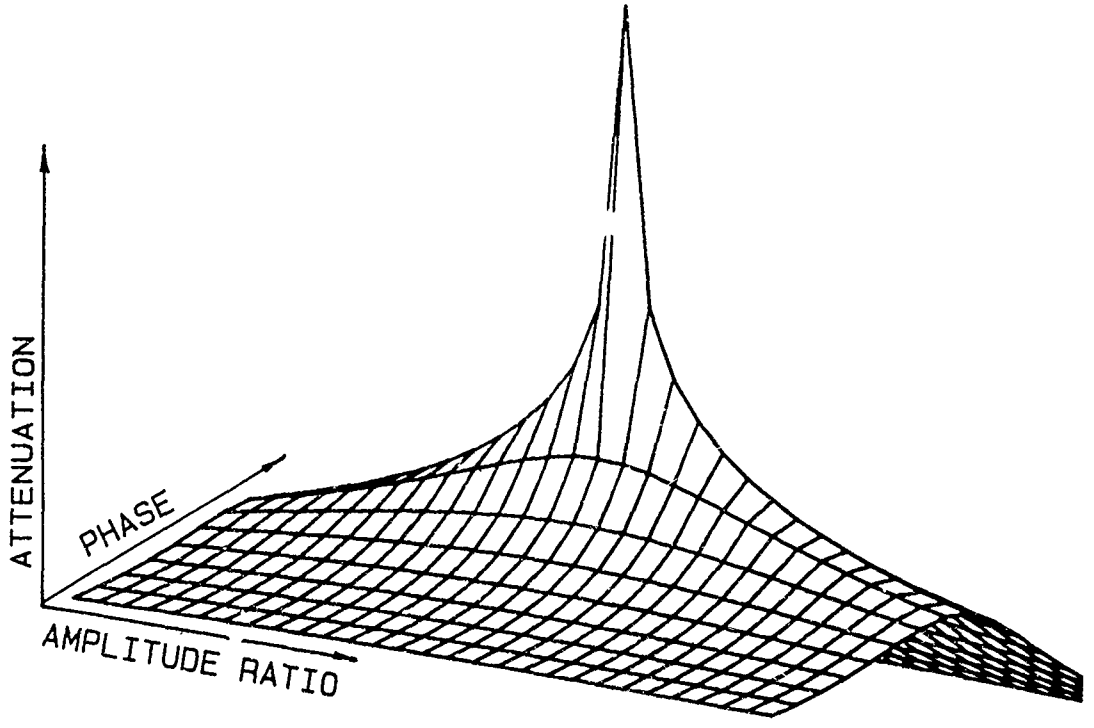


Figure 7. A plot of the sensitivity "surface" of suppression vs. amplitude and phase accuracy.

The analysis makes clear the need for high gain-bandwidth-product amplifiers, rigid oscillator construction to avoid mechanical resonances, and close proximity of the resonator and accelerometer to avoid phase shifts due to the finite velocity of propagation of the vibration wave between the resonator and accelerometer. Close proximity is also necessary in order to minimize acceleration level differences due to rotations.

Figure 8 is a simplified schematic diagram of the amplifiers used in the present work. Input to the charge amplifiers is from a three-axis piezoelectric accelerometer, having charge sensitivity of 12 picocoulombs/g. The 20 pF capacitor is supplied by the accelerometer cable. The 300 megohm resistor tailors the accelerometer low-frequency response and provides a path for bias current of the input FET. The gain amplifier may be switched from inverting to noninverting to provide the correct phase for the feedback signal for its axis. A voltage gain of about

40 was sufficient for the resonator used. The summing amplifier output is fed to each crystal electrode through 100 kilohm resistors.

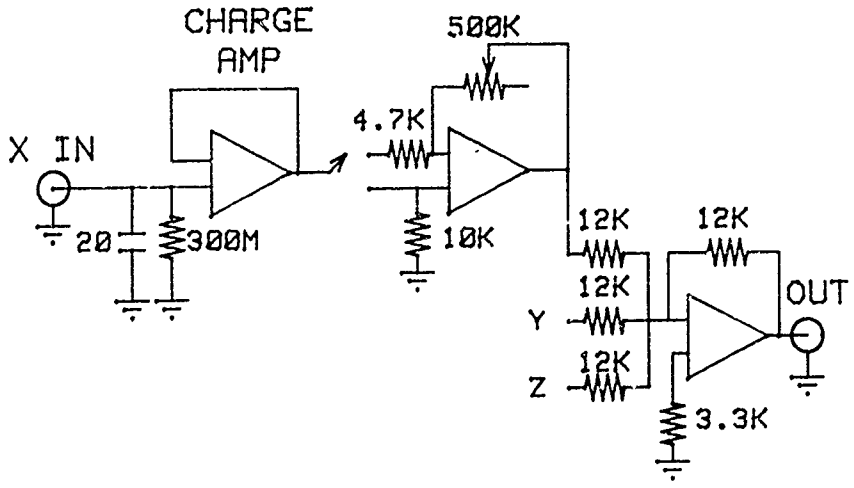


Figure 8. Simplified schematic diagram of the suppression circuit. The X-channel is identical to the Y and Z channels. The X, Y, and Z channels are summed in the last stage.

Figure 9 shows the reduction of phase noise close to the carrier, obtained by modifying a commercial oscillator to accept the suppression signal. The upper trace is the vibration-induced phase noise due to 0.01 g^2/Hz random vibration between 10 Hz and 250 Hz. The lower trace is the phase noise when suppression is applied, showing 35 to 40dB of suppression. The *'s at the lower left corner of the graph represent suppressed sinusoidal vibration at 5, 10, and 15 Hz. Sinusoidal excitation was used here because of limitations of the random-noise generator driving the shake table. The Figure shows, then, that 35 to 40dB of suppression is obtained over more than 5 octaves, from 5 to 250 Hz.

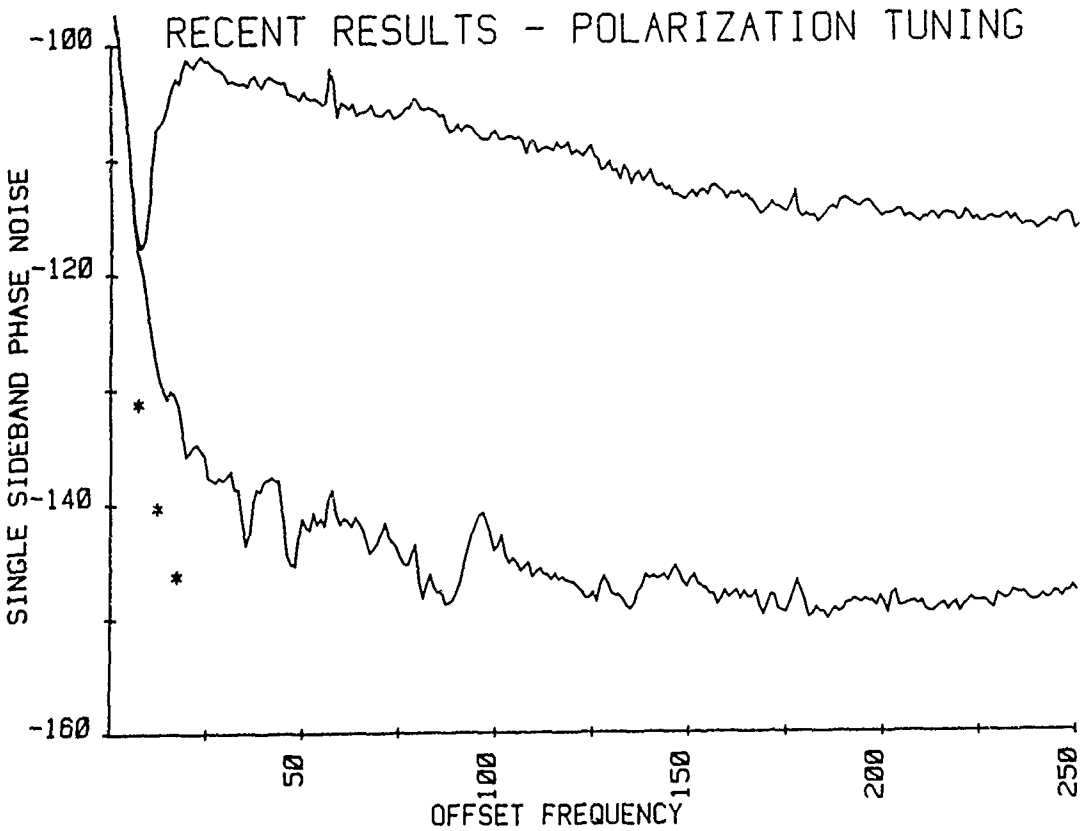


Figure 9. Reduction of vibration-induced phase noise using the circuit of Fig. 8.

Resonator Pairs

A second acceleration-compensation scheme, based on a series of patents by Army personnel,[7-11] is the use of matched pairs of resonators. It has been shown empirically that the acceleration sensitivity of a quartz crystal resonator has the properties of a vector. (This is not known a priori, since all other resonator characteristics exhibit tensor properties.) If two resonators are aligned such that the sensitivity vectors are antiparallel, the net sensitivity would be the algebraic difference between their magnitudes. Figure 10 shows the resultant

acceleration sensitivity of a pair of resonators. The magnitude of the vector on the individual resonators is 5 x 10^{-11}/g, the resultant of the pair is reduced by nearly a factor of 10 to 6 x 10^{-11}/g.

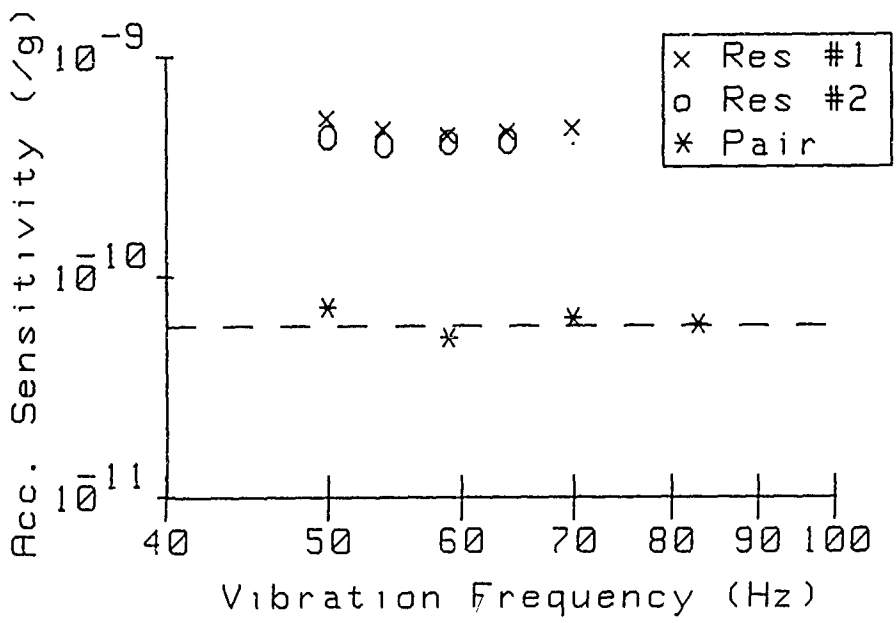


Figure 10. Experimental results of the acceleration sensitivity of a counter-aligned pair of resonators.

In order to reduce the vibration-induced sidebands to an insignificant level, the magnitude of the acceleration-sensitivity vector must be reduced by a factor of 100. That means that the two resonators in the pair must be matched to better than one percent. It is very difficult to find two resonators with acceleration sensitivity matched to that degree. In fact, it is difficult even to measure the sensitivity to within one percent error. It is, therefore, essential to find a method to adjust the relative magnitudes of the acceleration sensitivity of the individual resonators during manufacture of the oscillator.

A possible method has been proposed that has been tested passively, but not yet in an actual oscillator circuit. Theoretical analysis shows that such adjustment is possible, under certain circumstances.

The impedance of two resonators in series is

$$Z_{12} = \frac{-j\left(\frac{(S_1)(A_2)}{C_{o1}} + \frac{(S_2)(A_1)}{C_{o2}}\right)}{(\omega)(A_1)(A_2)} \qquad (1)$$

where

$S_1 = \omega^2 - 2\gamma_1 a\omega^2 - \omega_{s1}^2$
$S_2 = \omega^2 - 2\gamma_2 a\omega^2 - \omega_{s2}^2$
$A_1 = \omega^2 - 2\gamma_1 a\omega^2 - \omega_{a1}^2$
$A_2 = \omega^2 - 2\gamma_2 a\omega^2 - \omega_{a2}^2$
C_{oi} = shunt capacitance of resonator i; i= 1,2
γ_i = acceleration sensitivity
ω_{si} = series resonant frequency
ω_{ai} = anti-resonant frequency

The desired result is that C_{o1} and C_{o2} exist such that

$$\frac{dZ_{12}}{da} = 0 \qquad (2)$$

Equation 2 can be rewritten, using equation 1, as

$$\frac{1}{C_{o1}} \frac{\gamma_1(\omega_{a1}^2 - \omega_{s1}^2)}{A_1} + \frac{1}{C_{o2}} \frac{\gamma_2(\omega_{a2}^2 - \omega_{s2}^2)}{A_2} = 0 \qquad (3)$$

A useful situation occurs if $\omega_{a1} = \omega_{a2}$. In that case equation 3 can be rewritten to give

$$-\frac{\gamma_1}{\gamma_2} = \frac{C_{o1}^2 L_1}{C_{o2}^2 L_2} \qquad (4)$$

where L_1 is the motional inductance of resonator #1 and L_2 is the motional inductance of resonator #2.

This leads to the possibility of adding shunt capacitance to one of the resonators to compensate for unequal γ's. The difficulty arises because ω_a is a function of C_o. Therefore, equation 4 must be solved simultaneously with the condition

$$\omega_{s1}^2 \left(1 + \frac{C_1}{C_{o1}}\right) = \omega_{s2}^2 \left(1 + \frac{C_2}{C_{o2}}\right) \tag{5}$$

where C_1 and C_2 are the motional capacitances of resonators 1 and 2, respectively. The result for the shunt capacitances are

$$\begin{aligned} C_{o1} &= \left(\frac{C_1 C_2}{\Omega}\right)^{1/2} \left\{ \left(\frac{\omega_{s1}}{\omega_{s2}}\right)\left(\frac{C_2}{C_1}\right)^{1/2} - \left(\frac{\gamma_1}{\gamma_2}\right)^{1/2} \right\} \\ C_{o2} &= \left(\frac{C_1 C_2}{\Omega}\right)^{1/2} \left\{ \left(\frac{\gamma_1}{\gamma_2}\right)^{1/2} - \left(\frac{\omega_{s2}}{\omega_{s1}}\right)\left(\frac{C_1}{C_2}\right)^{1/2} \right\} \end{aligned} \tag{6}$$

where $\Omega = \frac{\omega_{s2}}{\omega_{s1}} - \frac{\omega_{s1}}{\omega_{s2}}$.

We are now in the process of testing this technique experimentally.

Conclusions

The experimental data that have been presented for the direct-to-crystal method shows that the acceleration sensitivity of a retrofitted oscillator can be improved by two orders of magnitude, at least between 5 and 250 Hz, and that this approach is viable. Future work will be concerned with designing an oscillator with this suppression technique in mind at the outset, with the goal of extending suppression to at least 2 kHz. To achieve better than 40 dB of suppression over a broad band, it is possible that this method will need to be combined with some degree of mechanical isolation, or with a resonator-pair.

The theoretical arguments that have been presented show that under certain circumstances pairs of resonators may be combined to yield an overall reduction of acceleration sensitivity.

Acknowledgements

The authors wishes to thank Mr. Dennis Bowman for the layout and construction of the amplifier circuitry, and Mr. Peter Thompson for performing the modifications of off-the-shelf oscillators.

References

1. R.L. Filler, "The Effects of Vibration on Frequency Standards and Clocks", Proc. 35th ASFC, pp 31-39, 1981.

2. R.L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators:A Review", Proc. 41st ASFC, pp 398-408, 1987.

3. A. Ballato, "Resonators Compensated for Acceleration Fields", Proc. 33rd ASFC, pp 322-336, 1979.

4. J.M. Przyjemski, "Improvement in System Performance Using a Crystal Oscillator Compensated for Acceleration Sensitivity", Proc. 32nd ASFC, pp 426-431, 1978.

5. C.K. Hruska, "On the Linear Polarization Effect with Alpha Quartz AT Plates", IEEE Trans. Sonics and Ultrasonics, Vol. SU-28, Mar 1981, pp 107-110.

6. J. Kusters, "The Effect of Static Electric Fields on the Elastic Constants of Alpha Quartz," Proc. 24th ASFC, 1970.

7. A. Ballato and J.R. Vig, "Acceleration Resistant Combination of Opposite-Handed Piezoelectric Crystals," US Patent No. 4,344,010, 1982.

8. A. Ballato and J.R. Vig, "Method of Fabricating Acceleration Resistant Resonators, Resonators So Formed," US Patent No. 4,365,182, 1982.

9. A. Ballato and J.R. Vig, "Method of Fabricating Acceleration Resistant Crystal Resonators," US Patent No. 4,409,711, 1983.

10. R.L. Filler, "Acceleration Resistant Crystal Resonators," US Patent No. 4,410,822, 1983.

11. J.R. Vig and F.L. Walls, "Acceleration Insensitive Oscillators", U.S. Patent No. 4,575,690, 1986.

12. R.L. Filler, "Acceleration Resistant Crystal Resonators", U.S. Patent No. 4,410,822, 1983.

Effects of a NBC Nutrient Solution on Physiological and Psychological Status during Sustained Activity in the Heat (U)

*Madeleine Rose, LTC, Ralph Francesconi, Dr., Leslie Levine, Ms., Barbara Shukitt, Ms., Ilse Munro, Ms., Louis Banderet, Dr., Armand Cardello, Dr., Philip Warren, Dr., Paula Poole, Ms., Peter Frykman, Mr., Michael Sawka, Dr.
U.S. Army Research Institute of Environmental Medicine
U.S. Army Natick Research, Development & Engineering Center
Natick, MA 01760-5007

INTRODUCTION

Current doctrine and the design of the M-17A2 mask restricts semi-solid and solid food intake by soldiers in the mission-oriented protective posture 4 (MOPP4) ensemble. Soldiers will be in a fasting state unless they can enter collective protection units to eat (1) or they are provided a nutrient containing solution which will allow them to use the current mask drinking system. It had been predicted that soldiers would be in the MOPP4 ensemble no more than 24 hours; however, due to the unpredictability of war and providing food supplies in wartime, there is a possibility of longer periods of food deprivation.

In 1982 a National Research Council (NRC) Committee (2) recommended a nutrient solution that would be suitable to sustain military personnel in the Nuclear, Biological, and Chemical (NBC) ensemble for 24 hours (NBC Nutrient solution). The solution was not envisioned to meet all nutrient needs but would allow the individual to function for 24 hours while drawing on body nutrient stores. The committee felt that the casualty rate from heat stress for individuals engaged in heavy work while wearing protective clothing would not be reduced by any nutrient solution while a sedentary individual would only need water. The NRC Committee recommended the NBC Nutrient solution to help maintain performance and prevent heat injury for soldiers working for sustained periods (24 hours) of moderate activity (300 watt workload, WBGT 21°C) where sufficient thermal exchange could occur. The NBC Nutrient solution would enhance physical performance by: replacing body water losses, providing glucose to prevent hypoglycemia and ketosis, replacing approximately 50% of the Na^+ loss, and providing calories to lessen muscle glycogen depletion.

ROSE, FRANCESCONI, LEVINE, SHUKITT, MUNRO, BANDERET, CARDELLO, WARREN, POOLE, FRYKMAN, SAWKA

Fasting and exercise can decrease the performance of a person attempting to work for extended periods in the heat (3-6). Fasting decreases physical (3-7) and mental (4,8,9) performance by lowering blood glucose levels, depleting muscle glycogen stores, altering body fluid homeostasis, etc. Increased heart rates, nausea, and side pain are symptoms of prolonged fasting (4,5). Prolonged severe exercise in the heat can lead to hyperthermia (10,11), blood electrolyte imbalances (12), hypoglycemia (13,14), and depletion of muscle glycogen (15,16). Hypohydration is a major factor affecting performance and recovery from exercise (17-18). Loss of more than 2% body weight due to dehydration can decrease plasma volume, increase blood osmolality, decrease stroke volume and cardiac output, increase heart rate, increase core temperature, reduce sweat rate, and reduce cutaneous blood flow (18,19-24). These physiological alterations can be dangerous for men fasting and exercising in the heat.

The objectives of this study were to determine the effects of the NBC Nutrient solution on the following variables: nutritional status; thermoregulation; endurance capacity; vigilance and alertness; solution acceptance, hunger, and thirst as a function of time; cognitive state and motor performance skills; and subjective reactions.

METHODS

The NBC Nutrient solution was tested without MOPP encapsulation but under environmental conditions and workload which the NRC committee predicted that the solution might be beneficial. Fluid intake sufficient to replace sweat loss was encouraged during the test to prevent dehydration from complicating the comparison of the NBC Nutrient solution to water.

Twelve male subjects between the ages of 18 and 35 years volunteered for this test. The data from one subject were not included in some of the physiological and psychological analyses because of an upper respiratory infection that affected temperature, heart rate, etc. Each subject was tested on two different fluids. The NBC Nutrient solution contained 93.7 kcal, 24.8 g carbohydrate (2.34%), and 24.1 mEq Na^+ (0.056%) per liter of solution with an osmolality of 160-180 mOsm/kg. The control solution was composed of water containing the same amount of coloring and flavoring as was added to the NBC Nutrient solution. Six subjects were tested at any one time, the assignment of the solutions was counterbalanced to minimize the effects of order of administration.

The training and 24-hour Heat Stress portions of the study were conducted in a large environmental chamber equipped with two 4-man treadmills. The following conditions were carefully controlled during the 24-hour Heat Stress tests: fluid consumption was encouraged to maintain euhydration (0.5 liter/hour sweat rate), 24 hours of sustained activity (45 min work/45 min rest), a mean metabolic rate during exercise

of about 400 watt (3.5 mph, level treadmill), and a hot/dry environment (37°C, 20% rh, wind speed=2 mph). During the training phase, the subjects were heat- and exercise-acclimated and trained in all tasks that would be performed during the 24-hour tests: body weight, hydrostatic weighing, maximal aerobic power, urine and blood sampling, rectal temperature, skin temperature, metabolic rate, sweating rate, and heart rate. The assigned clothing was T-shirts, shorts, socks, and tennis shoes (MOPP0). Psychological tests included: Food Acceptance and Subjective Self-Report Questionnaire, Vigilance, Cognitive, Motor Performance, and Self-Reported Symptoms and Mood States tests. When the soldiers voluntarily ended participation or at the 24th hour, a Food Acceptance and Subjective Self-Report post-test questionnaire was administered. Details of methods used are described elsewhere (25).

RESULTS AND DISCUSSION

The physical characteristics of 11 subjects who completed the study are presented in Table 1. The mean percent body fat (15.9±6.5%) was average for young men but the values ranged from very lean to slightly obese. The relative $\dot{V}O_2max$ (54.2±7.8 ml/kg/min) indicated above average aerobic fitness for the group with a range from somewhat low to very high aerobic fitness (42.1-63.5 ml/kg/min). The total experimental time spent in exercising and testing (endurance time) for the two trials were not significantly (NS) different (X_{NBC}=17.1±3.6 h, 29.8 miles; $X_{control}$=16.0±2.9 h, 28.0 miles). The reasons that were given for discontinuing the experiment were varied, but the major reasons were sore knees, foot problems, and blisters from walking on the treadmill.

Table 1. Physical characteristics of subjects that completed the NBC Nutrient Solution and control solution trials (n=11).

Variable	Mean±SD	Range
Age, years	23.5± 4.6	19.00- 32.00
Height, cm	176.2± 5.2	165.00-183.00
Weight, kg	78.7±12.7	59.65- 97.05
Body fat, %	15.9± 6.5	6.60- 27.50
$\dot{V}O_2max$, l/min	4.2± 0.5	2.96- 4.98
$\dot{V}O_2max$, ml/kg/min	54.2± 7.8	42.10- 63.50

Results indicated that efforts to maintain euhydration by forcing fluids were successful (Table 2). Fluid balance and body weight data showed that the subjects were hydrated but some weight loss did occur (1.35% NBC, 1.98% control). The sweating rate and fluid intake data

indicated that the subjects were receiving enough fluid to replace sweat. While the difference in fluid intake was not statistically significant, the fluid intake for the control trial was 10% less than for the NBC trial. However, the relatively high urinary output placed the subjects in a slightly negative fluid balance which was corroborated by their weight loss. Pre-test differences in body weight between the two fluid trials were not significant which indicated that the subjects were able to maintain their baseline weights. There was a slightly greater (NS) overall weight loss pre- to post-test during the control trial (1.60±1.19 kg, 1.98%) compared to the NBC Nutrient solution trial (1.05±0.78 kg, 1.35%).

Table 2. Fluid balance for subjects consuming the NBC Nutrient solution and a control solution.[1,2]

FLUIDS	CONTROL SOLUTION	NBC NUTRIENT SOLUTION	p
OUTPUT			
Sweating Rate (ml/h)	518.0±205.0	547.0±194.0	NS
Urine Output (ml/h)	228.6±139.5	281.3±186.1	NS
INTAKE			
Fluid Intake (ml/h)	663.1±195.9	738.9±249.3	NS
BALANCE (ml/h)	-83.5	-89.4	

[1]Excluding subject #6.
[2]Mean±SD

Both solutions were effective in maintaining hydration as indicated by the clinical chemistries. Most differences remained in the clinically normal range. Plasma sodium levels and urinary specific gravity manifested no significant differences at any of the respective sampling times between trials. Plasma sodium levels for the NBC Nutrient solution trial were slightly higher than for the control trial (NS) which may have been due to the sodium content of the NBC Nutrient solution. Significant increments in either of these variables may be indicative of hypohydration. While plasma osmolality was not different between trials for the first three sampling intervals, by 8 hours into the study and at all sampling times thereafter, plasma osmolality was significantly ($p<0.05$) elevated during the NBC Nutrient solution trial. It is possible that the nutrient composition of the NBC Nutrient solution provided sufficient molecules and ions to the circulation during this trial to maintain the slightly elevated osmolality. It again should be emphasized, however, that even at the sampling time at which significant differences were observed, absolute differences in osmolality were minimal and during both trials the values were within the range of normal for euhydrated subjects.

The data clearly demonstrate that under these conditions, the NBC Nutrient solution provided no thermoregulatory or cardiovascular advantage compared to the control solution during the first 12 h. This is evidenced in the data observed for rectal (T_{re}) and skin temperatures, sweating rate, heart rate (HR), and metabolic rate. If anything, there is a tendency for most responses to be slightly elevated during the NBC Nutrient solution trial compared to the control trial. There are two hypotheses that can possibly explain this trend. The slightly elevated (NS) T_{re} observed during the NBC Nutrient solution trial (Figure 1) may reflect the higher plasma osmolalities. Previous research has shown an association between increased body temperature and increased plasma osmolalities (26). The increased thermal strain may then have caused the higher HR (Figure 2). The second possible explanation for these slight (NS) differences relates the higher T_{re} and HR to the somewhat elevated metabolic rates. It is possible that during the NBC Nutrient solution trial the subjects felt better and were somewhat more active, and therefore expended slightly more energy. However, during the metabolic data collection periods, great effort was made to control for any extraneous activity that would influence the measurements. However, our results show that these small differences between the NBC Nutrient and the control solutions were not significant, nor did they approach significance.

Controversy exists as to the importance of carbohydrate drinks in long term exercise (27,28). These drinks are of benefit when consumed during moderate and high intensity anaerobic exercise (29-31) because they provide glucose which is the principal source of energy for this type of exercise. The beneficial effects of carbohydrate feeding during light-moderate intensity aerobic exercise, such as in the present study, may be negligible because the exercise is fueled by lipid oxidation with a very small demand for carbohydrate (32). Since the subjects in the present study were engaged in a low intensity activity (29% $\dot{V}O_2max$) with intermittent rest periods, muscle glycogen stores were probably not reduced markedly. The amount of glucose ingested from the NBC Nutrient solution may have been sufficient to keep muscle glycogen stores from being depleted. Evidence to support this view could be found in the data on urinary ketones. Ketones are not usually produced until body glycogen stores are nearly depleted. The subjects drinking the NBC Nutrient solution generally did not spill ketones into their urine whereas 7/11 subjects drinking the control solution produced ketones.

The most significant metabolic differences ($p<0.001$) seen between the two test solution trials occurred in the levels of blood free fatty acids (FFA). Higher values usually occur in a state of fasting, exercise, or malnutrition. The increasing FFA levels were expected for the subjects ingesting the control solution because of their fasting state (Figure 3). The lower FFA levels for the subjects during the NBC trial indicated that the glucose in the NBC Nutrient solution may have decreased the need for

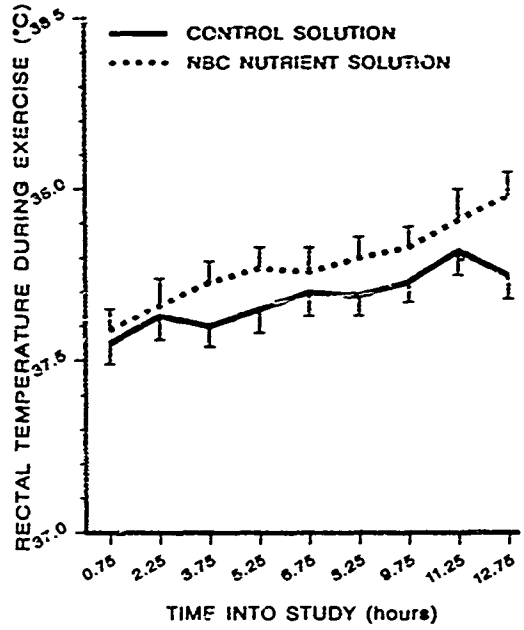


Figure 1. Rectal temperature during treadmill exercise in the heat.

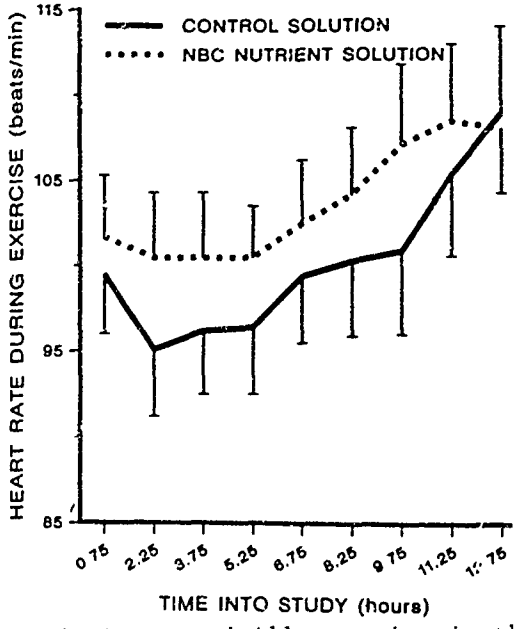


Figure 2. Heart rate during treadmill exercise in the heat.

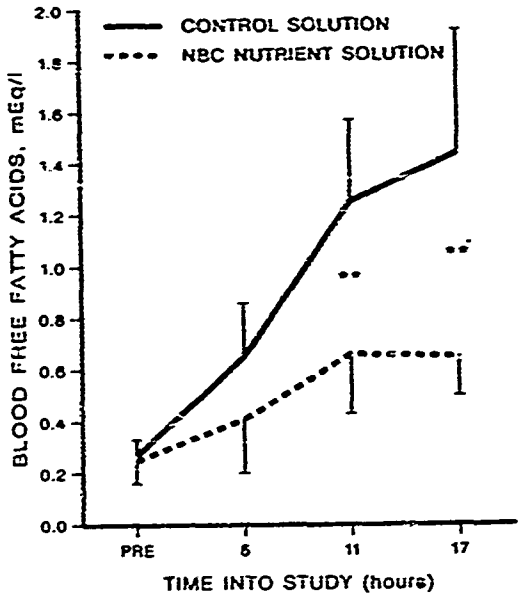


Figure 3. Blood free fatty acid levels during 24-h heat stress trials. Asterisks indicate post hoc significance levels at $p<0.01$.

fat oxidation to sustain energy requirements. This increased reliance on FFA for energy during the control trial was also reflected in the Respiratory Exchange Ratio (RER). The RER values were significantly different ($p<0.001$) between trials and over time (Figure 4). During the NBC Nutrient solution trial carbohydrate was being metabolized from the NBC Nutrient solution resulting in a higher RER whereas glucose was not exogenously available during the control trial. The ratio of carbon dioxide production to oxygen consumption (RER) averages 0.82 (23) under sedentary conditions when more carbohydrate is used than fat as the body's fuel. During steady-state submaximal exercise (as in the present study), the RER decreases indicating increased fat oxidation. Another indicator of increased fatty acid oxidation was the appearance of ketones in the urine (3,33) during the control trial. While drinking the NBC Nutrient solution, only two subjects showed urinary ketones and this occurred at the time that they withdr from the 24-h test (14th and 24th hours). During the control trial, 7 of the 11 subjects showed urinary ketones at some point after nine hours of testing. Of the 4 subjects who did not show any signs of ketones, 2 subjects had dropped out of the study by the 12th hour and the other 2 subjects had consumed the largest breakfasts (1100 and 1475 kcal) on the test day. The other subjects had consumed a third to a half that number of calories.

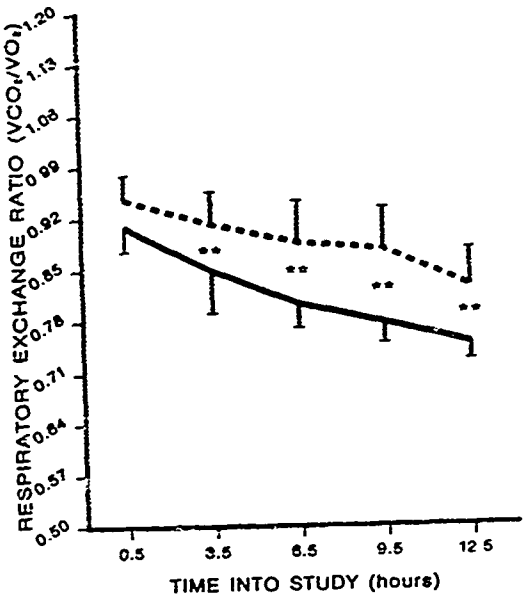


Figure 4. Respiratory exchange ratio (RER) changes during 24-h heat stress trials. Asterisks indicate post hoc significance at $p<0.01$.

The NBC Nutrient solution was expected to provide carbohydrate to counteract hypoglycemia (Figure 5). Carbohydrate is essential for proper functioning of the central nervous system (CNS). Under normal conditions and in short-term starvation, the brain uses blood glucose almost exclusively but has no stored supply of this nutrient (32,34). Symptoms of modest reductions in blood glucose include feelings of weakness, hunger, and dizziness which impair exercise performance (32). Sustained and profound low blood sugar can cause irreversible brain damage (32,34). As starvation or low carbohydrate intake progresses, the CNS switches from glucose to ketones as its primary fuel. Delaying the appearance of ketones may allow the CNS time to adapt to using large amounts of ketone bodies as fuel (35,36). If ketones are produced faster than they can be used, metabolic acidosis can develop and lead to hyperventilation, increasing drowsiness, and unconsciousness (37). Since the glucose levels did not drop below physiologically normal levels, it was not possible to assess the effects of either solution on hypoglycemia. However, the NBC solution was successful in delaying the appearance of ketones compared to the control solution.

There were no significant differences between the two solutions on the cognitive and motor performance tests; however, the final performance value for the control solution was less (NS) than that for the NBC

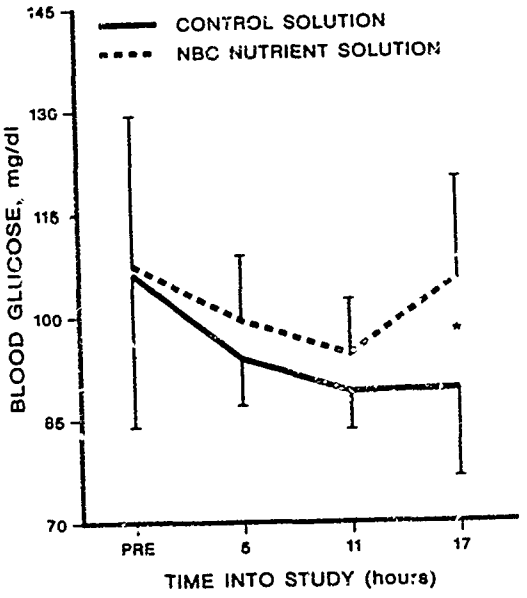


Figure 5. Blood glucose levels during 24-h heat stress trials. Asterisks indicate post hoc significance levels at $p<0.05$.

Nutrient Solution on seven of the nine cognitive tasks (third administration). It was not until the third administration on any of the cognitive tasks that values decreased from the pre-test scores. However, at this time (15-16 hours after the start of the test) only eight and four subjects remained in both conditions. Since cognitive performance did not decrease until the third administration, it is difficult to evaluate whether the NBC Nutrient Solution had a beneficial effect on performance. Another study in which the subjects endured longer or were placed under higher stress earlier would be needed to determine the value of the NBC Nutrient Solution for sustaining cognitive performance.

Statistical analyses of the two trials revealed no significant effects of either the NBC Nutrient solution or time for the four measures of vigilance that were obtained. It was anticipated that the combined effects of fatigue, hunger, etc. would produce a degradation of vigilance performance, and that the NBC Nutrient solution would have a positive effect on that degradation. While there is a hint that vigilance did decrease over time for both the control and the NBC Nutrient solution, the small number of subjects prevents any conclusions from being drawn.

Compared with the control solution, the NBC solution was associated with lower symptom intensity and improved mood (Figure 6). Overall, both the Total Symptom and the Total Mood Disturbance Scores were

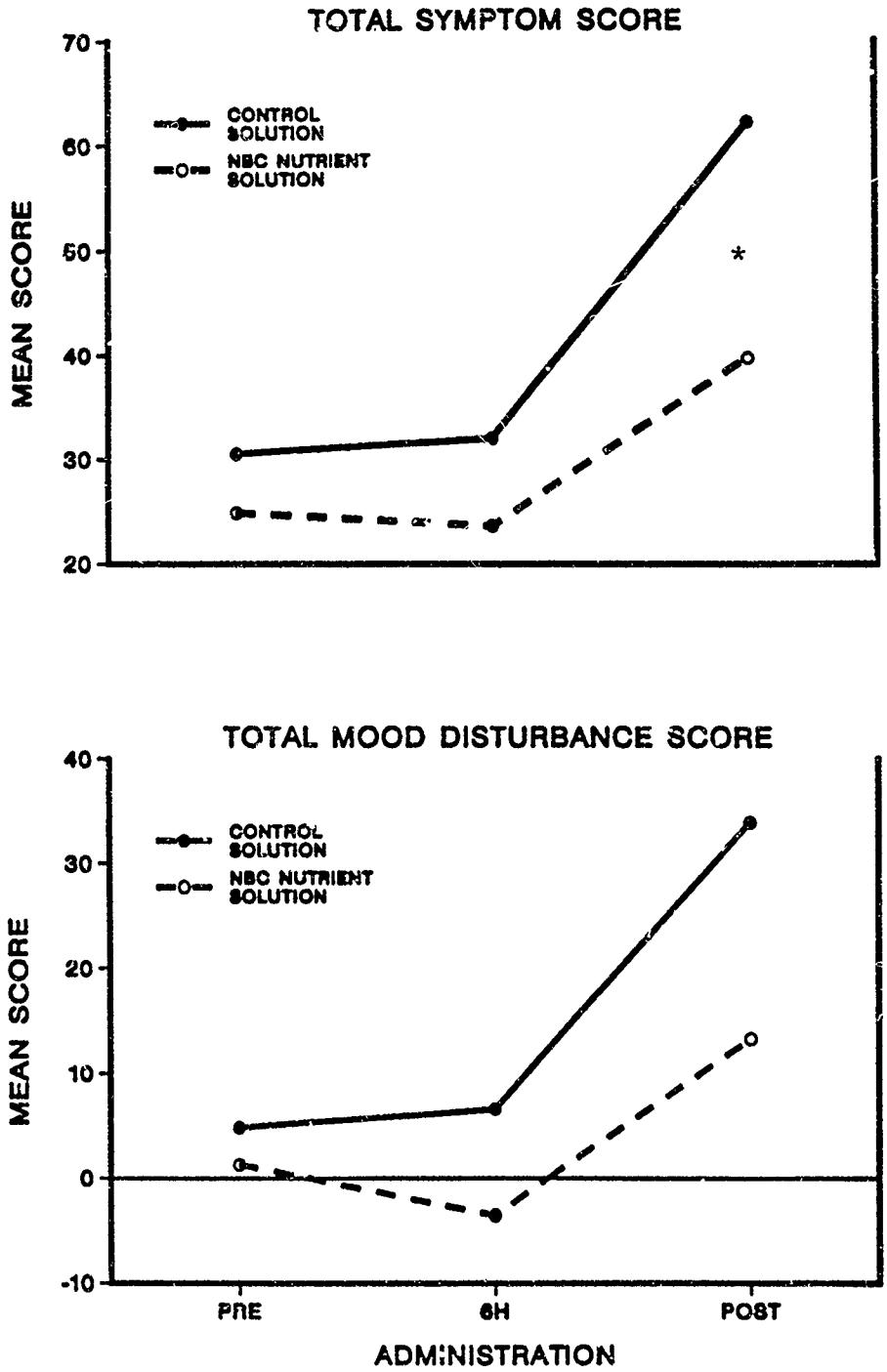


Figure 6. General symptom and mood effects. Asterisks indicate post hoc significance levels at p 0.05.

significantly ($p<0.05$) lower in the NBC solution condition. This effect was primarily due to changes that occurred after six hours. At the six hour administration, no significant differences from baseline were seen for either the symptom or mood measure. After six hours, there was still no significant change in mean intensity in the NBC solution condition while there was a significant increase in both measures ($p<0.01$) in the the control condition. At termination, the mean symptom score in the NBC solution condition was only 63% of that reported in the control condition and the mood disturbance score, which showed an even greater difference ($p<0.01$), was 39% of that seen in the control condition. It can be seen that there is considerable similarity between general changes in symptoms and mood. No significant effects related to the order in which the solutions were ingested were seen for either the symptom or mood score. After 6 hours, self-report measures showed that perceptual-motor, cognitive, emotional, gastrointestinal, and thermal problems were eliminated or attenuated when the NBC Nutrient solution was consumed. No harmful effects were reported. However, these findings do not constitute an unqualified endorsement of the solution. It is possible that the subjects were biased in favor of the NBC solution. The inadvertent omission of artificial sweetening from the control solution made it readily apparent to the subjects that they were consuming colored, flavored water. Nevertheless, the data do not suggest that the subjects were indiscriminately reporting higher levels of symptomatology and mood distress in the control condition.

It is difficult to generalize from the simulated test conditions to the NBC environment for which the solution is intended. Soldiers in running shoes and shorts do not experience all the same problems as those encapsulated in MOPP4. Breathing problems (painful breathing, difficulty breathing, shortness of breath), headache, and nausea were consistently greater in casualties (voluntary/medical terminations) than in survivors and are common complaints of soldiers in the MOPP4 configuration (38). The effect of the solution on respiratory problems could not be assessed during this study. Headache and nausea did show significant increases but since symptomatic treatment was offered by the medical monitor, it is possible that this obscured any differential effects of the NBC solution and control solution. Both the positive findings and the confounding factors suggest that a truly double-blind test of control vs the NBC solution be conducted in MOPP4.

Although the subjective evaluations do not show a major beneficial effect of the NBC Nutrient solution over the control solution, the NBC Nutrient solution was preferred over the control solution ($p<0.001$). There may have been some beneficial cognitive effects on perceived energy, psychological "lift", etc. The significant increase in hunger and decreases in self-reported alertness and strength over time are consistent with expected metabolic and physiological effects of consuming water or a calorie-containing fluid for an extended period of time while

treadmill walking in heat without prolonged rest. Although the data from the post-test questionnaire provide some evidence of a beneficial effect of the NBC Nutrient solution on perceived energy and providing a psychological "lift", the within-test data show only a general trend that the NBC Nutrient solution ameliorates these effects. The acceptance data from both the within-test and post-test questionnaires are clear in showing that subjects preferred to drink the NBC Nutrient solution over the control solution. Although preference itself may not be the most critical factor in developing a nutrient replacement fluid, preference is very frequently correlated with voluntary consumption which is critical under conditions of high heat stress and sweating.

In summary, under conditions simulating MOPP4 configuration at a WBGT of 21°C and moderate exercise, water and the NBC Nutrient solution were equally effective for the first 12 hours in maintaining hydration, heart rate, core temperature, skin temperature, vigilance, cognitive performance, and motor performance. The results of this study supported previous findings (11,18-24) that the amount of fluid ingested rather than the composition of the fluid was the important factor in maintaining body temperature and associated physiological processes during exercise in the heat. There were no significant differences between the two solutions and over time on cognitive, motor performance, and vigilance tests. The subjects generally performed better on the psychological tests while drinking the NBC Nutrient solution but no conclusions could be drawn because of the small number of subjects that continued past the 12th hour. The difference between the two solutions might not have been noticeable because the subjects were dropping out of the study prior to the onset of hypoglycemia and just as ketosis was developing. The NBC Nutrient solution was effective in improving mood and lessening symptom intensity after 6 hours of exercise in the heat. The data upheld the recommendations of the NRC Committee that water would be sufficient for the first 6 hours without a significant degradation in physiological and psychological performance during light activity at a WBGT of 21°C (70°F). The study provided further evidence that water could be sufficient for up to 12 hours if a meal had been eaten prior to encapsulation.

This study could not provide conclusive data after the 12th hour because of the decrease in the number of subjects; however, it is possible to speculate that the NBC Nutrient solution may have been more effective than water for the 12-24 hour time period. The delay in appearance of ketones, higher blood glucose levels, improvement in mood, decrease in symptom intensity, higher acceptability ratings, and the tendency for better performance on the vigilance, cognitive and motor performance tests indicated that the NBC Nutrient solution may be of some benefit to the soldier after 12 h of fasting. However, the results of this study did not indicate a clear superiority of the NBC Nutrient solution over plain water when consumed in amounts adequate to replace fluid loss. It would be difficult to generalize to soldiers encapsulated

in MOPP4 from soldiers in running shoes and shorts because they did not experience all the same problems. This study was designed to determine if the contents of the NBC Nutrient solution would enhance performance compared to water and it showed that when fluids were forced, water was as effective as the NBC Nutrient solution for the first 12 hours.

Providing glucose in dilute concentrations as in the NBC Nutrient solution prevents a hyperosmolar effect and allows for maximal fluid delivery to the body to prevent gastro-intestinal disturbances and dehydration. Any method of providing nutrients to a soldier in MOPP4 would cause some logistical and operational problems but it is recommended that every effort be made to provide nutrients to the soldier to prolong his ability to work.

This study was not designed to determine the effects of ad libitum consumption of fluids. Previous research showed that subjects tended to voluntarily dehydrate when allowed to drink ad libitum (18). The higher acceptability ratings ($p<0.01$) and the slightly higher consumption (NS) of the NBC Nutrient solution might improve fluid intake over that of water and therefore the ability to maintain hydration in soldiers who are not being forced to drink. It is recommended that a study be specifically designed to evaluate long term ad libitum consumption of the NBC Nutrient solution during prolonged work in the heat in MOPP4.

REFERENCES

1. Headquarters Dept of the Army. NBC Protection, Field Manual No. 3-4. Washington, DC: US Army Adjutant General Publications Center, 1985.
2. Conclusions and recommendations arising from a workshop, held June 3-4, 1982, to determine nutritional requirements of military personnel in protective clothing, Committee on Nutritional Requirements in Protective Clothing, Food and Nutrition Board, Commission on Life Sciences, National Research Council.
3. Taylor HL, Henschel A, Mickelsen O, Keys A. Some effects of acute starvation with hard work on body weight, body fluids and metabolism. J Appl Physiol 1954;6:613-623.
4. Henschel A, Taylor HL, Keys A. Performance capacity in acute starvation with hard work. J Appl Physiol 1954;6:624-633.
5. Consolazio CF, Nelson RA, Johnson HL, Matoush LO, Krzywicki HJ, Issac GJ. Metabolic aspects of acute starvation in normal humans: Performance and cardiovascular evaluation function. (Laboratory Report No. 298) Denver, CO: US Army Med Res Nut Lab, 1966.
6. Lategola MT. The effect of 5-day, complete starvation on cardiopulmonary functions of aerobic work capacity and orthostatic tolerance. Fed Proc 1965;24:590.
7. Taylor HL, Brozek J, Henschel A, Mickelsen O, Keys A. The effect of successive fasts on the ability of men to withstand fasting during hard work. Am J Physiol 1945;143:148-155.

8. Guetzkow H, Taylor HL, Brozek J, Keys A. Relationship of speed of motor reaction to blood sugar level during acute starvation in man, abstracted. Fed Proc 1945;4:28.
9. Kety SS. The general metabolism of the brain in vivo. In: Richter D, ed. Metabolism of the nervous system. New York: Pergamon Press, 1957:221-237.
10. Costill DL, Sherman WM, Fink WJ, Maresh C, Witten M, Miller JM. The role of dietary carbohydrates in muscle glycogen resynthesis after strenuous running Am J Clin Nutr 1981;34:1831-1836.
11. Costill DL, Kammer WF, Fisher A. Fluid ingestion during distance running. Arch Environ Health 1970;21:520-525.
12. Frizzell RT, Lang GH, Lowance DC, Lathan SR. Hyponatremia and ultramarathon running. JAMA 1986;255:772-774.
13. Pirnay F, Lacroix M, Mosora F, Luyckx A, Lefebvre P. Effect of glucose ingestion on energy substrate utilization during prolonged muscular exercise. Eur J Applied Physiol 1977;36:247-254.
14. Costill DL, Bennett A, Branam G, Eddy D. Glucose ingestion at rest and during prolonged exercise. J Appl Physiol 1973;34:764-769.
15. Karlsson J, Saltin B. Diet, muscle glycogen, and endurance performance. J Appl Physiol 1971;31:203-206.
16. Ahlborg B, Bergstrom J, Ekelund LG, Hultman E. Muscle glycogen and muscle electrolytes during prolonged physical exercise. Acta Physiol Scand 1967;70:129-142.
17. Williams MH. Nutritional aspects of human physical and athletic performance. Springfield, IL: Charles C. Thomas Publisher, 1976:44-75 and 169-207.
18. Strydom NB, Wyndham CH, van Graan CH, Holdsworth LD, Morrison JF. The influence of water restriction on the performance of men during a prolonged march. So Afr Medical J 1966;31:539-544.
19. Pitts GC, Johnson RE, Consolazio FC. Work in the heat as affected by intake of water, salt and glucose. Am J Physiol 1944;142:253-259.
20. Ladell WSS. The effects of water and salt intake upon the performance of men working in hot and humid environments. J Physiol 1955;127:11-46.
21. Wyndham CH. Heat stroke and hyperthermia in marathon runners. In: Milvy P, ed. The marathon: Physiological, medical, epidemiological, and psychological studies. New York: New York Academy of Sciences, 1977:128-138.
22. Keys A, Brozek J, Henschel A, Mickelsen O, Taylor HL. Human starvation. U. of Minn Press, 1951.
23. Consolazio CF, Johnson RE, Pecora LJ. Physiological measurements of metabolic function in man. New York: Blakiston Division, McGraw-Hill Book Co., 1963.
24. Saltin B. Aerobic and anaerobic work capacity after dehydration. J Appl Physiol 1964;19:1114-1118.

25. Rose MS, Franesconi RP, Levine L, Shukitt B, Munro I, Banderet L, Frykman P, Sawka M. Effects of a NBC nutrient solution on physiological and psychological status during sustained activity in the heat. Technical Report No. T25-87. Natick, MA: US Army Research Institute of Environmental Medicine, 1987.
26. Sawka MN, Young AJ, Francesconi RP, Muza SR, Pandolf KB. Thermoregulatory and blood responses during exercise at graded hypohydration levels. J Appl Physiol 1985;59:1394-1401.
27. Felig P, Cherif A, Minagawa A, Wahren J. Hypoglycemia during prolonged exercise in normal men. N Engl J Med 1982; 306:895-900.
28. Van Handel PJ, Fink WJ, Branam G, Costill DL. Fate of ^{14}C glucose ingested during prolonged exercise. Int J Sports Med 1980;1:127-131.
29. Hargreaves M, Costill DL, Coggan A, Fink WJ, Nishibata I. Effect of carbohydrate feedings on muscle glycogen utilization and exercise performance. Med Sci Sports Exerc 1984;16:219-222.
30. Costill DL, Miller JM. Nutrition for endurance sport: Carbohydrate and fluid balance. Int J Sports Med 1980;1:2-14.
31. Coyle EF, Hagberg JM, Hurley BF, Martin WH, Ehsani AA, Holloszy JO. Carbohydrate feeding during prolonged strenuous exercise can delay fatigue. J Appl Physiol 1983;55:230-235.
32. McArdle WD, Katch FI, Katch VL. Exercise physiology: Energy, nutrition, and human performance. 2nd ed. Philadelphia, PA: Lea and Febiger, 1986.
33. Drury DR, Wick AN, MacKay EM. The action of exercise on ketosis. Am J Physiol 1941;134:761-768.
34. Guyton, AC. Textbook of medical physiology. 6th ed. Philadelphia, PA: W.B. Saunders Co., 1981:971-972.
35. Owen OE, Morgan AP, Kemp HG, Sullivan JM, Herrera MG, Cahill GF. Brain metabolism during fasting. J Clin Inves 1967;46:1589-1595.
36. Cahill GE, Aoki TT. Partial and total starvation. In: Assessment of energy metabolism in health and disease. Columbus, OH: Ross Laboratories, 1980:129-134.
37. Bell GH, Emslie-Smith D, Paterson CR. Textbook of physiology and biochemistry. 9th ed. NY: Churchill Livingstone, 1976:195-196.
38. Munro I, Rauch TM, Tharion W, Banderet LE, Lussier AR, Shukitt B. Factors limiting endurance of armor, artillery, and infantry units under simulated NBC conditions. Proceedings, Army Science Conference, 17-19 June 1986;111:85-96.

Advanced Gunn Oscillators For Millimeter-wave Smart Munition Systems (U)

Raymond L. Ross, Mr., Arthur C. Paolella, Mr., *Raymond A. Ross, Mr.
Deborah A. Semasko, Ms., and Robert P. Moerkirk, Mr.
US Army Electronics Technology and Devices Laboratory
Fort Monmouth, New Jersey 07703-5000

INTRODUCTION

Millimeter-wave smart munition systems require low cost, efficient GaAs devices. For many systems, cost and performance considerations dictate the use of GaAs Gunn devices for VCO applications. However, current state-of-the-art GaAs Gunn oscillators are limited in both efficiency and output power. Historically, attempts to improve these low values have been mainly directed at improving the thermal efficiency of the device by various heat sinking techniques, i.e., plated heat sink, integral heat sink, diamond impregnation, and, most recently, by substrate removal. Little or no research has been directed at redesigning the electronic structures of the basic Gunn diode in an attempt to improve its thermal dissipation properties. One such approach would be to control the doping profile of the active region to account for nonuniform thermal dissipation in that layer.

For the first time, millimeter-wave GaAs Gunn structures have been grown by computer-automated molecular beam epitaxy (MBE) wherein the active region is graded in impurity concentration. By shaping the carrier concentration along the active n-layer, the deleterious temperature effects produced by standard flat profile designs can, essentially, be eliminated. A number of structures with varying profiles have been processed, packaged, and evaluated for Ka (26.5-40 GHz) and V (50-75 GHz) band operation. Output powers exceeding 300 mW at nearly 7 percent efficiency have been obtained in these device structures at Ka-band while 90 mW at 2.75 percent efficiency has been achieved at V-band. When compared with similar structures grown with flat active-region profiles, a 200 percent increase in output power and nearly a 300 percent increase in efficiency is realized. Operating at 5 to 6 V and 700 to 900 mA, these devices are similar to present devices and, therefore, are ideal candidates for retrofit into present systems requiring no change in power supply requirements or current design. In addition, the improvement in

device performance resulting from temperature-compensated impurity profiles can be expected to reduce prime power requirements for next-generation millimetric systems, as well as impacting new monolithic millimeter-wave circuit designs being developed for these systems.

PROFILE DESIGN CONSIDERATIONS

Standard GaAs Gunn diode structures are generally n++nn+ epitaxial layers grown on low resistivity n++ substrates wherein the n-active layer has a flat doping profile. When the device is biased in either polarity, heat sink negative (cathode) or heat sink positive (anode), mobility gradients develop as a result of heat which is generated mainly within the active region. A significant improvement in overall performance can be realized by designing doping profiles such that the resistivity ratio $\overline{\rho}/\rho_{max} = 1$ along the active region at the device operating temperature. As a result, the efficiency and available output power increases.

The basis for altering the carrier concentration along the n-active layer can be understood from the following first order considerations. The resistivity $\rho = (q\ \mu_n\ n)^{-1}$ is essentially dependent on mobility μ_n which is a strong function of temperature as shown in Fig. 1. Near the Debye temperature $\mu_n \alpha T^{-1}$ and since $n \simeq$ constant to 525°K, then $\rho \alpha T$.

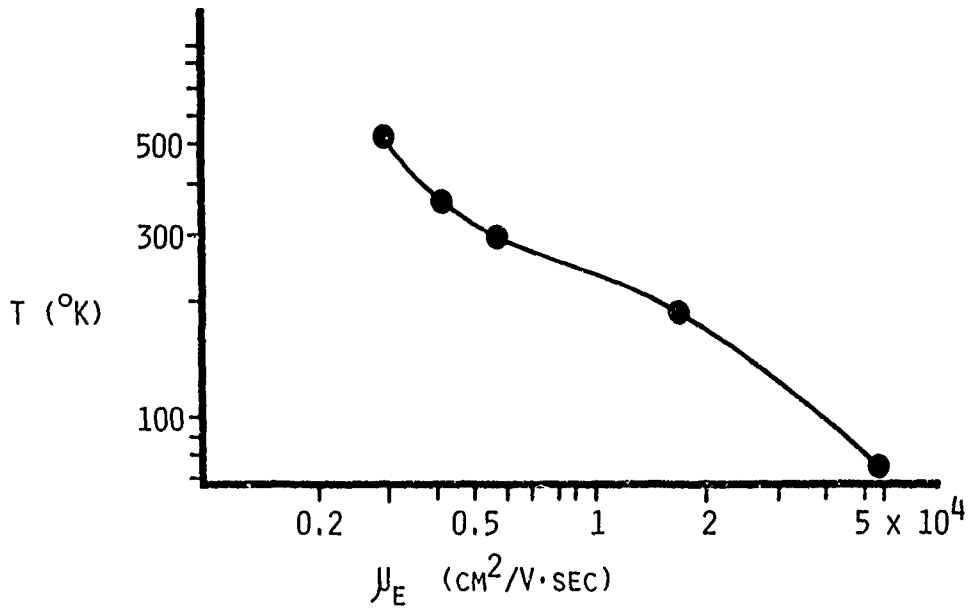


Fig. 1. Electron mobility versus temperature over the normal operating range of a GaAs Gunn device.

For a device with a uniformly doped active layer, the temperature profile in that layer varies exponentially as determined from solution of the heat flow equation.[1] Accordingly, the doping concentration is graded exponentially to compensate the variation of resistivity produced by the temperature gradient.

Referring to Fig. 2, the slope parameter α is defined as $\alpha = n_1/n_2$ where n_1 and n_2 are the carrier concentrations in the active region at the active-buffer and active-contact interfaces, respectively. As noted, a flat active layer ($\alpha = 1$) device can be operated with either bias polarity, heat sink-cathode or heat sink-anode. For such a structure, the maximum device temperature, T_{max}, which occurs at the active-buffer interface[1] can be reduced up to two thirds by operating the Gunn device with its anode toward the heat sink, i.e., heat sink positive (anode), compared to the conventional heat sink negative (cathode) configuration.[2] This reduction of T_{max} occurs because maximum heat is generated where the electric field is largest, namely, at the anode. With a positive heat sink, the anode location is closer to the heat sink and, therefore, in a more favorable location for heat extraction. Accordingly, improved performance can be realized in a flat active layer device simply by operating the device in the heat sink-anode configuration.

For most system applications, Gunn diodes are biased in the heat sink-cathode configuration For these applications, improved performance over flat active region devices can be realized by gradir · this layer such that $\alpha > 1$. Such devices must be operated with the substrate positive for

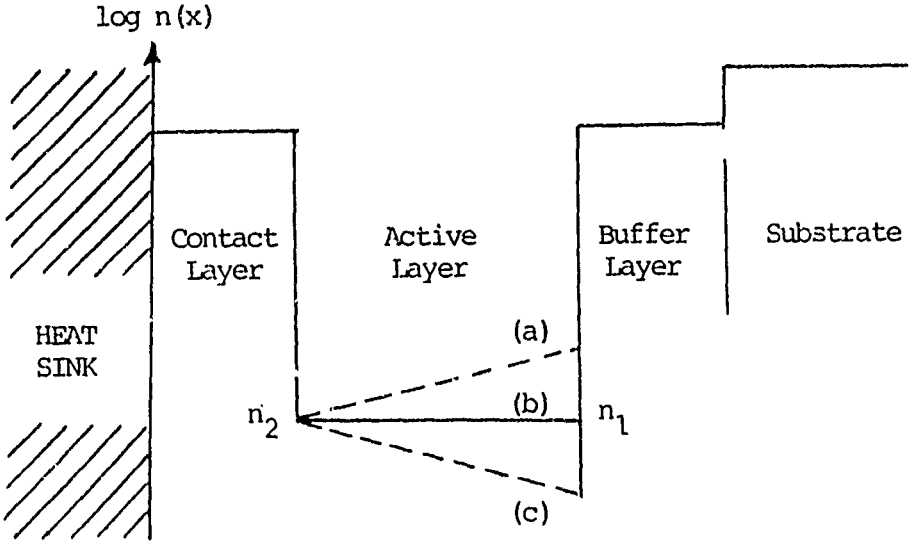


Fig. 2. Gunn diode doping profile with slope parameter $\alpha = n_1/n_2$: (a) graded active region, $\alpha > 1$; (b) conventional flat profile, $\alpha = 1$; (c) graded active region, $\alpha < 1$.

optimum output power and efficiency and to avoid device burn out. For this case, measured output power for devices having $\alpha>1$ always exceeded that for devices with $\alpha = 1$. Even further improvement of rf performance can be achieved by grading the active layer concentration n(x) such that $\alpha < 1$ and operating the device in the heat sink-anode configuration. This is the optimum case for the reasons explained; producing superior rf performance when compared with devices having flat ($\alpha= 1$) or positive slope ($\alpha > 1$) active region profiles.

FABRICATED DEVICE DETAILS

All GaAs epitaxial layers were grown in a Varian GEN-II MBE system (Fig. 3) which has been computer automated as described earlier.[3] Wafers were processed using both flip chip (Fig. 4a) and integral heat sink (IHS) technology (Fig. 5a) with initial mesa diameters of 0.003" to 0.004". The wafers were diced to form individual chips, 0.008" to 0.010" on a side, and then thermocompression bonded onto the pedestal of a standard picopill millimeter-wave package (Fig. 4b, 5b). Contact to the top surface was made via a gold maltese cross-shaped preform (Fig. 6) after which the device was in-package etched to exhibit a dc threshold current value of approximately 1.0 A. Final assembly consisted of hermetically sealing the package with a gold plated top cap.

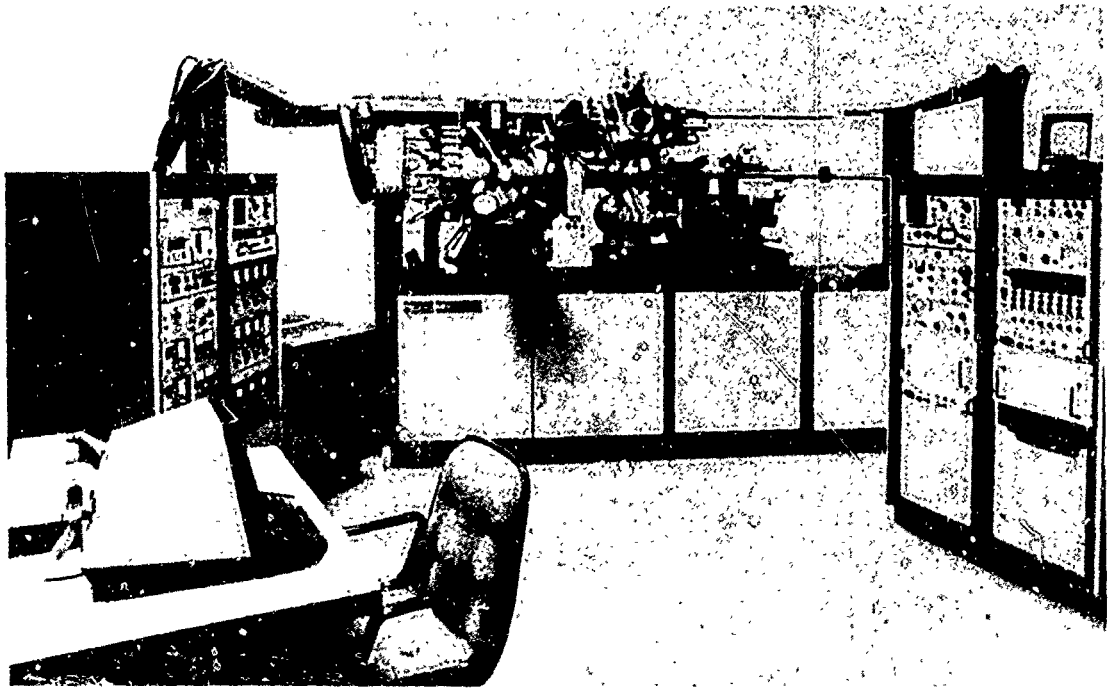


Fig. 3. Varian GEN-II molecular beam epitaxy system.

optimum output power and efficiency and to avoid device burn out. For this case, measured output power for devices having $\alpha>1$ always exceeded that for devices with $\alpha = 1$. Even further improvement of rf performance can be achieved by grading the active layer concentration n(x) such that $\alpha < 1$ and operating the device in the heat sink-anode configuration. This is the optimum case for the reasons explained; producing superior rf performance when compared with devices having flat ($\alpha= 1$) or positive slope ($\alpha > 1$) active region profiles.

FABRICATED DEVICE DETAILS

All GaAs epitaxial layers were grown in a Varian GEN-II MBE system (Fig. 3) which has been computer automated as described earlier.[3] Wafers were processed using both flip chip (Fig. 4a) and integral heat sink (IHS) technology (Fig. 5a) with initial mesa diameters of 0.003" to 0.004". The wafers were diced to form individual chips, 0.008" to 0.010" on a side, and then thermocompression bonded onto the pedestal of a standard picopill millimeter-wave package (Fig. 4b, 5b). Contact to the top surface was made via a gold maltese cross-shaped preform (Fig. 6) after which the device was in-package etched to exhibit a dc threshold current value of approximately 1.0 A. Final assembly consisted of hermetically sealing the package with a gold plated top cap.

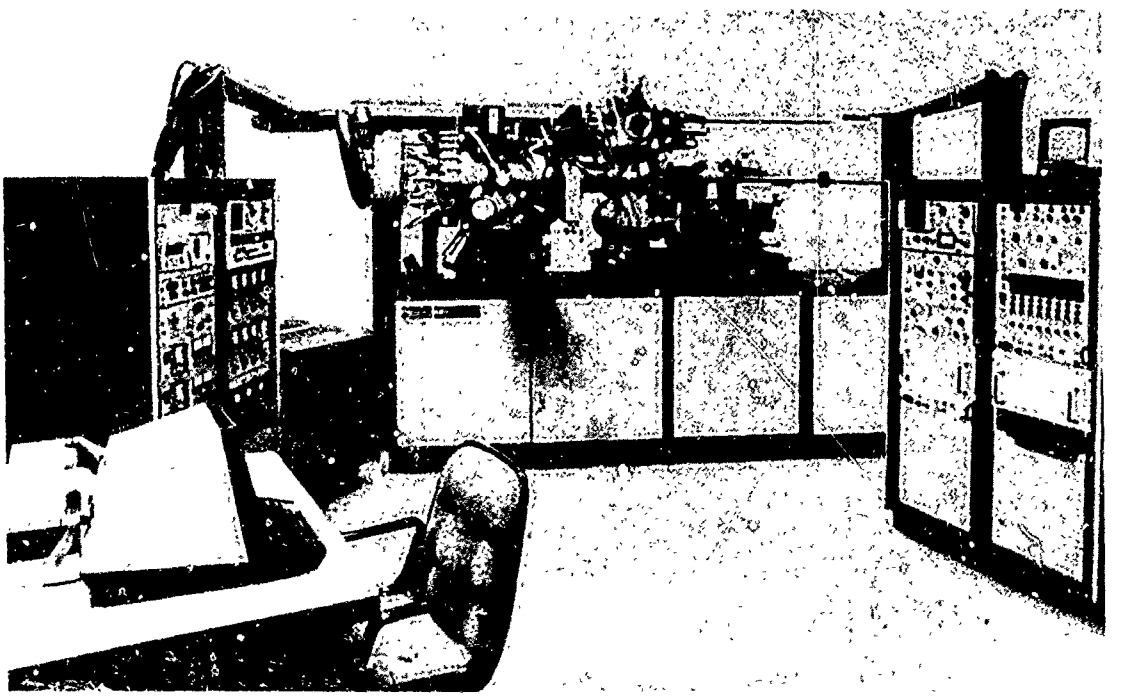
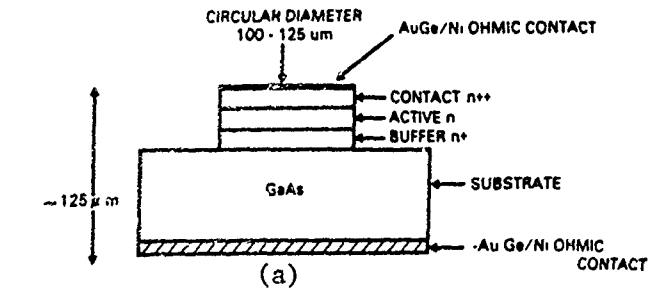


Fig. 3. Varian GEN-II molecular beam epitaxy system.



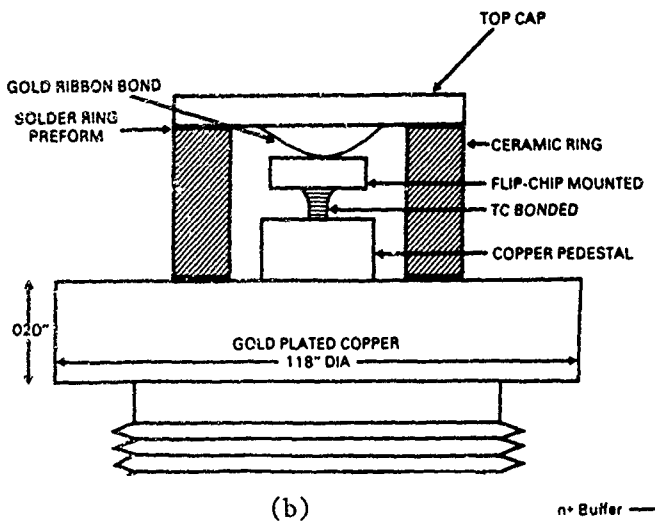


Fig. 4. GaAs Gunn device: (a) die for flip chip mounting; (b) flip chip device in millimeter-wave package.

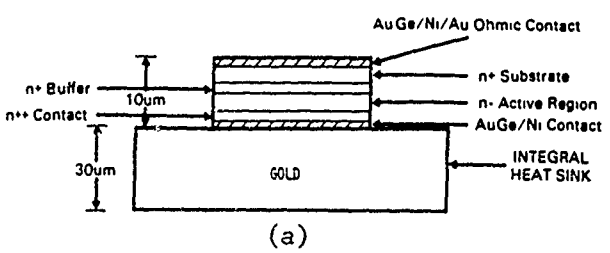
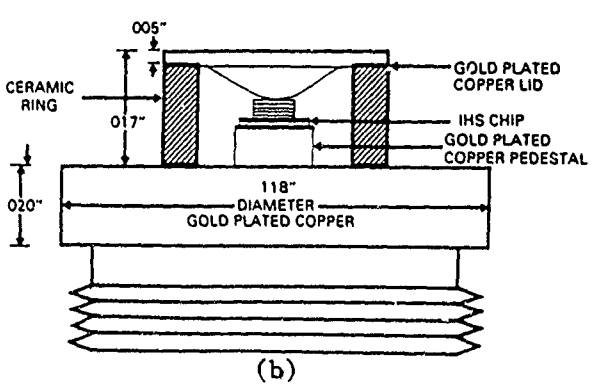


Fig. 5. GaAs Gunn device: (a) die with integral heat sink; (b) integral heat sink device in millimeter-wave package.



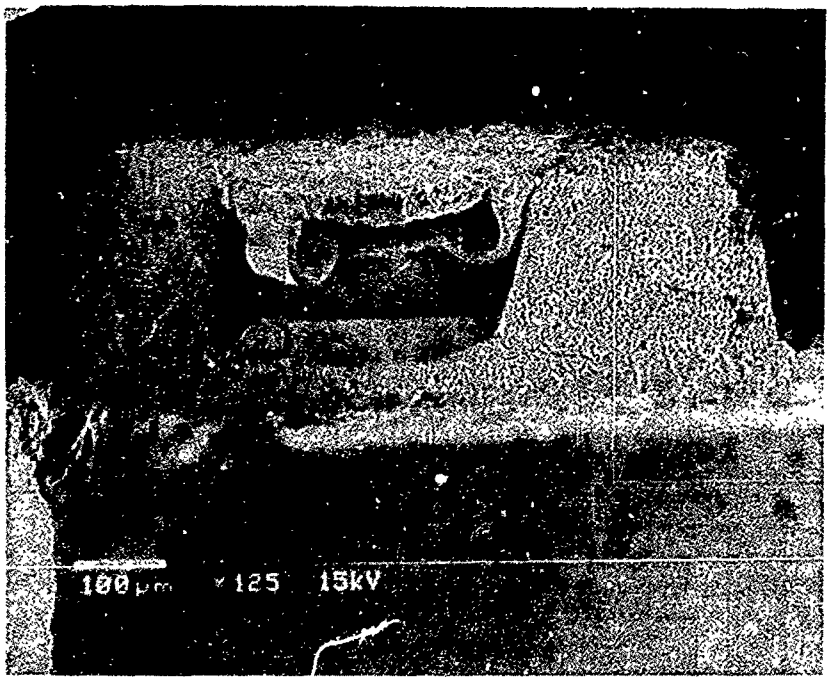
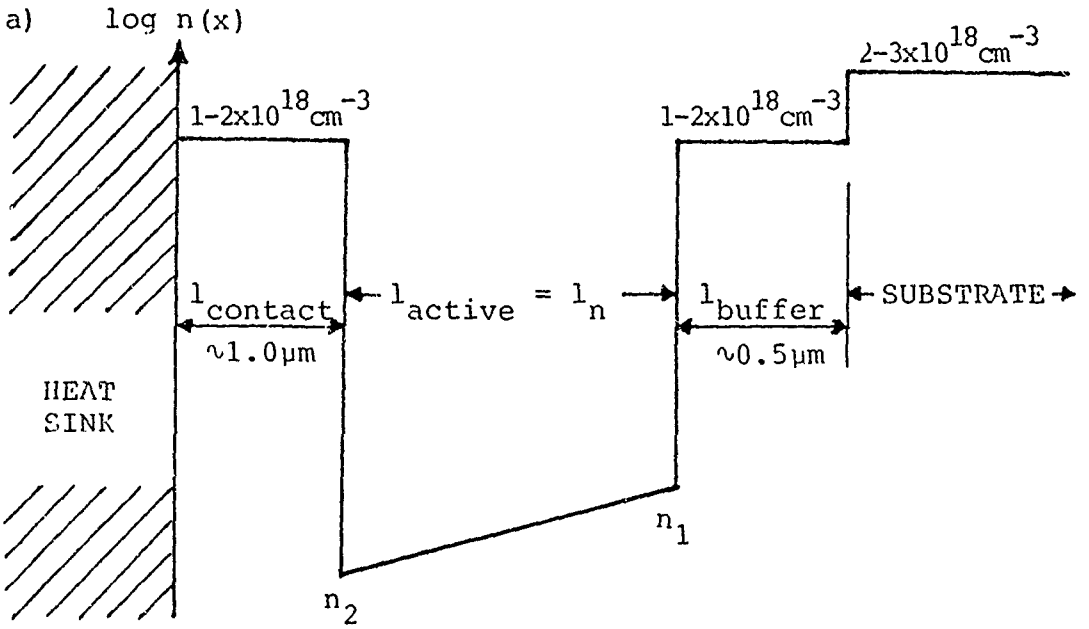


Fig. 6. SEM photo of GaAs Gunn device in millimeter-wave package.

MEASURED RESULTS

Selected devices were evaluated for rf performance in a full-height, inductive post-coupled, rectangular-waveguide cavity oscillator at Ka-band[3] and a full height, radial disk, second harmonic test oscillator at V-band[4]. Variation of post diameter permitted operation over the frequency band of interest while output power was maximized with an adjustable back short. Doping profiles and best measured rf results for various Ka-band Gunn devices having doping ratio $\alpha \geq 1$ are summarized in Fig. 7. For flat active region profiles ($\alpha = 1$) optimum rf performance, as obtained from devices fabricated from wafer G-43, produced a best output power of 220 mW with a 4.4 percent efficiency. Devices fabricated on wafer G-38 did not oscillate as the active region's carrier concentration was too high resulting in an n·l product which was nearly twice the recommended design value of $1.2\text{-}2.0\times10^{12}$ cm^{-2}. In comparison to the rf performance results for the flat profile devices, maximum output power for the exponentially graded structures was obtained from devices fabricated on wafer G-67, which had the largest slope parameter and an average active region carrier concentration approximating that of the best performing flat profile device, G-43-13. Typical output power from the G-67 group ranged from 280 to 320 mW; the highest being 325 mW operating at an efficiency of 6.7 percent. The highest efficiency of all devices tested was 6.9 percent; obtained from a device producing 290 mW and fabricated from wafer G-67. Although both G-66 and G-68 have nearly identical active layer lengths and slope parameters, their active layer concentration

a)



b)

Sample Number	n_1 (cm^{-3})	n_2 (cm^{-3})	l_n (μm)	$\alpha = n_1/n_2$
G-38	1.3×10^{16}	1.3×10^{16}	2.97	1.00
G-43	6.0×10^{15}	6.0×10^{15}	2.83	1.00
G-66	1.55×10^{16}	1.15×10^{16}	2.96	1.35
G-67	1.1×10^{16}	4.6×10^{15}	3.22	2.39
G-68	5.5×10^{15}	4.1×10^{15}	2.80	1.34

c)

Device Number	V_{th} (Volts)	I_{th} (Amps)	V_{op} (Volts)	I_{op} (Amps)	FREQ (GHz)	PWR (mW)	EFF (%)
G-38							
G-43-13	2.0	1.13	6.0	0.83	27.40	220	4.4
G-66-12	2.2	1.30	6.0	1.10	28.05	260	4.0
G-67-2	2.1	1.10	6.1	0.80	28.10	325	6.7
G-68-8	2.0	0.95	5.5	0.71	28.80	240	6.1

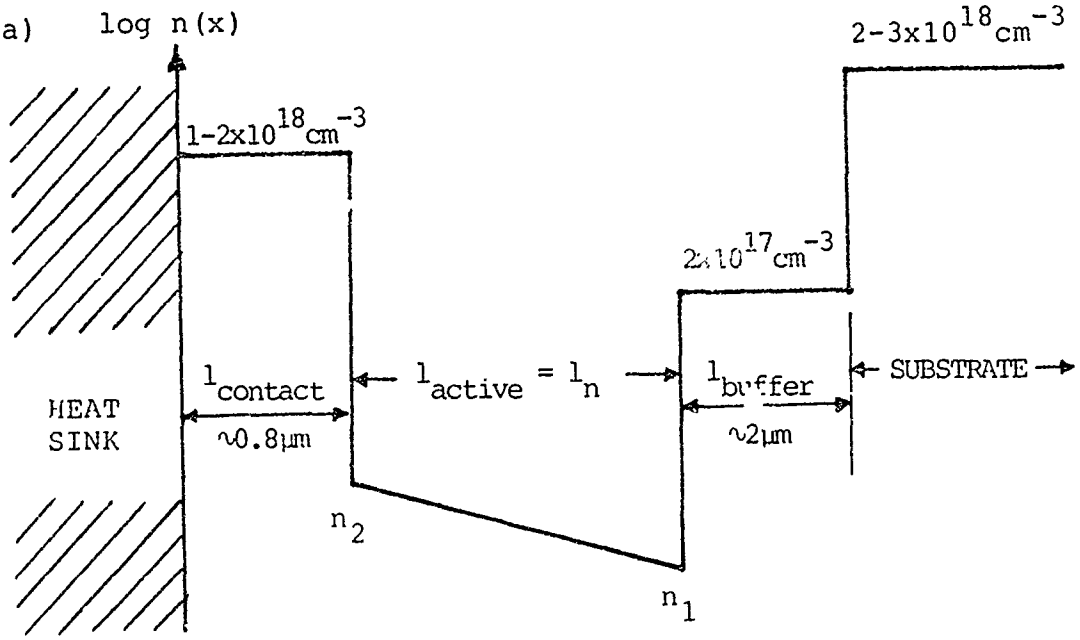
Fig. 7. GaAs bulk effect mmW (a) doping profile with slope parameter $\alpha = n_1/n_2$; (b) active region slope parameter values; (c) best performance results for Ka-band (26.5-40 GHz) flip chip mounted fundamental mode devices having flat and positive ($1.0 \leq \alpha \leq 2.4$) active layer slopes.

levels are significantly different, as shown in Fig. 7b. Of these two groups, G-66 devices, with an average active layer carrier concentration approximating that of the non-oscillating G-38 group, produced the greater power, while G-68 group devices with their lower average active region carrier concentration and rate of exponential grading, operated at greater efficiencies. Except for a small region, the efficiency of all graded region devices exceeded that of the best performing flat profile structure. Finally, as noted above, maximum output power and efficiency for flat profile structures occur when the device is operated in the heat sink-anode configuration. Indicative of the improved rf performance which can be achieved with this configuration is device G-43-14 which produced a 30 percent increase in output power at an efficiency of 5.2 percent when the device was biased heat sink positive.

Doping profiles and test measured rf results for various Ka and V-band Gunn devices having doping ratio $\alpha \leq 1$ are summarized in Fig. 8. As expected, simply operating the conventional flat profile G-44 structure in the heat sink-anode configuration produced a 20 percent increase in output power at an efficiency of 5.3 percent at Ka-band while performance at V-band resulted in a 50 percent increase in output power at an efficiency of nearly 1 percent over operation of the same device in the heat sink-cathode mode.

Maximum Ka-band output power of 345 mW at an efficiency of 6.8 percent was obtained at 31.2 GHz for a device from wafer G-111. This device also produced the greatest V-band (67.1 GHz) output power (90 mW) and efficiency (2.75 percent). Compared to the conventional flat profile G-44 devices operating heat sink-anode, wafer G-111 devices operating in the same mode produced 47 percent more output power with 28 percent greater efficiency at Ka-band while at V-band these figures increased to 100 percent more output power with over 200 percent greater efficiency. When compared to flat profile G-44 devices operating in the normal heat sink-cathode mode, wafer G-111 devices operating in the preferred heat sink-anode mode produced 77 percent more output power with 39 percent greater efficiency at Ka-band increasing dramatically to three times greater output power (200 percent increase) at nearly four times greater efficiency (293 percent increase) at V-band. Summarized in Fig. 9, these Ka and V-band rf performance figures establish the state-of-the-art in GaAs Gunn devices. It is to be noted that the maximum output power was always obtained from devices with the steepest active region doping gradient. In general, both G-111 and G-118 devices exhibited similar performance; however, the reliability of G-118 devices at cold temperatures (-54°C) exceeded that of G-111 devices. This is due, in part, to the lower active layer average doping concentration.

a)



b)

Sample Number	n_1 (cm^{-3})	n_2 (cm^{-3})	l_n (μm)	$\alpha=n_1/n_2$
G-44	9.0×10^{15}	9.0×10^{15}	2.60	1.000
G-111	6.7×10^{15}	1.3×10^{16}	2.50	0.5154
G-118	6.0×10^{15}	1.0×10^{16}	2.62	0.6000

c)

Device Number	V_{on}* (Volts)	V_{op}* (Volts)	I_{op} (mA)	FREQ (GHz)	PWR (mW)	EFF (%)
G-44/2/1-4	+4.26	+4.72	843	31.0	195	4.9
	+5.80	+6.20	691	67.0	30	0.7
G-44/2/1-4	-4.44	-4.82	920	30.9	235	5.3
	-5.45	-5.95	840	67.1	45	0.9
G-111/1/5-12	-4.60	-5.20	976	31.2	345	6.8
	-5.41	-6.33	518	67.8	90	2.75
G-118/3/5-12	-5.35	-5.80	688	31.1	310	6.4
	-6.20	-6.79	511	67.95	88	2.53

*Polarity of substrate relative to heat sink

Fig. 8. GaAs bulk effect mmW (a) doping profile with slope parameter $\alpha = n_1/n_2$; (b) active region slope parameter values; (c) best performance results for Ka (26.5-40 GHz) and V (50-75 GHz) band IHS constructed devices having flat and negative ($0.515 \leq \alpha \leq 1.0$) active layer slopes.

Device Profile	Device Number	Mode of * Operation	Percent Increase			
			Ka-band		V-band	
			Power	Eff	Power	Eff
(I) Flat (II) Flat	G-44/2/1-4 G-44/2/1-4	HS-C HS-A	20	8	50	29
(I) Flat (II) Graded	C-44/2/1-4 C-111/1/5/12	HS-A HS-A	47	28	100	205
(I) Flat (II) Graded	G-44/2/1-4 C-111/1/5/12	HS-C HS-A	77	39	200	293

*HS-C = Heat Sink-Cathode
HS-A = Heat Sink-Anode

Fig. 9. Percent increase in output power and efficiency at Ka (26.5-40 GHz) and V (50-75 GHz) bands of device (II) over device (I) as a function of device profile and mode of operation.

RELIABILITY STUDIES

Selected devices from wafers G-111 and G-118 (10 each) were tested prior to and after a seven day burn-in, under operating conditions. Both dc and rf performance remained essentially constant in all cases. Devices were temperature cycled over a -54°C to +85°C temperature range. Devices from G-118 all exhibited well behaved operation over this full mil-spec temperature range while G-111 devices had normal operation over a -20°C to +85°C range. Some G-111 devices failed at temperatures of -20°C or lower.

Five G-118 devices which survived initial seven day burn-in and temperature cycling over a -54°C to +85°C range were selected for long term performance and reliability studies. Devices from G-111 were omitted from these studies since failure occurred (short circuit) in some devices at low temperature. The devices under test were operated cw into a matched load in an oven environment at an elevated temperature of 110°C. At various intervals in time, the rf performance of the five oscillators was measured at room temperature, without circuit adjustment, for variation of threshold current, operating current, output power, frequency, and efficiency. Essentially no notable changes in these measured parameters have occurred after 10,000 hours of continuous operation at 110°C. Since no device failure or major degradation has occurred, it is not possible to

predict MTTF. Extrapolation of these results to +85°C yields a continuous operation of 50,000 hours while at room temperature (+25°C) this amounts to 1.7×10^5 hours, or 19.4 years!

CONCLUSIONS

Precise control of doping concentration and impurity profile by computer automated MBE permits the growth of nearly ideal GaAs Gunn structures. By grading the doping profile along the active n-region to offset mobility changes caused by self-heating, enhanced operation and improved device performance has been achieved at Ka and V-bands. The output power and efficiency of devices produced from this material exceed those obtained from conventional flat profile devices operated in both the heat sink-cathode and the preferred heat sink-anode modes establishing new state-of-the-art performance figures for GaAs Gunn devices. Repeated temperature cycling and accelerated life tests have proven the reliability of these structures. These are the first reported GaAs Gunn oscillators to operate over the full mil-spec temperature range while accelerated life test studies have indicated no device failure and no notable change in rf performance after an equivalent room-temperature operation of 19.4 years. Operating at voltages and currents of present commercial Gunn diodes, these devices are ideal candidates for retrofit into existing systems, requiring no change in power supply requirements or current design. In addition, the improvement in device performance resulting from temperature-compensated impurity profiles can be expected to reduce prime power requirements for next-generation millimetric systems, as well as impacting new monolithic millimeter-wave circuit designs being developed for these systems.

ACKNOWLEDGMENTS

The authors are grateful to the numerous individuals who have contributed to the research presented here. In particular, to Ms. S. Mason for growth of MBE materials, Mr. D. Knight and Mr. D. Worley under the supervision of Mr. S. Neylon for IHS device processing, to Ms. J. Findlay and Mr. D. Worley for device rf evaluation and Dr. John Ondria for technical discussions.

REFERENCES

1. Knight, S., "Heat Flow in n++-n-n+ Epitaxial GaAs Bulk Effect Devices", Proc. of the IEEE, pp. 112-113, January 1967.

2. Johnson, N. O., Olsson, K. O. I. and Wildheim, S. J., "Temperature in Gunn Diodes with Inhomogeneous Power Dissipation", IEEE Trans. E. D., Vol. ED-18, No 3, pp. 158-165, March 1971.

3. Paolella, A., Ross, R. L. and Ondria, J., "Advanced mm-Wave Sources by Automated MBE", Microwave Journal, pp 149-159, April 1986.

4. Ondria, J., "Wide-Band Mechanically Tunable and Dual In-Line Radial Mode W-Band (75-110GHz) CW GaAs Gunn Diode Oscillators," Proc. Seventh Biennial Cornell Electrical Engineering Conf. on Active Microwave Semiconductor Devices and Circuits, Vol. 7, Cornell University, Ithaca, New York, pp. 309-320, August 14-16, 1979.

Stimulated Brillouin Scattering Experiments With Flashlamp-Pumped Dye Lasers (U)

*Mr. Stephen D. Russell, Mr. John J. Ehrlich,
Mr. Stanley P. Patterson, and Dr. Raymond W. Conrad
U.S. Army Missile Command

INTRODUCTION

Flashlamp-pumped lasers, on the whole, suffer from beam qualities that go from bad to worse as the output energies are scaled from a few joules to tens of joules per pulse. This degredation can be caused in a flowing dye media by such things as turbulence, optical disturbances due to temperature gradients, and acoustic shock waves from flashlamps. Beam qualities of 50 to 100 times diffraction limited are common and spatial filtering, with significant energy loss, does not appear to be the best way to achieve a nearly diffration - limited beam.

In recent years, many types of lasers and optical systems have benefited from the use of non-linear optical phase conjugation to compensate for optical inhomogeneities similar to those found in flashlamp-pumped dye lasers. Of the various methods for achieving phase conjugation, SBS seems to hold the most promise for use with flashlamp-pumped dye lasers.

In order to achieve SBS threshold, a combination of spectral linewidth, beam quality (divergence), and power requirements must be met. The difficulty in obtaining these often mutually exclusive requirements from a single laser cavity probably explains why SBS has not been previously demonstrated with this type of laser. One possible solution to this problem is an oscillator/single-pass amplifier configuration.

EXPERIMENTAL DETAILS

The experimental setup was as shown in Fig. 1. The oscillator was a Phase-R Model 1400 coaxial flashlamp-pumped dye laser. It was configured with a 2 meter long cavity containing 3 turning/polarizing prisms and 2 etalons. The oscillator output was then expanded to fill a Candela FLDL-20 linear flashlamp-pumped dye laser. The laser output from this oscillator/single-pass amplifier system was then focused into a high

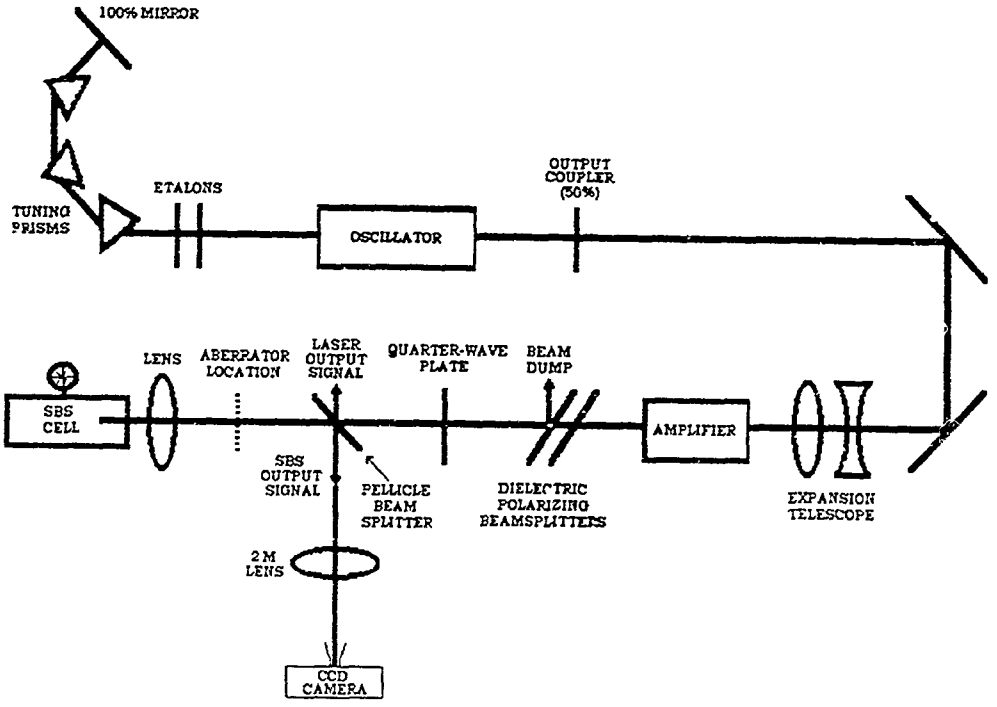


Fig. 1

Oscillator/Amplifier configuration. The SBS experiment was performed by observing the SBS output signal and comparing it with the laser output signal. To investigate the phase conjugate ability of the SBS return, an acid etched glass plate could be inserted (aberrator location) and the SBS return could then be compared to the return obtained when the lens/cell combination was replaced with an ordinary miror. The quarter waveplate/polarizers combination was intended to prevent energy from re-entering the laser system.

pressure methane gas cell. One quarter-wave plate and two dielectric polarizing beamsplitters were used in combination to try to prevent the SBS return from the gas cell from re-entering the laser system. A pellicle beamsplitter was placed just before the focusing lens so that energy measurements and beam profiles could be taken of the incident and SBS returned beams. Space was left between the beamsplitter and focusing lens for the occasional insertion of an aberrator to test for the phase conjugation property of wavefront correction.

All experiments were performed using Rhodamine 6G as the laser dye with wavelength operation of the oscillator set at 590nm. The 2 etalons were removable, thus allowing linewidths of 10GHz or 100GHz. Once the system was operational it was observed that the output of the amplifier faithfully followed the linewidth and beam quality of the oscillator.

RESULTS

SBS was successfully demonstrated in the methane cell and reflectivity data (energy out of cell vs energy into cell) was taken for various combinations of linewidth, focal lengths, and gas pressure. A portion of this data is presented in Figs. 2 and 3. Fig. 2 shows 10GHz and 100GHz operation at a constant focal length and pressure while Fig. 3 shows operation with two different focal lengths at constant linewidth and pressure.

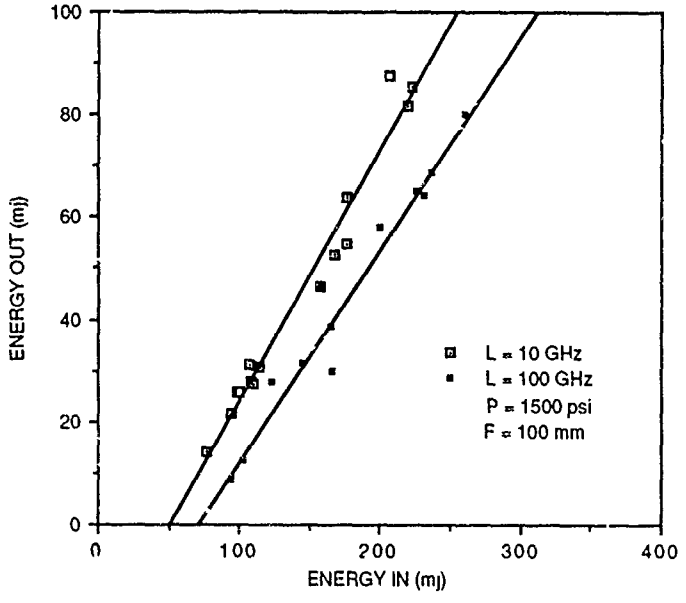


Fig. 2

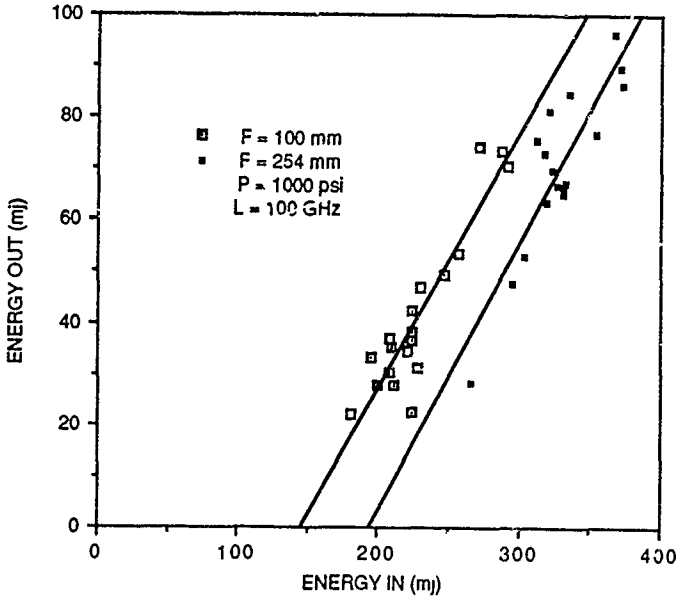


Fig. 3

Typical pulse waveforms taken are shown in Figs. 4-8. Figs. 4 and 5, taken when the SBS cell was absent from the setup, illustrate that the oscillator and amplifier pulse shapes were identical. Figs. 6 and 7, taken off of the beamsplitter, show input and output waveforms when the SBS cell was present and when the laser energy was well above the SBS threshold. A comparison of Fig. 6 with Fig. 8 indicates that the presence of the SBS cell caused the laser to continue to lase even after output from the oscillator had stopped. A closer inspection of this continued, apparently mode-locked lasing action (Fig. 9) revealed a period of 35 usec which was the same as the 35 usec round trip time for light traveling between the 100% oscillator mirror and the SBS cell. The continued lasing had two causes: first, the electrical inductance of the flashlamp circuit in the amplifier caused the amplifier flashlamp pulse to be longer than that in the oscillator; and second, laser energy from the SBS cell was getting back into the amplifier. The reason that this small amount of energy fedback into the amplifier was possibly due to a slight depolarization effect at some point. Thus, due to this feedback, a phase conjugate resonator was formed.

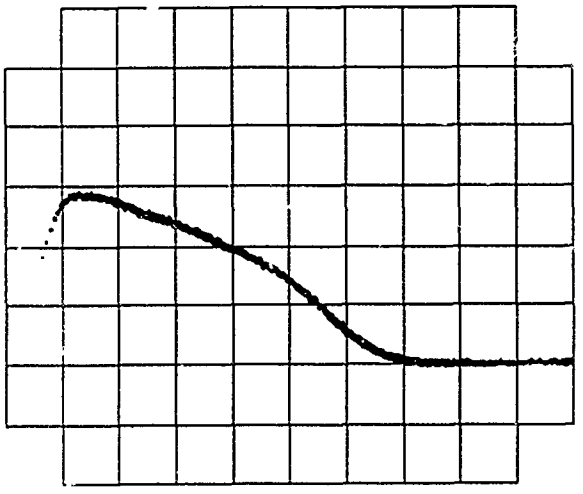


Fig. 4
Oscillator pulse. 100 nSec/Division

Fig. 5
Oscillator/Single pass amplifier pulse. 100 nSec/Division

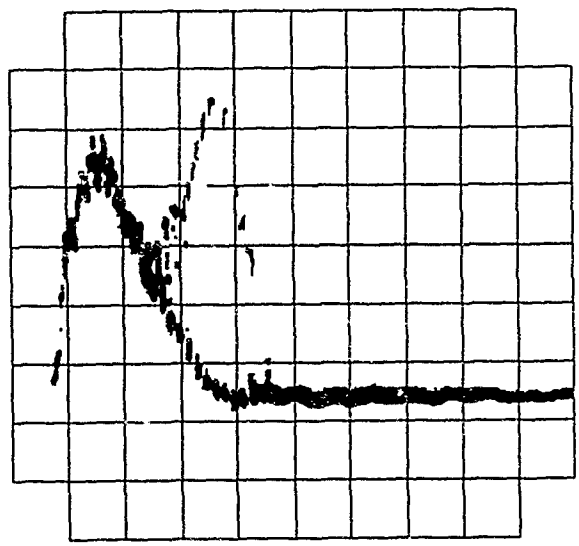


Fig. 6
Laser output waveform with SBS cell. 200 nSec/Division

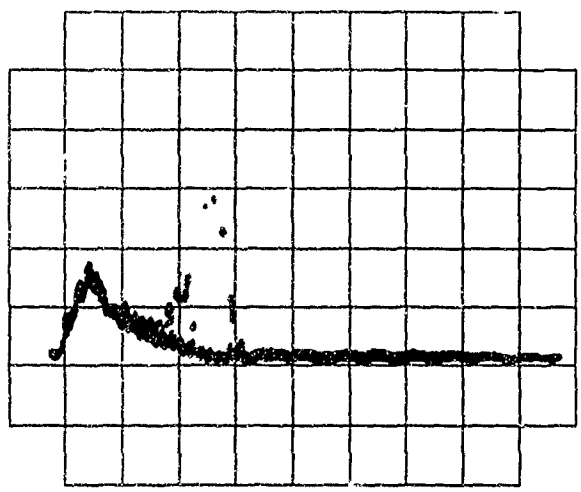


Fig. 7
SBS output waveform. 200 nSec/Division

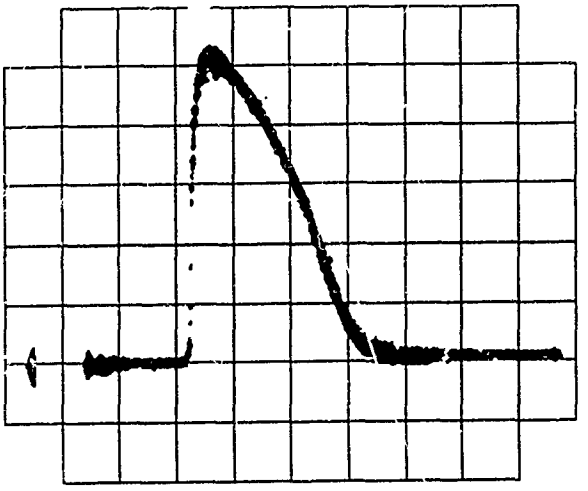


Fig. 8
Laser output waveform without SBS cell. 200 nSec/Division
This is the same as Fig. 5 except for time scale.

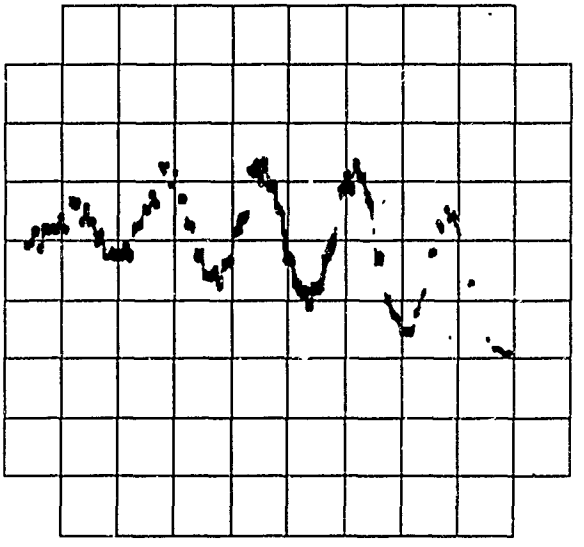


Fig. 9
Expanded view (20 nSec/Division) of the beginning
of the "mode-locking" action.

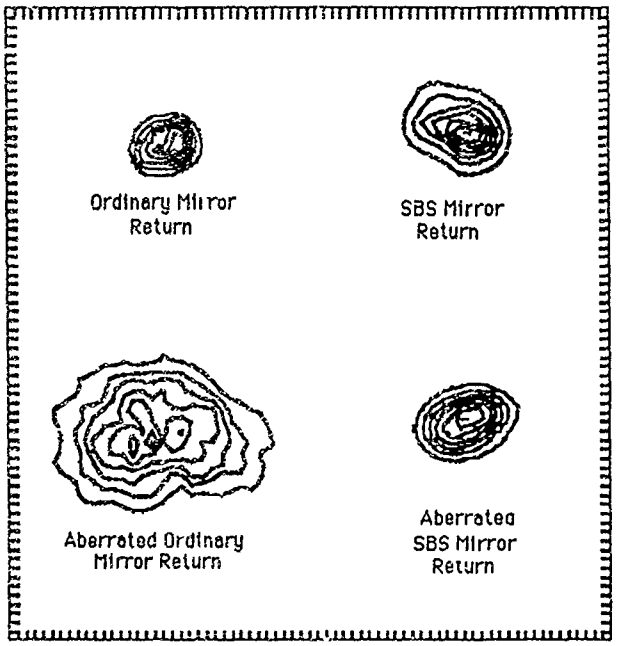


Fig. 10
Intensity contours taken with the CCD camera. The ordinary mirror return differs from the SBS mirror return in that the SBS cell and lens were replaced with an ordinary laser mirror. The spot size difference is explained in the test. Both of the aberrated returns show the effect of the insertion of an aberrator placed in front of each of the mirrors.

In Fig. 10 the phase conjugate ability to compensate double-passed aberrations is displayed. The intensity profiles between the two SBS mirror returns show no difference other than shot-to-shot variation, while the ordinary mirror return and aberrated ordinary mirror return are strikingly different. The spot size difference between the ordinary mirror return and the SBS mirror returns was due to the fact that the light beam incident on the mirror, SBS or ordinary, was slightly divergent. This yielded a slightly divergent return beam for the ordinary mirror and a slightly convergent return beam for the SBS mirror. Therefore, the SBS return beam reached a focus somewhat in front of the camera position.

An appreciation of phase conjugation can be found in the realization of the fact that when the SBS mirror is used, the phase plate aberrator is actually intracavity. A normal laser cavity would never lase with the presence of such an aberration plate.

CONCLUSION

Phase conjugation via SBS is possible with flashlamp-pumped dye lasers. By designing a PCR with dye lasers, it is possible to greatly improve the beam quality, and hopefully, as ongoing tests will show, to achieve higher output energies with large lamp systems. With the improvements already obtained, it appears that phase conjugation has already made a significant impact on the future outlook for high average power visible dye lasers.

☆U.S GOVERNMENT PRINTING OFFICE 1988 627 323/ 80066