

**TECHNICAL
LIBRARY**

AD-A056400



Army Science Conference Proceedings

20 - 22 June 1978

Volume I

Principal Authors A thru F

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

**Deputy Chief of Staff for Research,
Development & Acquisition
Department of the Army**



DEPARTMENT OF THE ARMY
OFFICE OF THE ADJUTANT GENERAL AND THE ADJUTANT GENERAL CENTER
WASHINGTON, D.C. 20314

HQDA Ltr 70-78-5

DRXRO-SO (M) (31 Aug 78)

15 September 1978

Expires 15 September 1979

SUBJECT: 1978 Army Science Conference, Volume I

SEE DISTRIBUTION

1. Inclosed for your information and use is Volume I of the 1978 Army Science Conference Proceedings. This volume contains the unclassified papers by principal authors A through F which were presented at the conference, 20-22 June 1978, U. S. Military Academy, West Point, New York.
2. Volumes II, III and IV of the Proceedings are being distributed separately.
3. Chiefs of Army Staff Agencies, major commanders, and heads of laboratories on the distribution list will insure that this volume is placed in the organization's Technical Library where it will be available for reference.
4. The expiration date does not apply to the 1978 Army Science Conference Proceedings which should be retained for reference purposes.

BY ORDER OF THE SECRETARY OF THE ARMY:

C. F. Briggs
C. F. BRIGGS
Colonel, AGC

Acting The Adjutant General

1 Incl
Vol I, 1978 Army Science
Conference Proceedings



DISTRIBUTION:

Director of Army Research
Office of the Surgeon General
Office, Chief of Engineers, ATTN: DAEN-RDZ-A
Office, Chief of Engineers, ATTN: DAEN-ZC
Office, Chief of Engineers, ATTN: DAEN-MCZ-S
Office, Chief of Engineers, ATTN: DAEN-CWM-R
The Army Library, ATTN: ANRAL-RS

COMMANDERS

US Army Materiel Development & Readiness Command, ATTN: DRCLDC
US Army Armament Research & Development Command
US Army Aviation Research & Development Command
US Army Armament Materiel Readiness Command
US Army Communications Research & Development Command
US Army Electronics Research & Development Command
US Army Mobility Equipment Research & Development Command
US Army Missile Research & Development Command
US Army Natick Research & Development Command
US Army Tank-Automotive Research & Development Command
US Army Communications & Electronics Materiel Readiness Command
US Army Missile Materiel Readiness Command
US Army Tank-Automotive Materiel Readiness Command
US Army Troop Support & Aviation Materiel Readiness Command
US Army Test & Evaluation Command
US Army Computer Systems Command
US Army Forces Command
US Army Ballistic Missile Defense Systems Command
US Army Training & Doctrine Command
US Army Intelligence & Security Command
US Army Electronics Materiel Readiness Activity
US Army Aircraft Development Test Activity
US Army Foreign Science & Technology Center
US Army Materials & Mechanics Research Center
US Army Cold Regions Test Center
US Army Tropic Test Center
US Army Ballistic Missile Advanced Technology Center
US Army Operational Test & Evaluation Agency
US Army Environmental Hygiene Agency
US Army Medical Research & Development Command
US Army Health Services Command
US Army Institute of Surgical Research
US Army Aeromedical Research Laboratory
US Army Medical Bioengineering R&D Laboratory
US Army Research Institute of Environmental Medicine
US Army Medical Research Institute of Infectious Diseases
US Army Tripler Medical Center
US Army Institute of Dental Research

SUBJECT: 1978 Army Science Conference, Volume I

US Army Research Institute for the Behavioral & Social Sciences
Walter Reed Army Institute of Research
Walter Reed Army Medical Center
Dwight D. Eisenhower Army Medical Center
Brooke Army Medical Center
Fitzsimons Army Medical Center
Madigan Army Medical Center
Letterman Army Institute of Research
Armed Forces Institute of Pathology
US Army Research Office
US Army Research & Standardization Group (Europe)
US Army Aberdeen Proving Ground
US Army Dugway Proving Ground
US Army Yuma Proving Ground
US Army Electronics Proving Ground
US Army White Sands Missile Range
US Army Watervliet Arsenal
US Army Rock Island Arsenal
US Army Signal Center & Fort Gordon
US Army Combat Surveillance & Target Acquisition Laboratory
US Army Atmospheric Sciences Laboratory
Redstone Scientific Information Center, ATTN: Chief, Documents
Harry Diamond Laboratories
Chemical Systems Laboratory
Fort Huachuca
Los Alamos Scientific Laboratory
US Army Engineer Waterways Experiment Station
US Army Cold Regions Research & Engineering Laboratory
US Army Engineer Topographic Laboratories
US Army Coastal Engineering Research Center
US Army Construction Engineering Research Laboratory

DIRECTORS

US Army Human Engineering Laboratory
US Army Materiel Systems Analysis Activity
US Army Ballistic Research Laboratory
US Army Logistics Center
US Army Concepts Analysis Agency
US Army Applied Technology Laboratory
US Army Acquisition Information Systems Agency
US Army Research & Technology Laboratories (USARTL), Ames Research Ctr
Applied Technology Laboratory, USARTL, Fort Eustis, VA
Structures Laboratory, USARTL, Langley Research Center, Hampton, VA
Propulsion Laboratory, USARTL, Lewis Research Center, Cleveland, OH
Lawrence Livermore Laboratory

Large Caliber Weapons Laboratory
Night Vision & Optics Laboratory
Benet Weapons Laboratory

PROJECT MANAGERS

Army Tactical Data Systems (ARTADS), Fort Monmouth
Remotely Monitored Battlefield Sensor Systems (REMBASS), Fort Monmouth
Army Tactical Communications Systems (ATACS), Fort Monmouth

SUPERINTENDENT

US Army Military Academy, ATTN: Technical Library
US Army Military Academy, ATTN: Acquisition Branch

COPIES FURNISHED:

Office, Under Secretary of Defense for Research & Engineering
Office, Deputy Director for Science & Technology, Defense Nuclear Agency
Chairman, Defense Science Board
Defense Advanced Research Projects Agency
National Security Agency
Institute for Defense Analysis
Defense Communication Engineering Center
Defense Communication Agency
Defense Science Board
Assistant Chief of Staff for Intelligence
Scientific Advisor, Code RD, HQ US Marine Corps
HQ, National Aeronautics & Space Administration (Code W)
National Bureau of Standards
National Science Foundation
National Academy of Sciences
National Institutes of Health
Office of Naval Research (Code 230)
Office of Naval Research (Code 102)
US Naval Research Laboratory (Code 2627)
Director, Naval Research Laboratory
Air Force Office of Scientific Research
Air Force Systems Command
Air Force Institute of Technology
Southwest Research Institute
Director, Mathematics Research Ctr, University of Wisconsin, Madison
Commandant:

National War College
Industrial College of the Armed Forces
Academy of Health Sciences
National Defense University
US Army Field Artillery School
US Army Reserve Components Personnel & Administration Center

P R O C E E D I N G S
O F T H E
1978 ARMY SCIENCE CONFERENCE
UNITED STATES MILITARY ACADEMY, WEST POINT, N.Y.
20 - 22 JUNE 1978

V O L U M E I

Principal Authors A through F

All experiments involving live animals that are reported in the Proceedings were performed in accordance with the principles of laboratory animal care as promulgated by the National Society of Medical Research

TABLE OF CONTENTS

PROCEEDINGS OF THE 1978 ARMY SCIENCE CONFERENCE

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Adams, N. L.	See von Bredow, J. D.	4	241
Allen, H. C. Cucksee, M. T.	The Effects of an Immiscible Binder Component on the Rheological and Mechanical Properties of a Composite Solid Propellant	1	1
Altstatt, L. B.	See Williams, H. L.	3	337
Alving, C. R. Steck, E. A. Hanson, W. L.	Improved Therapy of Leishmaniasis by Encapsulation of Antimonial Drug in Biodegradable Artificial Phospholipid Vesicles (Liposomes)	1	17
Arutunian, G.	See Gerhart, G. R.	2	17
Ashman, W. P.	See Sacco, W. J.	3	155
AuCoin, T. R. Ross, R. L. Wade, M. J. Savage, R. O.	Semi-Insulating Gallium Arsenide for Microwave and Millimeter-Wave Device Applications	1	25
Aulick, L. H. Wilmore, D. W.	Determination of Wound Blood Flow in the Thermally Injured Soldier	1	39
Ayoub, J. A.	See Rahman, A. R.	3	103
Bailey, C. E.	See Taschek, W. G.	3	295
Bailey, C. L. Eldridge, B. F. Hayes, D. E.	The Survival of St. Louis Encephalitis Virus in Overwintering Mosquitoes	1	55
Balbach, H. E. Novak, E. W.	Field Use of the Environmental Impact Computer System	1	65
Balsara, J. P.	See Kiger, S. A.	2	237

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Barber, V. C. Odom, E. C.	Deterioration and Reliability of Pavements	1	79
Bartosevich, J. F.	See Scovill, J. P.	3	225
Beatrice, E. S. Lund, D. J. Cours, D. Wampner, P. Slincy, D. H.	Project MILES: Biomedical Research and Coordination in Safe Field Exercises	1	95
Beatrice, E. S.	See Zwick, H.	3	395
Bell, R.	Life Expectancy of US Army Commercial Design Administrative Vehicles	1	109
Bender, E. S. Robinson, P. F.	Biomonitoring - A Final Method To Measure Pollution Abatement	1	121
Berg, N. J. Lee, J. N. Udelson, B. J.	Real-Time and Memory Correlation Via Acousto-Optic Processing	1	125
Boothe, R. R.	See Fahey, M. D.	4	35
Borowick, J.	See Fanuele, M. A.	4	51
Bowden, C. M.	See Howgate, D. W.	2	159
Brewer, R. J.	See Bruce, C. W.	1	157
Brown, C. D.	Gated TV Instrumentation System	1	141
Bruce, C. W. Pinnick, R. G. Brewer, R. J. Yee, Y. P. Fernandez, G.	In Situ Measurements of Aerosol Absorption with a Resonant CW Laser Spectrophone	1	157
Brusitus, J. M.	See O'Neill, T. R.	3	17
Buder, T. E.	See Gschwind, R. T.	4	69

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Buffa, A. J.	See Creedon, J. E.	4	1
Bulusu, S.	Explosion on a Single Molecule Level: A Conceptual Model Based on Ionization and Fragmentation of TNT Under Electron Impact	1	173
Burdick, C. K. Patterson, J. H. Mozo, B. T. Camp, R. T.	High-Frequency Hearing Loss Incurred By Exposure to Low-Frequency Noise	1	187
Burgess, E. W.	See Salomon, L. L.	3	167
Buser, R. G.	See Rohde, R. S.	3	141
Cady, E. M.	Intensive Tropic Function Testing	1	201
Camp, R. T.	See Burdick, C. K.	1	187
Canham, J. E.	See Zwick, H.	3	395
Capellos, C. Iyer, S.	Energetic Transient Species Formed Via Electronic Excitation of s-TNB and s-TNT	1	215
Carleton, H. D.	Military Use of Bulk Explosives	1	231
Carter, F. L.	See Salomon, L. L.	3	167
Carter, J. L.	See Creedon, J. E.	4	1
Caslavsky, J. L. Viechnicki, D. J.	Laser Rods Fabricated from AMMRC Grown Nd:YAG	1	247
Chow, S. T. Pupich, J. J.	FLIR Image Enhancement by Automatic Low Frequency Gain Limiting	1	261
Christensen, C. R. Upatnieks, J. Guenther, B. D.	Coherent Optical Correlation In Real Time for Missile Terminal Guidance	1	275
Chulay, J. D.	See Haynes, J. D.	2	95

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Colangelo, V. J. Lessen, G. P.	Physical and Mechanical Relationships in Electro Slag (ESR) Steel	1	291
Collett, E.	A Four-Channel Polarimeter to Measure Nanosecond Laser Pulses	1	305
Cook, C. C.	See Helbert, J. N.	2	119
Cours, D.	See Beatrice, E. S.	1	95
Creedon, J. E. Buffa, A. J. Carter, J. L. Levy, S. McGowan, J. W. Schneider, S.	Brassboard Modulator for HELS	4	1
Croisant, W. J. Nielsen, P.	Analysis of a Nonlinear Electromagnetic Field Penetration Problem	1	323
Crosley, D. R.	Laser-Induced Opto-Acoustic Pulses in a Flame	1	337
Cucksee, M. T.	See Allen, H. C.	1	1
D'Agostino, J. A. Dixon, R. E. Elmer, F. J. Rohde, R. S.	Fourier Spectroscopy of Artillery Gunflash	4	17
Davis, L. K. Horn, W. J.	The Influence of Large Runway Surface Roughness on Aircraft Response	1	353
Deitz, P. H.	See Gschwind, R. T.	4	69
DeSantis, G.	See White, R. M.	3	325
DeSimon, L. B.	See Huget, E. F.	2	173
Desjardins, R. E.	See Haynes, J. D.	2	95
Diggs, C. L.	See Haynes, J. D.	2	95

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Dixon, R. E.	See D'Agostino, J. A.	4	17
Dowgiallo, E. J.	Hybrid Power Source for Vehicular Propulsion	1	363
Drysdale, W. H. Kirkendall, R. D. Kokinakis, L. D.	Sabot Design for a 105mm APFSDS Kinetic Energy Projectile	1	377
Eaton, R. Lee, W. D.	Electrical Conductivity of Selected Graphite Intercalation Compounds in the Range $4K < T < 300K$	1	393
Eccleshall, D.	See Temperley, J. K.	4	223
Ehrlich, J. J.	See Howgate, D. W.	2	159
Eitelman, S. D.	See McCauley, J. W.	2	401
Eldridge, B. F.	See Bailey, C. L.	1	55
Elmer, F. J.	See D'Agostino, J. A.	4	17
Erickson, W. C.	See Keenan, R. E.	2	209
Erline, T. F.	Highly Survivable Truss Type Tail Boom	1	409
Erline, T. F.	See Haskell, D. F.	4	85
Essenwanger, O. M. Stewart, D. A.	Fog and Haze in Europe and Their Effects on Performance of Electro-Optical Systems	1	425
Evans, T. D.	See Law, H. Y. H.	2	299
Ewanizky, T. F.	Unstable Resonators for Army Laser Designators	1	441
Ewing, E. P. Takeuchi, A. Shirai, A. Osterman, J. V.	Scrub Typhus: Mechanism of Infection at the Cellular Level	1	453

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Fahey, M. D. O'Reilly, G. T. Boothe, R. R. Spaulding, W. G.	Quiet Radar Theory and Tests	4	35
Fanuele, M. A. Borowick, J. Kazules, A. J. McCray, J. A.	Lightweight Foliage Penetration Radar	4	51
Fernandez, G.	See Bruce, C. W.	1	157
Fine, B. J. Kobrick, J. L.	Human Performance Under Climatic Stress and the Fallacy of the "Average" Soldier: Potentially Serious Implications for Military Operations in Extreme Climates	1	459
Fishburn, B.	See Lu, P. L.	2	371
Fitzpatrick, R.	See Hynes, T. V.	4	103
Fowler, B. W.	Mie Theory for Non Spherical Particles	1	475
Freese, C. E. Neal, D. M. Lenoe, E. M.	Proof Test Procedures for Ceramic Missile Radomes	1	491
Freibergs, E.	See Klohn, K. L.	2	251
Frey, R. B.	See Howe, P. M.	2	145
Friedman, E.	See Niemela, J.	4	169
Froehlich, H. L.	See von Bredow, J. D.	4	241
Galanti, C. J.	See Pleckaitis, C. A.	3	89
Gambino, L. A.	An Experimental Digital Interactive Facility	2	1
Gerhart, G. R. Arutunian, G.	Pulsed Holographic Analysis of Large Vibrating Vehicle Components	2	17

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Gilbert, A. L. Giles, M. K.	Novel Concepts in Real-Time Optical Tracking	2	31
Giles, M. K.	See Gilbert, A. L.	2	31
Goldman, R. F.	First Battle in the Heat: Physiological Logistics for Success	2	47
Gooley, W.	See Salomon, L. L.	3	167
Greenberg, R. I.	See Merritt, J. A.	2	417
Gschwind, R. T. Deitz, P. H. O'Bryon, J. F. Rogers, H. R. Buder, T. E.	System Stimulation of STAFF	4	69
Guenther, B. D.	See Christensen, C. R.	1	275
Hafer, A. S.	See Kirby, R. L.	4	113
Hafer, T. F.	See Kirby, R. L.	4	113
Hall, C. L.	See Haynes, J. D.	2	95
Halpin, B. M.	See McCauley, J. W.	2	401
Hannigan, J. F.	Direct Electronic Fourier Transforms (DEFT) for Camouflage Signature Measurement (CSM)	2	63
Hanson, W. L.	See Alving, C. R.	1	17
Harris, O. R.	The Strange Behavior of Electromagnetic Waves in Conducting Magneto-Dielectric Media	2	79
Hartwig, G. W.	See Haskell, D. F.	4	85
Haskell, D. F. Kregel, M. D. Hartwig, G. W. Erline, T. F.	PATHAT Gun System - Concept and Feasibility	4	85

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Haut, M. J.	See Williams, H. L.	3	337
Hayes, D. E.	See Bailey, C. L.	1	55
Haynes, J. D. Chulay, J. D. Diggs, C. L. Desjardins, R. E. Hall, C.	Cultivation of Human Malaria Parasites	2	95
Heberlein, D. C.	Hardening of Countermine Structures	2	103
Helbert, J. N. Cook, C. C. Poindexter, E. H.	High-Speed Electron-Beam Litho- graphic Resists for Micron and Submicron Integrated Circuits	2	119
Hoad, D. R. Young, W. H. Meyers, J. F.	Velocity Measurements About a NACA 0012 Airfoil with a Laser Velocimeter	2	131
Hockmeyer, W. T.	See Kovatch, R. M.	2	283
Horn, R. E.	See Klohn, K. L.	2	251
Horn, W. J.	See Davis, L. K.	1	353
Houle, M. J.	See Long, D. E.	2	327
Howe, P. M. Frey, R. B.	Catastrophic Reaction of Compart- mentalized Ammunition - Causes and Preventive Measures	2	145
Howgate, D. W. Bowden, C. M. Ehrlich, J. J.	MIRADCOM Program in Swept-Gain Superradiance	2	159
Hudson, P. H.	See Litt, E. S.	4	129
Huget, E. F. DeSimon, L. B.	A Coating Agent for Promotion of Metal-to-Porcelain Bonding	2	173

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Hynes, T. V. Fitzpatrick, R. Parsons, G. R. Plumer, J. R.	High Energy Laser-Hardened Trans- parent Windscreens for Army Aircraft	4	103
Hynes, T. V.	See McCauley, J. W.	2	401
Iyer, S.	See Capellos, C.	1	215
Jack, W. H.	See Kirby, R. L.	4	113
Jacobs, H.	See Klohn, K. L.	2	251
Joachim, C. E.	Tunnel Destruction State-of-the- Art	2	183
Johnsmeyer, W. F.	See O'Neill, T. R.	3	17
Johnson, D. J.	See Williams, H. L.	3	337
Joncich, D. M.	A Comparison of the Actual and Predicted Performance of a Solar Assisted Space Heating System	2	195
Juriaco, I. P.	See Randers-Pehrson, G.	3	119
Kazules, A. J.	See Fanuele, M. A.	4	51
Keenan, R. E. Erickson, W. C.	Use of Computers in Mold Design	2	209
Kelley, N. J.	See Rahman, A. R.	3	103
Kennedy, P. J.	The Mechanism of Corrosion Inhi- bition by Dinonylnaphthalene- sulfonates	2	223
Kerschensteiner, M.	See Sarver, E. W.	3	183
Kiger, S. A. Balsara, J. P.	Response of Shallow-Buried Structures to Blast Loads	2	237
Kinsler, R. E.	See Kirby, R. L.	4	113

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Kirby, R. L. Kinsler, R. E. Mullen, O. L. Jack, W. H. Kruse, L. R. Wilson, R. D. Hafer, A. S. Hafer, T. F. Vanderbeek, F. J.	A Vulnerability Analysis of Candidate Engines for the XM-1 Tank	4	113
Kirkendall, R. D.	See Drysdale, W. H.	1	377
Klayman, D. L.	See Scovill, J. P.	3	225
Klohn, K. L. Horn, R. E. Freibergs, E. Jacobs, H.	Silicon Waveguide Line Scanning Antenna for Millimeter Waves	2	251
Knapp, S. C.	See Knox, F. S.	2	267
Knox, F. S. Wachtel, T. L. Knapp, S. C.	Mathematical Models of Skin Burns Induced By Simulated Postcrash Fires As Aids in Thermal Protective Clothing Design and Selection	2	267
Kobrick, J. L.	See Fine, B. J.	1	459
Kokinakis, L. D.	See Drysdale, W. H.	1	377
Kovatch, R. M. Wellde, B. T. Hockmeyer, W. T.	Research on Immunization Against African Sleeping Sickness	2	283
Kregel, M. D.	See Haskell, D. F.	4	85
Kruse, L. R.	See Kirby, R. L.	4	113
Kukafka, I.	See Litt, E. S.	4	129
Law, H. Y. H. Evans, T. D.	A Markov Model for Reliability, Availability and Maintainability Policy Guideline of an Airmobile Combat System	2	299

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Lee, B. L. Lewis, R. W. Sacher, R. E.	Environmental Effects on the Mechanical Properties of Glass Fiber/Epoxy Resin Composites - Effect of Static Immersion in Water on the Tensile Strength of Crossply Laminates	2	313
Lee, J. N.	See Berg, N. J.	1	125
Lee, W. D.	See Eaton, R.	1	393
Lenoe, E. M.	See Freese, C. E.	1	491
Lessen, G. P.	See Colangelo, V. J.	1	291
Levy, S.	See Creedon, J. E.	4	1
Lewis, R. W.	See Lee, B. L.	2	313
Litt, E. S. Kukafka, I. Hudson, P. H. McCormick, M. J.	The Expendable Set-On Communica- tions Jammer	4	129
Long, D. E. Houle, M. J.	A Method for the Rapid Evaluation of Hazards from Toxic Waste Deposits	2	327
Lowen, G. G. Tepper, F. R.	Dynamics of Pin Pallet Runaway Escapement	2	343
Lozar, C. C.	Establishing Habitability Factors for the Design of Office Environments	2	357
Lu, P. L. Slagg, N. Fishburn, B.	The Effects of Physical and Chemi- cal Processes on Two-Phase Detonations	2	371
Lund, D. J.	See Beatrice, E. S.	1	95
Maass, J. A.	MESH (Multiple Electronically Synapsing Hierarchy)	4	139

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Martin, D. L.	A Nonlinear Constitutive Relationship for Composite Propellants	2	385
Mason, C. J.	See Scovill, J. D.	3	225
Maurits, W. J.	See Sarver, E. W.	3	183
McCauley, J. W. Halpin, B. M. Eitelman, S. D. Hynes, T. V.	Radar Absorptive Material from Industrial Effluent	2	401
McCormick, M. J.	See Litt, E. S.	4	129
McCray, J. A.	See Fanuele, M. A.	4	51
McGarrigle, R. E.	See von Bredow, J. D.	4	241
McGowan, J. W.	See Creedon, J. E.	4	I
Merritt, J. A. Meyer, H. C. Greenberg, R. I. Tanton, G. A.	The Production of Decaborane-14 from Diborane by Laser Induced Chemistry	2	417
Meyer, H. C.	See Merritt, J. A.	2	417
Meyers, J. F.	See Hoad, D. R.	2	131
Miller, M. C.	Method to Experimentally Determine the Aerodynamic Pressure Distribution on Spinning Bodies	2	433
Mitchell, T.	See von Bredow, J. D.	4	241
Mozo, B. T.	See Burdick, C. K.	1	187
Mullen, O. L.	Evaluation of 105mm HEAT, HEP and KE Projectiles Against Masonry Structures	4	151
Mullen, O. L.	See Kirby, R. L.	4	113

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Narva, M. A.	Development of a Systematic Methodology for the Application of Judgmental Data to the Assessment of Training Device Concepts	3	1
Neal, D. M.	See Freese, C. E.	1	491
Nielsen, P.	See Croisant, W. J.	1	323
Niemela, J. Friedman, E.	Analysis of POS/NAV Requirements for Airborne RSTA/EW Systems	4	169
Novak, E. W.	See Balbach, H. E.	1	65
O'Bryon, J. F.	See Gschwind, R. T.	4	69
Odom, E. C.	See Barber, V. C.	1	79
O'Neill, T. R. Brusitus, J. M. Taylor, D. L. Johnsmeyer, W. F.	Evaluation of Dual-Texture Gradient Camouflage Pattern	3	17
O'Reilly, G. T.	See Fahey, M. D.	4	35
Osterman, J. V.	See Ewing, E. P.	1	453
Owens, F. J. Sharma, J. Wiegand, D. A.	Molecular Studies of the Mechanism of Shock Initiation of Solid Explosives	3	31
Parsons, G. R.	See Hynes, T. V.	4	103
Patterson, J. H.	See Burdick, C. K.	1	187
Perkins, J. C.	Development of a Systematic and Quantitative Method for Evaluating Instructional Material: Instructional Material Adequacy Guide and Evaluation Standard (IMAGES)	3	47
Peters, D. A.	See Schrage, D. P.	3	209
Peterson, E. G.	See Salomon, L. L.	3	167

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Pflegl, G. A.	See Simkins, T. E.	3	249
Pickwell, G. V.	See Vick, J. A.	3	309
Pinnick, R. G.	See Bruce, C. W.	1	157
Pitt, D. M.	Army Helicopter Flight Simulation, A Research and Development Tool for the Future	3	61
Platou, A. S.	Aeroballistics of Corkscrew Projectiles	3	77
Pleckaitis, C. A. Galanti, C. J. Santanelli, A. S. Stech, G.	Integrated Avionics Control System (IACS)	3	89
Plumer, J. R.	See Hynes, T. V.	4	103
Poindexter, E. H.	See Helbert, J. N.	2	119
Pupich, J. J.	See Chow, S. T.	1	261
Rahman, A. R. Kelley, N. J. Ayoub, J. A. Westcott, D. E.	New Energy Saving Techniques for the Production of Dense Foods	3	103
Randers-Pehrson, G. Juriaco, I. P.	Computer Aided Self-Forging Fragment Design	3	119
Reitmeyer, R. A.	Computer Aided Design, Design Auto- mation and LSI; Keys to High- Performance Military Electronics	3	129
Riessler, W. A.	See Sattler, J. P.	3	195
Ritter, K. J.	See Sattler, J. P.	3	195
Robinson, P. F.	See Bender, E. S.	1	121
Rogers, H. R.	See Gschwind, R. T.	4	69

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Rohde, R. S.	See D'Agostino, J. A.	4	17
Rohde, R. S. Buser, R. G.	Thermal Blooming and Air Breakdown Interaction for Pulsed High Energy Lasers	3	141
Rorabaugh, D. T.	Improved Ceramic-Metal Composite Vehicular Armor	4	191
Ross, R. L.	See AuCoin, T. R.	1	25
Sacco, W. J. Ashman, W. P. Swann, C. L. Sturdivan, L. M.	Trauma Indices and Applications	3	155
Sacher, R. E.	See Lee, B. L.	2	313
Salomon, L. L. Peterson, E. G. Burgess, E. W. Gooley, W. Carter, F. L.	Characterization of Obscuring Smokes in the Field	3	167
Santanelli, A. S.	See Pleckaitis, C. A.	3	89
Sarver, E. W. Maurits, W. J. Kerschensteiner, M.	Methodology Development for the Identification and Analysis of Trace Contaminants Found on Army Installations	3	183
Sattler, J. P. Worchesky, T. L. Ritter, K. J. Simonis, G. J. Riessler, W. A.	Infrared Diagnostics for Near- Millimeter Wave Sources	3	195
Savage, R. O.	See AuCoin, T. R.	1	25
Scanlon, R. D.	See Simkins, T. E.	3	249
Schneider, S.	See Creedon, J. E.	4	1
Schrage, D. P. Peters, D. A.	Comparison of the Effect of Struc- tural Coupling Parameters on Flap- Lag Forced Response and Stability of a Helicopter Rotor Blade in Forward Flight	3	209

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Scovill, J. P. Bartosevich, J. F. Mason, C. J. Klayman, D. L.	A New Class of Antimalarial Agents: 2-Acetylpyridine Thiosemicarbazones	3	225
Sharma, J.	See Owens, F. J.	3	31
Shaw, W. E.	Applications of a Gated Imaging System in Evaluation of Laser Designator Performance	3	235
Shirai, A.	See Ewing, E. P.	1	453
Simkins, T. E. Scanlon, R. D. Pfleigl, G. A.	Radial and Tranverse Response of Gun Tubes to Traveling Ballistic Pressure	3	249
Simonis, G. J.	See Sattler, J. P.	3	195
Slagg, N.	See Lu, P. L.	2	371
Sliney, D. H.	See Beatrice, E. S.	1	95
Smyth, C. C.	Internal Cockpit Reflections of External Point Light Sources for the Model YAH-64 Advanced Attack Helicopter	3	263
Spaulding, W. G.	See Fahey, M. D.	4	35
Stark, W. J. Wyatt, W. T. Strayer, R. E.	Validation of High-Altitude EMP Environments Based on Reported Damage During Operation Fish Bowl	4	203
Stech, G.	See Pleckaitis, C. A.	3	89
Steck, E. A.	See Alving, C. R.	1	17
Stewart, D. A.	See Essenwanger, O. M.	1	425
Strange, J. N.	The Essex Program: A Study of the Effects of Underground Low-Yield Nuclear Weapons Employed in a Tac- tical Warfare Scenario	3	279

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Strayer, R. E.	See Stark, W. J.	4	203
Sturdivan, L. M.	See Sacco, W. J.	3	155
Swann, C. L.	See Sacco, W. J.	3	155
Takeuchi, A.	See Ewing, E. P.	1	453
Tanton, G. A.	See Merritt, J. A.	2	417
Taschek, W. G. Bailey, C. E.	High Energy Metal Hydride Fuel Cell Power Source	3	295
Taylor, D. L.	See O'Neill, T. R.	3	17
Temperley, J. K. Eccleshall, D.	Transfer of Energy from Charged Asymmetric Transmission-Line Pairs	4	223
Tepper, F. R.	See Lowen, G. G.	2	343
Udelson, B. J.	See Berg, N. J.	1	125
Upatnieks, J.	See Christensen, C. R.	1	275
Vanderbeek, F. J.	See Kirby, R. L.	4	113
Vick, J. A.	See von Bredow, J. D.	4	241
Vick, J. A. Pickwell, G. V.	Toxicity and Treatment of Sea Snake Envenomation	3	309
Viechnicki, D. J.	See Caslavsky, J. L.	1	247
von Bredow, J. D. Vick, J. A. Adams, N. L. McGarrigle, R. E. Froehlich, H. L. Mitchell, T.	Protection Against Exposure to Multi-Lethal Concentrations of Nerve Agents	4	241
Wachtel, T. L.	See Knox, F. S.	2	267
Wade, M. J.	See AuCoin, T. R.	1	25
Wampner, P.	See Beatrice, E. S.	1	95

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Wellde, B. T.	See Kovatch, R. M.	2	283
Westcott, D. E.	See Rahman, A. R.	3	103
White, R. M. DeSantis, G.	The Impact of Female Anthropometry on the US Army	3	325
Wiegand, D. A.	See Owens, F. J.	3	31
Williams, H. L. Johnson, D. J. Haut, M. J. Altstatt, L. B.	The Effect of Certain Aromatic Compounds on Enzymes Involved in Heme Synthesis	3	337
Williamson, C. E.	Strong and Specific Interactions of Some Incapacitating Phenothi- azines with Nucleic Acids	3	349
Williamson, G. R.	Steel Fibers as Web Reinforcement in Reinforced Concrete	3	363
Wilmore, D. W.	See Aulick, L. H.	1	39
Wilson, R. D.	See Kirby, R. L.	4	113
Worchesky, T. L.	See Sattler, J. P.	3	195
Wuebben, J. L.	Simulation of Drifting Snow in a Hydraulic Flume	3	379
Wyatt, W. T.	See Stark, W. J.	4	203
Yee, Y. P.	See Bruce, C. W.	1	157
Young, W. H.	See Hoad, D. R.	2	131
Zwick, H. Beatrice, E. S. Canham, J. E.	Laser Bioeffects; Low Level Effects; Impact on Army Laser Systems	3	395

THE EFFECTS OF AN IMMISCIBLE BINDER COMPONENT ON
THE RHEOLOGICAL AND MECHANICAL PROPERTIES OF
A COMPOSITE SOLID PROPELLANT (U)

*HENRY C. ALLEN and MARJORIE T. CUCKSEE
US ARMY MISSILE RESEARCH AND DEVELOPMENT COMMAND
REDSTONE ARSENAL, ALABAMA 35809

The Improved HAWK antiaircraft missile is powered by the M112 rocket motor. This motor operates in boost and sustain modes derived from concentric booster and sustainer composite solid propellant grains (Figure 1). The booster propellant is relatively fast burning to provide the thrust level necessary to accelerate the missile to intercept velocity, while the sustainer propellant burns slowly over a much longer time to maintain missile velocity and maneuverability during the mission. The booster grain is fully bonded to the sustainer grain which in turn is fully bonded to the steel motor case by means of an adhesive case liner.

The chamber pressure of Improved HAWK motors during firing is governed by the ratio of burning surface area to the area of the nozzle throat. The internal perforation of the booster grain is shaped to a predetermined surface area. If additional area is exposed to combustion, an overpressure condition results which can burst the motor case. Another type of aberration which can defeat the mission is that flame may reach the steel case prematurely and burn a hole through it, causing the motor to lose thrust and become unstable in flight. A condition which can produce both of these malfunctions is that in which a normally bonded propellant interface is unbonded.*

*The term "unbond" denotes a condition in which there is no bond where one was intended, whether or not the intended bond ever existed. The term "debond" denotes a condition where a bond existed but is now broken.

The stresses which tend to debond the grains from one another or from the liner are due to temperature variations in the motor environment coupled with coefficients of thermal expansion for the propellants which are an order of magnitude greater than that of the steel case. As the motor is cooled, strains are induced in the propellant with concomitant stresses on the bond interfaces. If the stresses exceed the bond strength, a debond occurs.

Several thousand Improved HAWK motors were produced without a significant unbond problem. Then, rather abruptly, a substantial number of unbonded motors appeared in one production lot. Both booster-sustainer and sustainer-liner unbonds were found, some of which were extensive in area. Because prior production had not exhibited the problem, it was assumed that a change had occurred in either motor processing or propellant raw materials.

IDENTIFICATION OF THE ANOMALOUS MATERIAL

Tests made on propellants removed from unbonded motors showed that sustainer propellant mechanical properties were significantly different at low temperatures in comparison to prior production. However booster propellant properties were normal; this isolated the problem to the sustainer propellant. Two materials are used in the sustainer propellant which are not components of the booster propellant. One of these is a solid oxidizer (nitroguanidine) while the other is a liquid polyester, poly-neopentyl glycol azelate (NPGA), a component of the propellant binder. Preparation and testing of propellants containing various lots of these two components showed that NPGA was responsible for the inferior properties found in propellants from unbonded motors.

PROPELLANT COMPOSITION AND PROCESSING

HAWK sustainer propellant consists of solid oxidizer particles imbedded in a rubbery polyurethane binder. The polyurethane binder controls the structural properties of the propellant and serves as a fuel during propellant combustion. Three hydroxy-functional liquid polymers comprise the major constituents of the rubber precursor: (1) a difunctional polyether (B2000), (2) a trifunctional polyether (TP4040), and (3) the polyester NPGA which has been found to have an effective functionality of 1.7. Isodecyl pelargonate is added to facilitate processing and to plasticize the cured propellant. The hydroxy-functional polymers are cured by reaction with hexamethylene di-isocyanate (HDI) under catalysis by ferric acetyl acetonate to form the polyurethane rubber.

In propellant manufacture, the liquid polymers and isodecyl pelargonate are blended to form a submix. The powdered solid oxidizers are then added until a homogeneous paste of high viscosity is formed. HDI and cure catalyst are added to start the cure reaction and the propellant is poured into rocket motor cases and test specimen containers before the cure reaction increases the viscosity beyond processable limits. Cure is complete after several days at a temperature of 43°C.

CHARACTERIZATION OF NPGA

Contacts with the NPGA manufacturer revealed that recent lots of the polymer, beginning with the lot that went into unbonded motors, were produced in a new facility. However, these lots met all specification requirements for use in HAWK motors (Table 1).

Samples of old and new lots of NPGA were characterized by gel permeation chromatography (GPC) to determine whether molecular weight changes had resulted from the facilities switch. The results (Figure 2) showed that the average molecular weight of the new lot was significantly higher than that of the old lot and that the molecular weight distribution was broader in the new lot, increasing the higher molecular weight fraction. However, the low molecular weight region which contains most of the hydroxyl groups was not decreased.

The effective functionality of NPGA was determined by finding the isocyanate/hydroxyl (NCO/OH) ratios required for incipient gelation in a mixture of the binder polymers cured with HDI. The results (Table 2) showed the new lot to have a significantly lower effective functionality than the old.

The viscosities of the new and old lots of NPGA were measured at 25°C and 43°C; the ratios of viscosities at these temperatures were compared as an indication of whether there was a difference in chain branching. It was found that the ratios were not significantly different (Table 2). This indicated that the higher molecular weight of the new lot was due to chain extension rather than branching, which was in agreement with the functionality data.

EFFECTS OF NPGA DIFFERENCES ON SOLUBILITY

The new lot of NPGA produced a decidedly cloudy submix indicating incomplete solubility. Submix formulations were prepared with old and new lots of NPGA and the quantities of the insoluble fractions were determined by centrifuging samples in graduated tubes.

In the submix formulation associated with the unbonded motors, the new lot of NPGA produced an immiscible fraction equal to 39.8% of the NPGA added. The old lot of NPGA in the same formulation gave an immiscible fraction equal to 31.8% the NPGA added. Samples then were prepared to the submix formulation associated with the old lot NPGA, and the insoluble fraction of the old lot of NPGA amounted to 24% of the NPGA added. Thus it became clear that factors other than NPGA variability were affecting its solubility.

The major difference between submix formulations associated with the old and new lots of NPGA was in the relative amounts of TP4040 and B2000 in each formulation. TP4040, being trifunctional, strongly influences the binder cross-link density and thus the modulus of the propellant. Consequently, the amount of TP4040 is deliberately varied with different lots of raw materials as a means of controlling propellant modulus within specified limits.

Submix samples were prepared with TP4040 contents ranging from 2% to 6% (based on propellant weight), and the insoluble fractions of the old and new lots of NPGA were measured. The data are shown in Figure 3. Two things were immediately obvious from these data: the insoluble fraction differed for the old and new lots of NPGA at all TP4040 levels; and the insoluble fraction for both lots was a strong function of TP4040 content.

COMPOSITION OF THE INSOLUBLE FRACTION

The insoluble fraction from a submix sample made with the new lot of NPGA was analyzed on the GPC. The results are given in Table 3 with comparative data for neat NPGA. It was seen that both the number average and weight average molecular weights were much higher for the insoluble fraction than for neat NPGA. This showed that the high molecular weight fraction of NPGA was the immiscible fraction in the submix and explained why the new lot of NPGA, with a larger high molecular weight fraction (Figure 2), was less soluble than the old lot.

The compositions of the insoluble fractions were determined spectrophotometrically using samples of known composition for calibration. The results are shown in Table 4; approximately two-thirds of the insoluble fraction was found to be NPGA. Since the insoluble fraction of the submix with the new lot amounted to 39.8% of the NPGA added, and 69% of that fraction was NPGA, the amount of NPGA alone not in solution in the submix was 27.5% of the quantity added. Similar treatment of the data for the old lot showed 21.3% of the quantity added to be insoluble. These data reflect solubility of NPGA with

2% TP4040 (propellant basis) in the submix, as was the case with the new lot of NPGA. Taking into account that the submix formulation associated with the old lot of NPGA contained 3.1% TP4040, reference to Figure 3 and Table 4 shows that 16.1% of the old lot of NPGA was insoluble as actually used. Thus the comparison of solubilities in the actual production submixes reveals that 27.5% of the new lot of NPGA was insoluble while only 16.1% of the old lot was not in solution.

EFFECT OF IMMISCIBLE FRACTION ON PROPELLANT RHEOLOGY

A plausible cause for motor unbonds was that the uncured sustainer propellant might not adequately wet the booster grain and chamber liner during motor processing because of the immiscible material. Since the quantity of the insoluble fraction of the submix was a function of TP4040 content, a means was available to determine how the amount of insoluble fraction affected uncured propellant flow characteristics. Propellant mixes were made containing various amounts of TP4040. Viscosity data were taken with a viscometer capable of applying a range of shear stresses to the sample. Viscosities were measured at 43°C immediately following a fixed mixing cycle. Measurements were made at high and low shear stresses. The results are shown in Figure 4 for propellants made with the new lot of NPGA.

The data showed that insoluble fractions amounting to less than approximately 20% of the NPGA produced low propellant viscosities at both levels of shear stress. The ratio of viscosities at the two shear levels was small at this solubility, indicating near-Newtonian flow behavior. As the insoluble fraction was increased, the viscosity at infinite shear increased only slightly, but the low shear viscosity increased rapidly. The ratio of viscosities increased correspondingly, indicating an increasing degree of thixotropy in the flow behavior. The ratio of viscosities, which was 1.6 at an insoluble fraction of 18%, was 9.6 at a 40% insoluble fraction. Taking appropriate data from Figures 3 and 4, a viscosity ratio of 1.7 is predicted for propellant with old NPGA and 9.6 for propellant with the new lot. Control data from production mixes agreed reasonably well with these findings, showing a viscosity ratio of 1.7 for propellants made with the old lot of NPGA, and 4.5 for propellant with the new lot.

EFFECT OF THE IMMISCIBLE FRACTION ON PROPELLANT MECHANICAL PROPERTIES

Certain mechanical properties of the propellants are specified control points for Improved HAWK motor production. One of these is the initial modulus (E_0) at 25°C, which should be in the range from

500 to 800 psi. The level of the trifunctional polymer TP4040 is the primary means permitted for controlling this property. The amount of TP4040 required with a given set of raw materials is determined experimentally by preparing and testing a series of small mixes with different TP4040 contents.

The new lot of NPGA and other raw materials associated with it in production required that 2% TP4040 be used in the sustainer propellant, which is the minimum amount permitted by the propellant specification. Even so, the modulus at 25°C was just under the maximum limit, effectively preventing the use of more TP4040 to reduce the insoluble submix fraction in production mixes.

Stress analysis of the Improved HAWK motor showed that the stresses on the bonded interfaces were a function of propellant modulus. Modulus and temperature are inversely related, so the greatest stresses occur at the lowest temperatures. Indeed, motor unbonds were found only after exposure to temperatures below -30°C. Mechanical property data on sustainer propellants removed from dissected motors showed that the new lot of NPGA and its associated raw materials had produced propellant with significantly higher modulus at low temperature than the old lot.

Experimental 6-kg propellant mixes were made with the old and new lots of NPGA at TP4040 levels of 2.0%, 3.5%, and 5.0%. After cure, the propellants were tested for mechanical properties at 25°C and -40°C. Die-cut tensile specimens with a gauge length of 2.70 in. were strained at a rate of 2.0 in./min. The properties are listed in Table 5; the moduli are shown graphically in Figure 5.

The most striking feature of the data is that the -40°C modulus decreased with increasing TP4040 level, while the 25°C modulus increased as expected. Inspection of the -40°C stress-strain curves showed that the curve shape changed with TP4040 content (Figure 6). This effect was attributed to the insoluble submix fraction, with the -40°C modulus responding in a manner somewhat analogous to the viscosity at low shear stress.

Another feature in the data was puzzling on first inspection: the new lot of NPGA, with a lower effective functionality, gave higher 25°C moduli at all TP4040 levels than the old lot. This was explained by the fact that the insoluble fraction was chiefly NPGA and B2000 with functionalities of 1.66 and 2.0, respectively. Removal of these lower functionality materials through immiscibility increased the average functionality of the remainder, thus increasing the 25°C modulus. When the new, less soluble lot of NPGA was introduced, the 25°C modulus

of the propellant increased. In response, the TP4040 level was reduced to stay within the specified modulus range, which in turn further decreased the solubility of the NPGA. Simultaneously, the -40°C modulus increased markedly, placing heavy stresses on the bonded interfaces.

Discussion of the foregoing data with stress analysts led to the conclusion that it would be preferable to accept higher 25°C moduli to reduce the -40°C moduli. Subsequently, the maximum modulus limit at 25°C was raised from 800 to 1000 psi, permitting the use of 4% TP4040. This greatly reduced the -40°C modulus, which increased the bond safety margin calculated from the stress analysis. No unbonded motors have been found among several hundred produced since this change was incorporated.

PROCESS CHANGES TO ACHIEVE SOLUBILITY

Despite the success achieved through raising the TP4040 content by waiving the 25°C modulus requirement, it remained very desirable to eliminate all effects of an immiscible binder component on propellant processing and mechanical properties. Because the motor is in production internationally, major changes to material specifications were not considered as viable options. Only process changes were viewed as acceptable approaches to increasing the solubility of NPGA.

A review of the chemical kinetics of the cure reactions suggested that changing the order of certain reactions might be effective. NPGA has primary hydroxyl groups which are far more reactive toward HDI than the secondary hydroxyl groups of TP4040 and B2000. Therefore, the first principle products of the submix-HDI reaction would be HDI-capped NPGA and chain-extended NPGA. Because high-molecular weight NPGA was shown to be the insoluble portion, any chain extension would aggravate the problem. Therefore, mixtures which contain both NPGA and HDI at their maximum concentrations should be avoided.

The approach to minimizing NPGA chain extension was to withhold the NPGA from the mix until a substantial amount of reaction had occurred between HDI and the other polymers. This reduced the amount of difunctional HDI which would be available to chain extend the NPGA, and simultaneously changed the solvent nature of the submix by introducing urethane linkages. Thus the NPGA, when added late, was introduced into a different chemical environment than in the regular procedure.

The technique of late NPGA addition had a pronounced effect on propellant processing and mechanical properties. Propellant viscosity at the end of the mix cycle was drastically reduced and thixotropy

was essentially eliminated. Viscosity data are shown in Table 6 for the same formulation made by the two processes, using the new lot of NPGA. The very low viscosities resulting from late addition of NPGA would be expected to eliminate the possibility of poor wetting by the sustainer propellant, and provide a number of other processing advantages as side benefits.

The effects of late NPGA addition on mechanical properties are shown in Figures 7 through 10 as a function of temperature. Figure 7 shows that late addition greatly reduced the initial modulus of the propellant across the temperature range. The 50% reduction at 25°C was attributed to a reduction in average binder continuum functionality resulting from greater solubility of the NPGA. This reduction in modulus is the desirable direction of change, because it would permit more TP4040 to be used without exceeding the 25°C modulus limit. At -40°C the relative reduction in modulus was even greater than at 25°C. This was interpreted to be a result of a smaller insoluble submix fraction. As previously discussed, low temperature modulus is a determinant of bond stresses; the low value resulting from late addition was decidedly advantageous. Figure 8 shows the effect of late NPGA addition on propellant tensile strength. The reduction in tensile strength was the expected result of the reduced modulus but the values remained well above specification requirements. The response of tensile strength to temperature was not significantly different for the two processing techniques. Figure 9 shows the effect of late addition on the ratio of tensile strength to modulus. In the investigation of motor unbonds this property proved to be an indicator of the tendency of the propellant to debond. It was seen that late NPGA addition markedly increased this ratio over the temperature range of interest. Figure 10 shows the effect of late addition on propellant strain capability. Again, a very large difference was seen for the two processing techniques. A portion of the increase resulting from late NPGA addition is attributed to the reduction in modulus shown in Figure 7. However, the magnitude of the difference was too great to attribute to this one factor. A large part of the increase was thought to be due to improved binder properties resulting from the NPGA being in solution. Late addition of NPGA is being evaluated further in anticipation of incorporating this process change into the production procedures for Improved HAWK motors.

ACKNOWLEDGMENTS

The authors express their appreciation to the Aerojet Solid Propulsion Company for portions of the data and to Dr. James Carver of the Propulsion Directorate for the GPC characterizations.

TABLE 1. SPECIFICATION FOR NPGA AND ANALYTICAL DATA FOR OLD AND NEW LOTS

Characteristic	Specified Limits	Old Lot	New Lot
Water (%)	0.04 maximum	0.028	0.030
Hydroxyl No.	52.5 to 57.5	56.8	53.9
Acid No.	1.5 maximum	0.40	0.39
Tin (ppm)	50 maximum	5	5
Volatiles (%)	0.5 maximum	0.09	0.10
Refractive Index	1.466 to 1.468	1.46670	1.46670
Viscosity, 25°C (cps)	9000 to 14,000	12,100	12,400
Reactivity	400 to 2100	800	1100

TABLE 2. NPGA CHARACTERISTICS NOT INCLUDED IN SPECIFICATION REQUIREMENTS

	Old NPGA	New NPGA
Number (average molecular weight)	1904	2080
Weight (average molecular weight)	3369	3910
Polydispersity	1.77	1.88
Effective Functionality	1.72	1.66
Viscosity at 25°C	2.80	2.84
Viscosity at 43°C		

TABLE 3. GPC DATA FOR THE INSOLUBLE SUBMIX FRACTION AND FOR NEAT NPGA (NEW LOT OF NPGA, 6% TP4040 IN SUBMIX)

	Insoluble Fraction	Neat NPGA
Number (average molecular weight)	3556	2080
Weight (average molecular weight)	6643	3910
Polydispersity	1.87	1.88

TABLE 4. COMPOSITION OF THE INSOLUBLE SUBMIX FRACTIONS (2% TP4040 IN SUBMIX)

	Old NPGA	New NPGA
NPGA (%)	67	69
Isodecyl pelargonate (%)	16	15
TP4040 (%)	5	2
B2000 (%)	12	14
Weight (average molecular weight)	4605	4581
Hydroxyl No.	30.4	32.4
Submix hydroxyl No.	39.0	39.9

TABLE 5. MECHANICAL PROPERTIES OF SUSTAINER PROPELLANT AT VARIOUS TP4040 LEVELS

Temperature	TP4040 (%)	Property	Old Lot	New Lot
25°C	2.0	Stress (psi)	142	118
		Strain (%)	66.1	63.2
		Modulus (psi)	680	800
25°C	3.5	Stress (psi)	154	138
		Strain (%)	62.7	58.9
		Modulus (psi)	820	910
25°C	5.0	Stress (psi)	167	162
		Strain (%)	56.6	52.6
		Modulus (psi)	940	1060
-40°C	2.0	Stress (psi)	360	340
		Strain (%)	78.3	75.4
		Modulus (psi)	3560	5600
-40°C	3.5	Stress (psi)	415	391
		Strain (%)	74.8	75.8
		Modulus (psi)	2890	3820
-40°C	5.0	Stress (psi)	450	452
		Strain (%)	71.6	62.1
		Modulus (psi)	2510	2840

TABLE 6. VISCOSITY DATA FOR PROPELLANT MIXES MADE WITH REGULAR AND LATE ADDITION OF NPGA*

	Viscosity (kilopoise)	
Shear Stress	Regular Process	Late Addition of NPGA
5000 dynes/cm ²	22	1.0
Infinite	4.8	0.6

*New lot of NPGA

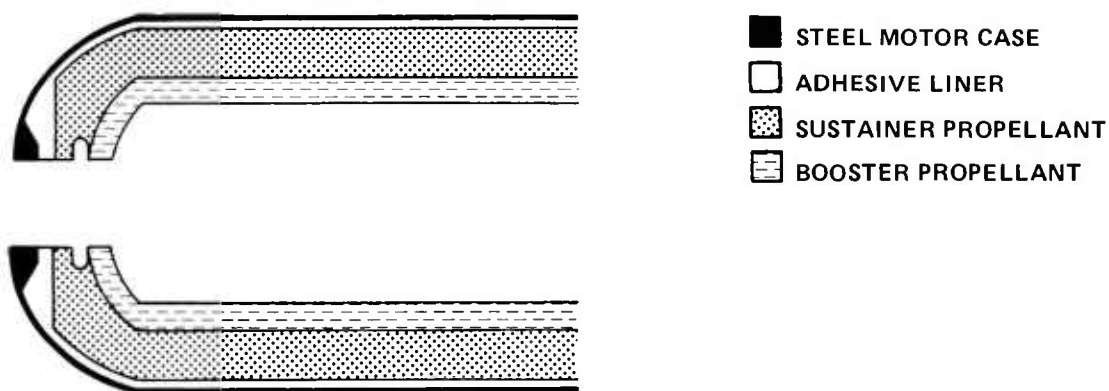


Figure 1. Schematic drawing of the forward end of the Improved HAWK rocket motor.

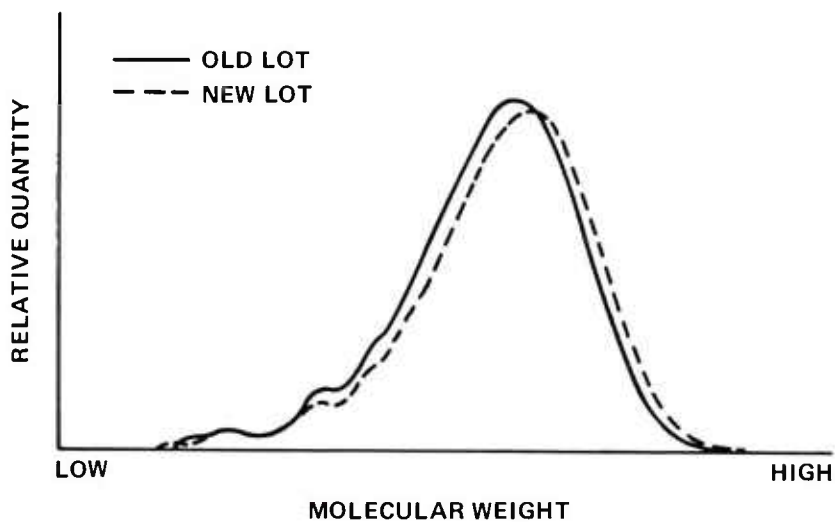


Figure 2. GPC analysis of old and new lots of NPGA.

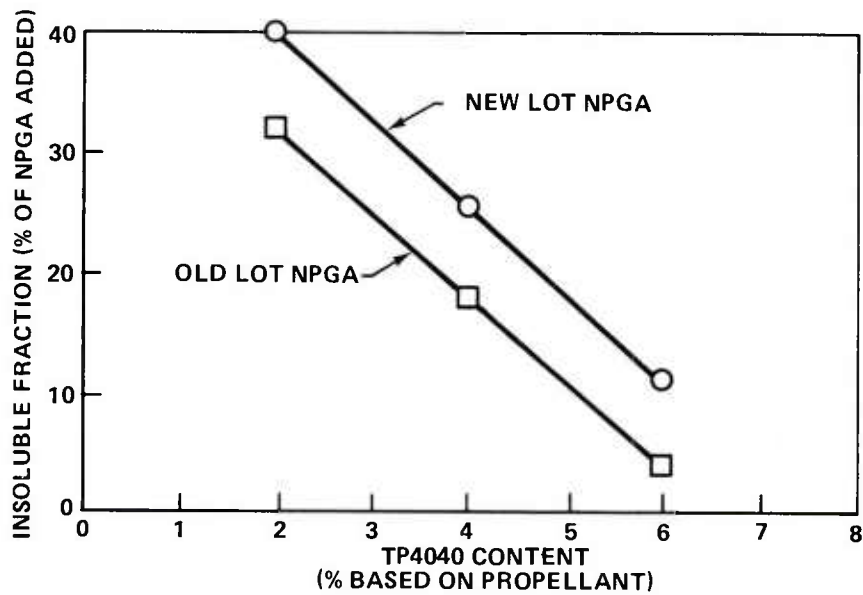


Figure 3. Insoluble fraction of submix versus TP4040 content of submix, for old and new lots of NPGA.

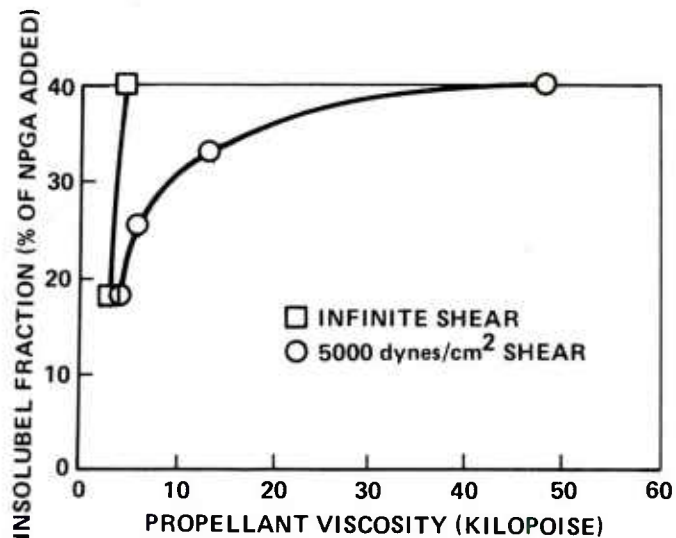


Figure 4. Effect of the quantity of insoluble fraction on propellant processing viscosity (new lot of NPGA)

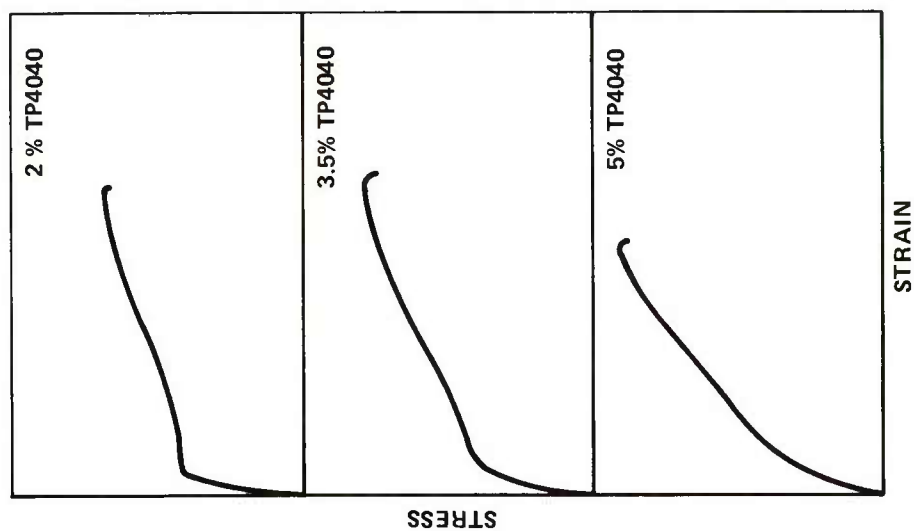


Figure 6. Effect of TP4040 content on the stress-strain behavior of sustainer propellant at -40°C.

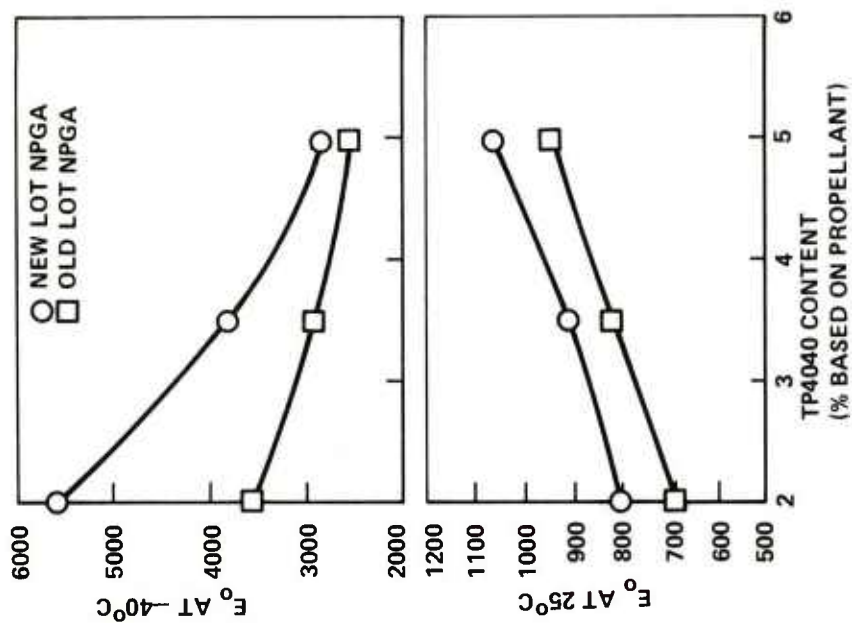


Figure 5. Effect of TP4040 content on the moduli of sustainer propellant at 25°C and -40°C.

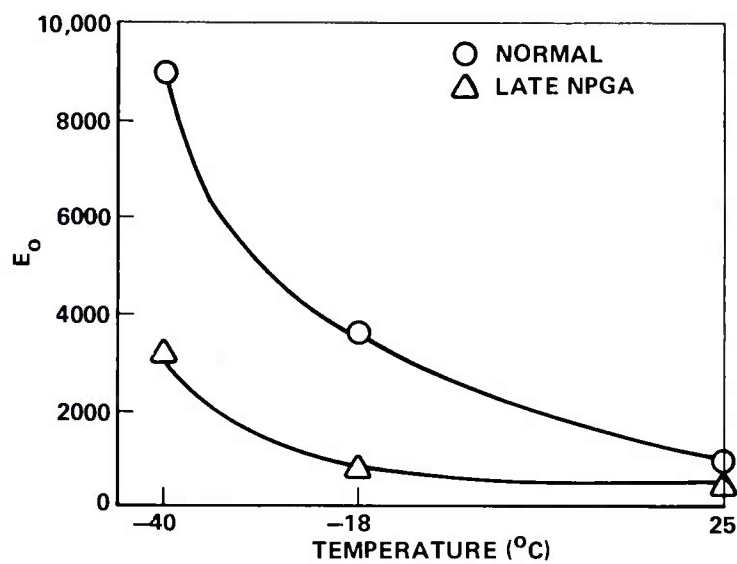


Figure 7. Effect of late NPGA addition on initial modulus.

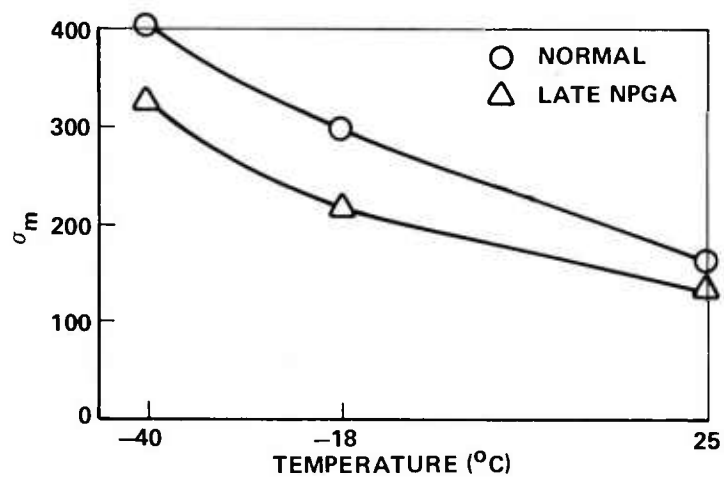


Figure 8. Effect of late NPGA addition on tensile strength.

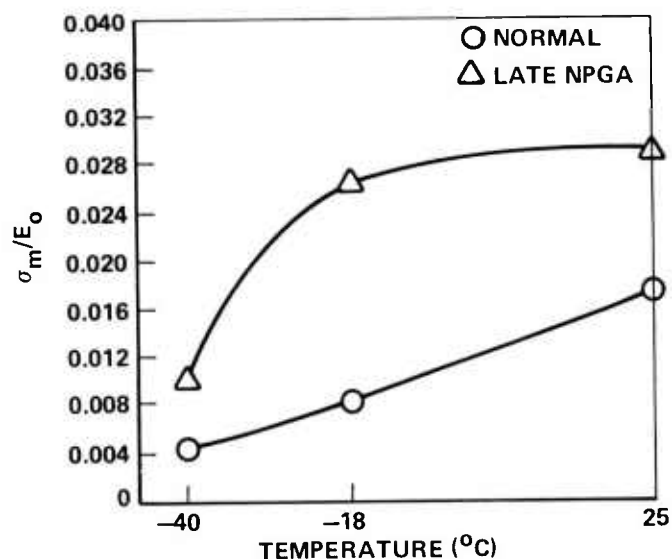


Figure 9. Effect of late NPGA addition on stress/modulus ratio.

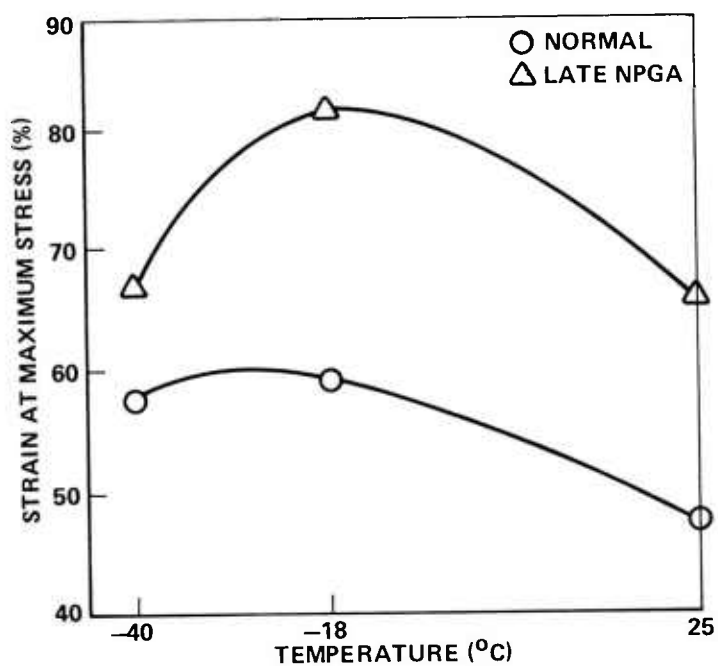


Figure 10. Effect of late NPGA addition on strain capability.

IMPROVED THERAPY OF LEISHMANIASIS BY ENCAPSULATION OF ANTIMONIAL
DRUG IN BIODEGRADABLE ARTIFICIAL PHOSPHOLIPID VESICLES (LIPOSOMES)(U)

*CARL R. ALVING, M.D., LTC, MC
EDGAR A. STECK, Ph.D. AND WILLIAM L. HANSON, Ph.D
WALTER REED ARMY INSTITUTE OF RESEARCH,
WASHINGTON, D.C. 20012, AND COLLEGE OF VETERINARY MEDICINE,
THE UNIVERSITY OF GEORGIA, ATHENS, GA 30602

INTRODUCTION

Leishmania, which are hemoflagellate protozoa, are important pathogenic intracellular parasites which cause diseases resulting in cutaneous, mucocutaneous, or visceral (kala azar) manifestations. The parasites reside chronically in phagocytes of the reticulo-endothelial system (1-4). At least 12 million people are infected with various forms of Leishmania (4). Leishmaniasis is highly infectious and represents a significant potential military health problem. Leishmania donovani was first demonstrated in smears taken post-mortem from the spleen of an English soldier in 1900 in India (2). Epidemics have occurred among soldiers fighting in the forest, as in Paraguay in the Gran Chaco war (3). In recent years, armed forces that have operated in endemic areas, such as Colombian Army, the Israeli Army in the Sinai, and the U.S. Army in Panama, have had high attack rates in certain units. The disease is endemic in the Middle East, Africa, India, China, Asian USSR, Central and South America, and other tropical and subtropical regions throughout the world. The illness may be severe, lingering, and may be recurrent despite therapy with antimonial compounds, the drugs of choice (1,4).

Treatment of leishmaniasis is hampered, and doses are limited, by the serious toxicities of antimonials (1). We have developed a novel approach to treatment of leishmaniasis which takes advantage of the localization of the organism in phagocytic cells. Our technique consists of injection of liposomes containing antimonial drugs. Liposomes are artificial, biodegradable membranes comprised of

lipids, including phospholipids and other lipids that can be obtained from natural or synthetic sources (5,6). As shown schematically in Fig. 1, the membranes are in the form of closed concentric spheres

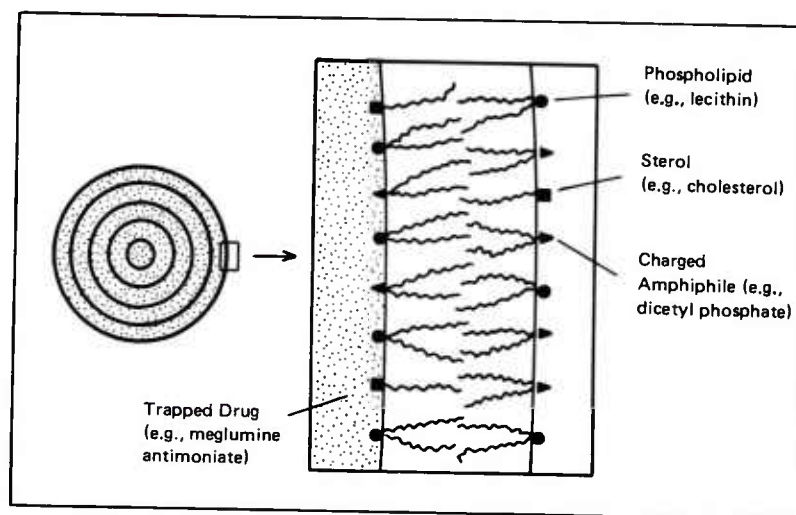


Figure 1. Schematic illustration of a liposome

separated by aqueous interspaces. Various drugs can be trapped in the internal aqueous regions of the particles (7-10). Upon intravenous injection the liposomes automatically and rapidly (within minutes) "home" to the same cells that contain the Leishmania, namely the macrophages in the reticuloendothelial system (7-10). The drug is gradually released in a high localized dose in the vicinity of the parasite. In this paper we describe the effect of liposome-encapsulated antimonial drug on the treatment of experimental leishmaniasis in hamsters.

MATERIALS AND METHODS

Testing of suppressive effects of drugs in an experimental infection was performed by a slight modification of previously published methods (11,12). Young 50-70 g golden hamsters (Mesocricetus auratus) were injected intracardially with the Khartoum strain of Leishmania donovani. Each injection contained 10^7 amastigotes obtained from spleens of donor hamsters.

After either 3 days, 10 days, or 17 days post-inoculation of

parasites, test drugs in three different dosage levels were administered intracardially daily, under "blind" experimental conditions, for four consecutive days. Intracardial injections were used because intravenous administration is not practical in hamsters, and intramuscular or intraperitoneal routes were not effective. Eight or nine animals were used for each experimental group. One day after the last treatment injection, all of the animals of each group were sacrificed. The livers were removed, and the total numbers of parasites per liver were determined from impression smears. Percent of parasite suppression at each dosage level was calculated by comparison to a parallel control group consisting of 6-8 infected, untreated animals that had been injected intracardially with corresponding volumes of normal saline.

Lipids were purchased from the following sources: phosphatidyl choline (Sigma Chemical Co., St. Louis, MO); cholesterol (Calbiochem, La Jolla, CA); dicetyl phosphate (K and K Laboratories, Plainview, NY). Meglumine antimoniate (Glucantime R) was purchased from Rhodia, Inc., New York, NY. Sodium stibogluconate (Pentostam R) powder was generously supplied as a gift by Dr. R.A. Neal, Wellcome Foundation Ltd., Beckenham, Kent, England.

Liposomes were prepared from a mixture of dipalmitoyl phosphatidylcholine (Sigma Chemical Co.), cholesterol (Calbiochem), and dicetyl phosphate (K and K Laboratories) in molar ratios of 2/1.5/0.22. The lipids, in chloroform, were dried in a pear-shaped flask on a rotary evaporator, followed by one hour under high vacuum in a desiccator. A small amount of acid-washed 0.5 mm glass beads was added, and followed by addition of a sufficient quantity of either 0.15 M NaCl or 0.308 M Meglumine Antimoniate (obtained from Rhodia, Inc. as Glucantime R) such that the phosphatidyl choline was 10 mM with respect to the final aqueous dispersion. The liposomes were swollen by shaking for two minutes on a Vortex mixer.

They were washed three times by diluting in 10 volumes of 0.154 M NaCl and centrifuging at 20,200g for 10 min at 22°. The final pellet was suspended with sufficient 0.15 M NaCl so that the phospholipid was about 10-80 mM with respect to the aqueous suspension. An aliquot (0.4 ml) was shaken with 2.1 ml of water and 2.5 ml of chloroform to disrupt the liposomes; the chloroform was washed twice with 2.5 ml of water; and the combined aqueous phases were sent an analytical laboratory (Galbraith Laboratories Inc., Knoxville, TN) for quantitative antimony analysis. The washed liposomes trapped ca. 2-11% of the antimonial drug present in the original swelling solution.

RESULTS

Fig. 2 compares the efficacies of encapsulated and unencapsulated meglumine antimoniate (Glucantime[®]) in suppression of leishmaniasis. It is evident that the encapsulated drug was superior to the unencapsulated drug, or to liposomes alone. Calculations from the data of Fig. 2, based on the amount of drug required to cause 50% suppression, suggested that the liposome-encapsulated drug was approximately 350 times more effective than the drug alone.

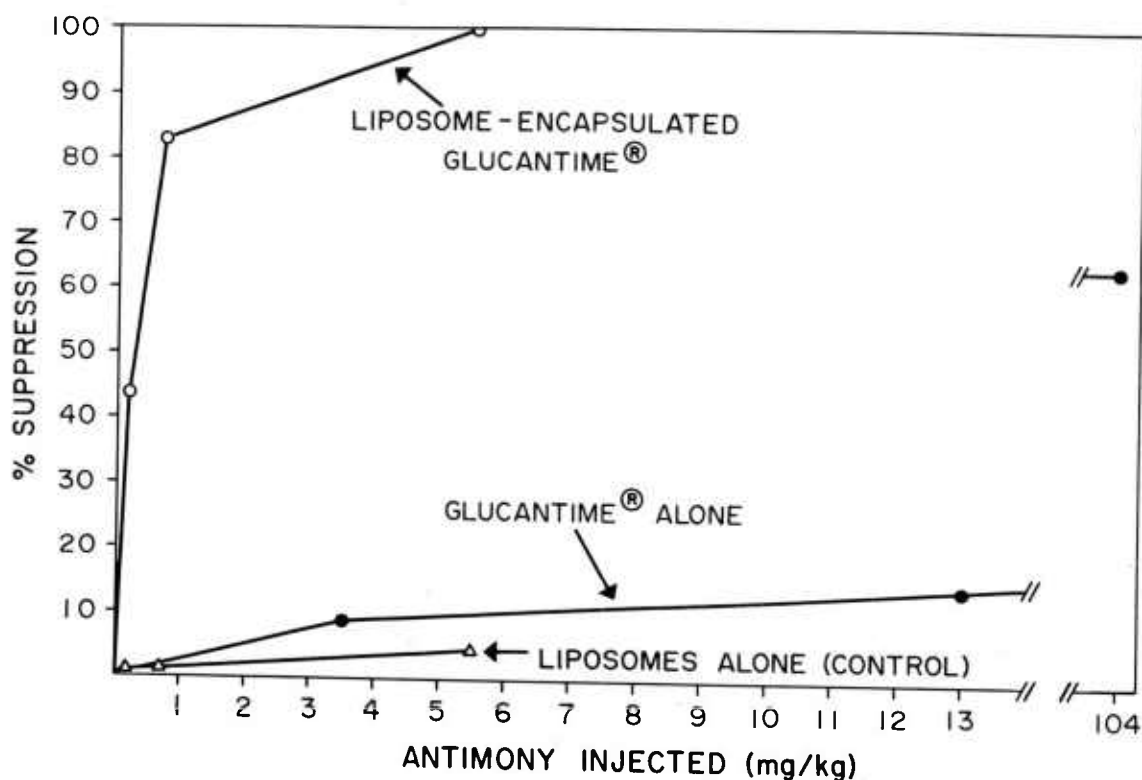


Figure 2.

Suppression of Leishmanial Infection by Liposome-Encapsulated Antimonial Drug. The hamsters had been infected for 10 days prior to treatment, and the indicated doses were administered daily for four consecutive days. In the control the equivalent volumes of liposomes swollen in normal saline were given instead of liposomes swollen in antimonial.

The data of Fig. 2 were obtained with animals whose treatment had been initiated 10 days following infection. Table 1 shows that the length of infection influenced the efficacy of therapy. Infections were compared that were 3, 10 or 17 days in duration before starting therapy. The superior effectiveness of liposome-encapsulated meglumine antimoniate was greater in a long-term (17 day) infection than it was in shorter term (3 or 10 day) infections. It should be noted that in Table 1 more than 100 times as much unencapsulated, compared to encapsulated, drug was used.

TABLE 1. Influence of Length of Infection on Efficacy of Treatment

Treatment Used	Time between infection and start of treatment (days):		
	3	10	17
	%Suppression		
Liposomes containing Glucantime ^R (1 mg/kg/day for 4 days)	99.8	82.8	61.3
Glucantime ^R alone (104 mg/kg/day for 4 days)	99.8	63.7	18

Two different antimonial drugs, meglumine antimoniate (Glucantime^R) and sodium stibogluconate (Pentostam^R) were compared in a long-term (17 day) infection (Table 2). The doses required to produce 50% suppression (SD₅₀) were determined. Under the conditions used, the liposome-encapsulated drugs were enhanced 700-900 times compared to the unencapsulated drugs (Table 2).

DISCUSSION

Prior to the introduction of currently used drugs, mortality in visceral leishmaniasis was higher than 90%; now it is reported at a still high level of 2-5% (3). The only drug available in the United States for treating all forms of leishmaniasis is sodium stibogluconate, and it has the status of an Investigational New Drug (IND) (1, 4). In the present study we demonstrate that encapsulation of sodium

TABLE 2. Relative Efficacies of Two Liposome-Encapsulated, or Unencapsulated, Antimonial Drugs in Treatment of a 17 Day Leishmanial Infection.^a

Therapeutic Agent	SD ₅₀	Enhancement Factor
Glucantime ^R	175	—
Liposome-Encapsulated Glucantime ^R	0.24	729
Pentostam ^R	450	—
Liposome-Encapsulated Pentostam ^R	0.52	865

a

The SD₅₀ is defined as the amount of drug antimony (mg/kg/day for 4 days) required to cause 50% parasite suppression. The enhancement factor is the SD₅₀ ratio of liposome-encapsulated drug vs. unencapsulated drug. In this experiment, dimyristoyl, rather than dipalmitoyl, phosphatidyl choline was used.

stibogluconate in liposomes resulted in approximately a 900-fold increased efficacy compared to unencapsulated drug (Table 2). The differences were more marked in long-term "chronic" than in short-term "acute" infections (Table 1). Chronic infections are the types most likely to be encountered in patients at the time of therapy. Meglumine antimoniate is widely used in other countries (1,4), and its effectiveness also was greatly enhanced (more than 700-fold) by encapsulation in liposomes (Table 2).

Antimonial agents are the drugs of choice in leishmaniasis, and in case of treatment failure another toxic drug, amphotericin B, often is used as a last resort (1,4). Antimony belongs to the same periodic group as arsenic, and toxicities, particularly to the heart, kidneys and liver are similar to those of arsenicals (1,13). Although the pentavalent antimonials, such as meglumine antimoniate and sodium stibogluconate, are a great improvement over trivalents, the pentavalents still are highly toxic, particularly in high or prolonged dosage, or in undernourished individuals (1). The doses of antimonials that can be given to patients are limited by potential toxicity, and

the ratio of (effective dose/toxic dose) is not very high. Although we do not have definitive evidence yet, we anticipate that encapsulation of antimonial drug in liposomes, and rapid uptake of liposome by reticuloendothelial cells, will minimize systemic toxic antimonial effects, especially those due to acute cardiomyopathy and toxic nephritis. Toxicity also may be minimal because, based on our animal model, less than 0.15% of an ordinary therapeutic dose may be used for equivalent results. In the doses approved for clinical use (for example, under the conditions of the IND for sodium stibogluconate) treatment failures are common. Because of the greatly enhanced efficacy of liposome-encapsulated drugs, it may be possible to overcome the problem of treatment failure by raising the dosage without increasing the risk of severe toxicity.

SUMMARY

We describe a novel technique for treating leishmaniasis by encapsulation of antimonial drugs in liposomes. The liposomes travel in the bloodstream to the same cells in which the Leishmania organism lives, namely the phagocytes of the reticulothelial system in the liver and spleen. The systemic toxicities of the antimonial agents, which ordinarily are substantial, presumably would be minimized by encapsulation in liposomes and by rapid uptake of liposomes by phagocytes. In an animal model, consisting of experimental visceral leishmaniasis in hamsters, we found that liposome-encapsulated antimonials were 350 - 900 times more effective than unencapsulated drugs in suppressing the infection.

ACKNOWLEDGEMENT

Supported in part by United States Army Medical Research and Development Contract No. DAMD-17-C-5011. The authors are indebted to Mrs. Mary Watson for preparation of the manuscript.

REFERENCES

1. Steck, E.A. The Leishmaniasis. In: Progress in Drug Research, Vol. 18, (E. Jucker, Ed.) Birkhäuser Verlag, Basel, pp 290-351, 1974.
2. Faust, E.C., Beaver, P.C., and Jung, R.C. The Leishmania Parasites of Man. In: Animal Agents and Vectors of Human Disease, Lea and Febiger, Philadelphia, pp. 34-64, 1968.
3. Biagi, F. Leishmaniasis-Introduction, Kala Azar, Cutaneous and Mucocutaneous Leishmaniasis. In: Tropical Medicine, 5th ed. (G.W. Hunter, III, J.C. Swartzwelder, and D.F. Clyde, Eds.), W.B. Saunders,

*ALVING, STECK and HANSON

Co., Philadelphia, pp. 411-429, 1976.

4. Mahmoud, A.A.F., and Warren, K.S. Algorithms in The Diagnosis and Management of Exotic Diseases. XXIV Leishmaniasis. J. Inf. Dis. 136: 160-163, 1977
5. Bangham, A.D. Membrane Models with Phospholipids. Prog. Biophys. Mol. Biol. 18: 29-95, 1968.
6. Paphadjopoulos, D. Phospholipid Membranes as Experimental Models for Biological Membranes. In: Biological Horizons in Surface Science (L.M. Prince, and D.F. Sears, Eds.) Academic Press, NY, pp. 159-225, 1973.
7. Gregoriadis, G. The Carrier Potential of Liposomes in Biology and Medicine. New Eng. J. Med. 295: 704-710, 765-770, 1976.
8. Tyrrell, D.A., Heath, T.D., Colley, C.M., and Ryman, B.E. New Aspects of Liposomes. Biochim. Biophys. Acta. 457: 259-302, 1976.
9. Poste, G., Papahadjopoulos, D., and Vail, W.J. Lipid Vesicles as Carriers for Introducing Biologically Active Materials into Cells. In: Methods in Cell Biology, Vol. 14 (D.M. Prescott, Ed.), Academic Press, NY, pp. 33-71, 1976.
10. Fendler, J.H., and Romero, A. Liposomes as Drug Carriers. Life Sci. 20: 1109-1120, 1977.
11. Hanson, W.L., Chapman, Jr., W.L., and Kinnamon, K.E. Testing of Drugs for Antileishmanial Activity in Golden Hamsters Infected with Leishmania Donovanii. Int. J. Parasitol. 7: 443-337, 1977.
12. Alving, C.R., Steck, E.A., Hanson, W.L., Loizeaux, P.S., Chapman, Jr., W.L., and Waits, V.B. Improved Therapy of Experimental Leishmaniasis by Use of a Liposome-Encapsulated Antimonial Drug. Life Sci. 22, 1978 (In Press).
13. Harvey, S.C. Heavy Metals. In: The Pharmacological Basis of Therapeutics, 5th ed. (L.S. Goodman, and A. Gilman, Eds.) Macmillan, NY, pp. 924-945, 1975.

In conducting the research described in this report, the investigators adhered to the "Guide for the Care and Use of Laboratory Animals," as promulgated by the Institute of Laboratory Animal Resources, National Academy of Sciences, National Research Council.

SEMI-INSULATING GALLIUM ARSENIDE FOR MICROWAVE AND
MILLIMETER-WAVE DEVICE APPLICATIONS (U)

THOMAS R. AUCOIN, MR., RAYMOND L. ROSS, MR.,
MELVIN J. WADE, MR., AND ROBERT O. SAVAGE, MR.
USA ELECTRONICS TECHNOLOGY AND DEVICES LABORATORY (ERADCOM)
FORT MONMOUTH, NEW JERSEY 07703

INTRODUCTION

A wide variety of semiconductor devices utilizing gallium arsenide (GaAs) are currently under development by the Army for use in secure communication and improved surveillance systems. GaAs is a semiconductor material characterized by a very high electron mobility, a direct band gap of 1.43 eV, and a high intrinsic resistivity ($\sim 10^8$ ohm-cm). The low field mobility of electrons in GaAs is one of its greatest attributes and offers high frequency operation in devices such as the field effect transistor (FET). The GaAs FET, a critical component in emerging military systems, is experiencing a rapidly expanding use in oscillator, mixer, logic element, power amplification, and low noise/high gain applications. This device threatens to replace virtually all low noise traveling wave tubes at frequencies from 4 to 30 GHz; is an attractive substitute for difficult to manufacture devices such as Gunn diodes for 4 to 8 GHz operation; and, monolithically fabricated as logic gates, can function at speeds to 10 GHz. In general then, GaAs FET's will eventually replace device configurations of more complex circuitry. However, its full potential has not been realized, largely due to the chronic material problems experienced by device manufacturers.

It is commonly held that the quality of semi-insulating substrates available to GaAs FET manufacturers is one of the major barriers to achieving high performance and reliable devices (1). Recent studies relating substrate quality to device performance stress that the principal problem appears to be the outdiffusion of acceptors such as chromium (added for compensation) or those caused by point

AUCOIN, ROSS, WADE and
SAVAGE

defects in the crystal lattice. In order to prevent this diffusion, one must either grow a buffer layer or reduce the acceptor impurity levels contained in the starting materials and/or resulting from the growth process. Also, the formation of crystal dislocations and vacancies must be minimized during the growth of the GaAs boule. These material problems directly manifest themselves in the GaAs FET's electrical parameters and reliability, particularly its drift in gain characteristics. Not only do the devices exhibit both short- and long-term drift, but each device appears to have its own unique drift signature. This drift phenomena seems to be caused by two mechanisms; surface depletion effects, and electrically active impurities and traps at the interface between the substrate and the active device layer. The solution to problems associated with impurities, vacancies, and traps within the semi-insulating substrates is the major objective of this research effort.

CRYSTAL GROWTH

Gallium arsenide, a III-V compound semiconductor, has the zinc blende structure consisting of two interpenetrating face centered cubic (fcc) sublattices. One fcc sublattice describes the position of the Ga atoms and the other, As atoms. The technology involved in the growth of GaAs is considerably more complicated than that employed for silicon. In the case of GaAs, one is dealing with more complicated binary phase equilibria and a highly volatile component, arsenic. Exact stoichiometry of the GaAs compound must be maintained during the growth process in order to achieve high mobility and crystal perfection. This requires precise control of the arsenic vapor pressure in the chosen growth system. The bulk compound is normally formed by the reaction of arsenic vapor with Ga metal at elevated temperatures in sealed quartz ampoules as shown in Figure 1. Typically, an As reservoir contained at one end of the ampoule is heated to 600 C. This generates approximately 1 atm of arsenic pressure in the system, a prerequisite for maintaining the stoichiometry of a GaAs melt (2). The arsenic vapor is reacted with gallium metal (~ 1260 C) contained in a quartz or pyrolytic boron nitride boat located at the other end of the ampoule. After the Ga has been completely reacted, single-crystal growth may be initiated by programmed cooling (gradient freeze) or by physically moving the ampoule (horizontal Bridgman) through proper temperature gradients. This indirect approach for the compounding and growth of GaAs is used because of the high vapor pressure of As at its melting point and at the melting point of GaAs, ~ 20 atm at 817 C, and ~ 60 atm at 1238 C, respectively.

Another approach to the growth of bulk GaAs, which is gaining favor, is the liquid encapsulated Czochralski technique (LEC CZ)

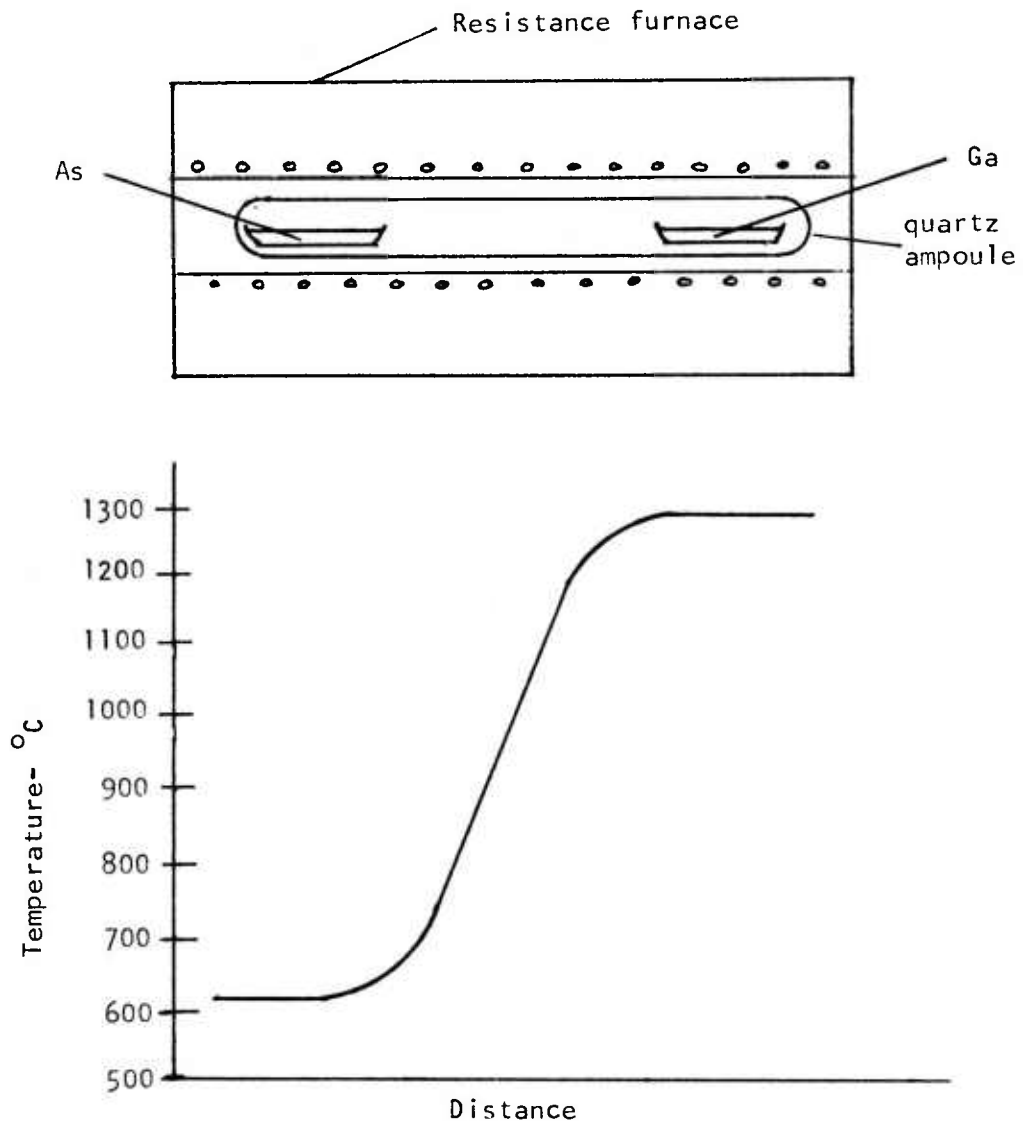


Figure 1. GaAs compounding furnace and temperature profile.

shown in Figure 2. In this method, the vaporization rate of arsenic from molten GaAs is reduced by placing a molten layer of a nonreactive encapsulant (boron oxide) on the melt surface. An inert gas pressure, which is higher than the vapor pressure of the melt, is then maintained over the molten boron oxide layer. A rotating seed crystal contacting the molten GaAs is then slowly withdrawn through the liquid encapsulant while the temperature is regulated to obtain the desired diameter crystal. Although this method has the advantages of being able to prepare large diameter crystals in a relatively short period of time, a major drawback is the fact that the GaAs feed material must be compounded in quartz ampoules. This two-stage process has an inherently higher residual donor concentration relative to gradient freeze or horizontal Bridgman grown GaAs due to the increased handling. A one-step process would be of great importance to microwave device manufacturers in that higher purity, large diameter substrates could be more economically produced.

The use of quartz as a container for As, Ga, and molten GaAs, and as a reaction vessel can lead to appreciable silicon contamination (3). Since silicon can act as a shallow donor or acceptor in GaAs (4), incorporation of a very small amount can preclude one from obtaining high resistivity material. Various schemes have been used to reduce the dissociation of the quartz components in order to suppress silicon contamination. These include pyrolytic boron nitride liners (5), deliberate introduction of oxygen (6), and addition of gallium oxide (7). All schemes in use still add impurities which detrimentally affect mobility, resistivity, thermal conductivity, and crystal perfection. Chrome, a midgap acceptor, is normally added to compensate shallow donors in GaAs, thereby making it semi-insulating ($>10^8$ ohm-cm). There are, however, drawbacks to its use. If the GaAs material has too many shallow donors ($> 5 \times 10^{16}/\text{cm}^3$), not enough chrome can be added to compensate the carriers because of its small distribution coefficient in GaAs ($\sim 10^{-4}$).

The addition of a large amount of chrome to the melt can cause formation of precipitates in the growing crystal (8). Also during epitaxial processing and device operation, chrome has been observed to migrate to the substrate/epi-layer interface and act as a charge trap (9). The in-situ compounding and subsequent crystal growth of semi-insulating GaAs, which we describe in this article, is a one-step process which completely eliminates, for the first time, all dependence on quartz components. This approach will lead to reduced silicon and oxygen impurity levels and thus reduce or eliminate the present levels of chrome doping required for compensation.

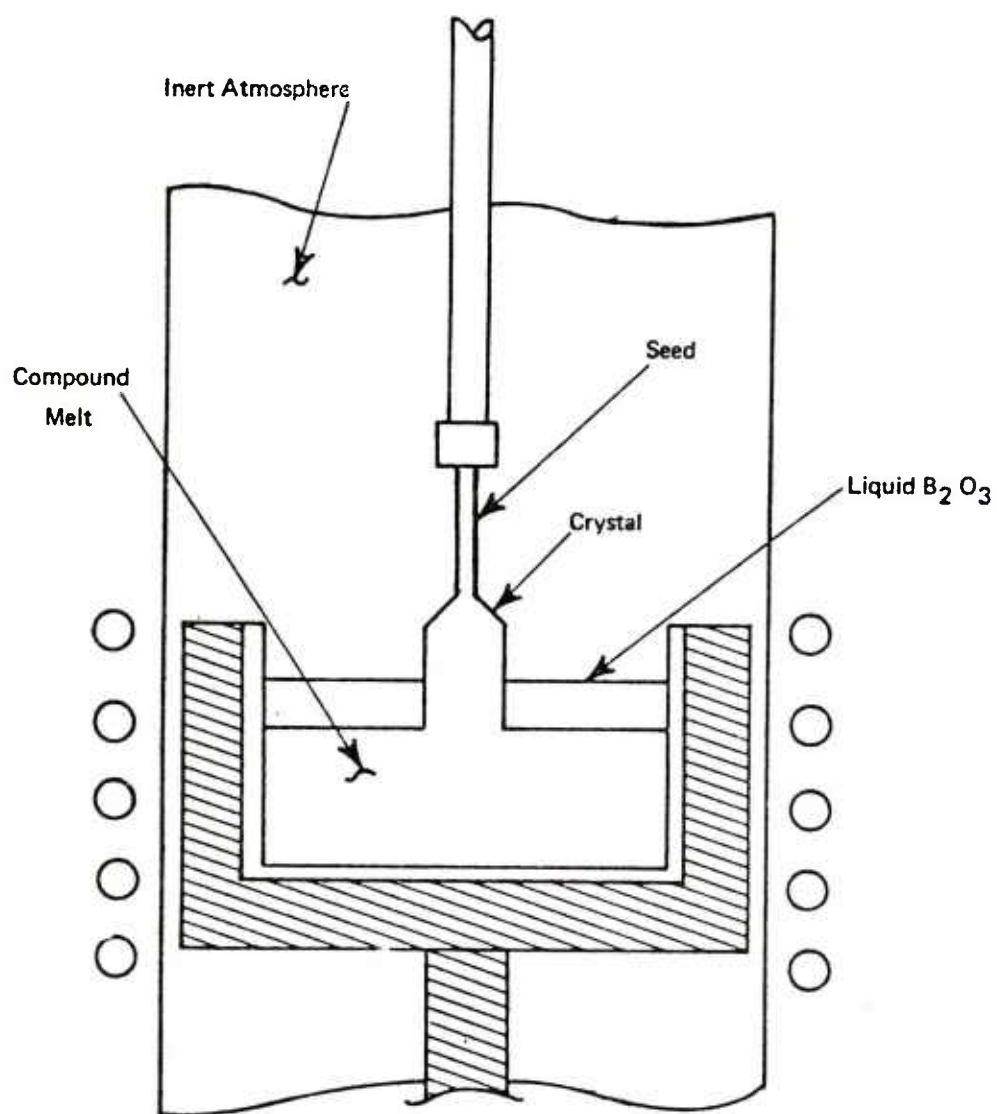


Figure 2. Liquid encapsulated Czochralski pulling technique.

The technique which we have developed for the growth of high purity semi-insulating GaAs consists of reacting elemental gallium and arsenic under a molten encapsulant. The charge was contained in pyrolytic boron nitride crucibles at nitrogen pressures to 100 atm. A specially designed high pressure Czochralski crystal pulling system was used for maintaining the high pressure nitrogen gas ambient and for growth of the bulk GaAs single crystals (Figure 3). This system contains provision for the following: rotation of the crystal and/or crucible; raising and lowering of the crystal and/or crucible; remote control of temperature, gas flow pulling and rotation rates; TV process monitoring; cryo fore-pumping; vac-ion pumping to 10^{-5} torr; hydraulic lifting of the chamber; and is designed for operation at pressures to 135 atm with neutral, oxidizing or reducing ambients. Other facilities have been installed to support this system with high-purity water, high-purity high-pressure nitrogen gas, vacuum drying, and a dust-free fume hood with rinse tanks.

The in-situ compounding and LEC CZ crystal growth of GaAs was carried out in the following manner. Stoichiometric quantities of high-purity gallium (Aluisse 99.9999%⁺) and arsenic (Canyonlands 21st Century) up to a total of 400 grams were weighed and placed in pyrolytic boron nitride (PBN) crucibles approximately 5 cm in diameter. The PBN crucibles were cleaned prior to loading by etching in a 1:1 solution of electronic grade HCl and de-ionized water for 10 minutes. After etching, the crucibles were rinsed with de-ionized water, methanol, and vacuum dried overnight at 200 C. The crucibles after being loaded were placed in carbon coated graphite or tantalum susceptors located inside the Czochralski pulling chamber. A 450-kHz radio frequency induction unit was used for heating. A dehydrated pellet of boron oxide encapsulant (Yamanaka Chemical Ind. Ltd.), weighing approximately 30 grams and 3.5 to 4.8 cm in diameter, was placed on top of the charge as shown in Figure 4. When molten, this pellet provided an encapsulant depth of approximately 1.2 cm.

The moisture content of the boron oxide encapsulant has been found to be critically related to the growth of high quality, twin-free single crystals of GaAs. If the boron oxide contains an excessive amount of moisture, many bubbles are continuously formed by the reaction of this moisture with the GaAs melt. These bubbles form at the GaAs melt/liquid encapsulant interface and subsequently rise to the surface releasing arsenic. These bubbles can promote twinning of the crystals and lead to nonstoichiometric arsenic deficient GaAs. Although the pellets of boron oxide appear completely dehydrated, they still contained a considerable amount of moisture, and use as received was unsatisfactory. An oil-free, high-vacuum baking system was constructed to further dehydrate the pellets. Heating the boron oxide

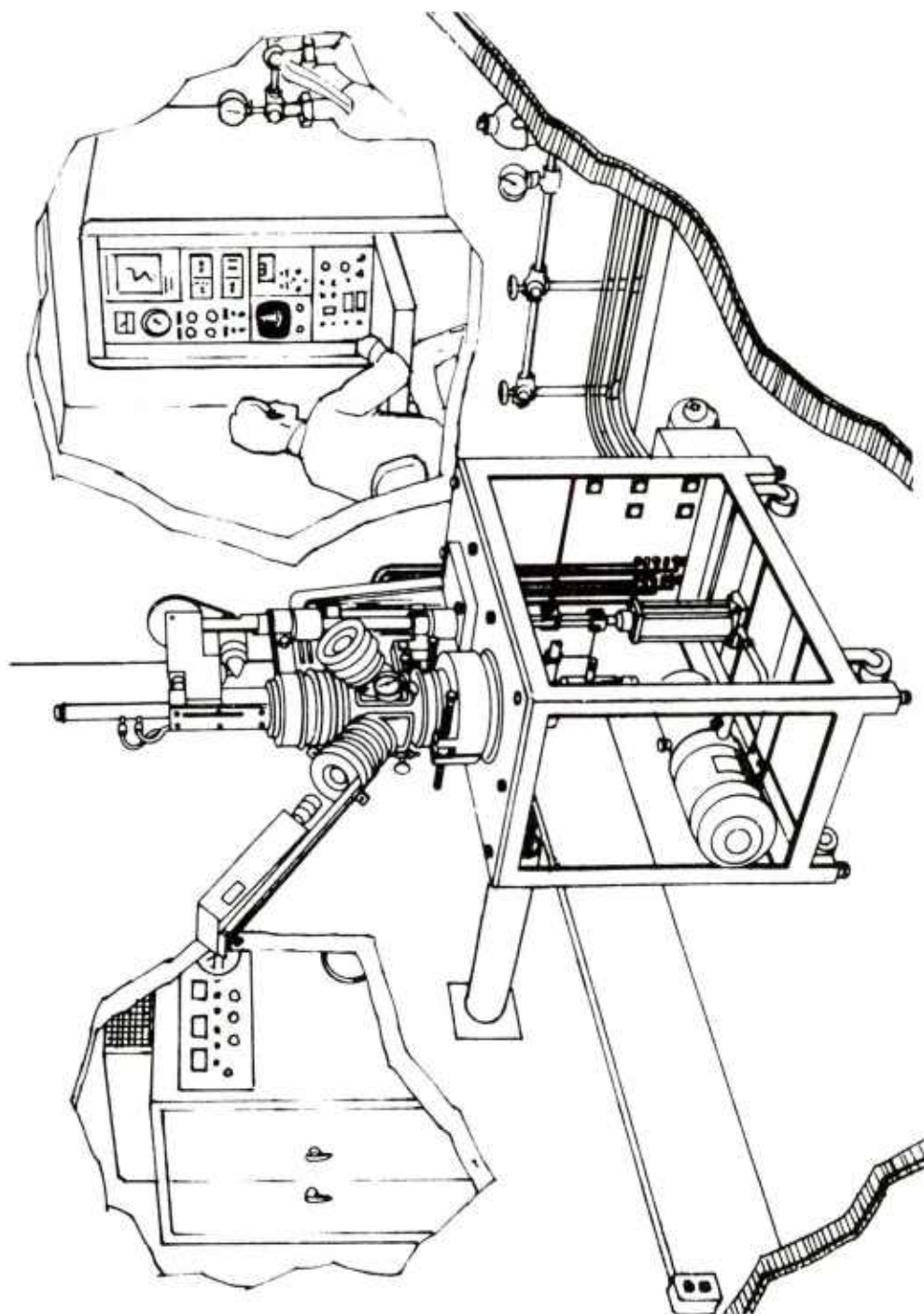


Figure 3. High pressure Czochralski crystal growth system.

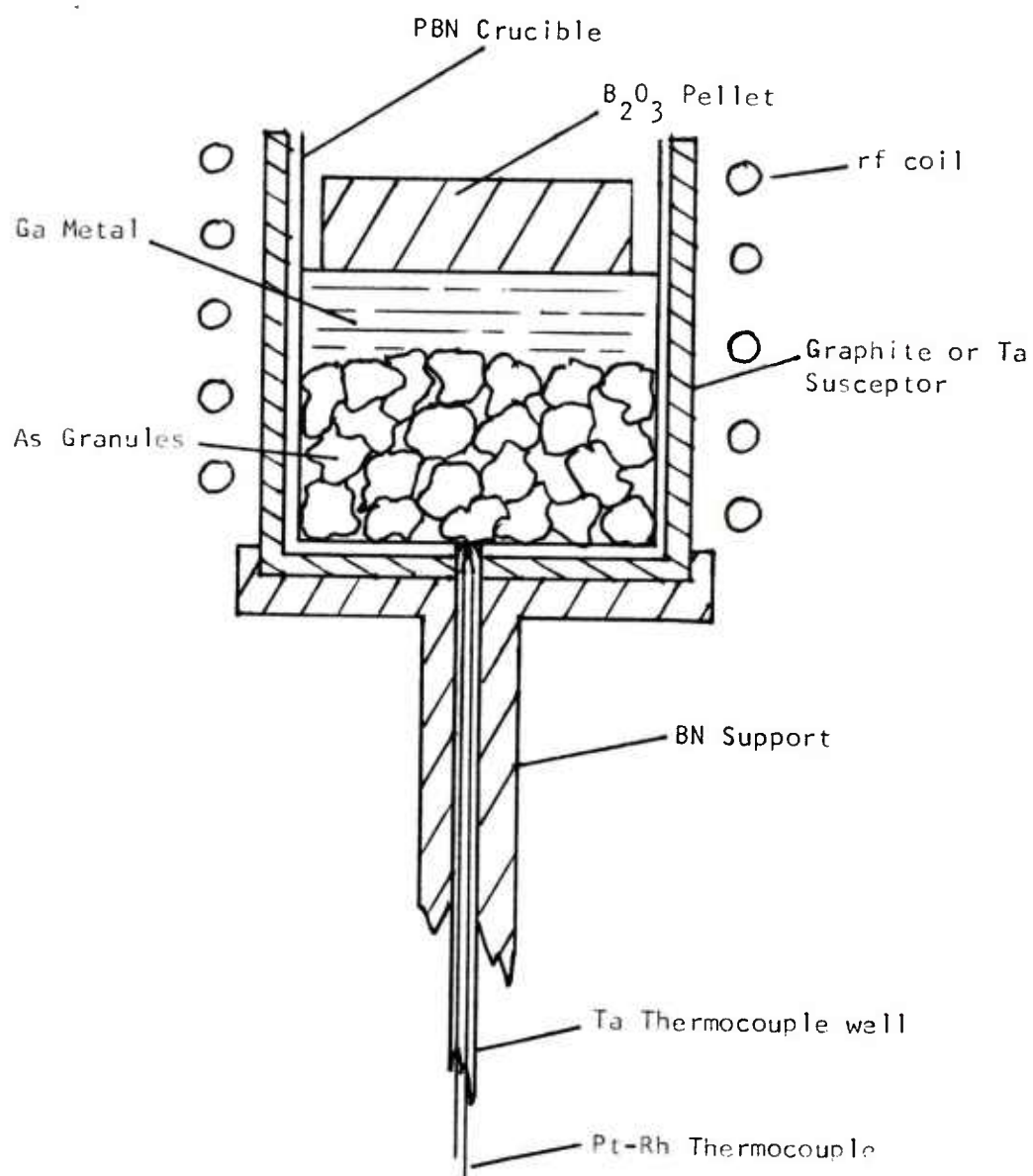


Figure 4. Experimental arrangement for in-situ compounding of GaAs.

AUCOIN, ROSS, WADE and
SAVAGE

to 1,000 C at 10^{-7} torr in induction heated platinum/gold crucibles was found to give very satisfactory results. After cooling to room temperature, the pellets were quickly removed and placed on top of the charge in the pulling chamber.

After the growth station was established (susceptor, crucible, gallium, arsenic, boron oxide, etc.) and the chamber closed, the system was evacuated to 10^{-2} torr by a cryopump and to 10^{-5} torr by a vac-ion pump. The system was then slowly heated to 325 C to remove any residual moisture and to vaporize arsenic oxides. If the temperature exceeded 325 C, significant amounts of arsenic were found to have been lost by volatilization prior to compounding. After several hours, the system was then backfilled with high purity nitrogen gas to 3.3 atm and the temperature increased to 450 C. During this step, the boron oxide was melted and flowed into the voids. The high purity nitrogen gas used in these runs was obtained by vaporizing liquid nitrogen at 3.3 atm and then charging storage tanks at 100 atm with an oil-free, high-pressure booster pump. After the boron oxide was completely melted, the nitrogen gas pressure in the pulling chamber was increased to the pressure to be used during GaAs compounding. This was varied from 40 to 80 atm and was found to greatly influence the yield of usable GaAs. At 40 atm, a GaAs yield of 90% was obtained with the remaining arsenic lost from the crucible. At 80 atm nitrogen overpressure, a GaAs yield of 99% was obtained. The reaction could be directly observed with a built-in TV monitor and also by a rapid temperature rise when approaching 700 C. This achievement represents the first time GaAs has been synthesized in a silica- and carbon-free environment and, therefore, should exhibit very low levels of oxygen contamination. Samples are being analyzed to ascertain the levels of these trace impurities and oxygen.

Single crystals of GaAs were grown from the in-situ compounded material directly after synthesis by the LEC CZ technique as previously shown in Figure 2. All crystals were grown in the $\langle 111 \rangle$ B direction (arsenic face towards the melt). Seed material was obtained from Laser Diodes, Inc. Nitrogen ambient gas pressures were used from 1.3 to 7.7 atm absolute during LEC CZ growth. At pressures to approximately 2 atm, a significant amount of arsenic was being lost and severe thermal etching of the crystal's surface was observed. When the pressure was increased to 2.3 atm and above, the arsenic loss was seen to significantly decrease. At nitrogen overpressures of 3 atm and above, thermal etching (arsenic loss from the crystal's surface) was considerably reduced.

Growth rates were generally between 1 to 1.5 cm per hour. Both concurrent and countercurrent rotation of crucible and crystal were

used at rates of 5 and 15 rpm respectively. The rotation direction of the crystal and crucible seemed to have a direct bearing on the shape of the melt/crystal interface, concurrent rotation producing a more convex interface.

ELECTRICAL MEASUREMENTS

The measurements of resistivity, mobility and Hall coefficient were performed by using the van der Pauw method (11) with an automated dc system (Figure 5). The van der Pauw technique is unique in that it requires only four contacts located anywhere on the periphery of a uniformly thick sample of arbitrary shape. However, the gain realized by reduced sample contacting and geometrical constraints is offset by the necessity that current and voltage leads be interchanged between different pairs of contacts. Various sample configurations have been devised to minimize the requirement that the electrical contacts be of a finite size and located at the sample's circumference (12). In this work, the conventional "clover-shaped" sample geometry was replaced by a "Greek-cross" structure which has been shown to be a valid van der Pauw configuration provided the arm length is equal to or greater than the arm width (13).

The resistivity, ρ , in ohm-cm is given by

$$\rho = \frac{\pi t}{\ln 2} \left(\frac{R_{AB,CD} + R_{BC,DA}}{2} \right) F \quad [1]$$

$$\text{where } R_{AB,CD} = \frac{V_{AB}}{I_{CD}}, \quad R_{BC,DA} = \frac{V_{BC}}{I_{DA}} \quad [2]$$

in units of ohms, t is the sample thickness in centimeters and F is the van der Pauw factor which is a dimensionless quantity dependent only upon the ratio of $R_{AB,CD}/R_{BC,DA}$ and is defined by the transcendental equation

$$\cosh \left(\frac{\ln 2}{F} \frac{R_{AB,DC}/R_{BC,DA} - 1}{R_{AB,CD}/R_{BC,DA} + 1} \right) = \frac{1}{2} e^{\ln 2/F} \quad [3]$$

The Hall mobility, μ , in square centimeter per volt second is given by

$$\mu = 10^8 \frac{t}{B\rho} \Delta R_{AC,BD} \quad [4]$$

where B is the applied magnetic field in gauss and $\Delta R_{AC,BD}$ is the change in resistance when the magnetic field is applied perpendicular

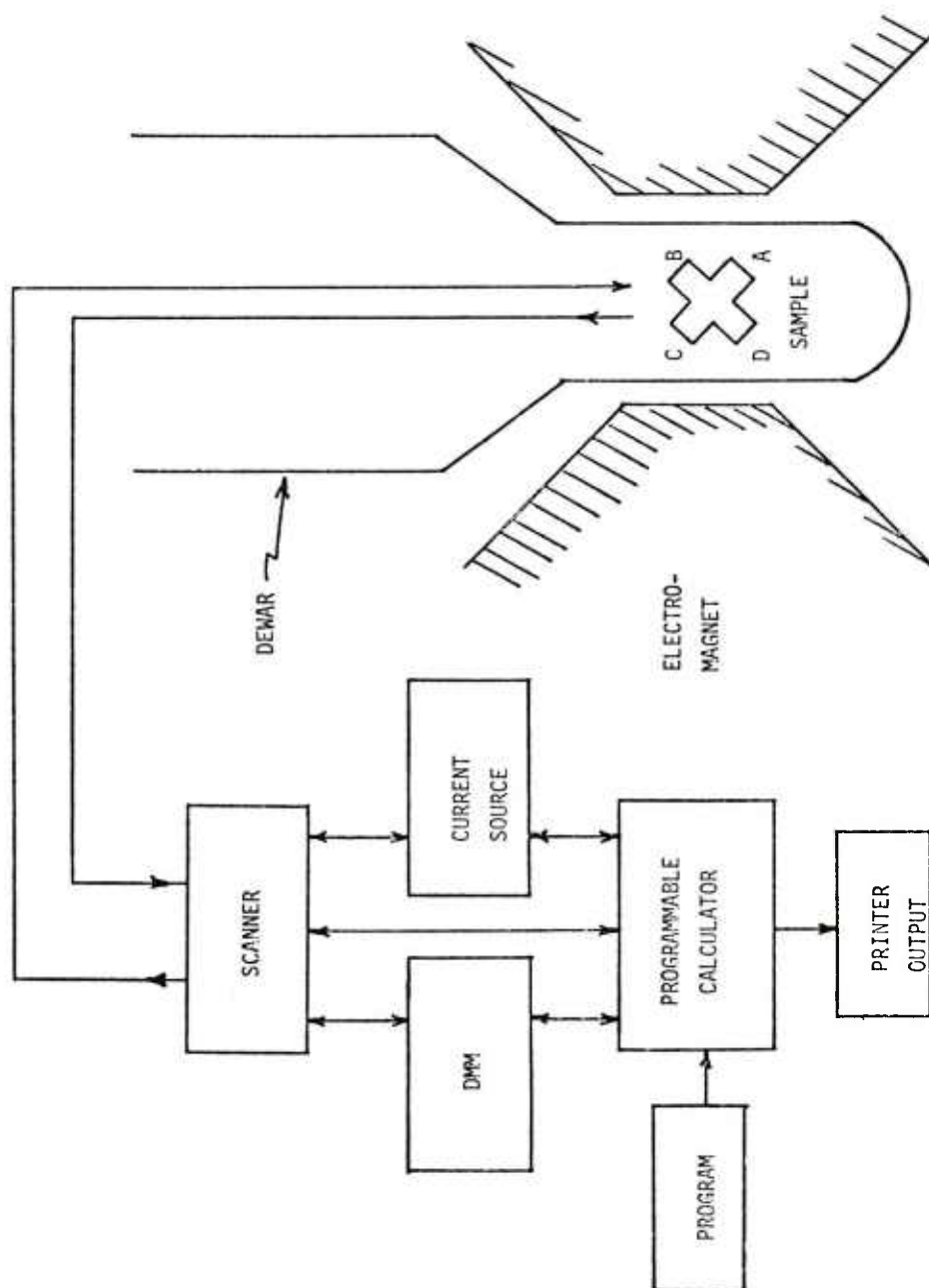


Figure 5. Automated van der Pauw Hall and resistivity system.

AUCOIN, ROSS, WADE and
SAVAGE

to the sample. The carrier density, n , in cm^{-3} is calculated from ρ , μ , and the electronic charge, q , in coulombs, using

$$n = 1/\rho q \mu \quad [5]$$

while the Hall coefficient, R_H , in $\text{cm}^3/\text{coulomb}$ is found from

$$R_H = 1/nq = \rho \mu \quad [6]$$

The interchanging of current and voltage leads, reversing of the direction of current flow and calculation of parameters expressed by Eqs. [1-6] are accomplished by an automated dc system as shown in Figure 5. The scanner, under control of the programmable calculator, sequentially connects and switches the appropriate current and voltage leads from the sample to the current source and digital voltmeter. Measurements of voltage and current are made under forward and reverse current flow. The resistances given by Eq. [2] and contained in Eq. [4] are then determined by the programmable calculator. Knowing the ratio of $R_{AB,CD}$ to $R_{BC,DA}$, a relaxation method incorporating an interactive procedure to obtain successive approximations to the first derivative of Eq. [3] is utilized by the programmable calculator to determine the van der Pauw factor, F . Having determined F , the calculations of ρ , μ , n and R_H proceed with results being tabulated by the output printer. Measurements at both 77 K and 300 K are performed in this manner. For the Hall parameters a dc magnetic field of 5 to 10 kG is applied in both directions.

The "Greek-cross" structure is fabricated by mounting the GaAs substrates on a glass microscope slide, covering the material with a metal mask with subsequent removal of excess material by careful sand-blasting. Electrical contacts are made by alloying small indium pellets at 350 C in a forming gas atmosphere by use of a "Penzac" type furnace.

SUMMARY

The gallium arsenide field effect transistor, a critical component in a wide variety of military communication and surveillance systems, has not achieved its full potential in part due to poor and unpredictable quality semi-insulating substrates. Trace impurities such as silicon, carbon, and oxygen, incorporated during compounding and subsequent crystal growth are believed to be the major causes of the substrate problem. A new modification of the liquid encapsulated Czochralski crystal growth process has been developed which employs in-situ compounding of GaAs from its elements at ambient gas pressures to 80 atm. This approach represents the first time GaAs has been

synthesized and subsequently grown by a one-step process in a silicon- and carbon-free environment. This technique, employing high-purity gas ambients, and oil-free, high-vacuum bake-outs of the starting materials, should lead to the lowest background levels of oxygen, silicon, and carbon obtainable in semi-insulating GaAs grown by the liquid encapsulated Czochralski process. Also, the present requirement for chrome compensation should be reduced or eliminated, dislocation densities decreased, substrate properties made reproducible, and the thermal conductivity improved. Preliminary data on the first GaAs boules grown by this in-situ compounding approach indicates high resistivity material is consistently being grown ($> 10^5$ ohm-cm) without intentional compensation.

ACKNOWLEDGMENTS

The authors wish to thank Drs. C.G. Thornton, J.A. Kohn, and F. Rothwarf for their technical suggestions, encouragement, and support, which were instrumental in the formulation and implementation of this GaAs research program.

REFERENCES

1. G. McCoy, US Air Force Report, "Gallium Arsenide Microwave Device Technology," (1974).
2. L.R. Weisburg, F.D. Rosi, and P.G. Herkart, "Properties of Elemental and Compound Semiconductors," Interscience Publishers, New York, 1959, p. 31.
3. J.W. Harrison, Solid State Technology, p. 49 (Jan. 1973).
4. S.M. Sze and J.C. Irvin, Solid-State Electron. 11, 599 (1968).
5. E.M. Swiggard, S.H. Lee, and F.W. Von Batchelder, Proc. of Sixth Gallium Arsenide and Related Compounds Conference, (St. Louis), p. 23, published by Institute of Physics, Bristol, England, 1977.
6. J.F. Woods, N.G. Ainslie, J. Appl. Phys. 34, 1469 (1963).
7. H. Merkel and S. Leibenzender, US Patent 3,480,394, Nov. 25, 1969.
8. L.F. Eastman, private communication.

AUCOIN, ROSS, WADE and
SAVAGE

9. H.B. Kim, D.L. Barrett, G.G. Sweeney, and T.M.S. Heng, Proc. of Sixth Gallium Arsenide and Related Compounds Conference, (St. Louis), p. 136, published by Institute of Physics, Bristol, England, 1977.
10. L.J. van der Pauw, Philips Rsch Reports 13, 1-9 (1958).
11. L.J. van der Pauw, Philips Tech. Rev. 20, 220-224 (1958/1959).
12. J.M. David and M.G. Buehler, Solid-State Electron., to be published.

DETERMINATION OF WOUND BLOOD FLOW IN THE
THERMALLY INJURED SOLDIER

*LOUIS H. AULICK, MAJ, MSC
DOUGLAS W. WILMORE, DR.
US ARMY INSTITUTE OF SURGICAL RESEARCH
BROOKE ARMY MEDICAL CENTER
FORT SAM HOUSTON, TEXAS 78234

Extensive thermal injury is one of the most severe stresses experienced by individuals in contemporary society. Metabolic, circulatory, and thermoregulatory alterations characterize the major compensatory adjustments following injury and, to a large extent, determine the survival potential of the patient. The initial response to injury, termed "burn shock," usually lasts 24-48 hours and is associated with a marked depression in the patient's metabolic rate, body temperature, and circulation. Following successful resuscitation, the patient gradually becomes hypermetabolic and febrile. These increases in body temperature and metabolism vary with the extent of injury, reach a peak within the first two weeks, and then slowly return to normal with wound coverage and healing. Since these metabolic and thermoregulatory adjustments have been well documented by Wilmore and colleagues (1-4), the intent of these experiments is to characterize the associated cardiovascular changes which occur with thermal injury. Particular emphasis is placed on wound blood flow to include factors involved in its regulation, and the impact of wound perfusion on the total circulatory status of the burn patient.

Acute loss of blood volume precipitates the initial burn shock phase of injury (5). During this phase, the patient becomes hypotensive and cardiac output drops below normal. With volume expansion and increased total peripheral resistance, blood pressure returns toward normal, and cardiac output begins to recover. After 24-48 hours, blood pressure stabilizes at a normal level but cardiac output continues to climb, associated with a progressive increase in plasma volume and decreasing peripheral resistance. The extent of this rise in total body

circulation is generally related to the size of the surface wound and may reach levels 2-3 times normal in the more extensively injured patients. The magnitude of this circulatory response is best appreciated in the sleeping patient when the level of activity reflects the minimal requirements for life. Typically, a patient with a 50% total body surface wound will, in this basal state, maintain a heart rate of 120-140/min and a cardiac index in excess of $7 \text{ L/m}^2 \cdot \text{min}$ (6). For normal uninjured persons to generate comparable levels of total body blood flow, they would have to work at roughly one half of their aerobic capacity. For most people, this would correspond to a good brisk jog. While the well trained athlete may do this amount of exercise for several hours each day, the burn patient must sustain this hyperdynamic circulation for at least several weeks. The exact time course of this circulatory response to thermal injury has not been well defined but presumably follows a course similar to metabolism and body temperature and returns toward normal with wound healing.

Gump and associates (7), in 1970, were the first to partition this extra blood flow. They noted that, as in exercise, the resting cardiac index of burn patients was linearly related to oxygen consumption but, at every level of aerobic metabolism, burn patients had higher cardiac outputs than other febrile surgical patients. They also found that splanchnic blood flow in three severely burned patients was slightly increased but represented a smaller portion of the cardiac output than it did in normals or patients with postoperative infection. From this, they concluded that a large portion of the increased cardiac output of burn patients was directed to the periphery.

Later, Wilmore, et al (3), provided additional indirect evidence to suggest that most of the extra peripheral circulation was directed to the body surface. They found that burn patients maintained above-normal surface temperatures despite increased evaporative cooling of the wound and concluded that this elevated skin temperature could be the result of increased superficial blood flow. Additional support for this thesis was provided by demonstrating that the coefficient of core-to-skin heat conductance, an index of skin blood flow, was twice normal in burned patients. Although these indirect measurements did suggest an increase in superficial blood flow in burn patients, the influence of the burn wound on the distribution of this elevated peripheral blood flow remained unknown. To address this problem, a series of studies were designed to measure peripheral blood flow in injured and uninjured limbs of burn patients and to partition this flow into that directed to the burn wound, normal skin, and resting skeletal muscle of the limbs.

MATERIALS AND METHODS

Subjects

Over 60 thermally injured patients were studied. The size of injury ranged from 3-86% of the total body surface. Patients selected for study were 1) from 15 to 55 years of age and free of any disease prior to injury; 2) normotensive and hemodynamically stable after an uneventful resuscitation; 3) in a normal state of hydration with a hematocrit greater than 33, and without abnormalities in serum electrolyte concentration, osmolality, or pH; 4) free of systemic infection as determined by clinical symptoms and signs and daily blood cultures; 5) 6-27 days postinjury, after removal of the eschar, and before significant healing had occurred.

The wounds were treated by a variety of techniques. Most patients were treated by the exposure method and received topical applications of a silver sulfadiazine cream (Silvadene cream) to the injured surface, but several patients were treated with 11% mafenide acetate topical antibiotic (Sulfamylon cream). Some wounds were covered with dressings soaked in saline, 5% mafenide-saturated dressings, or 0.5% silver nitrate; two were treated with cutaneous allograft biological dressings. Although the effects of these treatments on limb blood flow are unknown, they do involve considerable manipulation of the patient, often resulting in discomfort. Therefore, to insure that each subject was well rested for the study, such procedures were minimized for at least 8 hours prior to the study.

Study Design

Two basic studies were performed. In the first, total limb blood flow was measured by a specially designed venous occlusion plethysmograph. In the second study, resting skeletal muscle blood flow was determined by a standard clearance technique.

The experiments designed to measure leg blood flow took place in an environmental chamber described previously (3). Room temperature was maintained at 30°C, and relative humidity ranged between 40 and 50%. Control subjects were studied in shorts or shorts and halter, and patients were similarly draped with light cotton towels. The post-absorptive subject was moved to the study room in the early morning and placed in bed. Water was given on request and, in some patients, intravenous infusion of 0.04 M sodium chloride solution was maintained to insure normal hydration. Nine copper-constantan thermocouples were attached to the skin at the same sites for all subjects (dorsum of foot, lateral and posterior calf, posterior and anterior thigh, dorsum

of hand, forearm, abdomen, and low back). Leg skin temperatures were monitored from both legs, using five additional thermocouples, in patients with asymmetric leg burns. In those patients treated in dressings, the thermocouples were placed on the wound under the dressing. A rectal probe was inserted to a depth of 10 cm from the external anal sphincter. Rectal and total body mean skin temperatures were monitored at 5-minute intervals throughout the study to insure that subsequent manipulation of the subject had no appreciable effect on body temperature.

After the subject had rested quietly for at least one hour, the steady-state mean leg skin temperature was determined for the left leg, and the thermocouples were removed. This leg was then inserted into a soft pliable, water-impermeable boot and placed in a full-length plethysmograph. Water was added to the plethysmograph at a temperature equal to the predetermined mean leg skin temperature and maintained at this temperature throughout a 30-minute equilibration period and the following 8-10 blood flow measurements. Setting water temperature equal to that of the leg surface when exposed to the air maintained comparable rates of heat exchange across the limbs while in the plethysmograph and prevented either transient or prolonged changes in body temperature during leg blood flow measurements. A 2- to 3-minute interval separated the blood flow determinations.

The validity (8), simplicity, and noninvasive character of venous occlusion plethysmography make this the optimal approach to the study of peripheral blood flow in critically injured patients. The plethysmograph used in this study, a rigid, rectangular box made of clear plexiglass, is illustrated in Fig. 1. To facilitate its use in injured limbs, it can be disassembled into three sections, a thigh plate and attached boot, a trough section with mesh sling to support the leg, and a full length top. The patient's leg was slipped through a tailor made opening in the thigh plate and into a large loose fitting, polyvinyl boot. The boot served to form a freely expandable, watertight seal between the limb and the plethysmograph, preventing fluid exchange across the burn wound and minimizing contamination. The boot and thigh plate were advanced to the proximal thigh and the leg placed in the mesh sling of the plethysmograph. The three sections of the plethysmograph were then locked together and the box filled with water equal to the mean skin temperature of the leg under study.

Venous occlusion was accomplished by rapid inflation of a 10 centimeter wide tourniquet cuff placed as high on the upper thigh as possible. Occlusion pressure was varied for each subject to obtain a

maximal rate of limb swelling. With venous occlusion, the limb swells, and the change in limb volume causes water to rise in a chimney located on top of the plethysmograph. The increase in column hydrostatic pressure is converted to an electrical signal, amplified and recorded.

The plethysmograph was calibrated periodically with the leg in place. The volume of the limb within the plethysmograph was determined by subtracting the volume of water in the plethysmograph from its known capacity. The eight to 10 flow measurements were averaged and leg blood flow expressed in ml/100 ml leg volume per minute.

In a separate set of experiments, resting skeletal muscle blood flow was determined in 10 burn patients (mean burn size = 42.5% total body surface, range 25.5-82%). The criteria for patient selection were the same as described earlier. All subjects were confined to bed for a minimum of one hour prior to muscle blood flow measurements. The actual studies took place in the Nuclear Medicine Clinic of the hospital where the ambient temperature was 25-27°C. Since this was slightly below thermal neutrality for resting burned patients, patient comfort was achieved by covering them with light cotton blankets. All subjects rested supine throughout the 20-30 minute study. Only those patients who rested quietly during the actual test, without any leg and/or foot movements, were included in this study.

Blood flow in the tibialis anterior muscle of the lower leg was measured by ^{133}Xe clearance as described by Lassen, et al (9). Basically, radioactive xenon gas (^{133}Xe) is dissolved in sterile 0.9% NaCl solution to a concentration of 0.5-1.0 mc/ml, and 0.1 ml of this solution is injected 1-2 cm into the tibialis anterior muscle through a 25 gauge hypodermic needle. The disappearance of ^{133}Xe is monitored for 20-30 minutes postinjection by a collimated scintillation probe placed directly over the injection site. Simultaneous measurements were performed on both legs in four control subjects and five patients. Muscle blood flow (MBF) is calculated from the tangent to the logarithmic curve of ^{133}Xe washout.

$$\text{MBF (ml/100 g muscle}\cdot\text{min)} = \lambda \cdot 100 \cdot \frac{-dQ/dt}{Q(t)}$$

where λ is the partition coefficient for ^{133}Xe (amount of tracer in one gram of tissue/amount of tracer in one ml blood) and $\frac{-dQ/dt}{Q(t)}$

describes the clearance rate of ^{133}Xe relative to the amount present at time t ; 0.70 was the value utilized for λ in both control subjects and patients.

RESULTS

Total leg blood flow was essentially normal in the uninjured legs of burn patients (Table 1), but increased in a curvilinear manner with the size of leg burn and approached a plateau of 8.0 ml/100 ml·min as the extent of leg injury exceeded 60% (Fig. 2). Resting skeletal muscle blood flow, however, was normal in burn patients (Table 2).

DISCUSSION

These leg blood flow measurements(10,11) clearly demonstrated that peripheral blood flow is increased during the hypermetabolic-hyperdynamic phase of thermal injury and strongly suggest that a major portion of this extra blood flow is directed primarily to the burn wound.

Increased wound blood flow was supported by two additional observations. For example, in the third degree wound, which is associated with superficial vascular thrombosis, leg blood flow was near control levels shortly after injury. Neovascularization is a constant feature of wound repair, and the development of granulation tissue in the open wound corresponds structurally with a dense, superficial, vascular bed. With time, therefore, blood flow increased to reach levels predicted from the size of limb burn, presumably associated with the formation of a richly vascularized wound bed. In contrast, partial-thickness injury does not ablate the superficial vascular bed and blood flow was elevated in these limbs as soon as circulatory volume was restored. A second example of increased wound perfusion developed subsequent to leg blood flow measurements before and after excision to fascia of a 82.5% leg surface wound. With the removal of the burn tissue, leg perfusion dropped from 5.27 to 3.33 ml/100 ml·min.

The increased rate of wound perfusion is, in part, responsible for the elevated surface temperatures of these patients. This is most evident in two groups of patients with and without leg burns but with comparable core and total body mean skin temperatures (Table 1). Since these two groups maintained similar levels of body heat content, the higher surface temperatures of the burned legs can only be the result of higher superficial blood flows. Considering the increased evaporative cooling which occurs from the burn wound, the impact of increased superficial blood flow in maintaining an elevated surface temperature is even more impressive.

These peripheral blood flow measurements in burn patients make it possible to characterize the shift in limb perfusion following thermal injury (Fig. 3). Stolwijk (12) has partitioned resting leg blood flow as follows: 29% to muscle, 46% to skin, and 25% to the remainder. Assuming little or no change in either muscle, fat, or bone perfusion, blood flow to the skin of an unburned leg, in a typical patient with a 50% total body surface burn, increases by no more than 7% above normal. If this same burn patient had a wound covering 50% of his leg, limb blood flow would approach 8 ml/100 ml leg volume per minute. Since the size of limb burn has no effect on resting skeletal muscle blood flow, 80% of total leg flow in this patient would be directed to the surface with the vast majority going to the burn wound.

What are the factors responsible for this redistribution of peripheral blood flow following thermal injury? Leg blood flow in uninjured legs of burn patients is normal, in spite of the elevated rectal temperature. A comparable degree of hyperthermia in resting normal man would result in a 4-5 fold increase in leg blood flow(13). Control levels of blood flow to the uninjured limbs suggest that these febrile patients vasoconstrict normal skin. Since such cutaneous vasoconstriction is an appropriate heat conservation response during fever, why does the wound remain dilated? The selected elevation in wound blood flow cannot be explained by a complete loss of intrinsic vascular smooth muscle tone. This was demonstrated by the capacity of severely burned legs with high basal flows to vasodilate further when limb surface temperature was increased (Fig. 4) (10). The burn wound appears to be "functionally" denervated, however, since it fails to vasodilate when its temperature is held constant and the patient's core temperature elevated 0.4-0.5°C by external heating (Table 3)(14). This loss of neurogenic vasomotor control of the burn wound, when normal skin vasodilates appropriately (Fig. 5), is most likely the combined result of 1) actual physical disruption of sympathetic vasomotor nerves at the time of injury, and 2) the presence of local inflammatory and metabolic factors which interfere with neuromuscular transmission in vessels which retain their innervation. Possible chemical vasodilators identified in the burn wound include such inflammatory products as histamine, kinins, and various prostaglandins (15,16,17). The anaerobic nature of granulation tissue is evident by increased rates of lactate production in the injured extremities (6). This metabolic environment with the associated release of lactate in the wound may also contribute to local vasodilation and interfere with extrinsic vasomotor influences. Actual physical denervation of wound microvasculature is most apparent after inflammation subsides and the wound is completely healed. While local chemistry should approach that of normal skin with wound closure, reflex vasomotor control to

the healed wound is still markedly reduced. The most likely explanation for this lag in reinnervation of the burn wound is the evidence that regeneration of sympathetic nerves in granulation tissue is slow and vascular reinnervation often incomplete, particularly if split-thickness skin grafts are utilized (18,19).

Regardless of the precise mechanisms responsible for the loss of neural control of wound vasculature, this functional denervation will allow local environmental factors to exert a greater influence on wound perfusion. Consequently, the control of wound circulation becomes less like that of normal skin and more like that of other critical tissues (heart, brain, and active skeletal muscle), the blood flows of which vary as a function of local metabolic conditions rather than as part of integrated total body thermoregulatory or baroreceptor reflexes.

What is the impact of this high obligatory wound blood flow on the distribution of total body circulation in the burn patient? The estimated cardiac index of a "typical" patient with a 50% total body surface burn wound exceeds $7.0 \text{ L/m}^2 \cdot \text{min}$. Based on actual measurements of splanchnic (7), limb (10), muscle (11) and renal (unpublished data) blood flow, and assuming no change in brain circulation but a rise in coronary perfusion proportional to change in total flow, one may partition the cardiac index of this hyperdynamic patient (Fig. 6). This estimate describes a general shift in the distribution of total body circulation toward the periphery with the major portion of the extra flow going to the wound. This increase in superficial flow is, however, well within maximum levels of cutaneous flow observed in resting-normal man under severe heat stress (20).

The dominance of local wound effects on the distribution of peripheral blood flow in patients with marked systemic alterations in metabolism and thermoregulation demonstrates the priority of the wound in the body's homeostatic response to injury. High obligatory wound blood flow is the major cause for the marked increase in total body circulation following thermal injury. Our calculations indicate that as much as 50-60% of the extra blood flow is directed to the burn wound. This increased superficial blood flow severely reduces the capacity of the patient to conserve body heat and maintain the febrile state. The associated increase in body heat production necessary to compensate for this insulative defect places additional metabolic demands on the cardiovascular system. In other words, the major portion of the increase in total body circulation either goes directly to the burn wound or develops indirectly as a consequence of the effects of

this high obligatory wound blood flow on the thermal balance of the patient.

The combined magnitude and duration of these circulatory adjustments to thermal injury represent a severe test of the body's cardiovascular system. The capacity to generate and sustain the required circulatory activity is a major determinant of the patient's ability to survive the initial injury and ultimately heal the wound.

REFERENCES

1. Wilmore, D. W.: Hormonal responses and their effect on metabolism. *Surg. Clin. N. Amer.* 56:999-1018, 1976.
2. Wilmore, D. W., Long, J. M., Mason, A. D., Jr., et al: Catecholamines: mediator of the hypermetabolic response to thermal injury. *Ann. Surg.* 4:653-669, 1974.
3. Wilmore, D. W., Mason, A. D., Jr., Johnson, D. W., and Pruitt, B. A., Jr.: Effect of ambient temperature on heat production and heat loss in burn patients. *J. Appl. Physiol.* 38:593-597, 1975.
4. Wilmore, D. W., Orcutt, T. W., Mason, A. D., Jr., and Pruitt, B. A., Jr.: Alterations in hypothalamic function following thermal injury. *J. Trauma* 15:697-703, 1975.
5. Unger, A., and Haynes, B. W., Jr.: Hemodynamic studies in severely burned patients. *Surg. Forum* 10:356-361, 1959.
6. Wilmore, D. W., Aulick, L. H., Mason, A. D., Jr., and Pruitt, B. A., Jr.: Influence of the burn wound on local and systemic responses to injury. *Ann. Surg.* 186(4):444-458, 1977.
7. Gump, F. E., Price, J. B., and Kinney, J. M.: Blood flow and oxygen consumption in patients with severe burns. *Surg. Gynec. Obstet.* 130:23-28, 1970.
8. Landowne, M., and Katz, L. N.: A critique of the plethysmograph method of measuring blood flow in the extremities of man. *Am. Heart J.* 23:644-675, 1942.
9. Lassen, N. A., and Lindberjerg, M. O.: Measurement of blood flow through skeletal muscle by intramuscular injection of Xenon-133. *Lancet* 1:686-689, 1964.

10. Aulick, L. H., Wilmore, D. W., Mason, A. D., Jr., and Pruitt, B. A., Jr.: Influence of the burn wound on peripheral circulation in thermally injured patients. *Am. J. Physiol.* 233(4):H520-H526, 1977.
11. Aulick, L. H., Wilmore, D. W., Mason, A. D., Jr., and Pruitt, B. A., Jr.: Muscle blood flow following thermal injury. Annual Research Progress Report, US Army Institute of Surgical Research, Brooke Army Medical Center, Fort Sam Houston, Texas, 1 July 1976 - 30 September 1977. Presented at the American Physiological Society Fall Meeting, Hollywood, Florida, October 1977.
12. Stolwizk, J. A. J.: Mathematical model of thermoregulation. In Physiological and Behavioral Temperature Regulation, Springfield, Charles C. Thomas, 1970, pp. 703-721.
13. Wyss, C. R., Brengelman, G. L., Johnson, J. M., Rowell, L. B., and Niederberger, M.: Control of skin blood flow, sweating and heart rate: role of skin vs. core temperature. *J. Appl. Physiol.* 36:726, 1974.
14. Aulick, L. H., Wilmore, D. W., Mason, A. D., Jr., and Pruitt, B. A., Jr.: Depressed reflex vasomotor control of the burn wound. Annual Research Progress Report, US Army Institute of Surgical Research, Brooke Army Medical Center, Fort Sam Houston, Texas, 1 July 1976 - 30 September 1977.
15. Arturson, G.: Prostaglandins in human burn wound secretions. *Burns* 3:112, 1978.
16. Anggard, E., and Johnson, D. -E.: Efflux of prostaglandin in lymph from scalded tissue. *Acta Physiol. Scand.* 81:440, 1971.
17. Edery, H., and Lewis, G. P.: Kinin-forming activity and histamine in lymph after tissue injury. *J. Physiol. (London)* 169: 568, 1963.
18. Clark, E. R., Clark, E. L., and Williams, R. G.: Microscopic observations in the living rabbit of the new growth of nerves and establishment of nerve-controlled contractions of newly formed arterioles. *Am. J. Anat.* 55:47, 1934.
19. Fitzgerald, M. J. T., Martin, F., and Paletta, F. X.: Innervation of skin grafts. *Surg. Gynec. Obstet.* 124:808, 1967.
20. Rowell, L. B.: Human cardiovascular adjustments to exercise and thermal stress. *Physiol. Rev.* 54:75-159, 1974.

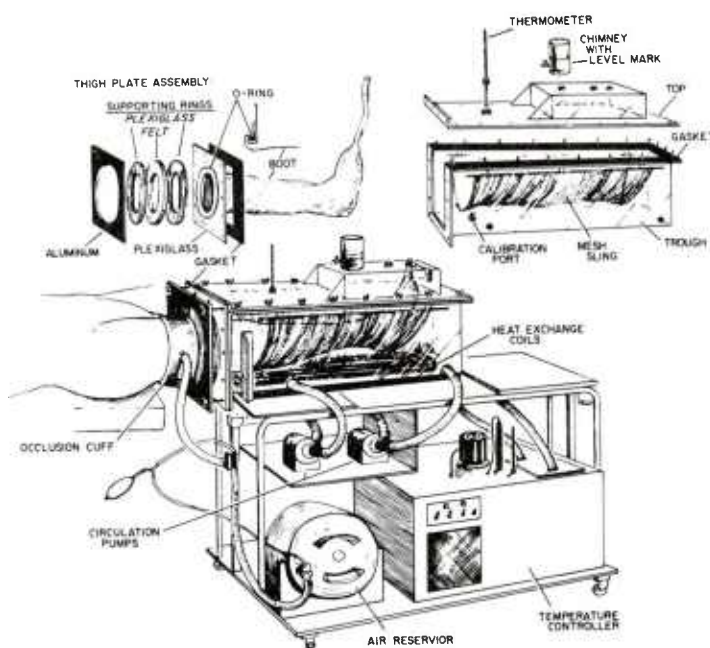


Figure 1. Leg plethysmograph. This design provides a safe, comfortable, noninvasive approach to the study of peripheral blood flow in critically injured patients.

Temperature ($^{\circ}\text{C}$)	Controls	No Leg Burn	Leg Burn
Rectal	37.0 ± 0.1	37.9 ± 0.2	38.0 ± 0.1
Mean Skin	34.5 ± 0.1	35.4 ± 0.3	35.8 ± 0.2
Mean Leg Skin	33.9 ± 0.2	34.6 ± 0.3	35.4 ± 0.2
Leg Blood Flow (ml/100 ml \cdot min)	2.74 ± 0.16	3.13 ± 0.21	5.89 ± 0.68

Table 1. Local effects of leg burn wound on limb blood flow and surface temperatures.

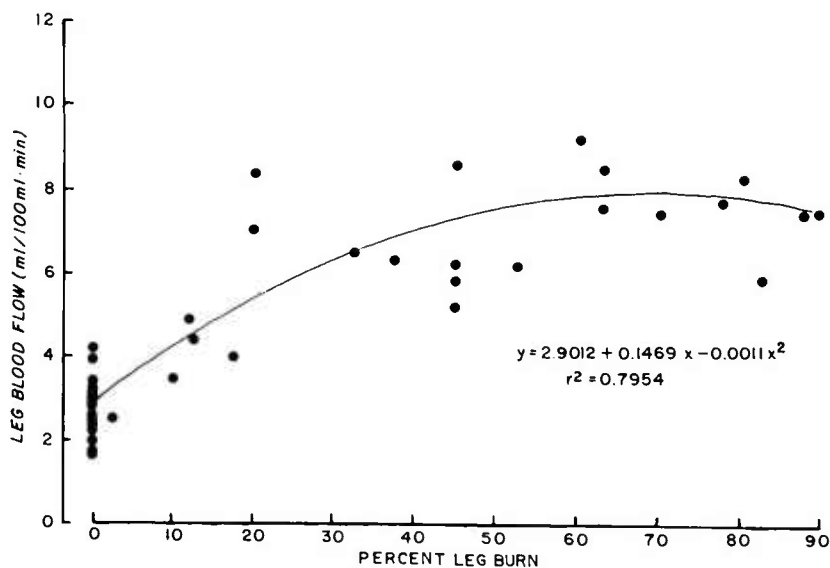


Fig. 2. The effect of local injury on leg blood flow. Leg blood flow increases with the extent of limb burn, suggesting that the extra blood flow is directed to the surface wound.

CONTROLS	3.29 ± 0.24
SMALL LEG BURNS	3.47 ± 0.50
LARGE LEG BURNS	3.57 ± 0.28

Table 2. Resting skeletal muscle blood flow (ml/100 g·min) following thermal injury. The local presence of a burn wound has no effect on muscle perfusion. Values represent group means ± S.E.M.

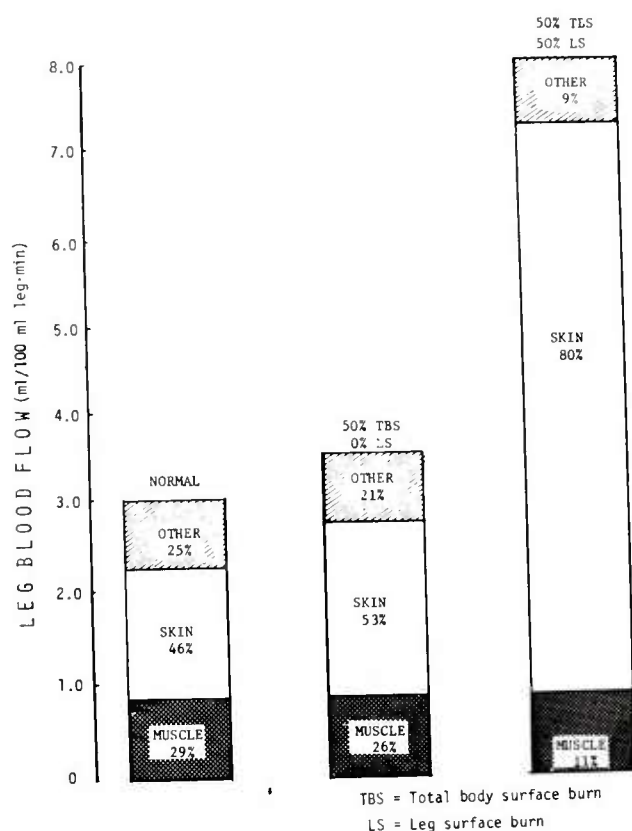


Figure 3. The estimated distribution of peripheral blood flow following a 50% total body surface (%TBS) burn. Leg blood flow is essentially normal in the uninjured limbs (0% LS - % leg surface burn) but increases markedly when 50% of the limb surface is burned (50% LS). The vast majority of this increase in peripheral blood flow is directed to the wound.

SUBJECTS	ΔT_{re}	LEG BLOOD FLOW (ml/100 ml·min)		
		BEFORE	AFTER	% INCREASE
CONTROLS	+ 0.4°C	2.93	4.56	56
SMALL LEG BURNS (0-2.5%)	+ 0.6°C	3.86	6.34	64
LARGE LEG BURNS (37.5-70%)	+ 0.5°C	7.17	7.73	8

Table 3. Changes in leg blood flow following 30 minutes of external heating. Leg surface temperature is held constant.

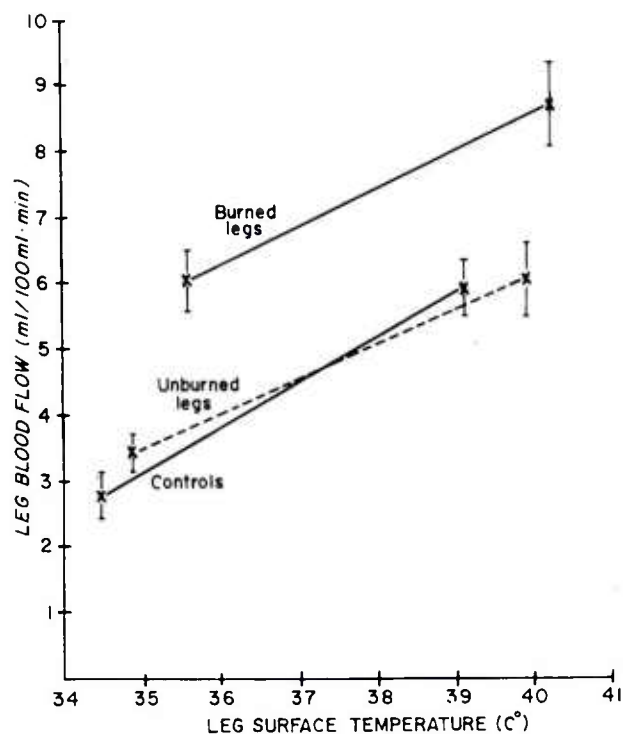


Fig. 4. Peripheral vascular response to local heating. Leg surface temperature was increased 5°C for 30 minutes. Burn wound vasculature has some basal tone which responds appropriately to changes in local temperature.

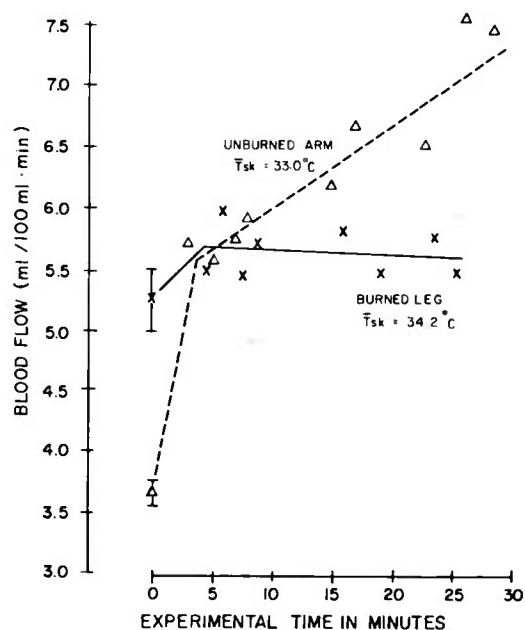


Fig. 5. Peripheral vascular response to indirect heating. Limb temperatures held constant and core temperature increased 0.5°C by 30 minutes of radiant heat. Reflex cutaneous vasodilation only evident in the uninjured limb.

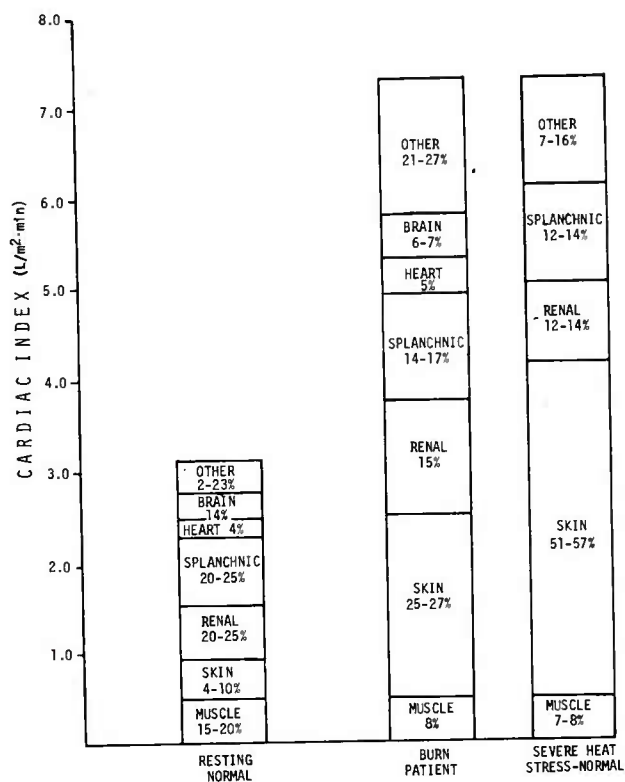


Fig. 6. Estimated changes in distribution of the cardiac index following a 50% total body surface burn. The major increase in skin blood flow is directed to the burn wound. The greater skin blood flow in heat stressed normals occurs from a fully dilated cutaneous vascular bed while unburned skin in the patient is vasoconstricted.

THE SURVIVAL OF ST. LOUIS ENCEPHALITIS
VIRUS IN OVERWINTERING MOSQUITOES

*CHARLES L. BAILEY, MAJ, MSC
BRUCE F. ELDRIDGE, COL, MSC
DAVID E. HAYES, MR.
WALTER REED ARMY INSTITUTE OF RESEARCH
WASHINGTON, D. C. 20012

INTRODUCTION

St. Louis encephalitis (SLE) virus leads all other arbo-viruses in causing human disease in the United States, and numerous epidemics have occurred since its initial detection in 1933. The largest being in 1975 when almost 2,000 cases were reported. The virus is maintained in nature during the summer and fall by a mosquito-bird-mosquito cycle. At least 3 mosquito species, Culex pipiens, C. nigripalpus and C. tarsalis have been incriminated as vectors. The seasonal distribution of human disease coincides with high infection rates in mosquitoes and birds during late summer and early fall, and human infections tend to be more prevalent in urban areas--especially in the eastern US where the peridomestic mosquito C. pipiens is the principal vector.

Although details of the summer and fall virus transmission cycle are relatively well known, the mechanism by which the virus persists during the interenzootic winter season has remained a mystery. In temperate regions, C. pipiens hibernate as inseminated adult females. This fact, along with the fact that this mosquito is one of the primary vectors of SLE virus during the summer and fall prompted us to test the hypothesis that SLE virus is maintained in these overwintering mosquitoes. This hypothesis has not gained wide acceptance in the past because of the belief that blood-feeding in pre-hibernating Culex mosquitoes is drastically reduced or suspended

so that their chance of taking a viremic blood meal before overwintering is exceedingly small or nonexistent (1). The usual sequence of events which follow blood-feeding in C. pipiens is maturation and oviposition of eggs. Techniques have been developed (2) whereby it can be determined if a mosquito has laid eggs (parous) or not (nulliparous). Previous investigations (3) have shown that parous females are seldom found among collections of overwintering mosquitoes, and those which are found are usually observed in early winter and are presumed not to survive until spring. However, ovarian diapause (a condition of the ovaries resulting in the taking of a blood meal without development of eggs) in laboratory reared C. pipiens has been demonstrated (4). In these cases, females, experimentally induced to hibernate, took full blood meals, but in most instances, ovarian development did not follow. If this situation occurs in nature, such females, upon dissection, would show no evidence of a previous gonotrophic cycle (i.e., blood meal). Nulliparity, then, would not be proof of the absence of a pre-hibernating blood meal. This study was designed to first look for the presence of SLE virus in hibernating mosquitoes and to conduct laboratory tests to explain how these mosquitoes may have taken a pre-hibernation viremic blood meal and yet remained nulliparous.

MATERIALS AND METHODS

Virus Isolation From Hibernating Mosquitoes

The source of the mosquitoes tested for presence of virus, was abandoned ammunition bunkers of several former U.S. Army forts. Collections were made in 4 mid-Atlantic states during January and February of 1976 and 1977 (Fig. 1). The bunkers consisted of a series of rooms 3 to 5 m wide, 5 to 10 m long and 2 to 3 m high. Construction material was either steel-reinforced concrete or clay bricks. The rooms where mosquitoes were collected had condensation on both walls and ceilings.

During the 1976 studies 1,116 mosquitoes were aspirated from the damp walls and ceilings of the bunkers and transported to the laboratory where they were retained for various lengths of time. Following one day in an insectary programed for a daily photoperiod of 16 hrs daylight, 8 hrs darkness and a constant temperature and relative humidity of 26°C and 80%, respectively, 715 female mosquitoes were identified and stored at -70°C in pools of 10 mosquitoes each for virus assay. The remaining 401 mosquitoes were held for 7 days in the insectary and then pooled. No SLE isolates were obtained

from the 1976 material.

Our failure to isolate SLE virus from overwintering Culex mosquitoes collected during 1976 prompted changes in the way we handled mosquitoes for the 1977 studies (Table 1). Previous experiments with Japanese encephalitis virus (5) have shown a marked reduction in virus multiplication in mosquitoes maintained at low temperature. These data suggested that any virus present would probably be in low concentration, possibly below detectable levels, and that incubation at high temperature might be necessary for detection of virus in mosquitoes from overwintering sites. We similarly speculated that a post-hibernation blood meal might enhance the chance of recovering virus. After the holding periods in the insectary, a chicken* was placed in the mosquito cage overnight to serve as a blood meal source. The blood-engorged mosquitoes were segregated from unfed mosquitoes and maintained in separate cages until blood digestion was complete, then assayed for virus. Each mosquito pool was triturated in tissue grinders with 2 ml medium 199-Hanks salts supplemented with 20% heated fetal bovine serum, NaHCO_3 , penicillin (500 units/ml) and streptomycin (500 $\mu\text{g}/\text{ml}$). After centrifugation for 30 min. at 475 x g in a refrigerated centrifuge, each supernatant was inoculated intracerebrally into a litter of 3-5 day old mice (0.02 ml/mouse). The remainder of each mosquito suspension was stored at -70°C to be used as necessary for virus reisolation attempts. Moribund or dead mice observed during a 14-day period were frozen at -70°C . All such suspect isolates were passaged by injecting a 20% brain suspension, in the medium described, into a second litter of mice. Results of the pre-bleedings of the chickens used to furnish blood meals for the mosquitoes eliminated any possibility that these chickens served as a virus source. SLE isolates were identified using a combination of complement-fixation (CF) and plaque reduction neutralization test (PRNT). Prototype strains used as reference SLE virus, included the Parton strain, and an isolate made from a 12 September 1975 light trap collection of Culex sp. during the SLE outbreak in Prince Georges County, MD.

*In conducting the research described in this report, the investigators adhered to the "Guide for the Care and Use of Laboratory Animals," as promulgated by the Institute of Laboratory Animal Resources, National Academy of Sciences, National Research Council.

Blood-feeding In Pre-hibernating Mosquitoes

Two characteristics of pre-hibernating adult C. pipiens females are a reduction in the blood-feeding drive and the failure of ovarian follicles to mature following a full blood meal. This latter characteristic has great importance to our understanding of the role of overwintering Culex mosquitoes in the winter maintenance of certain arboviruses. Evidence is presented which offers an explanation of how pre-hibernating female C. pipiens mosquitoes may take a viremic blood meal, yet not undergo ovarian development.

The C. pipiens mosquitoes used in these experiments were from a selfmating colony established in 1975 from hibernating females collected from Ft. Mott, Salem Co., New Jersey. Larvae for experimental material were reared in our insectary using conventional techniques, and at a temperature of 27°C and a daily photoperiod of L:D 16:8 (hr light: hr dark per 24 hr period). Since it has been established that the pupal stage is the photoperiod sensitive stage in C. pipiens, random distribution into experimental groups was made at the time of pupation. The treatment scheme used in this study is diagrammed in Fig. 2.

RESULTS AND DISCUSSION

SLE Virus Isolation

A summary of the mosquitoes collected, the number which took a blood meal and the viruses isolated during the winter of 1977 is presented in Table 1. Two SLE virus isolations were obtained from pools of 10 mosquitoes each that were collected (after the coldest periods of the winter) on 26 January and on 22 February, at Ft. Washington, MD and Ft. Mifflin, PA, respectively. Both virus isolates were obtained from mosquitoes given a post-hibernation avian blood meal. Although by no means conclusive evidence that virus isolation was contingent upon the blood meal, we did have 112 non-bloodfed mosquito pools in 1976 and 53 in 1977 in which no SLE viruses were isolated. This is compared to 2 isolates from a total of 62 bloodfed mosquito pools in 1977.

Demonstration of Increased Pre-hibernation Blood-feeding and Decreased
Ovarian Diapause

The experiment reported here was designed to represent a model of fall populations of pre-hibernating female C. pipiens subjected to varying periods of warm temperatures ("Indian" summers). Our objective was to test the hypothesis that fall fluctuations in temperature would result in some females taking blood meals without the reversal of ovarian diapause. Such females would develop fat bodies, remain nulliparous, and be prepared for hibernation. The results of the blood-feeding trials and subsequent determination of ovarian development are shown in Table 2. At the time of the first blood-feeding trial only 5% of the short photophase females took a blood meal compared with 78% of the long photophase group. Of those taking full blood meals, 80% of the short photophase group had ovaries which failed to mature, compared with 0% of the long photophase group. After 3 days of warming, the percentage of females taking blood had risen to 45%, but 40% of these had undeveloped ovaries. Even after 4 days, when blood-feeding took place in 74% of the females, lack of ovarian development occurred in 25% of those taking blood. After 5 days there was no differences between the two groups. We feel that the results of these experiments provide a biologically plausible model for the natural occurrence of blood-feeding in pre-hibernating C. pipiens.

SUMMARY

Two strains of SLE virus were isolated from natural populations of hibernating C. pipiens mosquitoes collected after the coldest part of the winter in Maryland and Pennsylvania. These data suggest that these mosquitoes can be considered to be winter maintenance hosts for SLE virus. There are two possible explanations for the presence of virus in these mosquitoes. One is that they became infected through transovarial transmission, the other is that these females took a pre-hibernation viremic blood meal. We have provided an experimental basis for the latter. A significant number of females subjected in the laboratory to environmental conditions (e.g., short photophases and cool temperatures) known to produce physiological changes related to hibernation and then warmed to 25°C for various periods of time showed an increase in blood-feeding drive but still displayed ovarian diapause. Such warming periods may be analogous to situations which are common in nature ("Indian" summers). We suggest that these experimental conditions may represent the mechanism by

which SLE virus overwinters in hibernating C. pipiens. This suggests a possible strategy for the control of potentially infected vectors in the vicinity of U.S. military installations and other concentrations of human activity. Spraying of abandoned ammunition bunkers, concrete drainage tunnels, and other similar manmade structures which serve as overwintering sites for C. pipiens could attack both vector and virus at the weakest point of their life cycles.

REFERENCES

1. Bellamy, R. E., Reeves, W. C. Scrivani, R. P. 1958. Relationships of Mosquito Vectors to Winter Survival of Encephalitis Viruses II. Under Experimental Conditions. Am. J. Hyg., Vol. 67: 90-100.
2. Detinova, T. S. 1962. Age-grouping Methods in Diptera of Medical Importance with special reference to some vectors of malaria. WHO Monogr. Ser., Vol. 47: 1-216.
3. Slaff, M. E. and Crans, W. J. 1977. Parous rates of overwintering Culex pipiens pipiens in New Jersey. Mosq. News, Vol. 37: 11-14.
4. Eldridge, B. F., 1966. Environmental control of ovarian development in mosquitoes of the Culex pipiens complex. Science, Vol. 151: 826-828.
5. LaMotte, L. C. Jr., 1963. Effect of low environmental temperature upon Japanese B encephalitis virus multiplication in the mosquito. Mosq. News, Vol. 23: 330-335.

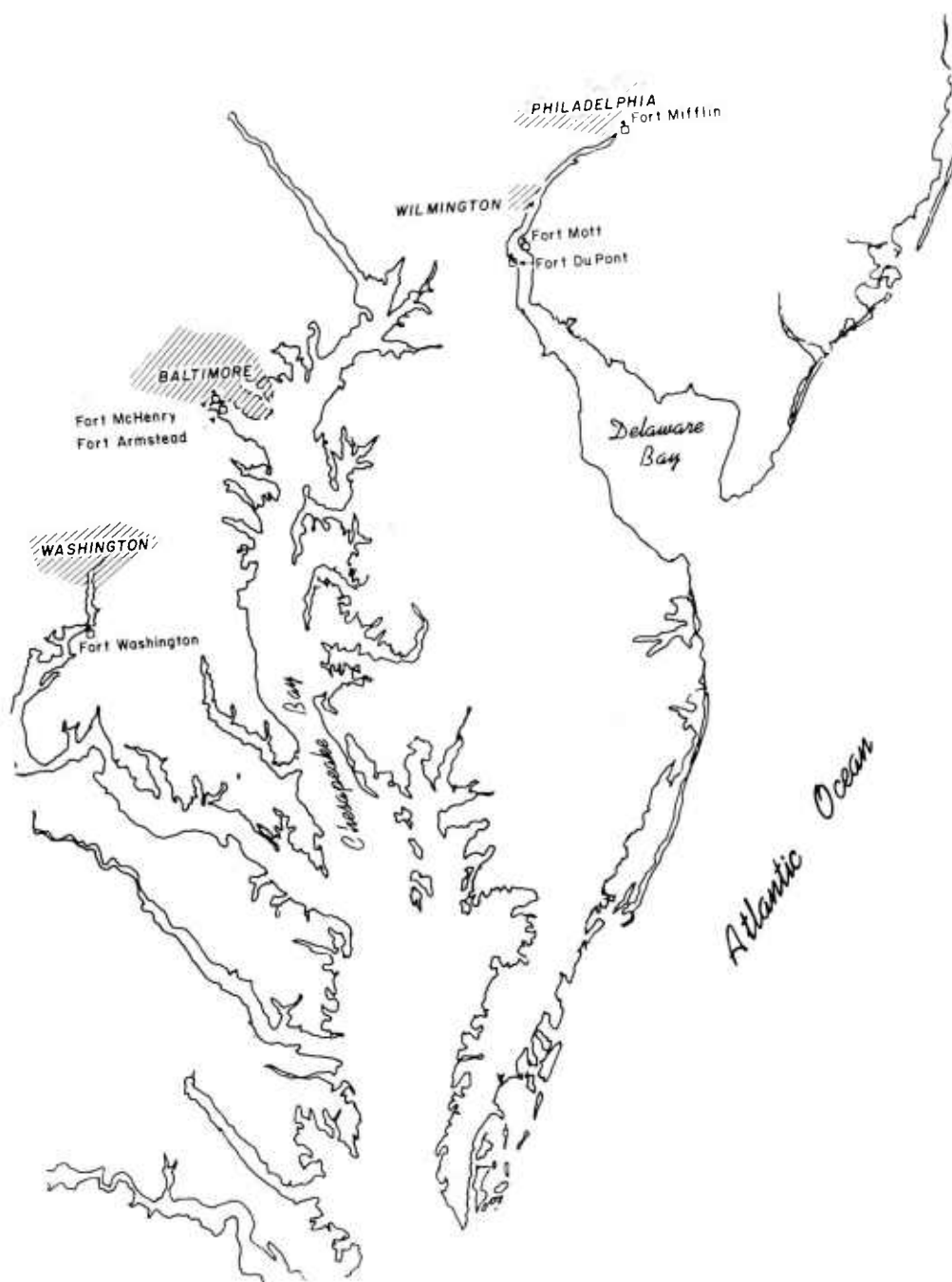


Fig. 1. Location of former U.S. Army forts where overwintering mosquitoes were collected.

PHOTOPERIOD AND TEMPERATURE REGIME

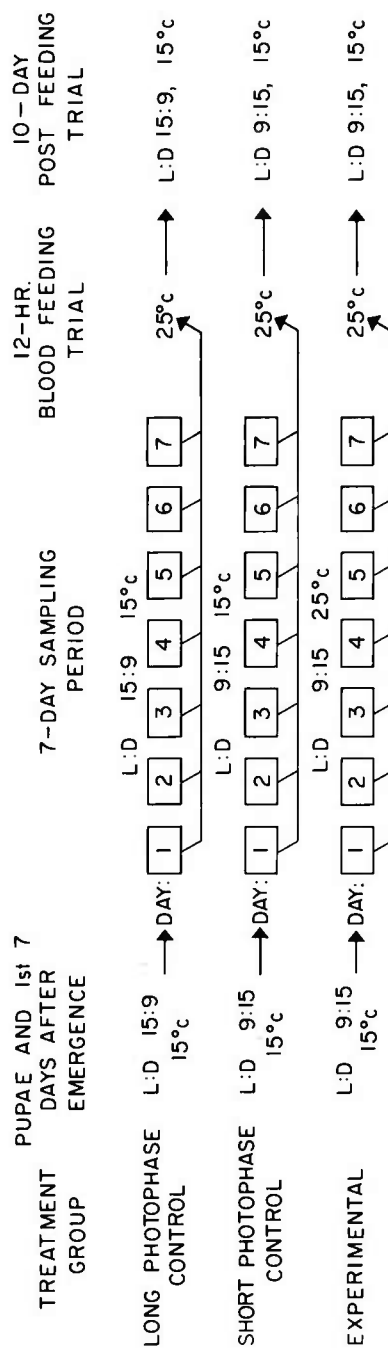


Fig. 2. Diagram of experimental treatment schedule.

Table 1. Hibernating *Culex pipiens* collected at abandoned Army Forts located in Pennsylvania, New Jersey and Maryland during the winter of 1977

Location	Mosquito Collection			Laboratory Handling					
	Date	Number Collected	Days held in insectary Before Blood-feeding	Blood Fed		Non Blood Fed		No. Isolations	No. Isolations
				No. Pools Tested	No. Isolations	No. Pools Tested	No. Isolations		
Ft Washington, MD	4 Jan	97	21	3	0	7	0	0	0
Ft McHenry, MD	5 Jan	9	7	0	0	1	0	0	0
Ft McHenry, MD	11 Jan	3	0	0	0	1	0	0	0
Ft Armistead, MD	12 Jan	33	20	2	0	1	0	0	0
Ft Washington, MD	26 Jan	215	20	14	1	7	0	0	0
Ft Mifflin, PA	22 Feb	406	15	25	1	15	0	0	0
Ft Mott, NJ	23 Feb	226	14	10	0	12	0	0	0
Ft Mott, NJ	1 Mar	173	20	8	0	9	0	0	0
Totals		1162		62	2	53		0	0

Table 2. Bloodfeeding and ovarian development in female *Culex pipiens* exposed to diapause-inducing conditions (L:D 9:15, 15°C) from pupal stage to 7th day of adult life, then warmed to 25°C for from 1 to 7 days.

Day of Adult Life	Experimental Group L:D 9:15, 15°C 25°C			Short Photophase Control L:D 9:15, 15°C			Long Photophase Control L:D 15:9, 15°C		
	Percent took full blood meal ¹	Percent with undeveloped ovaries ²		Percent took full blood meal ¹	Percent with undeveloped ovaries ²		Percent took full blood meal ¹	Percent with undeveloped ovaries ²	
7	5	80.0		7	71.4		78	0.0	
8	18	61.1		6	83.3		70	0.0	
9	21	52.4		5	60.0		79	0.0	
10	45	40.0		8	75.0		73	0.0	
11	74	25.7		6	50.0		82	0.0	
12	61	1.6		9	55.5		77	0.0	
13	42	0.0		6	66.6		76	0.0	
14	71	0.0		5	80.0		-	-	

¹Each group comprised of 100 females

²Of those taking full blood meal

FIELD USE OF THE
ENVIRONMENTAL IMPACT COMPUTER SYSTEM

HAROLD E. BALBACH, Ph.D.*
EDWARD W. NOVAK, Ph.D.
CONSTRUCTION ENGINEERING RESEARCH LABORATORY
CHAMPAIGN, IL 61820

INTRODUCTION AND BACKGROUND

Planners of Army activities such as military construction, operation of installations, and conduct of training exercises have considered the environment when making decisions for decades. These considerations usually took the form of adapting the action to the demands of the local setting. Only rarely has the primary goal of an Army military activity been to alter the environment radically. The requirements of the National Environmental Policy Act of 1969 (NEPA) have refocused this environmental consideration. The perspective of environmental "consideration" has also changed. In response to AR 200-1, which adopts these requirements to the Army, a much more rigid, systematic consideration of the environment must be performed.

The requirement to prepare an Environmental Impact Statement (EIS) which discussed those actions which were significant enough to potentially alter the environment required an openness not common among government agencies. Further, the original legislation did not specify exactly what was to be examined, nor did it discuss the form that the EIS should take. Subsequent Executive Orders and the President's Council on Environmental Quality guidelines for the preparation of such documents helped answer some of these questions. The form of the EIS and general scope of coverage was now established. This did not, however, tell government agencies, including the Army, HOW to examine the environment. Neither did any of these supplementary regulations really indicate WHO should carry out the study

leading to an EIS, although it was declared that the approach should be "interdisciplinary" (1). An interdisciplinary approach requires that persons knowledgeable of the biophysical and socioeconomic environment being affected and the project being assessed be used in the assessment process.

At this time, few government agencies or departments had the environmental complement of such an interdisciplinary group available. One common source of such expertise was the engineering consulting firm. Planners and decision makers were used to dealing with these consultants, and many contracts for EIS preparation were, and are, placed with such firms. Such private businesses were probably no more knowledgeable than government personnel with respect to the specific environment in question, but they had the capability of adding temporary personnel in almost any specialty in immediate response to a particular need. This mode of response was not generally available to government agencies, and is still not usually permitted.

How did these firms attack the problem of the content of an EIS? They did it in much the same way it was done within agencies--an ad-hoc or "off the top of the head" procedure. Lacking guidance as to content or depth of coverage, the EIS became, over the course of the first few years of NEPA, a weighty tome, or even multiple volumes. These hundreds and thousands of pages were expensive to produce, and took months or years to complete, and still did not address the primary points of an EIS adequately. Fifty percent or better of the information included in such documents contained useless information. So many agencies had important projects in abeyance, though, that funds were made available for this work. Such massive documents became the norm for an EIS in many areas. Decision makers neither had the time nor the interest to wade through hundreds of pages to locate the precious little information useful in making a decision. But was it possible for an agency to produce a legally and morally sufficient EIS internally? Could much of this cost be eliminated by using agency personnel already familiar with the action?

DEVELOPMENT OF THE ENVIRONMENTAL IMPACT COMPUTER SYSTEMS

Rationale

In response to this type of question, the Construction Engineering Research Laboratory of the Corps of Engineers (CERL) began work, in late 1971, to develop procedures whereby Army personnel could prepare and review EIS's themselves, with a minimum of outside assistance. By this time the assessment called for by NEPA to determine if the action was potentially "significant" had become, in the Army as in many other agencies, a formal document in its own right. The Environmental Impact Assessment (EIA) was first suggested, then required, to follow the same format as the FIS. CERL's charge, then, included EIA's as well as EIS's, thus covering an extremely wide range of scope and significance required by the procedures eventually developed.

Within the Army, major commands place slightly varying burdens on field personnel to prepare formal, written assessments in the format required by AR 200-1. Whether written in this format or not, though, the considerations are supposed to be similar. It was especially strongly believed that these more preliminary examinations of environmental impact should be able to be done by Army personnel with a minimum input from outside specialists. Since written EIA's are ten times more frequent in the Army than are EIS's, savings in personnel time and costs are potentially even greater in the EIA than in the EIS process.

Several shortcomings were identified in the EIA/EIS procedures which had evolved without real direction. The direct dollar costs were felt to be excessive. Capital outlay by the Army for approximately 690 written EIA's and EIS's per year was conservatively estimated to average 12.4 million dollars per year. Time delays of 12 to 24 months were seen in many cases. Part of this time was needed for the contractor to undertake the detailed studies deemed necessary for the preparation of a complete document. Depending on the stage of completion of the project, however, cost escalations due to inflation were often several times the direct cost of the EIS preparation. For all these reasons, among others, the existing procedures, or lack of them, were identified as excessively costly.

Other problems concerned the coverage of the document, itself. The challenges to the accuracy of an EIS were always pointing out some area of consideration that had not been included. Each agency of the government seemed to be able to identify, in the EIS's of other agencies, serious shortcomings which affected them. The procedures developed as a result of CERL's research would have to be comprehensive enough to include all reasonable areas of impact. Further, this comprehensive product would have to be capable of being used by Army military and civilian personnel of widely varying backgrounds and educational attainment.

Approach

A basic premise was made by CERL researchers at the very beginning. This assumed that it was feasible to divide the actions of the Army into discrete activities, and that these activities could then be reaggregated as necessary to describe a particular action. A second major assumption was made that the "human environment" could be similarly divided into its characteristic parts, or attributes. Thus, if both premises were accepted, environmental impact could be examined as if it were a massive matrix, with activities along one axis, attributes along the other, and the impact of their coincidence at their intersection. In basic concept, there is a parallel to the matrix developed several years earlier by the Geological Survey (2), though the Army "matrix" was several times larger initially.

As developed, Army activities were grouped into 9 broad areas. Construction, Training, Operation and Maintenance (of installations), Research, Development, Test & Evaluation, and Mission Change are examples of such areas. In all, approximately 900 different Army activities were developed. Examples of construction activities are shown in Table 1. The list is open-ended, and may be expanded if necessary to adequately describe an emerging area of military activity. The context of the continental United States was generally assumed when developing these activities, but they will apply well to almost any peacetime sphere of action.

Environmental attributes were also developed within a general context of the continental United States. The biophysical and socioeconomic environment was categorized into 700 attributes. These were placed in 13 areas, such as ecology (biology), water quality, air quality, economics, earth science, and noise. Examples of ecology

60 support operations	61 temporary roads
64 equipment fueling/maintenance	65 temporary food services
66 solid waste disposal	67 liquid waste disposal
71 exploration of site	73 clearing site
74 grubbing site	75 stumping site
76 draining site	80 demolition
90 removal and disposal	91 brush removal/disposal
92 tree removal/disposal	93 lumber removal/disposal
94 concrete removal/disposal	100 excavation
105 channeling	106 dredging
120 earthworks and burrowing	130 rock excavation and quarrying
166 caissons	167 cofferdams
168 dewatering	169 drainage
180 bituminous construction	185 paving-bitum
187 curing/sealing-bitum	262 pest/insect protection
285 furberishing(landscape-planting-seeding)	

Table 1. Activities, as included in EICS, which are involved in a small military construction project. Activities and numbers are the same as those on the matrices in Figure 1.

1 natural setting	2 game animals
3 game fish	4 rare or endangered species
5 increase in undesirable species	10 impacts on game animals
11 encroachment on natural habitats	12 threatened species

Table 2. Attributes included in EICS for Ecology for use in considering potential impacts. Attributes and numbers are the same as those on the matrices in Figure 1.

attributes are shown in Table 2. The "matrix" thus developed must consider at least 630,000 potential impacts generated by 900 activities on 700 environmental attributes. Since it was desirable to be able to consider each of these relationships in light of individual project specifications, geographic location, season of the year, type of units involved, etc., the number of variations present brings the total number of impacts which must be examined to several million.

Since no personnel could be asked to examine these several million relationships individually, the Environmental Impact Computer System (EICS) was developed (3). This system selects, from among the impacts covered, those applicable to the project, location, and magnitude of the action under consideration. The results are given to the user in the form of a set of matrices, and are scored by the "need to consider" that particular relationship (Fig 1). The severity of impact is not generally predictable through the basic EICS, however, but must be estimated by the preparer of the EIA/EIS. Approximation of the magnitude of an impact requires specific background information and mathematical models not generally available either to the Army or to other document preparers.

One quantitative model, the Economic Impact Forecast System (EIFS), has been prepared by CERL for Army use. This was possible only because nationwide data sources existed which could be used for the background. EIFS has proven especially useful because economic impact is a consideration in most actions, and is an area where quantification is traditionally expected.

In addition to the indication of the need to consider an area, supplemental parts of EICS give other information and guidance to the user. Each of the attributes is completely defined in a separate handbook (4) where examples of types of impacts and interactions with other attributes are also covered. An explanation of each type of impact is also retrieved, as are parallel suggestions for the avoidance or minimization of impact. If impacts are considered unavoidable in the practical sense, this judgment is given for use in the EIA/EIS. Current modules of EICS include, as well, reference to positive effects, especially those which are not intended or not obvious to the planner or engineer.

	ECOLOGY ATTRIBUTES									
	1 2 3 4 5 6 7 8 9 10									
	1	2	3	4	5	6	7	8	9	10
CONSTRUCTION ACTIVITIES	60								A	B
	61	A	A	A				A	B	B
	64			A						
	65					A				
	66					A				
	67			B		A				
	71	C	C				C	C		
	73	A	A		C	B	B	A	A	
	74			A		A				
	75			A		A				
	76		A	B		C	A		A	
	80	B	B				B	E		
	90	A		B						
	91					C				
	92					C				
	93					C				
	94	B		B				B		
	100	A	B	A		A	B	A	B	
	105					A				
	106					A				
	120	A	B	A		A	B	A	B	
	130			C						
	166			A		A				
	167			A		A				
	168			B		B	C			
	169					B				
	180			B						
	185	C			C			C	C	
	187			B						
	262		B	A					B	
	285	C	B	B		B				B

	ECOLOGY ATTRIBUTES					
	1 4 5 1 2					
	1	4	5	1	2	
CONSTRUCTION ACTIVITIES	60				B	C
	61	B			C	C
	65			A		
	66			A		
	67			A		
	71	C			C	
	73	B	C	B	A	H
	74			A		
	76			C		H
	80	B			B	
	90	B				
	91			C		
	92			C		
	94	B			B	
	100	B		A	B	C
	120	B		A	B	C
	168			B		
	169			B		
	285	C		B		C

Figure 1. EICS impact matrices for a small construction project. Numbering of activities and attributes corresponds with numbers in Table 1 and Table 2. Score of "A" represents greatest need to consider, "B" and "C" show lesser need. Figure 1A includes all potential impacts on the attributes. Figure 1B illustrates the reduction in potential impacts when filter questions have been answered to reduce extraneous impact presentation.

FIELD APPLICATION OF EICS

For which sorts of actions was the EICS intended to be applied? The comprehensiveness of the system was designed with the larger project in mind. Especially logical is its use on actions suspected or known in advance of assessment to be wide-ranging or environmentally controversial. In practice, a decision is often made in advance of formal assessment that a full-fledged EIS will be needed to answer questions that have already arisen. Particularly controversial, in the military activity context, is the ongoing re-alignment of installations, functions, and personnel. EICS has been utilized for analysis of these problems, and has generally been of considerable utility in pointing out considerations in areas not within the personal expertise of the team conducting the initial study.

Application has also been made to major construction projects, including troop housing, family housing, and test site proposals. Other applications have been to installation ongoing operations assessments, training program reorganizations, and a variety of smaller construction projects. Applicability to construction projects is even more logical when the planner is choosing among several possible sites for a project. EICS assists in pointing out potential problems associated with each site. The planner may then assess the sites, and rank them in order of acceptability. Or, EICS may indicate, after examination of the output, that several sites are equally acceptable from an environmental point of view. This relieves the planner of some concerns at this stage, or reduces the uncertainty factor greatly.

User Procedure

What procedure does the planner, or other preparer of an EIA or EIS follow to use the EICS? Since the system has been released for interim use while research is still taking place to develop associated modules of the systems, CERL processes requests for output at this time. The EICS user manuals for construction (5) and for mission change, operations & maintenance and training (6) contain detailed instructions to the user. Questions are asked in the manuals concerning aspects of the proposed action, its location, and the surrounding environment. These questions are designed so that they

may normally be answered by non-specialists following a visit to the site, examination of topographic maps of the vicinity, and brief consultation with installation personnel possessing specialized input. In any case, if a question should be unanswerable, input may be made without that answer. Input forms, with the question answers, are then mailed to CERL according to instructions in the manuals.

What is the purpose of these questions? They represent a way to save effort for the assessment preparer through reducing the number of potential considerations presented by EICS. Earlier, we pointed out that EICS was capable of relating Army activities to the environment in several million ways. No person can be expected to examine them all. The computer, however, does so very rapidly, and relatively inexpensively. Information supplied by the system user on the input form allows the computer to select from among these relationships the tiny fraction most applicable to the action under consideration. Thus, this user may be shown the few hundred--or, in a larger action, few thousand--potentially important impacts. Figure 1 represents actual EICS output showing the difference between an unfiltered and a filtered matrix.

What must the user do with those potential impacts? Ideally, they will serve to plan and focus the examination which must take place before a document is prepared. EICS output does not do the writing of the EIA/EIS. A person, or group, must still do so. They must determine if the activity predicted to cause a problem will take place at all. If not, then it will cause no impacts. Does the attribute which is predicted to be impacted exist at that site? The preliminary "filter" questions cannot cover all possible site characteristics, so certain of these attributes may not be present on the actual site. If they are not present, then no impact can be postulated. In this manner, the potential impacts presented by EICS may be reduced to the most likely and serious. These are then discussed as the main body of the EIA/EIS. The knowledgeable human at the end of the chain is still the ultimate filter determining the problems which are important.

Applications to Date

Who has used EICS? First, Army civilian and military personnel at all levels have used all presently available modules of the system. These include installation facility engineer and training

directorates personnel, as well as those in major command headquarters and in various research organizations. EICS has also been applied by contractors retained by the Army to prepare assessments and statements by specifying in the contractor's scope of work that EICS output should be examined to identify all major potential impact areas before site-specific information is collected and before an ad-hoc group decides indiscriminately the major impact areas to be investigated. Also, the EICS output has been used, prior to retaining a contractor, to identify major problem areas of the investigation and then specify these in the scope of work. A potential for savings in the Army exists through employing either of the above methods as compared to the traditional general specifications.

What cost-savings may be expected through use of EICS? Our experience indicates that they may be considerable in some cases, and small in others. The largest savings may be projected for those larger projects where the nature of the activity and its impacts are generally unfamiliar to those preparing the EIA/EIS. The "focusing" effect, here, is estimated to result in cost savings of about 50%. Time savings may also be evident. Contractors familiar with the system may wish to make its use a part of their contract, and this has been done in several cases. Fewer economies are apparent when the system is applied to very small or very large projects. When small, or routine, activities are covered by EICS, only a fraction of the potential effects will be determined to be likely or probable in any one case. In such cases it may be acceptable for one accession of EICS to be made for a category of actions taking place in one general setting, and then consulted from time to time when brief assessments are prepared. When very extensive actions having very severe impacts over wide areas are assessed, most of the cost of preparing the EIS may be field studies which determine just what the baseline conditions are. If such studies account for most of the contractor's costs, then the savings to be derived from focusing on certain impacts are a lesser proportion of the total contract cost.

FIELD APPLICATION OF EIFS

The estimation of economic impact, through EIFS, was the first area where quantified estimates of impact were attempted. It also is being maintained temporarily by CERL. Requests for its use may be directed to CERL or to the environmental offices within TRADOC or

FORSCOM headquarters. This system is interactive, with telephone connections being possible by offices with appropriate computer terminals. After access is gained, the user is queried about the numbers of personnel affected, their salary, their status as civilian or military, changes in installation procurements, if any, and other pertinent input information as described in DA PAM 200-2 (7). An estimate of the magnitude of economic impact on the local economy is then returned within seconds to the user. Any region of the United States may be so examined, and the user aggregates the region to fit the nature and location of the project. This subsystem has received very wide usage by the Army, and other DOD and government offices, for hundreds of possible actions, including thousands of alternatives to them. Current costs are only a few dollars per use, a savings of several hundred to several thousand dollars over other techniques. In fact, EIFS has made possible routine examination of the economic impact of every action including those previously believed too small to bother assessing. Total cost savings to the Army is estimated at well over .5 million dollars per year.

SUMMARY AND CONCLUSIONS

Lessons Learned

What has been learned by field use of computer systems for environmental assessment? First, the value of such systematic examination of environmental impacts has been shown many times over. Almost any written EIA or EIS may be made more complete and more responsive to the legal and moral commitments required of the Army by legislation, executive orders, and Army regulations. Personnel without proper educational background or knowledge of the project and the affected environment, or ad-hoc groups organized to put together a quick document rarely have the outlook necessary to utilize EICS initially. Such a group will probably best relate to a handbook method such as DA PAM 200-1 (8). Contractors and consultants are often able to make excellent use of EICS products, and may be able to realize time and cost savings through its use. Improvement in the quality and completeness of the resulting EIA/EIS may be the single most obvious result of its use. As more persons become trained, however, and familiar with the system, savings will be realized in personnel time as well, both in length of time required and in the numbers of

hours spent in analysis of EICS output and preparation of the EIA/EIS document.

Have problems arisen? The basic difficulty we have seen is that EICS does too good a job of bringing up potential impacts. Those persons who have treated the EIA/EIS process as an exercise in paper shuffling, with little specific consideration of the purpose of the assessment, frequently see presentation of this large number of potential problem areas as an added burden. If personnel are few, and already burdened with other duties, it may not be possible for them to do a better job of assessing actions. If it is possible for their supervisors to believe in doing a better, but not necessarily faster, job, then EICS certainly allows this. Improvement of quality and thoroughness is clearly the most common effect of EICS application, assuming personnel so tasked are allotted adequate time to do a good job.

Because of simplicity and ease of use, rapid access, and availability of regionally specific economic baseline information, the EIFS continues to be very useful for virtually all EIA/EIS's requiring economic consideration regardless of the turnaround time required to complete a document.

Keeping Abreast

What may be expected to change in the future? The nature of Army military activities may be expected to change to accommodate new weapons systems and new tactics. EICS can certainly accommodate such changes, and allow assessment of the peacetime training effects of the new procedures. All activities of the government, including the military, will certainly be expected to manage the real property allocated to them in a responsible way. Public scrutiny of Army activities can only become more intense. Development of tools such as the Environmental Impact Computer System and its associated components will assist Army personnel and consultants to predict, and therefore be able to minimize, those adverse environmental effects which might potentially interfere with a particular Army mission. In addition, legal requirements are constantly changing. Environmental legislation is being reviewed and new provisions and interpretations of NEPA are being prepared. Development and adaptation of EICS and other systems to meet newly arising actions must continue if this goal is to be assured.

REFERENCES

1. Congress of the United States. National Environmental Policy Act of 1969. PL 91-190, Jan 1, 1970. 83 Stat. 852 (42 USC 4321, et seq.). Section 102(2)A.
2. Leopold, L. B., F. E. Clarke, B. B. Hanshaw and J. R. Balsley. *A Procedure for Evaluating Environmental Impact*, United States Geological Survey (USGS) Circular 645 (USGS, 1971).
3. Jain, R. K., T. A. Lewis, L. V. Urban and F. E. Balbach. "Environmental Impact Assessment Study for Army Military Programs," TR D-13 of the CERL, 1973 (available from NTIS No. AD#771062).
4. Environmental and Energy Systems Division, Construction Engineering Research Laboratory. Environmental impact computer system attribute descriptor package: reference document. Technical Report E-86 of CERL, 767 pp, 1976 (NTIS AD-A024 303/OWP).
5. Urban, L. V., F. E. Balbach, R. K. Jain, E. W. Novak and R. E. Riggins. "Computer-Aided Environmental Impact Analysis for Construction Activities User Manual," TR E-50 of the CERL, 1975 (available from NTIS AD-A008 988/8WP).
6. Riggins, R., and E. W. Novak. "Computer-Aided Environmental Impact Analysis for Mission Change, Operations and Maintenance and Training Activities: User Manual," TR E-85 of the CERL, 1976 (available from NTIS ADA 022698).
7. Webster, R. D., R. A. Mitchell, R. L. Welsh, E. Shannon and M. L. Anderson. "The Economic Impact Forecast System: Description and User Instructions," TR N-2 of the CERL, 1976 (available from NTIS AD-A27 139/5WP and as Department of the Army Pamphlet 200-2).
8. Jain, R. K., L. V. Urban and G. S. Stacey. Department of the Army Handbook for Environmental Impact Analysis, Technical Report E-59 of the CERL, 1974 (available from NTIS ADA006241, also available as Department of the Army Pamphlet 200-1 from GPO D101.22:200-1).

DETERIORATION AND RELIABILITY OF PAVEMENTS

VICTOR C. BARBER, MR.

*EUGENE C. ODOM, MR.

U. S. ARMY ENGINEER WATERWAYS EXPERIMENT STATION
VICKSBURG, MISSISSIPPI 39180

PART I: INTRODUCTION

Pavement design has traditionally been set apart from maintenance and repair considerations. Early design procedures, by omission, presumed that pavements performed some service up to a point where sudden failure occurred as a result of some predetermined quantity of traffic. This "failure" was some finite definition of the pavement condition. However, it has been intuitively obvious that the pavement deterioration, or damage, occurred upon initiation of traffic and gradually increased to some point where conditions were unsatisfactory. It was also obvious to the designer that this unsatisfactory condition varied from one facility to another depending upon the needs and desires of the user.

More recently, members of the pavements community of engineering have sought after methodology for quantifying deterioration of pavements in various modes and for more properly quantifying failure of a pavement. These achievements have been considered essential to obtain the highest possible benefit from a pavement throughout its entire life. This concept has been termed "life-cycle management," which can be described as the management of a pavement from its conception to the end of its life. The term then should include planning, design, maintenance, repair, and some control of usage.

In order to effect life-cycle management, essentially every trade-off must be optimized. Particularly, a design must be at minimum cost with respect to deterioration. Additionally, the design

must be aimed directly at the level of reliability desired as a procedure for minimizing design redundancy or design insufficiency.

In the development of a pavement design method the U. S. Army Corps of Engineers (CE) adopted the California Bearing Ratio (CBR) tests for defining material strengths and developed the CE design method for road and airfield pavements (1,2). This semi-empirical method was selected in a time of military need for its ease of application by troops in the field. Through the years, the CE design methods for rigid, flexible, and other pavement types have undergone several modifications and are still in use today. These methods are deterministic in that they provide for design of new pavements with respect to a specific failure criterion, and do not provide for analysis of gradual deterioration in the respective modes, or "system drift." This implies sudden failure of pavements upon application of some computed quantity of traffic applications.

Major research programs have been undertaken in the past decade to advance CE design procedures and to improve overall design and management capabilities. In the early 1970's the Waterways Experiment Station (WES) initiated research programs to improve the CE design procedures as well as to improve design procedures of other agencies.

Among these programs was a program of study that was approved and funded in 1974 to develop life-cycle procedures for pavements based upon pavement deterioration and statistical reliability (3,4). This research is aimed at quantification of deterioration of pavements and assessment of reliability. This program of study, which is currently in progress, is partially sponsored by the U. S. Department of Agriculture Forest Service (FS). The FS is participating in the study as a result of their need to assess damage to roads caused by logging operations (4). Data available, along with data that are being collected, are expected to provide for analysis of damage caused by various types of vehicles. This capability will provide the basis for development of a differential cost analysis procedure to aid in assessment of maintenance cost to private sector timber industry.

The Department of Defense, through the CE, is participating in the effort as a result of determination of the need not only to predict deterioration of roads in a military scenario, but due to the need to assess pavement reliability in military, tactical, and logistical operations. This capability is a major step in the direction of overall systems reliability that includes ground and air vehicles as well as the mediums (pavements) upon which they operate.

The needs of these governmental agencies, based upon financial as well as operational criteria, then have given rise to this research effort to develop procedures for life-cycle management using deterioration and reliability concepts. This paper describes the first phase of this research program.

Objective

The objective of this study was to investigate the hypothesis that effective pavement life-cycle management can be achieved through utilization of deterioration and reliability concepts. This hypothesis was evaluated through a study to develop methodology to predict deterioration and to determine pavement reliability, respectively. In order to accomplish this hypothesis investigation, several intermediate objectives were set forth as follows:

- a. Utilize the surface rutting mode of deterioration to develop a pilot deterioration prediction procedure.
- b. Further develop the rutting prediction procedure into a stochastic reliability assessment system.
- c. Combine the deterioration and reliability models into a deterioration and reliability analysis procedure for use in life-cycle management. This procedure is a pilot procedure that incorporates the rutting mode of deterioration.
- c. Provide a base for expanding these developments to include other modes of deterioration.

Accomplishment of these objectives establishes a technological base for the development of an effective life-cycle management procedure.

Scope of work

The scope of work was designed to effectively provide for accomplishment of the objectives. The initial efforts consisted of a search of existing data to determine whether enough rutting data were available to provide a basis for development of the deterioration and reliability models. The data were analyzed and, being found tentatively satisfactory, were utilized for this purpose.

Literature was reviewed and studies were made to determine the most suitable method of analysis of the available data. Upon selection of methodology for data analysis, a major portion of the

research effort consisted of analysis of data and comparison of existing data to that being accumulated in ongoing field evaluations.

As deterioration and reliability models came forth from the analysis, computer programs were developed to provide for computerized operation of the various models. Ultimately, deterioration and reliability models for the rutting mode were developed for unsurfaced, gravel-surfaced, and two- and three-layer flexible pavements, respectively. These models were then combined to provide for deterioration and reliability analysis as well as for differential damage analysis where mixed traffic occurs. This system is termed the Differential Analysis System (DAS).

PART II: DATA COLLECTION AND CURRENT STUDIES

The CE design method has required revalidation and revision since its adoption due to the ever-changing nature of traffic. Therefore, it has been necessary for the CE to conduct prototype tests over the years to provide data to serve as a basis for revalidation and change. The net result of such an extended series of tests has been the accumulation of myriad prototype pavement performance data. The data are necessarily in terms usable by the early CE design methods, namely rut depths, thickness of layers, strengths of layers and subgrades in terms of CBR, and vehicle characteristics which include number and configuration of wheels, tire pressures, and wheel loads.

The end result of the aforementioned data search was that a large sample of high-quality rutting data was available that applied to two- and three-layer flexible pavements, as well as to gravel-surfaced and unsurfaced facilities. The data were initially termed suitable if the variables were within appropriate ranges and were all included and deliberately recorded. Criteria for rejection of certain data were based on range, reaction of test pavements, and engineering judgment with respect to overall suitability.

The data derived from this data search are used in the development of the constitutive deterioration and reliability models. The availability of such a quantity of data on deterioration in terms of rutting on CE-designed pavements is the basis for the selection of rutting and CE design parameters to conduct this basic research on deterioration and reliability. With additional data, the model may be expanded to other deterioration modes and even to other design procedures according to the needs of other governmental agencies.

In order to effectively achieve the overall goals of the pavement deterioration program, an inter-agency agreement (3) was entered into by the CE and FS to develop ways to predict deterioration and assess reliability. The agreement not only called for the pilot studies to develop methodology using existing data, but also instituted a large-scale field testing program to accumulate actual deterioration data pertinent to FS and CE roads for use in validation of this procedure for employment by FS and CE agencies. The modes of deterioration considered paramount for this study were rutting, roughness, slipperiness, cracking, and surface loss on aggregate-surfaced roads. Currently, approximately 50 test sites have been selected throughout the United States at various regional locations in areas having suitable design and traffic features to provide a deterioration environment. Test sections have been established at these test sites, and various testing procedures are conducted on the pavement structure to measure deterioration in the different modes.

To determine type and quantity of traffic that brings about deterioration, the traffic is monitored by a combination of electrical counters and visual observation samples. These field tests have been in progress for approximately two years. The combination of tests and traffic monitoring, when successfully pursued over a period of approximately six years, is expected to provide abundant data for deterioration and reliability system validation.

PART III: DEVELOPMENT OF DETERIORATION MODELS

Initial data analysis

The primary objective in the analysis of existing data was the establishment of the rate of rut depth (RD) change as a function of the independent variables. In this instance, where up to eight independent variables were involved, it was recognized early that the method of data analysis would be critical in terms of time, cost, and overall results of the analysis.

Regression analysis procedure

The procedure of analysis ultimately selected for use is an orderly method of developing mathematical relations from sets of data using multiple regression analysis. This procedure was developed by Lu, Lytton, and Moore (5), although the basic methods have probably been employed by others due to its direct approach to the analysis of data. The procedure utilized is termed a two-step constrained regression procedure. The first step of this method is essentially

a selection regression procedure using a multiplicative model to obtain the approximate exponents of the independent variables. The second step determines the coefficients of linear combinations of the products. The final model is selected based upon four factors:

a. As simple an expression as can reasonably describe what is actually taking place in terms of the dependent variables and reflecting the effect of the independent variables.

b. High correlation of the model.

c. Small prediction error.

d. Satisfaction of physical constraints.

The variables involved in the regression equations are tabulated as follows:

a. $RD \equiv$ rut depth in inches.

b. $P \equiv$ equivalent single-wheel load in pounds.

c. $t_p \equiv$ tire pressure in pounds per square inch.

d. $R \equiv$ number of load repetitions.

e. $t_i \equiv$ layer thickness in inches.

f. $C_i \equiv$ layer strength in CBR.

The total number of variables changes from one pavement type to another depending upon the number of layers that make up the pavement. The procedure described below is in general terms and applicable to all of the pavement types.

The final step in equation development is determining the rut depth as a function of the other variables. The equations thus derived take the form:

$$RD = f(P, t_p, R, t_i, C_i) + E_1$$

where E_1 is the regression error due to lack of fit of the new equation, stochastic variation of the variables and climatic variations, and other factors not explained in the variables.

Development of regression models

The existing data were used to develop models for rutting in the manner described in the preceding paragraphs. Experience and familiarity in the area of CE pavement design procedures provided some insight into the probable behavior of the variables, and therefore some transformations were attempted in the case of many of the variables. The thickness equations currently in use by the CE (6) were studied to determine the most likely form that some of the variables might assume.

The final forms of the rutting models resulting from the regression analysis of the four categories are listed below. Included with each model are the standard error (SE) and the coefficient of correlation r .

a. Unsurfaced facility model:

$$RD = 0.110 \left[\frac{P^{0.4925} t_p^{0.8548} R^{0.5018} (\log t)^{0.4293}}{c_1^{1.9773} c_2^{1.2015}} \right]$$

where $SE = 0.399$, $r = 0.9403$

b. Gravel-surfaced facility model:

$$RD = 0.1741 \frac{P_K^{0.4707} t_p^{0.5695} R^{0.2476}}{(\log t)^{2.0020} c_1^{0.9335} c_2^{0.2848}}$$

where $SE = 0.294$, $r = 0.9117$

c. Two-layer flexible pavement model:

$$RD = 1.9431 \left[\frac{P_K^{1.3127} t_p^{0.0499} R^{0.3240}}{[\log (1.25 t_1 + t_2)]^{3.4204} c_1^{1.6877} c_3^{0.1156}} \right]$$

where $SE = 0.411$, $r = 0.8779$

d. Three-layer flexible pavement model:

$$RD = 0.03117 \frac{P_K^{1.5255} t_p^{0.0897} R^{0.3450} C_1^{-0.7617} C_2^{-0.5505} C_3^{-0.3089}}{[\log (1.25 t_1 + t_2)]^{0.8847} (\log t_3)^{1.1674}}$$

where SE = 0.444, r = 0.8418

The variables influencing the rut depth for the unpaved categories are listed in Table 1, while the variables in the flexible pavement models are listed in Table 2.

TABLE 1.--VARIABLES FOR UNSURFACED AND GRAVEL-SURFACED FACILITY RUTTING MODELS

Index	Variable
1	RD = rut depth (in.)
2	P = equivalent single-wheel load (ESWL) (lb)
2a	P _K = equivalent single-wheel load (ESWL) (kip)
3	t _p = tire pressure (psi)
4	t = thickness of top layer (in.)
5	*C ₁ = CBR of top layer
6	*C ₂ = CBR of bottom layer
7	R = repetitions of load or passes

- * C₁ < C₂ for unsurfaced facilities
 C₁ > C₂ for gravel-surfaced facilities

PART IV: DEVELOPMENT OF RELIABILITY MODELS

Development procedure

The deterioration models for rut depth analyses of the four respective pavement types as developed previously are necessarily deterministic models. They represent not necessarily the best fit of all the data, but instead a "good fit" as indicated by the error and correlation values. To further expand the applicability and utility of the models, statistical reliability concepts are invoked to

TABLE 2.--VARIABLES FOR THREE-LAYER FLEXIBLE
PAVEMENT MODEL

Index	Variable
1	RD = rut depth (in.)
2	P = equivalent single-wheel load (ESWL) (lb)
2a	P_K = equivalent single-wheel load (ESWL) (kip)
3	t_p = tire pressure (psi)
4	t_1 = thickness of asphalt pavement (in.)
5	t_2 = thickness of base (in.)
6	C_1 = CBR on top of base
7	t_3 = thickness of subbase (in.)
8	C_2 = CBR on top of subbase
9	C_3 = CBR on top of subgrade
10	R = repetitions of load or passes

account for the variability of all input data and to predict rutting in terms of expected rut depth and variance from that expected rut depth. This concept provides for models for each pavement type that utilizes a total description of the input variables in terms of means and variances of a set of values for each variable. This provides for an accounting of the variability of pavement properties in a statistical manner instead of accepting only changes of values of measured parameters. The statistical determination of the rut depth in terms of expected value and variance provides not only for better analysis of rut depth in terms of an expected rut depth and a probable deviation from that value, but also for accounting for material variability in a statistical manner. The primary benefit of using total input data and determining the expected values and variances of a probability density function on rut depth is the ability to evaluate the reliability of the pavement. The reliability of a facility can be determined in terms of the probability that a given rut depth will occur under given circumstances. It follows then that such a system can be an excellent evaluation tool as well as a design tool. The great advantage as a design procedure is the capability to adjust reliability or conservatism to a desired value and selection of the design parameters to suit those conditions. In order to address

these concepts of reliability and the accounting for material variability, it was necessary to develop stochastic models for the definition of a probability density function of the rut depth.

If rut depth (RD) is considered a continuous, random variable with some probability density function, $f(RD)$, the expected value of $H(RD)$ is defined as

$$E[H(RD)] = \int_{-\infty}^{\infty} H(RD) f(RD) dRD$$

The expected value of RD is the mean or average of RD, which is termed μ_{RD} , or $\mu_{RD} = E(RD)$. The variance of RD as a variable is denoted by σ_{RD}^2 and is defined to be:

$$\sigma_{RD}^2 = E[(RD - \mu_{RD})^2]$$

Occasionally, taking the expected value of a complicated function can be a difficult process as in the case of the rut depth models previously shown. In order to overcome these difficulties, the expected value was approximated by taking a Taylor's series expansion and truncating all but the first three terms:

$$f(RD) = f(RD - \Delta RD) + f'(RD - \Delta RD)\Delta RD + 1/2 f''(RD - \Delta RD)\Delta RD^2 + \dots$$

If ΔRD becomes $RD - \mu_{RD}$, then the final generalized form can be expressed as follows:

$$E[f(RD)] = f(\mu_{RD}) + 1/2 f''(\mu_{RD})\sigma_{RD}^2$$

Taking the variance of the rut depth models was also a painstaking operation, further complicated by the forms imposed upon some of the variables. The Taylor's series expansion was again applied. The variance of the rut depth models is denoted by $V(RD)$ and is expressed as:

$$V(RD) = [f'(\mu_{RD})]^2 \sigma_{RD}^2 + 1/4 [f''(\mu_{RD})]^2 \sigma_{RD}^4 + \sigma_{lof}^2$$

in which σ_{lof}^2 is the variance of lack of fit.

The expected value and variance models of each of the four pavement types have been developed using the first three terms of a Taylor's series expansion and have been programmed for computer solution as part of the analysis model.

Determination of reliability

As has been previously stated, reliability as defined and used in this study is the probability that the rut depth will not exceed some predetermined value subject to conditions that are expressed by the independent variables. If a normal distribution on the dependent variable is assumed, the probability that the rut depth will not exceed some maximum value RD_A can be calculated using the equation

$$P = \frac{RD_A - E(RD)}{\sqrt{V(RD)}}$$

where $E(RD)$ is the expected value of the rut depth as determined from the appropriate model and $V(RD)$ is the variance of the rut depth as similarly determined. The value P represents the variable Z of the cumulative distribution function $F(Z)$. In order to determine reliability R , or that is, the area under the distribution curve defined by $E(RD)$ and $V(RD)$ and to the left of the maximum rut depth (RD_A), enter a normal distribution function table from a statistics handbook with a value of P (or Z) determined previously and determine the area under the distribution ($F(Z)$), which is the reliability R . As an illustrative example, assume the following values:

$$RD_A = 3 \text{ in.}$$

$$E(RD) = 2 \text{ in.}$$

$$V(RD) = 1 \text{ in.}$$

then $P = \frac{RD_A - E(RD)}{\sqrt{V(RD)}} = \frac{3 - 2}{\sqrt{1}} = \frac{1}{1} = 1$. Entering a normal distribution table with a value of $Z = 1$, it can be seen that $R = F(Z) = 0.8413$. Or, it can be said in this case that there is a probability of 0.8413 that the rut depth will not exceed a predetermined maximum value of 3 in.

PART V: DIFFERENTIAL ANALYSIS SYSTEM

The rutting models previously described have been entered on a computer program to form the base of the Differential Analysis System (DAS), which provides for a utilization of the rutting models to determine the rate of deterioration and/or reliability of any of

the four types of facilities described in terms of rutting. The term "differential" has been given to the computational system to emphasize the fact that differences in results caused by changes in input can be determined by the user in any assessment of damage caused by various vehicle types or the effects of changes in the structure. The DAS provides for one automatic iteration of the computational processes. Differential analysis as described above can be achieved simply by repeated iterations of the system while changing any variable or variables desired. The system as shown is adequate for limited use where the various models apply and is adequate to develop the original hypothesis that life-cycle management can be achieved through deterioration and reliability concepts.

The second stage of the DAS will provide for optimization of rutting and reliability values with respect to constraints imposed through the input of values or ranges of values representing the variables. Additional optimization will be achievable in terms of design and rehabilitation costs as well as selection among the four pavement options. As data pertinent to other pavement types, especially rigid pavements and selected hybrid pavements, become available, models will be included for their analysis.

Range and distribution of variables

The DAS is considered to be applicable to design and evaluation problems where normally encountered values of the variables are utilized. The range of applicability of any computational system is constrained by, if not limited to, the boundary conditions of data upon which it is based.

Utilization of the DAS

There are numerous applications of the DAS that can be made by those concerned with life-cycle management or any aspect thereof. As is the case with any similar system, validation is required to render the DAS directly applicable to specific locales having unique conditions. This version is, of course, in terms of rut depth as the dependent variable and major item of analysis. Validation of the DAS not only can achieve local applicability, but can incorporate the other deterioration modes as a data base is made available. Field data collection, for validation purposes, that is currently under way has been described previously.

Although the specific applications of the DAS are numerous and necessarily depend upon user needs, some of the more pertinent applications are described as follows:

a. Design and evaluation: The DAS is directly applicable to CE design and evaluation problems in the same sense as are current criteria due to similarity in data bases and results. The added features of the DAS are namely the ability to (1) modify limiting failure criteria (rut depth), (2) adjust conservatism to any desired degree by imposing a required degree of reliability, and (3) determine the reliability of a facility.

b. Optimization: Iterations of the DAS while making changes in appropriate variables can provide for optimization of a design with respect to cost, reliability, serviceability, layer properties, and materials.

c. Differential analysis: Iterations of the DAS provide directly for the analysis of the effects of different quantities and magnitudes of loads. The equivalent single-wheel concept (6) makes this possible by providing the capability to incorporate various vehicle configurations. This feature provides a quantitative basis for assessment of damages caused by various categories of vehicles and, when used on a relative basis would not require locality validation of the DAS.

d. Planning: The DAS can be considered an effective stochastic-type planning tool for quantitative estimation of future maintenance and repair needs as well as time-to-maintenance estimation. This feature, in connection with such procedures as CPM and PERT, can be used to effectively program work loads and expenditures.

e. Military operations: The tactical and logistical operations that could benefit from use of the DAS are too numerous to mention in detail. Such considerations as optimization of construction capabilities by constructing facilities having only a required reliability, and using facility reliability concepts to aid in tactical planning and maneuvers are key considerations that could be better quantified using DAS concepts.

The use of the DAS as it exists herein necessarily includes use of the CBR method of strength evaluation, which is in itself not a true physical material parameter. This feature is not to be considered a deterrent, however, to prospective users bound to other design procedures. Material strength parameters can be stated in any suitable terms where a sufficient data base exists for validation. True material parameters, such as Poisson's ratio (μ) and elastic properties (E), in various forms can be utilized and would provide for a more rational approach to the overall operation. Other methods

of portraying strength can be used where data are available. In all cases, any bias is removed in the actual correlation indicated during the validation stage and should be the basis of judgment as to whether a particular procedure is employed.

PART VI: CONCLUSIONS

The conclusions drawn as a result of this study are as follows:

a. The hypothesis that effective pavement life-cycle management can be achieved through the use of deterioration and reliability concepts has been investigated and proven.

b. Models were developed that effectively portray the deterioration of a facility and assess its reliability in terms of the rutting mode of deterioration.

c. The deterioration and reliability models show high correlation and small residual error and, therefore, when combined to form the heart of the DAS, should provide for effective rut depth prediction and reliability assessment.

d. The DAS can be used for the purposes indicated in the section entitled "Utilization of the DAS."

e. The DAS, as a first-generation system, provides a technological basis for development of a complete life-cycle management system for all modes of deterioration pertinent to all pavement types through expansion and validation as data are made available.

f. The DAS can be used in its present form for differential analysis on roads where damage incurred by various vehicle types must be determined as a basis for cost assessment.

REFERENCES

1. Ahlvin, R. G. et al., "Developing a Set of CBR Design Curves," Instruction Report 4, Nov 1959, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss.
2. Ahlvin, R. G. and Brown, D. N., "Combined CBR Criteria," Technical Report No. 3-495, Mar 1959, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss.

3. Forest Service-Federal Aviation Administration Inter-Agency Agreement, 1974, unpublished.
4. Barber, V. C., Odom, E. C., and Patrick, R. W., "Pavement Deterioration Analysis for Design and Evaluation Systems," Technical Report S-76-15, Dec 1976, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss.
5. Lu, D. Y., Lytton, R. L., and Moore, W. M., "Forecasting Serviceability Loss of Flexible Pavements," Research Report No. 57-1F, Nov 1974, Texas Transportation Institute, Texas A&M University, College Station, Tex.
6. Hammitt, G. M., II, Hutchinson, R. L., Rice, J. L., Thompson, O. O., and Brown, D. N., "Multiple-Wheel Heavy Gear Load Pavement Tests, Volume IV: Analysis of Behavior Under Traffic," Technical Report S-71-17, Nov 1971, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss.

BEATRICE, LUND, COURS,
WAMPNER AND SLINEY

AD-A056408

PROJECT MILES: BIOMEDICAL RESEARCH
AND COORDINATION IN SAFE FIELD EXERCISES (U)

EDWIN S. BEATRICE, MD, LTC *
DAVID J. LUND
DEPARTMENT OF BIOMEDICAL STRESS,
LETTERMAN ARMY INSTITUTE OF RESEARCH
PRESIDIO OF SAN FRANCISCO, CA

DAVID COURS, LTC
PAUL WAMPNER
US ARMY PROJECT MANAGER FOR TRAINING DEVICES
ORLANDO, FL

DAVID H. SLINEY
US ARMY ENVIRONMENTAL HYGIENE AGENCY,
ABERDEEN PROVING GROUND, MD

The Multiple Integrated Laser Engagement System (MILES) is the core system for a family of laser engagement simulators which have the potential to revolutionize Army tactical unit training.

MILES devices are being developed for the M16 rifle, the Army's full family of machineguns, the VIKER, DRAGON, TOW, the main battle tanks (M60A1, A2, A3), and the M551 AR/AAV. Follow-on efforts will expand the MILES system into air defense weapons, helicopters, artillery, high performance aircraft, and enemy weapons systems. The prototype packages consisted of laser transmitters which simulated the effect of the weapons, the laser detector array which detects and decodes incoming laser signals, and hit indicating mechanisms which combine audio and visual signals to convey near misses and kills.

The system uses low power gallium-arsenide (GaAs) lasers. Each device is lightweight, and its addition to the base weapon will not affect the normal handling, accuracy, or performance of that weapon. For every weapons system involved, the laser transmitter will have a hit probability comparable to the weapon simulated, as

BEATRICE, LUND, COUPS,
WAMPNER AND SLINEY

well as duplicating the weapons' effects. An infantryman, for example, can "kill" another infantryman with his M16 device equipped rifle, but cannot disable a tank. Conversely, a tank can "kill" not only another tank, but also TOW crews and infantrymen. The key to this is distinct pulse codes for each weapon and discrimination logic in each detector.

By duplicating the ranges and simulating the lethality effects and characteristics of direct fire weapons, realistic, two-sided exercises can be conducted with reduced controller requirements and increased training value.

The MILES program will not only provide greater fidelity, it will extend tactical engagement simulation to full company team and battalion task force level training to include night operations. The MILES basis of issue for each active Army division will include enough devices to permit two battalion task force exercises to be conducted simultaneously. Operational testing of the MILES system is scheduled for August 1978 with production and initial issue beginning in 1979.

Proper operation of this system requires the laser to be purposely directed at personnel. Ocular exposure approaches certainty. Field deployment will therefore ultimately reflect the confidence of the user in his understanding of the ocular hazard of the injection diode laser. Cognizance of this fact led to a meeting in Orlando, FL, on 4 February 1976 with personnel from the US Army Environmental Hygiene Agency; the Project Manager for Training Devices, Orlando, FL, and the contractor, to discuss the planned engineering development model of the MILES laser system (1).

The meeting attendees concluded the proposed MILES system emitted optical radiation exceeding current protection standards (Class 1-System). In response to the recommendation that further biological research be performed, those groups involved in project MILES convened at the Letterman Army Institute of Research (LAIR) in April 1976 for a review of the biomedical effects of gallium-arsenide laser radiation. It was pointed out that it is not sufficient to simply test the MILES device against a biological system (the eye) in controlled laboratory situations. The assignment of this system to a "safe" category (Class 1) is controlled by existing regulation AR 40-46 (2) and TB MED 279 (3). A research program bearing on the provisions of these regulations was outlined.

The fundamental requirement was to produce data that would indicate a need to modify existing standards, which is based in part, on limited data. Experiments were designed to answer questions and provide data on the damage threshold for a single-pulse exposure at the GaAs wavelength (905 nm) and the effect of exposure to pulse trains at the MILES code and frequency. In addition, experiments were designed to study the effect of the non-circular irradiation geometry resulting from exposure to collimated GaAs radiation (4). The most direct method to provide these data would be through the direct evaluation of retinal lesions created by exposure to a pulsed GaAs system. However, the injection diode laser was an uncooperative source when one attempted to optically couple the emission onto the retina, to the point where no retinal lesion had been successfully produced with a laser operating under the variables required of Project MILES. These variables approach the limits of a single-junction, pulsed, room temperature, GaAs laser.

The ocular hazard at a wavelength of 860nm had previously been determined for 120 KHz Pulse Repetition Frequency (PRF) radiation from a cryogenically cooled GaAs laser diode(5). The experiment was sufficiently defined to predict a rectangular retinal irradiation geometry. The retinal burns were in all instances circular in shape when viewed ophthalmoscopically, and were frequently oval in shape in retinal flat preparation and histopathological section (6).

MATERIALS AND METHODS.

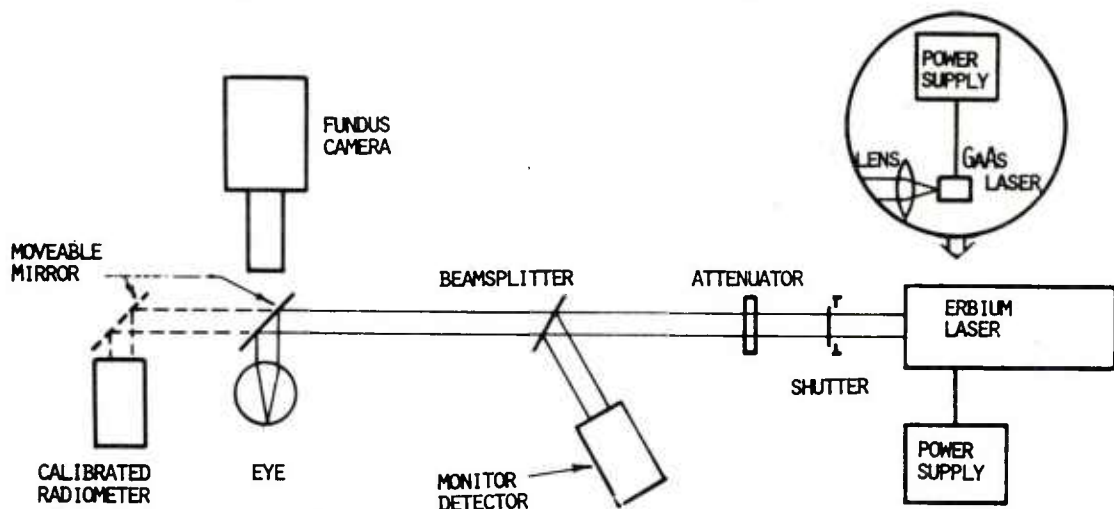


FIGURE 1: EXPERIMENTAL APPARATUS

BEATRICE, LUND, COURS,
WAMPNER AND SLINEX

Experiments were conducted to provide sufficient data for further interpretation of the biomedical effects data base, utilizing lasers which approximated the MILES device in spectral and/or temporal emission characteristics.

The apparatus used in these experiments is shown in figure 1. A shutter controlled the exposure sequence and attenuating filters reduced the beam energy to the desired level. A beamsplitter directed a portion of the energy into a detector which recorded the dose of each exposure. This detector was calibrated by reference to a radiometer each day the system was used. A goniometer mount provided rotation of the animal about the pupil of the eye to be exposed, allowing precise positioning of the exposures on the retina. An accurately repositionable mirror, which directed the beam into the eye, was moved to permit fundus camera observation of the retinal exposure sites.

Each laser, as required, was positioned in the exposure system. The divergence of the beam at the eye exposure position was measured so that an accurate estimation of the size and geometry of the retinal image could be made.

The animals used in these experiments were rhesus monkeys (*Macaca Mulatta*) weighing between 2 and 5 kg. Preanesthetic medication consisted of a sedative dose of phencyclidine hydrochloride (0.25 mg/kg) intramuscular and atropine sulfate (0.2 mg) subcutaneously. Anesthesia was induced with sodium pentobarbital (approximately 5 mg/kg) via the saphenous vein. A pediatric intravenous injection set was placed into the saphenous vein to administer fluids and to facilitate additional anesthetic. The pupils were dilated and sutures of 3-0 silk were placed in the upper eyelid to facilitate manipulation. While the eyes were open during the experiment, physiologic saline was used to maintain good corneal transparency.

The animals were positioned in the exposure system and the fundus examined via the Zeiss fundus camera. Any abnormalities were noted. Thirty-six to forty eight exposures were placed in a square array utilizing suprathreshold marker burns to accurately locate the rows and columns for subsequent examination.

Detailed ophthalmoscopic examination of the exposure sites was conducted at one hour post exposure. The criteria for damage were the presence of a lesion visible via this examination.

BEATRICE, LUND, COURS,
WAMPNER AND SLINEY

For most of the systems evaluated, a probit analysis was performed (7). The presence or absence of changes in appearance of the target area was noted for each irradiance studied. The ratio of observed responses to the total number of exposures at given doses was then plotted on logarithmic probability paper where the ordinate is probability (percentage observed) and the abscissa reflects the dose.

From the plot, the ED_{50} (effective dose required to produce an observable response 50 percent of the time) was obtained. Confidence intervals about the dose response curve were calculated. Because of the experimental design, the ED_{50} has greatest statistical significance and is often referred to as the "damage threshold."

ERBIUM: YLF EXPOSURE SERIES

This laser produced Q-switch pulses of 180-ns duration at a wavelength of 850 nm with a beam diameter of 1.6 mm.

Dose response data were obtained for two exposure conditions. Initial exposures used a laser beam divergence of 0.7 milliradian, producing a minimal retinal irradiation diameter. After an ED_{50} had been obtained for this condition, a +9 diopter lens was introduced into the beam to produce a 26.8 milliradian beam divergence resulting in a 400 μ m retinal irradiation diameter. The results of the 850-nm Erbium laser exposures are tabulated in Table I.

TABLE I

ED_{50} Dose for Q-Switch Erbium Laser Exposure - 850 nm

	ED_{50}
Minimum Retinal Irradiance Diameter	$12 \pm 3 \mu J$
400 μ Retinal Irradiance Diameter	$138 \pm 20 \mu J$

This experiment established the damage threshold for a single, short-duration pulse in the spectral region of the GaAs laser.

BEATRICE, LUND, COURS,
WAMPNER AND SLINEY

NEODYMIUM: YAG - PULSE TRAIN EXPOSURE SERIES

The laser was a continuously pumped acousto-optic Q-switched Nd:YAG laser. The wavelength was 1064 nm. Within the pulse repetition frequency range of these experiments, the pulse duration was 180 ns. The laser was pulsed continuously at the desired frequency and an external shutter was used to pass the desired number of pulses. The laser beam at the eye was 2mm in diameter and collimated to produce a minimum retinal irradiation area. These data are summarized in Table II.

TABLE II
Laser Exposure Data Nd Laser Pulse Trains - 1064nm

PRF		Number of Pulses					
		1	2	3	6	74	1000
100Hz	ED ₅₀ (μ J)	128.9	75.1	89.8			
1000Hz	ED ₅₀ (μ J)		80.1	51	55	16.4	10.1
3000Hz	ED ₅₀ (μ J)		60.6	43.6	30.4		

These data show a high degree of additivity for two pulses, with the additivity being greater for shorter interpulse spacing. The degree of additivity lessens following the second pulse. If the data for 1000Hz is plotted on log paper as total pulse train energy vs total exposure duration, the slope of the resulting line is nearly identical to the slope of existing continuous wave (cw) neodymium laser ED₅₀ data. This is at variance with the guidance of TBMED 279 which assumes a different time dependency of the safe level for the two exposure conditions. This variance encouraged the investigators to perform a literature search for all existing pulse train exposure data. Examination of these data, for pulse durations from 10 ns to 1 ms and pulse repetition frequencies from 1Hz to 10,000Hz reveal a consistent relationship between pulse train data and equivalent continuous wave data. This relationship is not adequately modeled by the procedures of TBMED 279.

GALLIUM ALUMINUM ARSENIDE (GaAlAs) EXPOSURE SERIES.

The source of radiation for this experiment was a cw GaAlAs stripe geometry laser diode, selected to have a maximum laser output of not less than 20 mW. The wavelength of maximum emission, measured in this laboratory, was 883 nm. The spectral bandwidth was not measured, but is reported by the manufacturer to be 2.5 nm.

BEATRICE, LUND, COURS,
WAMPNER AND SLINNEY

The laser was driven with a stable 385 milliamp source. A 5.5-mm focal length lens collimated the laser emission. The beam divergence was 0.65 milliradian by 4.3 milliradian.

A horizontal row of 12 suprathreshold retinal exposures were made with the erbium laser to produce location markers. Three rows of GaAlAs laser exposures were located below the marker row.

All exposures were of 30-second duration. The ED_{50} , determined from 185 exposures in six eyes, was 230 millijoules. The lower and upper 95% confidence limits were 202 millijoules and 262 millijoules, respectively.

MILES PROTOTYPE EXPOSURE SERIES.

The laser source used in this experiment was a prototype gallium-arsenide laser training device having two modes of operation. In the pulse-repetition-frequency (PRF) mode, the output consisted of a continuous train of 100-ns, 905-nm pulses at a repetition rate of 1,600 Hz. In the pulse code mode, the interpulse spacing is not constant and the average repetition rate is 132 Hz. The pulse energy and duration are the same in both operating modes.

In each eye a total of forty-eight exposures were made in a grid pattern for exposure durations from 1 sec to 90 sec. The total intraocular energy was .212 erg/pulse for the 1 watt laser and 1.64 erg/pulse for the 10 watt laser.

Evaluation of the retinal sites was made by funduscopy observation, intravenous fluorescein angiography, retinal flat preparation, and/or epon imbedded serial sections for light microscopy (Trypan blue, azure II staining). These analyses were carried out for immediate, 1 hour and 24 hour intervals after laser exposure.

These ocular exposures did not produce the type of retinal opacity which is typically seen by funduscopy after laser irradiation of the retina. Exposures in the 30-second PRF sequences were characterized by the development of a pale gray clouding within 10 seconds after initiation of the exposure while the laser continued to irradiate the retinal site. At the end of the 30-second interval, the exposure site measured approximately 350-400 microns and was darkened at the periphery with central diffuse clouding (8).

The incidence of observed retinal changes at various exposure levels is summarized in Table III.

TABLE III
Summary of MILES Prototype Laser Exposure Data

<u>Lasers</u>	<u>Mode</u>	<u>Exposure duration</u>	<u>TIE</u>	<u>Exposures/Changes</u>
1 watt	PRF	30 sec	1.0 mJ	100/67
		10 sec	0.34 mJ	25/18
		5 sec	0.17 mJ	7/5
		1 sec	0.034 mJ	10/0
1 watt	Pulse Code	30 sec	0.084 mJ	14/0
10 watt	PRF	30 sec	7.9 mJ	19/11
10 watt	Pulse Code	30 sec	.65 mJ	5/5

The subtle retinal changes were persistent (24 hrs after exposure). However, no fluorescein leakage or histological evidence of retinal alteration was confirmed in any of the exposure sites.

With the exception of the direct observation of retinal "clouding," none of the techniques (angiography, flat preparation, serial microscopy, and fundus photography) routinely used to determine the site or extent of the change demonstrated any retinal alteration.

DISCUSSION

Analysis of the results of these studies reveal a paradoxical situation. One result is a recommendation that the provisions of TBMED 279 be changed in a manner that allows the latest MILES device to be placed in Class 1 - safe laser (9). A second result is a body of biological data which indicates that retinal alterations are occurring at irradiance levels below those afforded by the MILES device. Note that there is an exact reversal of the condition which existed at the start of the study.

The ocular doses from the MILE prototype device which produced retinal "clouding" were below the existing TBMED 279 standard for pulse train irradiation. These doses approached those used in functional testing which produced measurable visual performance degradation.

This subtle retinal clouding was a different end point response criteria than was used in the body of dose response data upon which the safety standards are based. Those data used the presence of a visible retinal opacity or lesion as a response criteria, as did the data reported here for the erbium laser, GaAlAs laser and repetitive pulsed Nd Laser. The clouding phenomenon does have correlates in other retinal response observations. The discrete lesion produced by the cw GaAlAs laser was surrounded by retinal clouding similar in appearance to that produced by the MILES prototype device.

Neodymium single pulse and multiple pulse exposures made at 2-3 X ED₅₀ produced a bright central whitened area of less than 100µm diameter which faded within ten seconds while the typical retinal opacity developed. At lower levels, approximating the ED₅₀ level, a reproducible change in the reflectivity from the retina was observed directly preceding the development of a retinal burn. At levels below the ED₅₀ the transient reflectivity change was the only effect noted.

Retinal effects for the erbium and GaAs lasers gave the impression that the changes are superficial to the retinal pigment epithelium for near "threshold" burn levels. If indeed changes are occurring superficial to the level of the pigment epithelium, then subtle exposures may produce damage to neural elements of the retina including the photoreceptors themselves without pigment epithelial change. Further, the accepted experimental ED₅₀ level, from which safe levels are determined, may be forced to be revised to account for other than retinal opacity levels.

COMPARISON OF DATA WITH STANDARD

Figure 2 presents the ED₅₀s for ocular exposure to the cw GaAlAs laser, the 120 KHz GaAs laser and the 180 nsec pulse duration 850 nm Erbium laser. Also presented are the ED₅₀s for ocular exposure to the cw neodymium laser and the pulsed neodymium laser. The neodymium laser data represented a consistent set derived in one laboratory with a common laser and observer. The neodymium data throughout were for minimal retinal irradiance diameter. The pulsed neodymium data and the pulsed erbium data are derived in one laboratory by a common observer using well behaved lasers in identical dose delivery systems. These data are therefore representative of the ratio of damage thresholds for the two wavelength regions.

The 850-950 nm data is inconsistent. The cw data does not lie on a common line and the separation between wavelength regions is

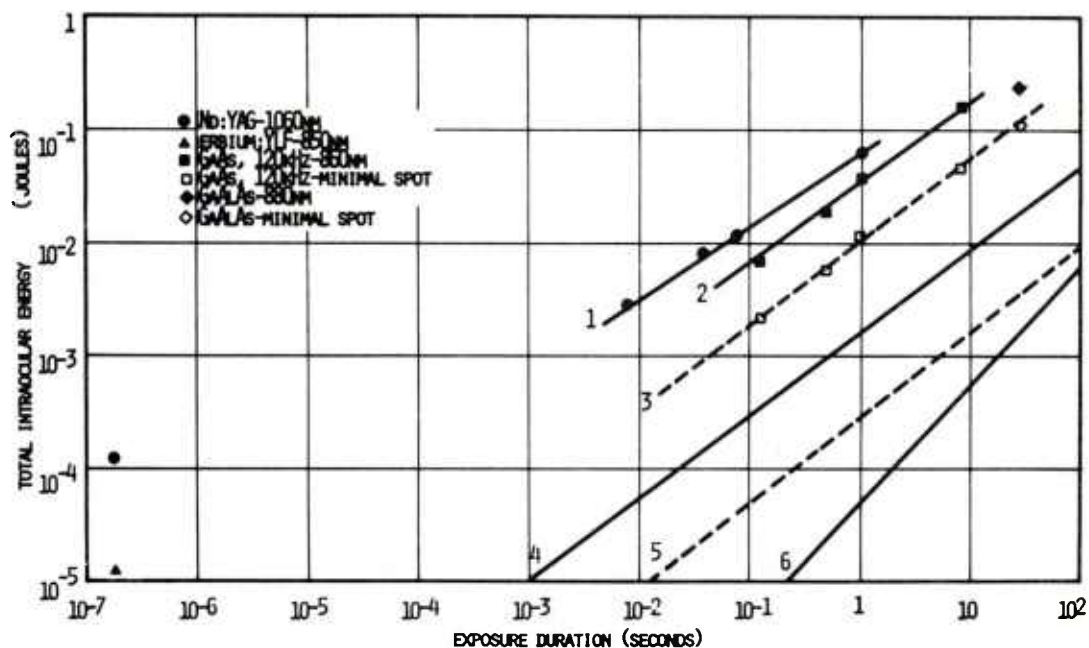


FIGURE 2: COMPARISON OF OCULAR DAMAGE THRESHOLDS
TO MAXIMUM PERMISSIBLE EXPOSURE LEVELS

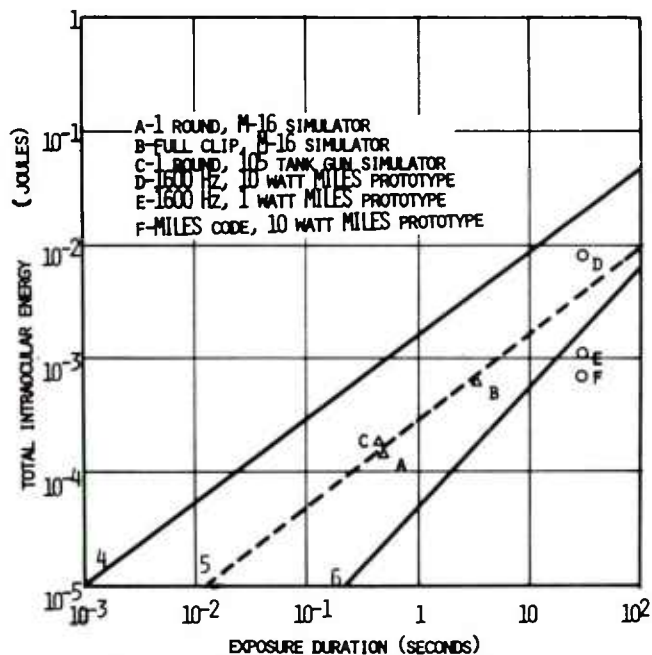


FIGURE 3: COMPARISON OF MILES DEVICE OUTPUT ENERGIES
TO MAXIMUM PERMISSIBLE EXPOSURE LEVELS

not the same as that obtained with the pulsed lasers. The 120 KHz GaAs and cw GaAlAs data are not for minimal image diameter. Extrapolation is possible from these data to the estimated threshold for worst case conditions. Data collected for a variety of pulse durations, wavelengths, and retinal irradiance diameters show that, on the average, (10)

$$1) \quad \text{Log } E = k - \log D$$

where E is the retinal radiant exposure ED_{50} and D is the retinal irradiance diameter. From this equation can be derived the relationship:

$$2) \quad \frac{TIE_1}{TIE_2} = \sqrt{\frac{A_1}{A_2}}$$

where TIE is the total intraocular energy and A is irradiated area. The 120 KHz laser had a beam divergence of 11 milliradian by 0.2 milliradian. It is highly unlikely that eye would be able to produce the approximately 3 micron spot predicted from 0.2 milliradian divergence; a dimension of at least 20 microns is probable. The cw diode produced a beam divergence of 4.8 milliradian by 0.6 milliradian. The eye would again produce at least 20 micron spot from the 0.6 millirad divergence. The 120 KHz diode irradiated a retinal spot of $165 \times 20 \mu\text{m}$ and the cw diode irradiated a retinal spot of $72 \times 20 \mu\text{m}$. From equation 2 the 120 KHz GaAs data must be reduced by a factor of 3.24 and the cw GaAlAs data reduced by a factor of 2.14. to present the worst case conditions (20 μ diameter spot). The estimated worst case line (line 3) is plotted on figure 2. This corrected data shows a consistent relationship to the neodymium data set.

Viewing multiple pulses is more hazardous than viewing a single pulse because of additivity of pulse effects. At question is the degree of additivity. Ocular damage thresholds were determined for exposure to pulse trains from a repetitively pulsed Nd laser. These measurements were designed to determine the additivity of the effectiveness of pulses as a function of pulse number and pulse separation. The consistent relationship between pulse train data and cw data led to the following formulation for computing the maximum permissible exposure to pulse trains.

$$3) \quad \text{MPE}^{\text{RP}}(T) = \text{CF MPE}(T)$$

BEATRICE, LUND, COURS,
WAMPNER AND SLINNEY

$MPE^{RP}(T)$ is the maximum permissible exposure expressed as total intraocular energy for the pulse train of duration T , $MPE(T)$ is the maximum permissible exposure for cw irradiation of duration T and CF is a correction factor which is dependent upon the pulse repetition frequency and the duration t of an individual pulse in the pulse train.

$$\begin{aligned} 4) \text{ for } t \text{ greater than } 2 \mu s \quad CF &= 3.5 PRFt^{-1} \\ \text{for } t \text{ less than } 2 \mu s \quad CF &= 1.8 \times 10^{-11} PRFt^{-1} \end{aligned}$$

This method has the advantage of applying the same margin of safety for pulse train exposures as is applied to cw exposures.

CF was derived by taking the ratio:

$$5) \quad CF = ED_{50}^{RP}(T) / ED_{50}(T)$$

$ED_{50}^{RP}(T)$ and $ED_{50}(T)$ are the damage threshold doses for pulse trains of total duration T and cw exposures of total duration T respectively.

Applying CF to the estimated worst case line for cw exposures yields a worst case (Figs 2, 3 line 4) for pulse train exposures. This line is shown for $PRF = 1,636 \text{ Hz}$, $t = 180 \text{ nsec}$, which are typical of MILES device parameters. The MPE for cw GaAs exposure (line 4) and the MPE^{RP} for pulsed GaAs exposure (line 5) derived by application of the CF for 1636 Hz , 180 nsec pulses are shown, as well as the MPE^{RP} obtained from TBMED 279 by application of C_p . (line 6) The emission energies of various MILES devices are shown for comparison. (Fig 3) In all cases, these energies are a factor of six below the projected ED_{50} levels and are approximately equal to the MPE^{RP} obtained by the CF method.

Another result obtained from the evaluation of pulse repetition data is:

$$6) \quad MPE^{RP}(T) = n^{3/4} Q MPE(t)$$

where n is the number of pulses in the pulse train, and $MPE(t)$ is the maximum permissible exposure for a single pulse. Q is a constant for any given pulse duration and wavelength, but its numerical value is not readily predictable. This relationship says that MPE^{RP} is essentially independent of PRF , but depends only on the number of pulses. Therefore, the results of the foregoing section are valid when applied to the average PRF of a non-uniformly spaced train of pulses.

BEATRICE, LUND, COURS,
WAMPNER AND SLINEY

SUMMARY AND RECOMMENDATION

These studies indicate that there is a method and a justification for modifying the portions of TB MED 279 that govern the computation of the MPE for pulse trains, and a recommended procedure for so doing has been forwarded. These recommended changes allow a MILES system which is Class I in nature, as defined by a revised TB MED 279.

However, these studies further point out the fact that there is much we do not understand about the interaction between laser radiation and the visual system.

Little information is available on experimental data derived from fundoscopic evaluation of retinal exposures which describe changes other than the presence of retinal opacity. It may be that other wavelengths produce subtle visible persistent alteration which can be observed by ophthalmoscopy. The proposed changes in the standard do not take into account repeated exposures at intervals on the order of minutes to hours or the results of chronic viewing of low "safe" level laser source. Until these results are available both the developer and user must understand the limitations of the current recommendations. The current biomedical research data supports the recommendation that for the field use of this family of low power GaAs devices, a safe field use can be relatively assured. However, in the mode of operation where fixed optics and probability of continuous or repeated exposure exist, such as at the maintenance depot or other repair facility, manuals should specifically caution the intended user about viewing the laser source.

IN CONDUCTING THE RESEARCH DESCRIBED IN THIS PAPER, THE INVESTIGATORS ADHERED TO THE "GUIDE FOR LABORATORY ANIMAL FACILITIES AND CARE," AS PROMULGATED BY THE COMMITTEE ON THE GUIDE FOR LABORATORY ANIMAL FACILITIES AND CARE OF THE INSTITUTE OF LABORATORY ANIMAL RESOURCES, NATIONAL ACADEMY OF SCIENCES - NATIONAL RESEARCH COUNCIL.

BEATRICE, LUND, COURS,
WAMPNER AND SLINEY

REFERENCES

1. MARSHALL, W. J. AND D. H. SLINEY: Preliminary theoretical Hazard Evaluation of the Engineering Development Model (ED) of the Multiple Integrated Laser Engagement Simulator (MILES) US Army Environmental Hygiene Agency Report No. 42-068-76, Feb. 1976.
2. Department of the Army: Regulation 40-46, Control of Health Hazards from Lasers and Other High Intensity Optical Sources. Washington, D. C., Hq., Department of the Army, 15 March 1974.
3. Department of the Army: Technical Bulletin, TBMED 279. Control of Hazards to Health from Laser Radiation. Washington, D. C. Hq., Department of the Army, 30 May 1975.
4. BEATRICE, E. S. AND D. J. LUND: Characteristics of Damage Produced by Non-circular Retinal Laser Irradiation. Letterman Army Institute of Research Report No. 31, Oct. 1976.
5. LUND, D. J., D. O. ADAMS, AND C. C. CARVER: Ocular Hazard of the Gallium Arsenide Laser. Letterman Army Institute of Research Report No. 30, Oct 1976.
6. ADAMS, D. O., D. J. LUND, AND P. SHAWALUK: The Nature of Chorio-retinal Lesions Produced by the Gallium Arsenide Laser. Invest. Ophthalmol 13:471-475, 1974.
7. FINNEY D. J.: Probit Analysis, Cambridge, Mass., University Press, 1972.
8. BEATRICE, E. S., D. J. LUND, M. CARTER, AND D. M. TALSMA: Retinal Alterations Produced by Low Level Gallium Arsenide Laser Exposure. Letterman Army Institute of Research Report No. 38, Feb 1977.
9. STUCK, B. E., D. J. LUND, AND E. S. BEATRICE: An Evaluation of Repetitive Pulse Data and Current Permissible Exposure Limits. Letterman Army Institute of Research Report (in print)
10. LUND, D. J. AND E. S. BEATRICE: Ocular Hazard of Short Pulse Argon Laser Irradiation. Letterman Army Institute of Research Report No. 42, Sept 1977.

LIFE EXPECTANCY OF U.S. ARMY COMMERCIAL
DESIGN ADMINISTRATIVE VEHICLES

*MR. RAYMOND BELL
US ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITY
ABERDEEN PROVING GROUND, MARYLAND 21005

INTRODUCTION: At the request of the Tank-Automotive Readiness Command (TARCOM), the Army Materiel Systems Analysis Activity (AMSAA) has initiated a study to reassess the useful life (years/miles) of the Army's commercial type administrative vehicles. Specifically, 122 different types of vehicles will be studied. Included among these commercial type vehicles are buses, sedans, panel trucks, dump trucks, ambulances, etc. In addition, it has been requested that the maintenance man-hour standards for these vehicles also be re-evaluated.

In this paper, a description of the study plan will be presented along with a discussion of the data base and methodology to be employed. In this connection, it will be shown how the existing Administrative Vehicle Management System (AVMS) data base will be used in developing maintenance cost models for the various vehicles under study. Further, in describing the methodology, it will be shown that the determination of the life expectancy of these administrative vehicles will be arrived at through an evaluation of the economic life of these vehicles supplemented by a Reliability, Availability and Maintainability (RAM) analysis of the vehicles.

STUDY REQUIREMENTS: In tasking AMSAA to carry out the administrative vehicle life expectancy study, TARCOM specifically requested that AMSAA update the life expectancies and maintenance man-hour standards contained in two Department of Defense Instruction (DODI) documents. DODI 4150.4 "Replacement and Repair Guidance, and Life Expectancies For Commercial Design Vehicles" contains a list of the life expectancies (year/miles) for 14 types of vehicles (see Table I). It is pointed out that the last update of these life expectancies

BELL

occurred in 1963 and thus an update of these lives was deemed necessary. In discussing these 14 types of vehicles, it was learned that 92 separate types of vehicles actually comprised the 14 vehicles listed in DODI 4150.4 (Table I) and that a life expectancy analysis was in fact required for each of the 92 vehicles. As an example of the division of the 14 vehicles listed in Table I into the 92 vehicles mentioned, the "Sedans, All" divides into five types: subcompacts, compacts, intermediate, regular and executive. In addition to the 92 vehicles to be studied, TARCOM requested that an additional 30 maintenance and service vehicles also be evaluated from a life expectancy viewpoint (see Table II).

An additional requirement that was included in the study was to evaluate the maintenance man-hour standards for commercial design vehicles contained in DODI 4151.10 "Maintenance Man-Hour Input Standards For Commercial (Transport) Design Motor Vehicles." This particular requirement was included in the study not only because the Department of Army (DA) requested an update of the values contained in DODI 4151.10 (see Table III) but because maintenance man-hour data will be accumulated during the course of the study.

STUDY PLAN: In order to accomplish the goals of this study (life expectancy determinations and updated maintenance man-hour standards), it is planned to determine the economic life of each of the vehicles being studied supplemented by a RAM analysis of these vehicles over the economic life span in order to determine if the life expectancy should be less than the economic life because of RAM considerations. The primary data requirement for this effort is maintenance data as a function of accumulated vehicle mileage for each of the 122 vehicles under study. Specifically, the types of maintenance data required are maintenance costs, man-hours utilized, mileage at which maintenance occurred, parts consumed, header data (vehicle type, line item number, description), date maintenance occurred and whether maintenance was of a scheduled or unscheduled variety. This type information will permit the generation of maintenance cost models as well as provide data for a RAM analysis. The maintenance cost models when combined with the present value (in FY 78 dollars) and the salvage values of the vehicles will thus provide all the data necessary for the economic life determination of the vehicles.

In order to obtain the primary data requirement (maintenance data as a function of accumulated mileage), AMSAA plans on accumulating this data through the currently existing Administrative Vehicle Management System (AVMS). AVMS data will be collected at six different sites (Fts. Benning, Lewis, Shafter, and Knox; Aberdeen Proving Ground (APG) and White Sands Missile Range (WSMR) beginning 1 October

BELL

1977 for one year. These sites were selected because they contain the largest quantities and have the widest diversification of types of administrative vehicles. It is anticipated that data will be collected on a total of approximately 6,000 vehicles at these sites.

DATA BASE: From past experience in the conduct of life expectancies studies for tactical vehicles, AMSAA immediately began a search for a data base from which maintenance data for administrative vehicles could be obtained. It was AMSAA's concept to obtain this data, if possible, from an existing data base rather than launching into an expensive data collection effort. In discussing the data requirements with Training and Doctrine Command (TRADOC) personnel, it was learned that the existing AVMS data base basically accumulates the desired data although not in the form required for the study. In the AVMS, each reporting installation compiles maintenance data on each of their administrative vehicles on a monthly basis but utilizes this data to supply an accumulated summary of the particular data element (man-hours, maintenance cost, etc.) since the beginning of the fiscal year to higher headquarters for review. There is no requirement in the AVMS to provide any of the data elements as a function of vehicle mileage as is required for a life expectancy study. It is also noted that in the current system the raw monthly maintenance data is disposed of after 90 days and further the mileage on the vehicle at the time of the maintenance action is not recorded.

Despite the shortcomings of the AVMS, from a life expectancy analysis standpoint, this data base was considered suitable for the study. To utilize the data base, each installation included in the study was requested to record the vehicle mileage at each maintenance action and rather than disposing of the monthly data, the study installations were requested to forward the data to the Materiel Readiness Support Activity (MRSA), Lexington, KY for computerizing. Since past data at these installations had been disposed of, except for the last 90 days which did not contain mileage entries, it was decided to initiate the life expectancy study with a new data collection effort beginning 1 October 1977. Another problem that had to be dealt with was that four of the installations (Fts. Benning, Knox, Lewis and Shafter) had their monthly maintenance data transcribed on magnetic tape while the other two installations (APG and WSMR) could only supply MRSA hand-written copies of the maintenance actions. Each of these two installations have been compiling about 800 maintenance actions a month.

MRSA acting as the central data receiving point not only has the large task of computerizing the hand-written APG and WSMR data but has to translate the tapes received from the other four installations

BELL

so that they could be used on their computer. It should be pointed out that each installation initially forwarded to MRSA an inventory tape of all administrative vehicles at each post so that the current mileages, ages, etc., of each vehicle (as of 1 October 1977) could be determined. During the course of the one year data collection period, MRSA will further review the data received from each installation for accuracy and completeness, assemble data by individual vehicles, sort maintenance actions into proper date and mileage sequences and consolidate data on magnetic tape.

Upon completion of the one year data collection effort, MRSA will forward the data to AMSAA for a determination if in fact, sufficient data has been collected for the life expectancy study. Assuming one year of data collection is sufficient, the data collected will form the basis of the determination of commercial administrative vehicle life expectancies and maintenance man-hour standards.

STUDY METHODOLOGY: The life expectancy of the vehicles being studied will be assessed by determining the mileage at which the average system cost per mile (costs associated with the acquisition, shipping and maintenance of the vehicle) is minimized (economic life). This cost analysis will be supplemented by an evaluation of the vehicle's Reliability, Availability and Maintainability (RAM) characteristics over the economic life span to establish if the vehicles life expectancy should be less than the economic life because of RAM considerations. In exercising this methodology, the procedure that will be employed will be to analyze the maintenance costs (scheduled and unscheduled) to determine how the costs are changing as the vehicles increase in mileage. This will result in the development of a maintenance cost function which will be combined with the vehicle investment costs to establish an average system cost function. The RAM characteristics will also be analyzed to determine how they are changing as the vehicles increase in mileage.

COST ANALYSIS: As noted above, the object of the cost analysis was to determine how the maintenance costs were varying as the vehicle mileage was increasing in order that the average system cost could be minimized. Thus, all the maintenance actions occurring with these vehicles will be costed in constant dollars (parts and labor) as a function of mileage. The analysis of this data will involve determining a continuous instantaneous maintenance cost curve (the instantaneous maintenance cost refers to the maintenance cost at a specific mileage). This curve will then be used to obtain the average system cost curve (the system cost refers to all costs associated with the procurement, shipment, and maintenance of a vehicle including such costs as the vehicle's acquisition price, administrative

BELL

expenses sustained, tooling costs, first and second destination charges, if any and maintenance costs). From the average system cost curve, the mileage at which the average system cost is at a minimum can be determined which represents the point where the overall average cost to the Army to procure, ship, and maintain the vehicle fleet is at a minimum. An example of the results of this type analysis is presented on Figure 1. As observed on the figure, the average system cost for the 2-1/2 ton truck was indicated to reach a minimum at 60,600 miles.

RAM ANALYSIS:

Unscheduled Maintenance Action Analysis. As noted above, data on maintenance actions (scheduled and unscheduled) is being collected during this study effort. This data is not being separated into those unscheduled maintenance actions resulting from a failure versus those actions not associated with a reliability failure. As a result, a reliability analysis based on the occurrence of unscheduled maintenance actions will be carried out rather than the usual reliability analysis based on failures. It should be noted, however, that a reliability analysis based on unscheduled maintenance actions provides a lower limit on a reliability failure analysis because if all unscheduled maintenance actions were in fact failures then the two analyses would be the same.

In analyzing the unscheduled maintenance actions, it is planned to develop a system Weibull failure rate function, i.e.,

$$r(t) = \lambda \beta t^{\beta-1} \quad t > 0, \lambda > 0, \beta > 0$$

where t = mileage on vehicle
 λ = scale parameter
 β = shape parameter

This function assumes that the probability that a vehicle will have an unscheduled maintenance action at mileage t is proportional to $r(t)$ and independent of the unscheduled maintenance action history of the system prior to t . This definition differs from the usual definition which states that the probability of an unscheduled maintenance action at mileage t is also proportional to $r(t)$ but conditioned on no unscheduled maintenance actions prior to t . The former definition applies to repairable systems whereas the latter definition does not. From this function, the probability that a vehicle with mileage t will complete an additional s miles without undergoing an unscheduled maintenance action (as determined by a non-homogeneous Poisson process) is

$$P(s/t) = e^{-\lambda(t+s)^{\beta}} + \lambda t^{\beta}$$

From this analysis, the probability of completing 75 miles without an unscheduled maintenance action as a vehicle is increasing in mileage through its economic life indication will be determined. An example of the results of this type analysis is shown on Figure 2.

Inherent Readiness Analysis. As with a reliability failure analysis, the determination of availability is normally based on failure data. For example, Inherent Availability (A_i) is normally defined as:

$$A_i = \frac{MTBF}{MTBF + MTTR}$$

where MTBF is the mean-time-between-failures and MTTR is the mean-time-to-repair.

As noted above, unscheduled maintenance actions rather than failure data will be available. Further, the data will provide information on the mean-man-hours-to-repair rather than the mean-time-to-repair. The mean-time-to-repair for a particular maintenance action could be less than the man-hours involved if two or more mechanics worked on a particular maintenance action. To utilize this data, however, to obtain an estimate of an availability statistic, one can determine the probability of a vehicle not undergoing active repair due to any unscheduled maintenance action when called upon to operate at a random point in time (Inherent Readiness) and this is given by the following expression:

$$R_i = \frac{MTBUMA}{MTBUMA + MMHTR}$$

where MTBUMA is the mean-time-between-unscheduled-maintenance-actions and MMHTR is the mean-man-hours-to-repair. It should be noted that the Inherent Readiness parameter is a lower bound on an Inherent Availability value, i.e., if all unscheduled maintenance actions were reliability failures and if no more than one mechanic ever worked on a maintenance action then the mean-man-hours-to-repair would be equivalent to the mean-time-to-repair and $R_i = A_i$. The results of this analysis will thus be used to determine if any degradation is occurring in the Inherent Readiness parameter as the vehicles are increasing in mileage through the economic life indication.

Maintainability Analysis. The object of this analysis will be to determine if the man-hours required for maintenance is changing as the vehicles increased in mileage and to provide data for a review of the current maintenance man-hour standards. In addition, a parts replacement analysis will consist of the following: (1) major component replacements as a function of mileage (engine, axles, differential and transfer case), (2) high cost parts (in excess of \$100) replacements, (3) ten most frequently replaced parts and (4) determination of the frequency of replacements for all vehicle parts.

Study Milestones. The conduct of this study is expected to take nearly two years. The study was formally initiated in August 1977 and is expected to be completed in June 1979. The study is essentially divided into three phases: (1) study feasibility phase (August 1977 - September 1977), (2) data collection phase (October 1977 - October 1978) and (3) analysis phase (November 1977 - June 1979). A detailed milestone schedule is indicated below.

MILESTONES:

Aug 77	Initial In-Process Review of Study Plan
Aug/Sep 77	Discuss Content of Installation Monthly Maintenance Report and Feasibility of Transferring Data to MRSA with Fts. Knox and Benning.
Sep/Oct 77	Visit Additional Data Collection Sites (APG, WSMR and Fts. Lewis and Shafter) to Discuss Implementation of Data Collection Effort.
Oct 77	Initiate Data Collection Effort.
Apr 78	Review Initial 6 Months of Data Collection.
May/Jun 78	Revise Existing Computer Programs on Basis of Initial Data Review.
Jun/Jul 78	Visit Data Collection Sites to Review Data Collection Effort.
Sep 78	Complete Data Collection.
Oct 78	Receive Data Tapes from MRSA.
Nov 78	Initiate Analysis of Data.
Apr 79	Complete Analysis of Data
May/Jun 79	Prepare Interim Report on Study Findings and Provide Briefings as Required.

BELL

TABLE I
COMMERCIAL VEHICLES
(LIFE EXPECTANCY YEARS AND MILES*)

<u>Description</u>	<u>Years</u>	<u>Miles</u>
Ambulance, All	8	60,000
Sedans, All	6	72,000
Station Wagon	6	72,000
Bus, Body on Chassis (BOC) (up to 37 passengers)	8	84,000
Bus, Body on Chassis (over 37 passengers)	10	150,000
Bus, Integral	12	300,000
Truck, 1/4 thru 3/4 ton (under 7,000 GVW)	6	72,000
Truck and Truck Tractor 1 thru 2 ton (7,000 thru 18,999 GVW)	7	84,000
Truck and Truck Tractor 2-1/2 thru 4 ton (19,000 thru 23,999 GVW)	8	84,000
Truck and Truck Tractor 5 thru 10 ton (24,000 thru 39,999 GVW)	10	150,000
Truck and Truck Tractor 11 ton and over (40,000 GVW and up)	12	300,000
Trailers and Semi Trailers	15	--
Motorcycle	5	30,000
Scooter 3W - Package Del	5	15,000

*Years or miles indicated whichever occurs first.

BELL

TABLE II
MAINTENANCE AND SERVICE VEHICLES

<u>NOMENCLATURE</u>	<u>LIN</u>
Ser Pltf Trk Mtd 4 x 2	S80048
Ser Pltf Trk Mtd 24,000 GVW	S80068
Ser Pltf Trk Mtd 4 x 4 28,000 GVW	S80070
Ser Pltf Trk Mtd 34,500 GVW	S80078
Ser Pltf Trk Mtd 4 x 2 28,000 GVW	S80088
Ser Pltf Trk Mtd 6 x 4 30 Ft. H	S80108
Trk, FB Tilt Frame	X45187
Trk, Hopper Coal	X48792
Trk, Hopper Coal 10 Ton	X48799
Trk, Maint LC 4 x 4 14/21,000 GVW	X53366
Trk, Maint LC 4 x 2 28/36,000	X53371
Trk, Maint LC 4 x 4 28/36,000	X53376
Trk, Maint LC 4 x 4 24/26,000	X53400
Trk, Maint LC 6 x 6	X53402
Trk, Maint 6 x 4 34,500	X53406
Trk, Maint Tele 3/4 Ton	X53572
Trk, Maint 4 x 4 5/9000 GVW	X53790
Trk, Maint Utility Panel	X53848
Trk, Maint Utility 7/10,000	X53851
Trk, Maint Utility 14/21,000	X53856
Trk, Mat'l Hdlg 21,000 GVW	X54428
Trk, Mat'l Hdlg 24,000 GVW	X54433
Trk, Mat'l Hdlg 4 x 2 32,000 GVW	X54445
Trk, Mat'l Hdlg Hi Lift	X54448
Trk, Ref Col R/H	X55820
Trk, Ref Col 4 x 2 24,000	X55832
Trk, Ref Col 4 x 2 28,000	X55837
Trk, Ref Col Comp 6 x 4	X55839
Trk, Ref Col 6 x 4 39,500 GVW	X55842
Trk, Ref Col 6 x 4 51,000 GVW	X55847

TABLE III
MAINTENANCE MAN-HOUR STANDARDS FOR
COMMERCIAL DESIGN VEHICLES

<u>Identification Vehicle Group</u>	<u>Direct Man-Hour Input Standards Per 1000 Miles</u>	<u>Average Annual Mileage</u>
A. Sedans.	2.0	14,000
B. Bus, Body on Chassis (BOC) (Up to & Including 29 Passengers).	7.0	12,000
C. Bus, Body on Chassis (BOC) (30-37) Passengers)	9.0	12,000
D. Bus, All 38 Passengers and Up	10.0	12,000
E. Station Wagon	2.2	16,000
F. Ambulances.	6.0	8,000
G. Truck, 1/2 Ton Pickup	3.5	10,000
H. Truck, Carryall: Truck, Sedan or Panel Delivery Trucks, Other 1/4 Ton Through 3/4 Ton	3.5	10,000
I. Truck and Truck Tractor, 1 Ton. . . .	3.0	9,000
J. Truck and Truck Tractor, 1-1/2 Tons: Truck and Truck Tractor, 2 Tons	8.0	7,000
K. Truck and Truck Tractor, 2-1/2 Tons.	7.5	7,000
L. Truck and Truck Tractor, 3 Tons - 4 Tons.	6.5	7,000
M. Truck and Truck Tractor, 5 Tons - 10 Tons	10.0	7,000
N. Truck and Truck Tractor, 11 Tons and Over.	12.5	9,000

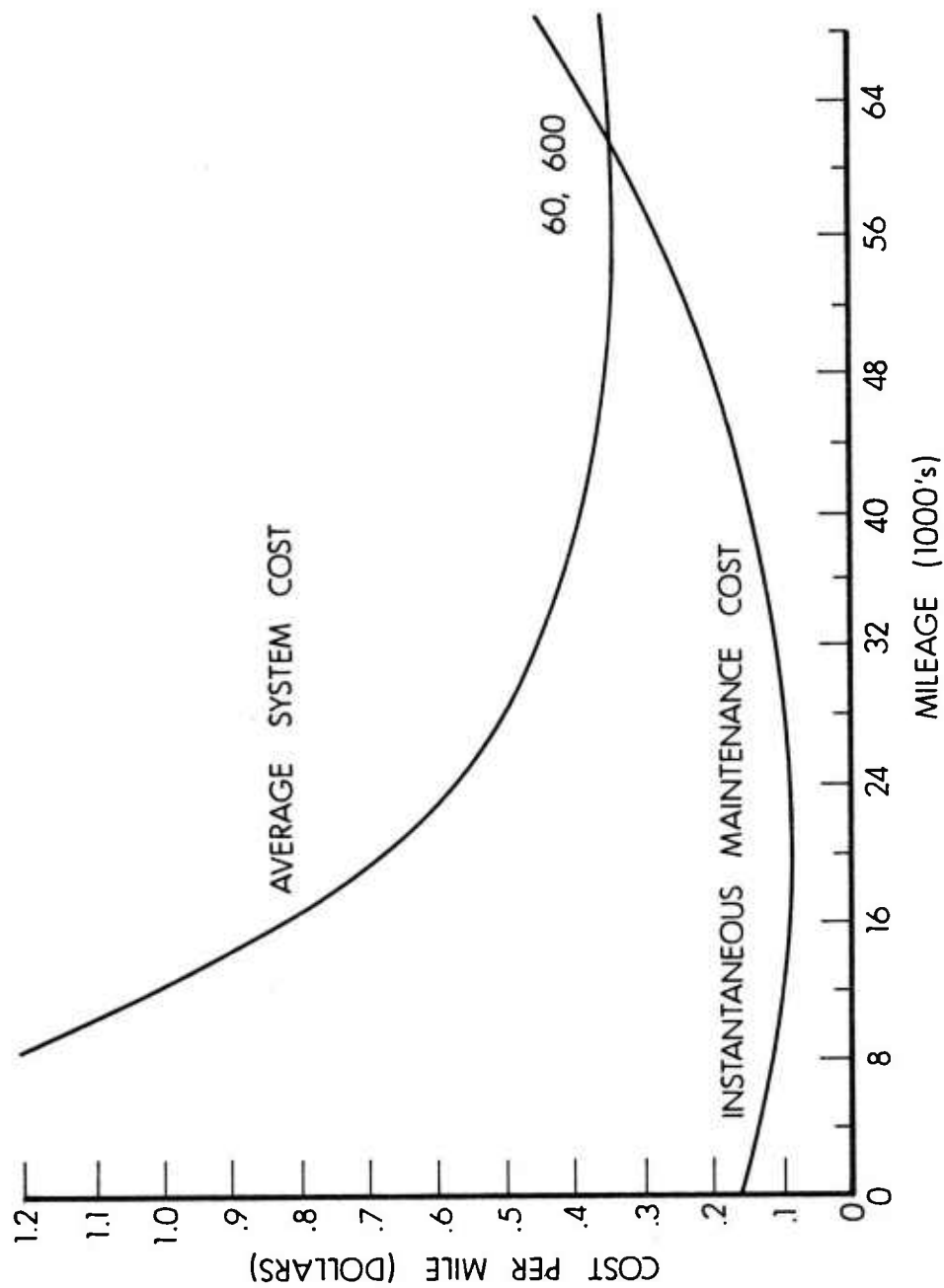


Figure 1. Cost Data - 2-1/2 Ton Truck.

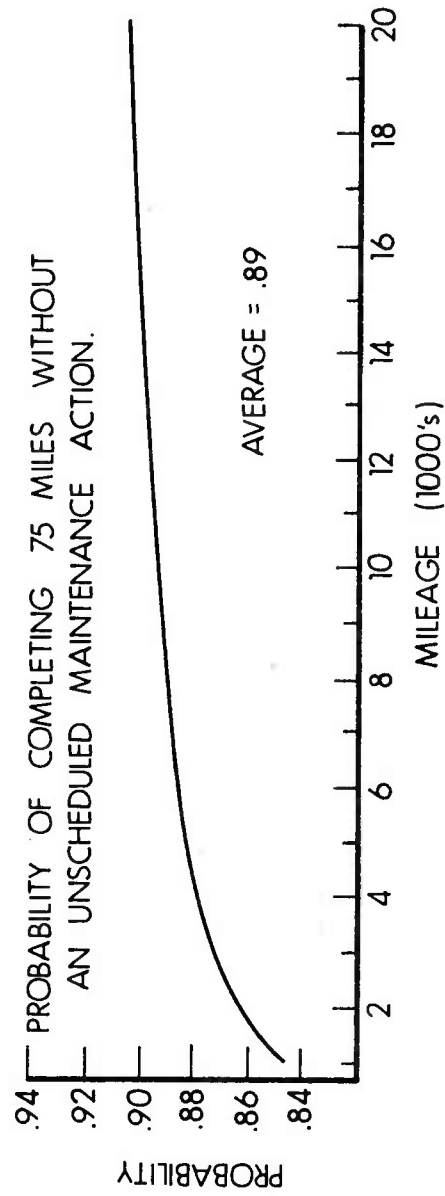


Figure 2. Reliability - 2-1/2 Ton Truck.

BIOMONITORING - A FINAL METHOD TO MEASURE
POLLUTION ABATEMENT

*EDWARD S. BENDER, MR.
PAUL F. ROBINSON, MR.
CHEMICAL SYSTEMS LABORATORY
U.S. ARMY ARMAMENT RESEARCH AND DEVELOPMENT COMMAND
ABERDEEN PROVING GROUND, MD 21010

The final goal of the Army's water pollution abatement program is to protect aquatic plants and animals by maintaining conditions favorable to life in the nation's waterways. Although most of this work involves monitoring chemical and physical characteristics of water, pollution is essentially a biological problem in that its primary effects are on living things. Therefore, it is appropriate to measure the success or failure of a pollution control system by monitoring the responses of aquatic plants or animals to the treated wastewater. Biological monitoring, or biomonitoring, of a treated effluent or waste stream thus provides the ultimate evaluation of pollution abatement. Legislators recognized the basic need for such monitoring by requiring in Public Law 92-500 (1972) that biomonitoring be conducted on all waste discharges. A biomonitoring system is now being tested at Radford Army Ammunition Plant (RAAP) in southwestern Virginia for just this purpose.

Biological monitoring is not a new concept. In the 19th century, coal miners used canaries to warn them when the air contained toxic gases. When the canary fainted, miners left the shaft for clean air above ground. Over twenty years ago, Rachel Carson in her book Silent Spring interpreted the death of birds and other wildlife as a symptom of the effects of pesticides on the environment. Each year the scientific literature contains numerous studies of organisms possessing either extreme tolerance or extreme sensitivity to environmental pollutants. Tolerant species are abundant in the

presence of a contaminant, while sensitive plants and animals may disappear even when only small quantities of a pollutant are present.

Any aquatic organism, whether a fish, a clam, or a worm, will respond in some measure to the collective effects of all factors in its environment. This response may be survival and prosperity or illness and eventual death. One response, as demonstrated by early physiologists, showed that fish will use more oxygen when they are under prolonged stress. Later it was found that the "breathing rate," that is, the rate at which water is pumped over the gills, changes when a fish is exposed to stress. We have combined the early physiological discoveries with contemporary technology to develop a system to measure that change.

The biomonitoring system consists of: a series of special tanks in which the "breathing rates" are measured; a wastewater distribution and dilution system; and a mini-computer to collect, store, and analyze the data. The entire system, with holding tanks and a diluter, is housed in a trailer, so that the system can be moved to any location.

Although the theory of the biomonitoring system is complex, the operation and maintenance are relatively simple. A portion of the plant effluent after final treatment is diverted to the water distribution system in the trailer. Water is delivered in a constant flow to eight test tanks each of which contains one fish. River water is delivered to four other tanks which serve as controls. The fish swims between two electrodes, and muscular contractions associated with "breathing" are measured by a change in potential, which is caused by a bioelectrical current from contractions of the jaw and opercular muscles. Before the testing begins, normal "breathing rates" are determined for each fish from measurements taken over a period of five days. During testing, "breathing rates" are determined for 15-minute intervals and each count is compared to normal "breathing rates" for that fish. When the fish are exposed to a sublethal toxic condition, the "breathing rate" changes. If six of the eight test fish have a significant increase in "breathing rate," an alarm is sounded.

The changes in "breathing rates" correlate directly to conditions which can occur when an acid spill, system upset, or other problem occurs on a manufacturing line. The alarm advises the plant manager that the quality of the waste stream has deteriorated and he must take appropriate action. Although the system cannot identify

the pollutant, it provides a warning more rapidly than conventional water quality monitoring and at a fraction of the cost.

Many Army production facilities have large containing ponds that can hold treated wastewaters for several days. After the alarm is sounded, the wastewater could be impounded in lieu of discharge. The pollutants can be identified by chemical analysis and additional treatment can be applied. Continuous information from the biomonitoring system provides a tool that enables management to recommend additional wastewater treatment when necessary.

In Europe and Africa, biomonitoring systems are being used to assay drinking water taken from large rivers polluted by industrial discharges. If fish, which are more sensitive to pollutants than man, show no signs of toxicity after exposure, the water is pumped into the drinking-water reservoirs.

Although the United States has no biomonitoring systems in practical use today, passage in 1977 of the Toxic Substance Control Act and Toxic Substances Act by the United States Congress should encourage their development. Under these laws all manufacturers will be required to demonstrate that their wastewaters do not have a toxic, mutagenic, or persistent effect upon the aquatic organisms in the receiving waters. Some industries are spending up to one million dollars to develop toxicological data for each compound present in their wastewater. The price for these data is high but, even so, the results may not apply to wastewaters of variable and complex composition such as those from ammunition plants. Therefore, a continuous evaluation is also needed. Ultimately, biomonitoring may become the final test of the adequacy of wastewater treatment and the method to safeguard environmental quality.

REAL-TIME AND MEMORY CORRELATION VIA
ACOUSTO-OPTIC PROCESSING (U)

*NORMAN J. BERG, PH.D, JOHN N. LEE, PH.D
BURTON J. UDELSON
US ARMY ELECTRONICS RESEARCH & DEVELOPMENT COMMAND
HARRY DIAMOND LABORATORIES
ADELPHI, MD 20783

The requirements of advanced radar systems, secure communications networks, and signal-warfare concepts place severe strains on the capabilities of presently available analog processing technologies. These applications require real-time processing with large bandwidth and dynamic range. The use of acousto-optic technology as an answer to these requirements appears very attractive. Three fundamental signal-processing schemes using the acousto-optic interaction have been investigated: i) real-time correlation and convolution, ii) Fourier and discrete Fourier transformation, and iii) programmable memory correlation. By combining previously known techniques with newly discovered phenomena, major advances in analog signal processing have been demonstrated. Time-bandwidth products in excess of 10,000 and linear dynamic ranges in excess of 50 dB have been achieved with real-time processors. Using the recently discovered acousto-photorefractive effect,⁽¹⁾ storage of surface-acoustic-wave (SAW) signals in lithium niobate (LiNbO_3) has been demonstrated. During the information storage time of up to several months, "live" signals can be acousto-optically correlated with the stored signals. This storage phenomenon can be used as the basis for a wide variety of new signal-processing architectures. A description of the experimental results for the real-time processors (correlators and Fourier transformers) and the memory correlator are given in sections II and III, respectively.

II. Real-Time Signal Processing Using Acousto-Optic Interactions
A. Background

The processing of signals using acousto-optic interactions is based on the diffraction of light by refractive index changes induced

in a piezoelectric material by a surface acoustic wave. Conservation of momentum and energy requires that the frequency of the light diffracted by the propagating surface acoustic wave be shifted by an amount equal to the acoustic frequency.⁽²⁾ The acousto-optic interaction is normally in the Bragg regime, where the incident light beam is incident at an angle, θ_B , to the SAW front. The sine of θ_B is given by

$$\sin \theta_B = \frac{\lambda_o}{2n\lambda_a} , \quad (1)$$

where λ_o is the wavelength of light in free space, λ_a is the acoustic wavelength in the material, and n is the index of refraction of the material. The Bragg regime permits maximum interaction efficiency, since constructive interference occurs only for the first order diffraction mode, all other modes being suppressed.

For Bragg interaction, the power ratio of diffracted light, I_1 , to incident light, I_o is ⁽²⁾

$$\frac{I_1}{I_o} = \sin^2 (1.4\ell\sqrt{M_w P_s}) , \quad (2)$$

where M_w is a figure of merit relative to the value for water, ℓ is the interaction length, and P is the acoustic power density. For sufficiently small P , a linear relationship exists between I_1 and P .

The method for achieving convolution via the acousto-optic interaction is indicated in Fig. 1. The amplitude of the light beam that has been diffracted by both acoustic waves, $A(t) \cos \omega_a t$ and $B(t) \cos \omega_a t$, is proportional to

$$[A(t) B(t) \cos (\omega_o - 2 \omega_a)t] , \quad (3)$$

where ω_o is the light frequency. This doubly diffracted beam is collinear with the undiffracted light of intensity I_o , and the two beams are heterodyned in a nonlinear mixer diode. The resultant output voltage is proportional to the product of eq. (3) with $\cos (\omega_o)$, and the component at the acoustic frequency ($2\omega_a$) is:

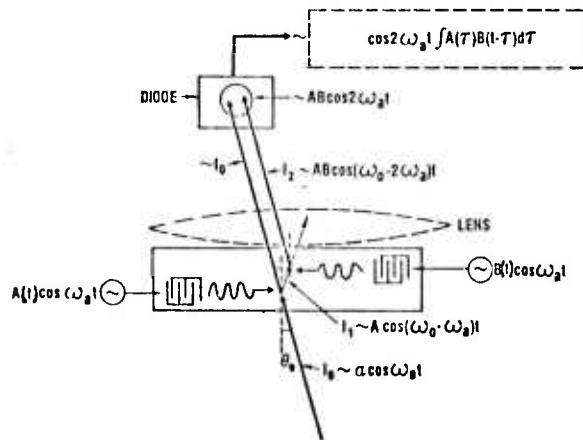


Fig. 1. Waveform Convolution Using Acousto-Optic Bragg Interaction

$$V_{\text{out}}(t) \sim A(t) B(t) \cos(2\omega_a) . \quad (4)$$

Eq. (4) gives the voltage due to a single light ray. The interaction region between transducers is fully illuminated, and the lens shown in Fig. 1 focuses the entire light beam onto the photo-diode. Recalling that the two signals are physically passing each other, it can be shown that the output voltage as a function of time is given by

$$V_o(t) = \int_{-\infty}^{\infty} A(\tau) B(t - \tau) d\tau , \quad (5)$$

which is equivalent to the convolution of $A(t)$ and $B(t)$.⁽³⁾ Further, the frequency of the output waveform is twice that of the input waveforms, and the output bandwidth is correspondingly twice the input bandwidth.

One of the primary advantages of using the acousto-optic interaction to obtain convolution (or correlation) as described above is that all nonlinear mixing occurs in the mixer diode. Under these conditions, the piezoelectric material is nondispersive, and the interaction is approximately linear according to Eq. (2). It will also be shown later that large dynamic ranges of signal amplitude are feasible.

The scheme described above uses a single SAW delay line to perform real-time convolution of two rf signals. Correlation is obtained directly if the two signals are symmetric waveforms, since in this case correlation and convolution are equivalent. Correlation of asymmetric waveforms in this scheme may be obtained by time inverting one of the rf inputs. However, the time-inversion process involves additional electronic complexity and problems of the quality of operational performance. As an alternative approach, a "two-crystal" correlator has been designed that uses two separate acoustic delay lines, one of bismuth germanium oxide (BGO) and the other of LiNbO_3 , placed side by side, as shown in Fig. 2. The incident light passes through the interaction regions of both lines. Since the acoustic wave velocity in BGO is slower (by about half) than the velocity in LiNbO_3 , an rf signal introduced into LiNbO_3 overtakes an rf signal that had been just previously introduced in the BGO. Both the BGO and the LiNbO_3 signals are launched in the same direction, thereby obviating time inverting one signal. If the LiNbO_3 signal entirely passes the BGO signal, the output at the photodetector corresponds to the correlation of the two waveforms. The maximum pulsewidth that can be completely correlated is equal to one third of the interaction time (or length) of the LiNbO_3 delay line. It is important to note that the two signals being correlated acousto-optically are in the form of surface acoustic waves in the LiNbO_3 and BGO. To obtain maximum correlation of two propagating signals, their wavelengths and spatial

extent must be the same in both crystals. To accomplish this, the input signal frequencies must be adjusted to be proportional to the ratio of the two SAW velocities, and the input pulse widths must be inversely proportional to this ratio. The correct pulsewidth ratios may be obtained by an auxiliary processing step using the same two-crystal acousto-optic setup.⁽³⁾

B. Experiment

A photograph of the setup used for our experiments is shown in Fig. 3. The krypton (Kr) laser at the far left has an output of 400 mw at 647.1 nm. Following the laser are lenses and a cylindrical mirror (at the far right) which transforms the circular laser beam into a uniform intensity sheet beam 15-cm wide by 60- μ m high at the position where the LiNbO₃ or BGO SAW delay lines are mounted. Following the SAW lines, a cylindrical lens and parabolic mirror are used to focus the light onto a PIN photodiode having less than 1-ns response time. The photodiode signal is then amplified by wide-band amplifiers.

Fig. 4 shows a closeup of a LiNbO₃ and a BGO line placed side-by-side in the "two-crystal" correlator configuration (the BGO is

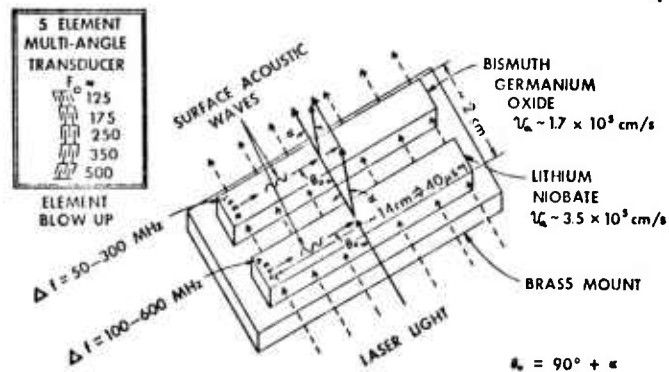


Fig. 2. "Two-Crystal" Waveform Correlation by Using Acousto-Optic Interaction in Adjacent BGO and LiNbO₃ Delay Lines.

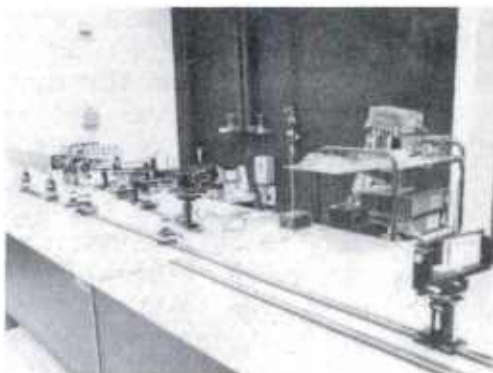


Fig. 3. Experimental Setup for Real-Time Acousto-Optic Signal Processing.

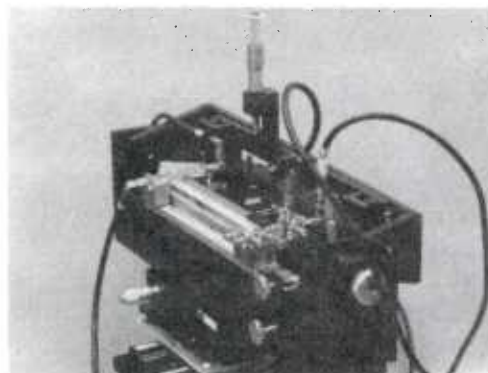


Fig. 4. LiNbO₃ and BGO Delay Lines Used for Acousto-Optic Signal Processing.

the darker crystal). The dimensions of each crystal were 15 x 1 x 1 cm. The LiNbO_3 used was Y-cut, Z-propagating, and the BGO used was 001-cut, 110-propagating. In both crystals, the SAW propagation direction was in the long dimension. To insure maximum acousto-optic coupling, it was necessary to obtain extremely good surface figure and polish on the crystals. The top surfaces of the crystals were polished to a figure of about one wave ($\lambda \sim 0.55 \mu\text{m}$) over the entire 15 x 1 cm area, with no visible defect marks under 100X magnification. Both sides of the crystals upon which the laser light impinges were polished so as to have no more than 0.5° deviation from perpendicularity with the top surface. Each of the two long edges were made as chip-free as possible. Chip length was no more than 25 μm , and the total length occupied by chips was less than 2% of the 15-cm length. The impinging Kr laser light beam was perpendicular to and filled the long dimension of the lines. The actual laser light intensity through the region of SAW propagation was approximately 10 mW.

Interdigitated finger transducers were used to convert the rf signals into SAW's. To achieve a large interaction bandwidth, a series of four transducers was used for this conversion. The center frequencies of these transducers in LiNbO_3 were 250, 355, 475, and 612 MHz. The center frequencies in BGO were lower by a factor equal to the ratio of the respective SAW velocities in LiNbO_3 and BGO (~ 2.075). The transducers were tilted slightly with respect to each other according to a scheme by Tsai.⁽⁴⁾ In this manner, the Bragg condition was satisfied simultaneously for all four center frequencies, thereby permitting a large instantaneous bandwidth. In addition, the transducers were translated with respect to each other in the direction of wave propagation to provide proper phase matching at the crossover frequencies.

Each device was operated initially as a convolver. For BGO and LiNbO_3 , input bandwidths of 140 and 250 MHz, respectively, were obtained. For both devices, the full design bandwidth was not achieved, because of very high insertion losses for the highest frequency transducer in each device and because of the response roll-off of the photodetector and amplifiers above 800 MHz. Interaction lengths of 65 and 40 μs , respectively, were obtained for the BGO and LiNbO_3 . These lengths corresponded exactly to the physical interaction lengths available in each device. The time-bandwidth product of these devices used as convolvers, obtained by multiplying the bandwidth with the interaction time, was 9100 for the BGO device and 10,000 for the LiNbO_3 device.

The acousto-optic correlator shown in Fig. 4 has been used successfully to correlate several waveforms of interest in radar applications. Fig. 5A shows the output obtained by correlating two mono-frequency square-wave pulses for a low-frequency correlator (LiNbO_3

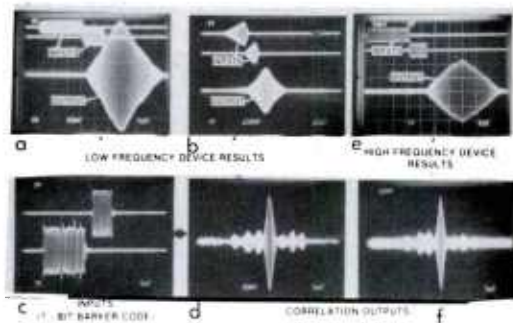


Fig. 5. Real-Time Correlation Output Obtained with Two Rectangular Waveforms (a) and (e) and with Seven-Bit Barker Code (d) and (f).

at 10 MHz and BGO at ~5 MHz), and Fig. 5E shows the corresponding result for the high-frequency correlator. The output pulsewidth is about four times longer than the short (LiNbO_3) input pulse, as expected. Fig. 5D and 5F show the results, in the low- and high-frequency devices respectively, of correlating the seven-bit Barker code⁽⁵⁾ shown in Fig. 5C. There is a strong central peak with three sidelobes on either side, as expected. The peak-to-sidelobe ratio appears to be just slightly greater than the theoretically expected 2.7:1 ratio.

The results of measurements of the dynamic range of both the low- and high-frequency acousto-optic correlators are shown in Fig. 6. Because of much lower transducer insertion loss at low frequencies, the low-frequency correlator exhibited a greater dynamic range. Extrapolating to a reference 1-MHz bandwidth, a dynamic range of about 84 dB is obtained. The output is linear over this entire range for variations of either the BGO or the LiNbO_3 input.

C. Applications

1) Signal-to-Noise Ratio Enhancement

One of the major functions of an acousto-optic correlator in radar applications would be to extract signals submerged in noise. It can be shown⁽⁶⁾ that the most effective means of extracting a signal from noise is by using a matched filter. A correlator is by definition an adaptive matched filter, i.e., its response is the equivalent to a matched-filter response for any waveform corresponding to the reference waveform. The time-bandwidth product of the filter is a

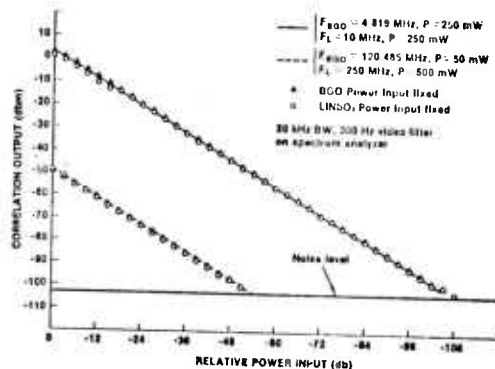


Fig. 6. Correlation Output versus Input Power, Demonstrating Linear Dynamic Range of Acousto-Optic Correlator.

direct measure of the signal-to-noise (S/N) enhancement. Hence, the relatively large time-bandwidth product, as well as the range of the instantaneous bandwidth, of the real-time acousto-optic processors described above make them very attractive candidates for advanced radar processing. Examples of the ability of the acousto-optic correlator to extract signals buried in noise are shown in Figs. 7 to 10. In Fig. 7 the output of the correlator is shown both for a noise-free and a noisy (S/N = -10 dB) monotone input signal. Figure 8 is a quantitative plot of output S/N versus input S/N for monotone inputs.

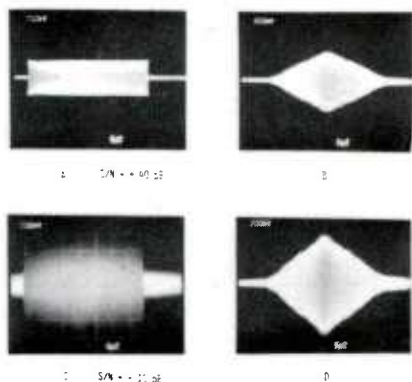


Fig. 7. Convolution Output with Noise-free Input (B and A) and with Noise-Degraded Input (D and C).

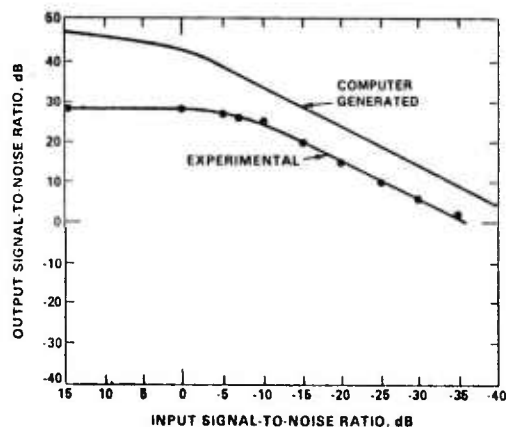


Fig. 8. Signal-to-Noise Enhancement as Obtained with a Digital Computer, Compared with Acousto-Optic Experimental Results.

The computer-generated result is based on an equivalent time-bandwidth product of 30,000, whereas the particular acousto-optic device had a time-bandwidth product of 7,500. Fig. 9 illustrates qualitatively the S/N enhancement for 100-MHz bandwidth, 20- μ s wide, linear FM chirps. Fig. 10 illustrates quantitatively the enhancement for these broadband signals. In all cases, a S/N enhancement on the order of 30 dB is possible.

2) Fourier Transformation

The ability to perform real-time Fourier analysis of rf signals is important in both communication and signal-warfare systems. Fourier transforms have been performed acousto-optically using the "chirp transform" method⁽⁷⁾, and a signal-processing architecture for performing discrete Fourier transforms acousto-optically has been devised.

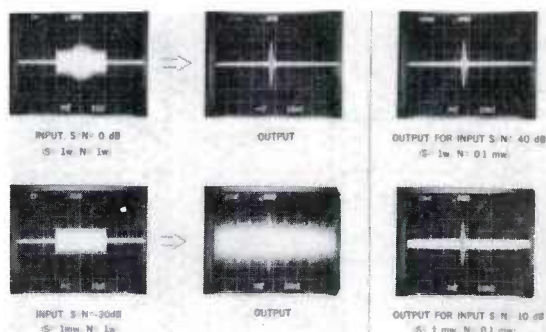


Fig. 9. Signal-to-Noise Enhancement with 100-MHz Bandwidth Linear FM Chirp Signals.

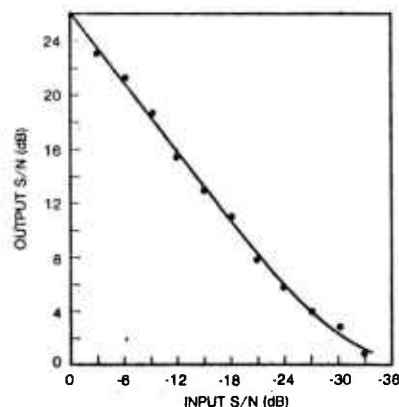


Fig. 10. Output Signal-to-Noise versus Input Signal-to-Noise for 100-MHz Bandwidth Signals.

a. Real-Time Chirp Transform

The Fourier Transform of a function $x(t)$ is defined as

$$X(f) = \int_{-\infty}^{\infty} e^{-i2\pi ft} x(t) dt . \quad (6)$$

If the substitution $-2ft = (f - t)^2 - f^2 - t^2$ is made in eq. (6) and the terms rearranged, the following is obtained:

$$X(f) = e^{-i\pi f^2} \int_{-\infty}^{\infty} [e^{-i\pi t^2} x(t)] [e^{i\pi (f-t)^2}] dt , \quad (7)$$

where the first bracketed term in the integral is identified as a pre-multiplication of the function $x(t)$ by a chirp. This premultiplied term is then correlated with a chirp. Finally, the correlation result is postmultiplied by a chirp. This entire sequence is referred to as the "chirp transform" implementation of the Fourier transform.

The acousto-optic devices described in section IIB were used to perform the correlation portion of this transform, and the required chirps were obtained from an impulse-excited, reflective-array-compressor SAW delay line. The results of this implementation are summarized in Figs. 11 to 14. The measured bandwidth of the transformer was about 80 MHz, as illustrated in Fig. 11 where the transformed output of a pulse containing three discrete frequencies is

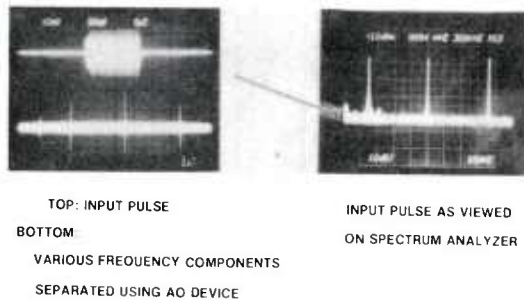


Fig. 11. Fourier Transform (lower trace) of Gated Signal (top trace) Containing Discrete Frequencies 18, 54, and 90 MHz.

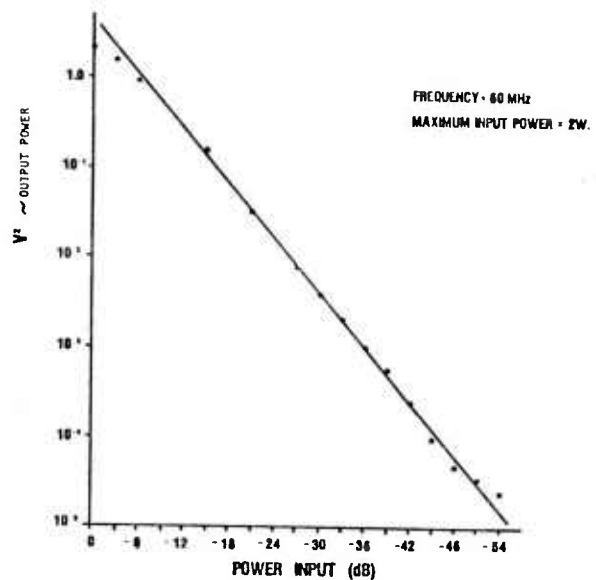


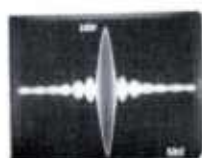
Fig. 12. Dynamic Range of the Acousto-Optic Fourier Transform Output.

shown. The dynamic range for the cw-gated signal exceeds 50 dB and is shown in Fig. 12. The structure of the sidelobes is strongly dependent on the spatial distribution of the light intensity of the incident laser beam. If a uniform intensity distribution is used, then a $(\sin x)/x$ sidelobe pattern results (Fig. 13). When illuminated with a Gaussian profile, the sidelobes were suppressed to about 27 dB below the main peak, as is shown in Fig. 14. The sidelobe amplitude is compared with a second, nearby, signal which is 25 dB down from the main signal.

These results are generally in accord with the best reported results obtained by using a pure SAW filter approach to achieve chirp transformation.⁽⁸⁾ The advantages of the acousto-optic approach, however, are twofold: i) the ease in applying weighting functions for sidelobe suppression, and ii) the capability of arbitrarily varying the chirp rate, which aids in locating signals.

b. Discrete Fourier Transform

An alternative to the chirp-transform architecture has been proposed which is especially useful when performing the discrete Fourier transform (DFT). The DFT is defined as



(EXPANDED VIEW)

OUTPUT FOR UNIFORM ILLUMINATION



(EXPANDED VIEW)

OUTPUT FOR TWO ADJACENT FREQUENCIES

HIGHER FREQUENCY ARROW 25 dB LESS INPUT POWER

Fig. 13. Sidelobe pattern ($\frac{\sin x}{x}$) of the Acousto-Optic Fourier Transform Obtained Using a Uniform-Intensity Laser-Beam Profile.

Fig. 14. Suppressed Sidelobe Pattern of the Acousto-Optic Fourier Transform Obtained Using a Gaussian Intensity, Laser Beam Profile; Pattern Compared with Higher Frequency Signal (arrow) at -25 dB.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi nk/N} . \quad (8)$$

If the substitution

$$nk = \frac{1}{4} \{ (k+n)^2 - (k-n)^2 \} \quad (9)$$

is made into eq. (8), the following results:

$$X_k = \sum_{n=0}^{N-1} x_n e^{\frac{-i\pi(k+n)^2}{2N}} e^{\frac{i\pi(k-n)^2}{2N}} . \quad (10)$$

The exponential factors in eq. (10) can be interpreted as two (chirp) waves propagating in opposite directions relative to the function to be transformed. This transform architecture is referred to as the triple-product convolver⁽⁹⁾ and can easily be realized using acousto-optic techniques. This architecture is illustrated in Fig. 15. The large sheet beam used in the previous implementations of the acousto-optic convolver is now replaced by a series of very narrow laser beams

where the amplitude of each laser beam is modulated (e.g., by an electro-optic crystal) in accord with the input digital data stream which is to be transformed. This structure is being constructed at the Harry Diamond Laboratories.

III. Acousto-Optic Memory Correlator

A. Acousto-Photorefractive Effect

A new photorefractive effect has been discovered whereby an index-of-refraction change (δn) in LiNbO_3 results from the interaction of high-intensity, short-duration laser pulses with propagating surface acoustic waves. This acousto-photorefractive effect can be used to implement an acousto-optic memory correlator. This effect is phenomenologically similar to a nonlinear "two-photon" photorefractive effect reported previously,⁽¹⁰⁾ but is not a simple extension of that effect. In the present experiment, an SAW pattern was stored as δn . The δn was found to be proportional to the rf amplitude, and increased sublinearly (0.7 power) with the number of laser pulses and as the 1.3 power of the incident laser energy density. The decay time varied from a few hours, when only green illumination (530 nm) was used, to several weeks, when combined green and infrared (IR) illumination (1060 nm) were used. The δn can be erased by exposure to ultraviolet radiation or by annealing at 250°C.

An acousto-optic memory correlator has been constructed wherein both the stored δn and the δn produced by a "live" propagating acoustic wave simultaneously modulate a low-power cw laser beam. The resultant detected signal is proportional to the correlation integral. The memory correlator operated at a center frequency of 10 MHz with a 1-MHz bandwidth. A large variety of complex signals, such as chirps and Barker codes, was stored and subsequently correlated. The stored signal strength was about 30 dB below that of the original live signal, and successful correlation with a live signal was achieved several weeks after storage.

A sketch of the setup used for our experiments is shown in Fig. 16. During the writing (storage) phase, illustrated at the left of the figure, a neodymium: yttrium aluminum garnet (Nd:YAG) laser was used in which the output was a single-mode (TEM_{00}) Gaussian pulse

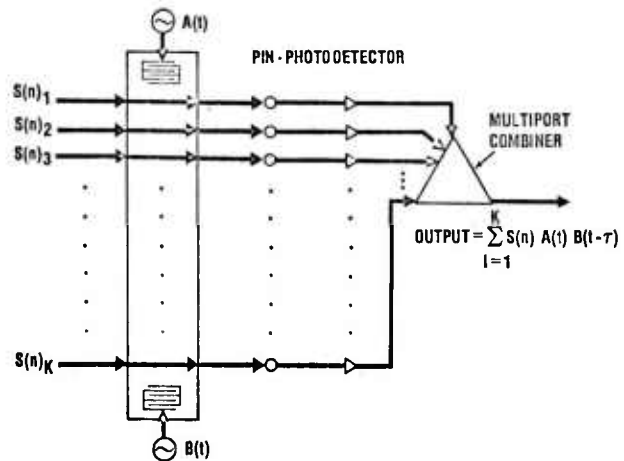


Fig. 15. Architecture for an Acousto-Optic Triple-Product Convolver.

having a 12-ns pulsewidth and an energy of 100 mJ/pulse. The 1060-nm IR output was converted to green (530 nm) by using either a CD*A temperature-tuned or a KDP angle-tuned doubler. By means of appropriate lenses, the resultant light was converted into a sheet beam about 1 cm wide and 250 μ m thick. This sheet beam was then focused along the top edge of the side of the LiNbO₃ crystal, so that it passed through the region of SAW propagation. Since many pulses were often used for storing a signal, the

launching of the acoustic waves was synchronized with the writing (Nd:YAG) laser pulse to within 3 ns of uncertainty. The synchronization was done to assure that the acoustic wave was in the same position during each laser pulse. The light beams used for both writing and reading were incident along the x-axis, perpendicular to the z-direction (optical c-axis) of acoustic propagation.

The reading phase, illustrated at the right of Fig. 16, was accomplished with a 10-mW cw He-Ne laser. Since the index-of-refraction pattern stored by the acousto-photorefractive effect was only a small ac component riding on top of a much larger ($\sim 10^3$) dc index change, the ac signal could not be measured by the means used conventionally for observing δn . Two methods were used for observing these changes. The presence of an ac index-of-refraction pattern could be verified qualitatively by observing the diffraction of light transmitted through this newly formed diffraction grating. Quantitative information was obtained by propagating a live acoustic wave in the crystal during illumination with the low-intensity cw reading laser. Both the stored and live signals modulated the light beam via the Bragg acousto-optic interaction. The doubly modulated light was focused onto a "square-law" detector diode the output of which is proportional to the correlation of the two acoustic signals.⁽³⁾

In the experiments described here, the δn was obtained in plates of Y-cut Z-propagating single-crystal LiNbO₃ which had been fabricated into SAW delay lines by the deposition of interdigitated-finger transducers at each end. The spacing between transducers was 7 cm, corresponding to a delay of about 20 μ s. The acoustic aperture of the

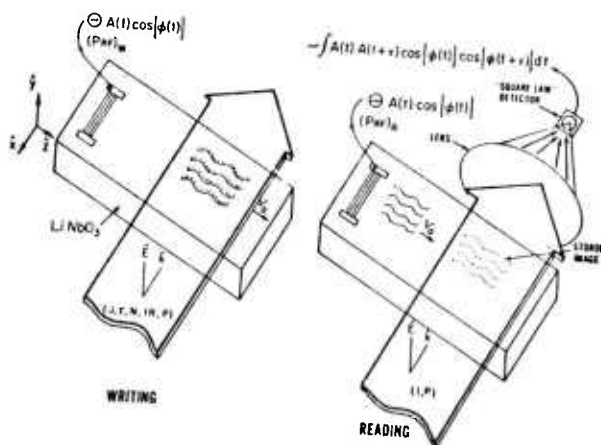


Fig. 16. Acousto-Optic Correlator in Both Writing (left) and Reading (right) Modes.

transducers was 1.5 cm, and their center frequency was 10 MHz with a 1-MHz bandwidth.

The variation of correlation output power ($\sim \delta n^2$)⁽²⁾ was measured as a function of number of laser pulses (N), incident laser energy density (J), and rf input power during storage (P). These results have been summarized into the graph presented as Fig. 17. The axes are the number of laser pulses (N) and laser energy density (J/cm²) per pulse, with lines of constant insertion loss plotted in the figure. Insertion loss is defined here with respect to the known output obtained from the "live" convolution of two 10-mW signals in an acousto-optic convolver. It should be observed from Fig. 17 that for a single laser storage pulse of 10³ mJ/cm², which is still below the damage threshold for LiNbO₃, the insertion loss is ~30 dB. Fig. 18

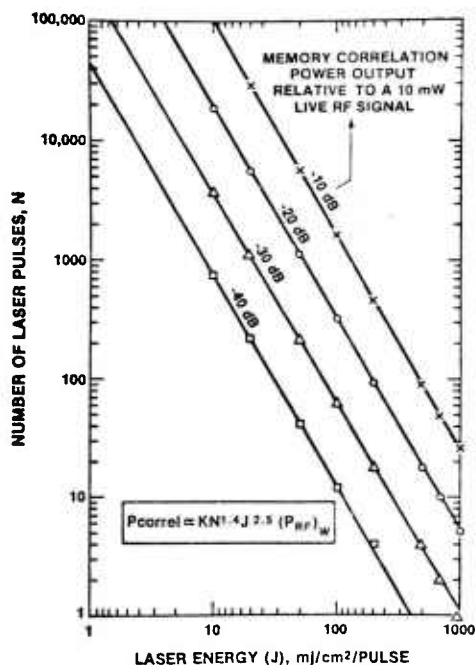


Fig. 17. Curves of Constant Output Power from Memory Correlator as Function of Number of Laser Pulses and of Laser Energy Density.

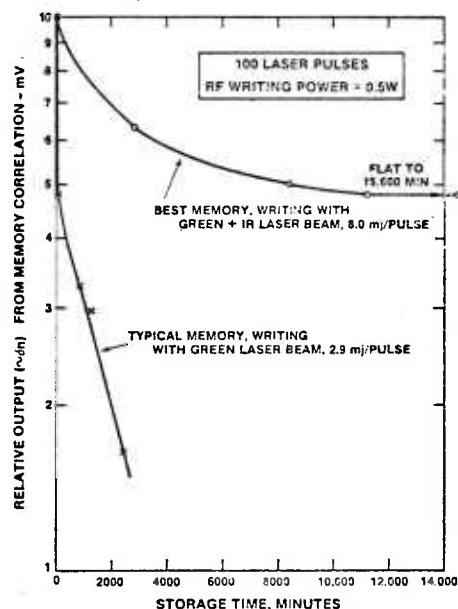


Fig. 18. Relative Correlator Output Power versus Storage Time for Two Cases, Green Writing Beam Alone and Combined Green and IR Writing Beam.

shows a plot of the decay of δn at room temperature as a function of time after storage for two cases: (1) green illumination only and (2) green with IR. The two cases show different time constants, the time constants for the second case being several orders of magnitude

larger than those for the first. We believe that the relatively flat regime obtained with the second case is probably the long-term natural decay time for charge in a relatively pure dielectric.⁽¹¹⁾

The physical basis of the acousto-photorefractive effect appears to be the movement of charge which results in a change in the local index of refraction via the electro-optic effect. Because of the magnitude of the effect, we speculate that the electric field accompanying the acoustic wave in the piezoelectric materials, rather than density changes alone, may be responsible for this movement of charge.

B. Applications

The acousto-photorefractive effect may be used to construct an acousto-optic memory correlator that is generically similar to an acousto-electric memory correlator reported previously.⁽¹²⁾ Successful correlation of live signals consisting of relatively complex waveforms has been accomplished with waveforms that were stored for as long as several weeks. These data are summarized in Figs. 19 and 20.

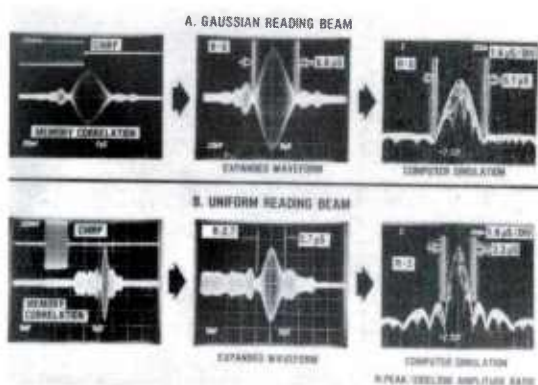


Fig. 19. Output Waveforms of Memory Correlator and Comparison with Computer Simulation for Linear FM Chirp by Using Both Gaussian and Uniform Reading Beams.

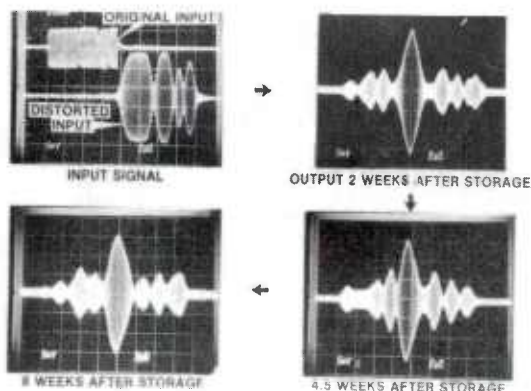


Fig. 20. Changes with Time of Output Waveforms of Memory Correlator for Seven-Bit Barker Code.

Fig. 19 illustrates the output obtained by correlating a live 10 MHz, 10- μ s wide rectangular pulse having a linear FM chirp of about 1 MHz against a similar previously stored signal. In Fig. 19A, the reading beam used had a Gaussian intensity distribution across the length of the stored image; Fig. 19B pictures were obtained for a uniform beam intensity. Comparison of the expanded output pulses with the computer simulations (shown at the right of Fig. 19) for both types of beam distribution demonstrates good agreement for both the null-to-null spacings and the peak-to-sidelobe ratio (R).

Fig. 20 indicates the observed changes of the output of the seven-bit Barker code correlation as a function of time. The slow

degradation of the stored pattern manifested itself in a distortion of the sidelobes; however, the main correlation peak remained relatively unchanged.

IV. Conclusions

Two new types of acousto-optic signal processors have been demonstrated. The first is a real-time correlator having a large time-bandwidth product ($\sim 3,000$). This device has been used to perform Fourier transformation (>50 dB dynamic range), as well as signal-to-noise enhancement (>30 dB). The second type of processor is a memory correlator in which signals may be stored for long periods of time, and in which live signals may subsequently be correlated with the stored signals. The feasibility of storing and correlating with complex, phase-coded waveforms for up to two months has been demonstrated. Both of these signal processors should find many applications in radar, communications, and signal-warfare systems.

References

1. N. J. Berg, B. J. Udelson, and J. N. Lee, Appl. Phys. Lett. 31 (1977), 555.
2. R. Adler, IEEE Spectrum, 4 (1967), 42.
3. N. J. Berg, B. J. Udelson, J. N. Lee, and E. Katzen, Appl. Phys. Lett. 32 (1978), 85.
4. T. S. Tsai, L. T. Nguyen, S. K. Yau, and M. A. Alhaider, Appl. Phys. Lett. 26 (1975), 140.
5. C. E. Cook and M. Bernfeld, Radar Signals, Academic Press, New York (1967), 245-247.
6. M. I. Skolnik, Introduction to Radar Systems, McGraw-Hill, New York (1962), Ch. 9.
7. L. R. Rabiner, R. W. Schafer, and C. M. Rader, IEEE Transactions on Audio and Electroacoustics AU17, No. 2 (1969), 86.
8. H. M. Gerard, P. S. Yao and O. W. Otto, Proc. IEEE Ultrasonics Symp. (1977), 77CH1264-1SU, p. 947.
9. J. M. Speiser and H. J. Whitehouse, Naval Ocean System Center, San Diego, CA, NUCTN 1355R (1974).
10. D. von Linde, A. M. Glass, and K. F. Rodgers, Appl. Phys. Lett. 25 (1974), 155.
11. J. P. Huignard, F. Micheron, and E. Spitz, Optical Systems for Photosensitive Materials for Information Storage, North-Holland, Amsterdam (1976), Ch. 16.
12. A. Bers and J. H. Cafarella, Appl. Phys. Lett. 25 (1974), 133.

GATED TV INSTRUMENTATION SYSTEM (U)

*C. DAVID BROWN, CPT, SC
US ARMY ABERDEEN PROVING GROUND
ABERDEEN PROVING GROUND, MD 21005

Weapon-pointing data for the hit-probability analysis of stabilized vehicle-mounted weapon systems have long been obtained by high-speed photographic techniques, which do not provide a real-time analysis capability, as film must be processed and requires excessive data-reduction time (film reading). Infrared (ir) tracking systems, also in use, provide no image and therefore are extremely difficult to set up and align. The Gated TV (GTV) system (Figure 1) consisting of a video camera, a tracker error processor (tracker) (TEP), a video monitor, and a digital or analog recorder, will derive weapon-pointing errors as digital numbers or analog voltages proportional to angular elevation-and-azimuth errors in real time. This output can be recorded or presented directly to automatic data-processing equipment (ADPE) to obtain weapon-pointing data as a function of time. This instrumentation provides the basis for measuring target position, lead angle, and target angular rate. Utilizing such information, stabilization and tracking studies of main and secondary weapon systems are readily accomplished.

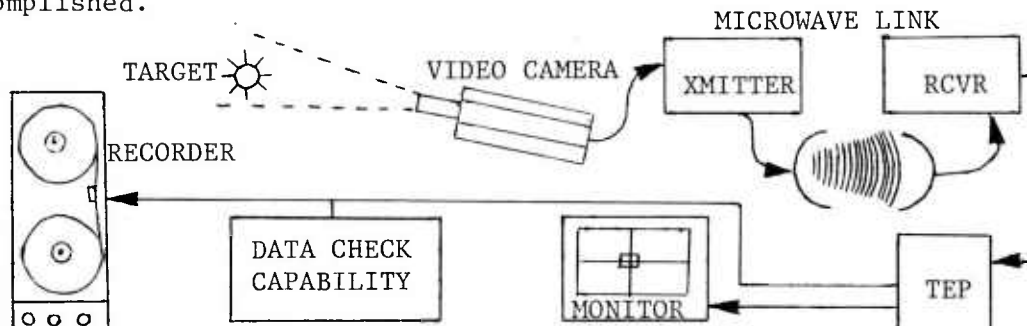


Figure 1 Block Diagram of GTV System

*BROWN

Since the error data are available in real time, the test director can monitor constantly, providing close control of the test and rapidly correcting situations that might result in the loss of valuable test items or critical data; moreover, a video monitor in the system provides a picture of where the weapon is pointing. Many costly retests are thus avoided, and better control yields more viable data.

CONCEPT OF SYSTEM OPERATION

The GTV system determines the pointing angles of the weapons and fire-control systems of combat vehicles. Figure 1 is a block diagram of this system. A video camera and lens are mounted on and boresighted with the gun tube of the combat vehicle. The video picture is transmitted by cable or microwave link to an instrumentation van containing the other system components. The received composite video signal is fed to the TEP, and an annotated video signal is fed from the tracker to a video monitor. The tracker converts the target position within the video picture to digital and analog values proportional in real time to the target displacement from a predetermined boresight. These values are then recorded on tape or sent directly to the ADPE.

The two most important components of the GTV system are the video camera and the video tracker. The video monitor is merely a means for viewing the data generation and monitoring the tracking phase of testing. Although these capabilities are invaluable for test control, data checking, and system installation and alignment, the video monitor plays no part in the error-signal generation. The digital and analog recorders are time-tested devices, and while it is essential to the operation of the GTV system that they function properly, they are simply data loggers and have nothing to do with the error-signal generation.

The video camera, long regarded as merely an imaging device, plays a primary role in the GTV system. Since the tracker input is the video signal from the camera, the camera is the transducer for the GTV system. The video camera must be stable, rugged, and linear, and have a wide dynamic range to ensure these characteristics in the overall system performance. The lens on the video camera must also possess the same characteristics. Changing the focal length of the lens changes the field of view (FOV) of the video camera. Since the tracker resolution is a fixed percentage (0.4% vertical) of the FOV, the resolution or accuracy of the system is a function of the focal length of the lens. For example, a 20-inch (508-mm) (focal length) lens on a 1-inch (25.4-mm) vidicon camera yields a system vertical FOV of 19.5 mils and an accuracy of ± 0.08 mil; an 80-inch (2032-mm) lens yields an accuracy of ± 0.02 mil but an FOV of only 4.8 mils. A trade off exists

*BROWN

between the desired accuracy and the FOV required to accommodate system errors and dynamics, lead angles, superelevation, etc.

The tracker accepts as its input the video signal from the camera. This video signal is a high-frequency analog-voltage stream of sequential left-to-right top-to-bottom (ignoring image reversal by the lens) scans of the camera image area. The tracker searches this input signal for a discontinuity: a voltage higher than an operator-established threshold, meaning a target brighter than the background; or a voltage lower than a threshold, meaning a target darker than the background. By identifying the target position in the input stream, the tracker determines the target position in the FOV. This operation is accomplished at the standard video-field (there are two fields to every 2:1 interlaced frame) scanning rate of 60 times per second. Since the standard video line-scanning rate is approximately 250 lines per field (500 per frame) in the active image area, maximum resolution is 1 line in 250, or 0.4% of the FOV. This resolution can be increased by higher scan rates; for instance, another standard scanning rate is 800 lines per frame, 400 per field, yielding a resolution of 1 line in 400 or 0.25% of the FOV. The trade off is that the higher scanning rate is somewhat specialized and the associated equipment is considerably more expensive and less available. There are applications, however, where the higher resolution is required, such as tracking through the gunner's sight.

The tracker is capable of edge or centroid tracking. In the edge mode the operator can select the leading edge of the target that he/she wants to be tracked. This mode is particularly useful for large targets that may extend out of the FOV. In the centroid mode the tracker digitally performs the mathematical operations:

$$\bar{X} = Xda / \int dA \qquad \bar{Y} = YdA / \int dA$$

to locate and track the center of area (A) of the target.

A tracking gate limits the processing of target information to an area automatically adjustable in size from 5% to 80% of the FOV. The gate can be manually positioned or, using an electronic servo loop, the gate will position itself such that the center of the gate coincides with the tracking point (edge or centroid) in the preceding video field; thus, in the current video field, the tracker will only search for the new target position in an area the size of the gate around the last known target position. While somewhat limiting frequency response with small gate sizes the gate greatly enhances the system noise immunity and tracking ability.

*BROWN

The tracker also generates a reticle display on the monitor, indicating the center of the FOV or the boresight reference point. This reticle aids in system setup and alignment. An additional feature allows the zero reference point to be moved from the center of the FOV and positioned by the operator.

The tracker has three different outputs. The video output produces the display on the video monitor, aiding system setup, alignment, and operation. The digital outputs are the chief error output signals, as all internal processing is digital, and digital data are the most desirable for further processing. These digital signals are 1's complement or 2's complement, 8 to 12 bit (depending on the resolution determined by the scan rate) binary numbers, indicating the target position in a cartesian coordinate fashion. The addition of analog-to-digital converters within the tracker generates alternate analog outputs varying between -5 and +5 volts, proportional to the target position. These analog outputs provide a convenient means of checking field data. Two digital displays on the front panel of the tracker also provide convenient readouts of the error signals.

The feasibility of the GTV system was to be tested by evaluating the system with the following criteria: the system should function as anticipated and provide at least the same amount of data as is derived from the high-speed photographic techniques and ir tracking systems; the system should be practical from a standpoint of ease of installation and operation; the system accuracy under field instrumentation conditions should be at least that demanded by the tracking and stabilizer testing for which it would be used.

LABORATORY TESTS

Not only was the GTV system a new instrumentation system, but the video camera was an entirely new form of transducer to be used for data collection. Thus the video camera and TEP, the main components of the GTV system, first underwent extensive laboratory evaluation designed to verify the operation and accuracy through a point-by-point analysis of predicted outputs versus actual outputs. Laboratory tests were designed to evaluate system static linearity, dynamic linearity, frequency response and phase lag, and mechanical stability in a vibration environment. The laboratory tests were designed to uncover potential problem areas before field testing and to provide a data base that would minimize the data to be evaluated during field testing. This would allow the chief concerns of the field testing to be ease of installation and operation, system durability, and field suitability.

*BROWN

The objective of the static system evaluation was to determine a correlation between the target position in the FOV and the output of the system. The evaluation of the complete system was conducted by testing the tracker for linearity and then adding the other components of the system, camera and monitor, and evaluating the system as a whole as each component was added.

To test the linearity of the tracker, a calibrated video-test-pattern generator was used to generate a video signal consisting of a linear dot array covering the entire FOV. The dots in the linear array were assigned numerical labels in a cartesian coordinate fashion. A linear regression analysis was performed on the tracker output values for the points as a numerical evaluation of tracker linearity.

An industry standard procedure exists for evaluating the linearity of video cameras. This procedure utilizes an EIA Standard Video Linearity Test Chart, a video bar-dot mixer, and a video monitor. The test pattern is placed so that it exactly fills the FOV of the video camera. The video output is run through a bar-dot mixer, where it is mixed with a calibrated electronic grid and is then fed to the video monitor. By noting the alinement (on the monitor) between the test pattern and the electronic grid, the camera linearity can be verified or adjusted to 1% or less of the picture area. During this test a new solid-state image-device video camera was examined. The camera used a charge-coupled device (CCD) for the image sensor, which was a 520 by 344 linear array of light-sensitive elements. The linearity of this camera was verified to well within 1% of the picture area. Because the picture on the CCD camera is formed, not by a sweeping beam but by fixed elements, this camera was considered to be inherently stable in linearity and picture aspect ratio. For this reason, this camera was used when possible during the evaluation of the GTV system; however, much testing was accomplished before the CCD camera was available.

The objective of the dynamic system linearity test was to make the same linearity evaluation as the static system linearity test, but this time to use a moving target. This test is an extremely valuable laboratory test because all test scenarios in which this instrumentation would be used would be dynamic. A linear array of 29 computer-addressable light emitting diodes (LEDs) (Figure 2) was placed completely within the FOV of a CCD solid-state video camera, connected to the tracker. The LEDs were lighted in succession by the computer and the tracker digital output was sampled by the computer and stored in the memory with the LED driving function for later processing.

*BROWN

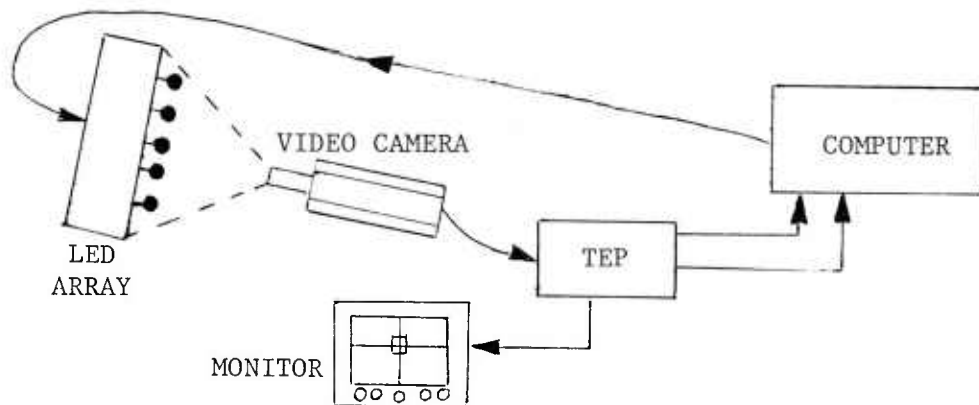


Figure 2 Dynamic Linearity and Frequency Evaluation Setup

Since the LEDs were arranged in a linear array, the tracker outputs could be compared to the known positions of the LEDs to verify the tracker dynamic linearity. The linearity of the tracker outputs was verified using the same linear regression analysis as in the static evaluation. Each LED was activated 30 times in succession and the array was swept in a sawtooth wave fashion so that 1000 data points (3.7 milliseconds apart) were taken. As discussed in following paragraphs, phase lag was evident in the tracker outputs during this dynamic testing. The linear regression routine accounted for this phase lag by iterating the regression subroutine, each time stepping the tracker output data back one sample interval, until the standard deviation was minimized, indicating the best fit. Because field data should never be discontinuous, as were the discrete LED positions, the linear regression analysis was performed on the time-displaced data, disregarding data points falling at the transitions in LED positions.

Because the tracking system is to be used to track moving or stationary targets in a dynamic (moving-vehicle, vibration, etc.) environment, the frequency response of the system is an essential parameter and could be a limiting factor for some applications. The term "break lock" is used when the speed of target movement in the FOV has exceeded the tracking ability of the TEP. Since the TEP inserts a gate and a target flag into the video signal going to the monitor, which follows the target when it is tracking properly, breaking lock is readily visible to the operator. Breaking lock often (and for extended periods during data acquisition) is extremely undesirable as data output during this time is meaningless.

Phase lag is also an important parameter, as there is a delay

*BROWN

between target acquisition and the location and presentation of the error outputs. This phase lag must be known where precise time correlation between target position and other data outputs is essential to the format of the test for which this information is being used.

A basic understanding of the theory of operation of the TEP leads to the following hypothesis of system frequency response and phase lag:

a. The output of the TEP is updated once per field or 60 times per second; therefore, the maximum theoretical frequency for proper track (meaning target-deflection reconstruction from the tracker output) of pure sine-wave deflection would be 30 Hz. Since this Nyquist sampling rate is theoretical and would not yield a desirable output for test data, it is suspected that the maximum frequency should be considered as substantially lower than 30 Hz, or as low as 10 Hz.

b. Because of the left-to-right top-to-bottom (of the reproduced picture) sweep of the video image device in the camera, target-acquisition time decreases as the target moves up and left in the FOV. The best case of time lag in target acquisition was when the target was in the lower right corner of the FOV (one sample) and in the upper left corner (succeeding sample). In this case the output lag would be the vertical retrace time plus two horizontal sweep times (two horizontal sweeps of the target are required for target acquisition) or approximately 627 microseconds. The worst case would be when the target moved from upper left to lower right. The time lag here would be the time required for two vertical scans minus two horizontal scans plus one vertical retrace, or approximately 34 milliseconds. Since all other forms of target movement are between these two extremes, with the target-movement frequency being low in most instances, the average lag of 17 milliseconds between target position and data output is probably a viable approximation, ignoring the persistence of the video image device or camera delays. Only the area within the gate is searched for target information, and the position is produced immediately following target acquisition in the edge track mode; in centroid track the target position is produced at the completion of sweeping the entire gate. This will produce differing results for rapidly varying signals.

c. Since the gate position is updated once per field and is calculated to center on the output of the preceding field, the gate will break lock if the target-deflection horizontal component during 1/60 second exceeds half the gate width or the vertical component exceeds half the gate height.

*BROWN

d. Because tracking and gate display is lost as soon as any edge of the gate goes out of the FOV, the target must remain in an area of height equal to the FOV vertical dimension minus the height of the gate and in width equal to the FOV horizontal dimension minus the width of the gate.

This evaluation utilized the same method and data as the dynamic linearity test. The sweeping LED array (Figure 2) provided a good, accurately controlled simulation of a moving target, and it was easy to compare the LED driving function with the tracker output to verify frequency response and phase lag. It was decided that frequency-response measurements should be taken with the entire target deflection within the gate and the gate in the manual position, to eliminate the variable of gate size, which is operator adjustable.

Figure 3 is a sample of several plots of the LED function and the GTV outputs versus time. These plots clearly indicated that the GTV outputs lag the LED positions by a generally constant time. This lag was arrived at analytically by taking the average number of data samples that the linear regression routine had to shift the GTV output data to achieve a best fit, and multiplying that number by the sample interval. The average shift was 9.37 samples; samples were taken 3.7 milliseconds apart, indicating that the average lag is 34.7 milliseconds, approximately two video-field times.

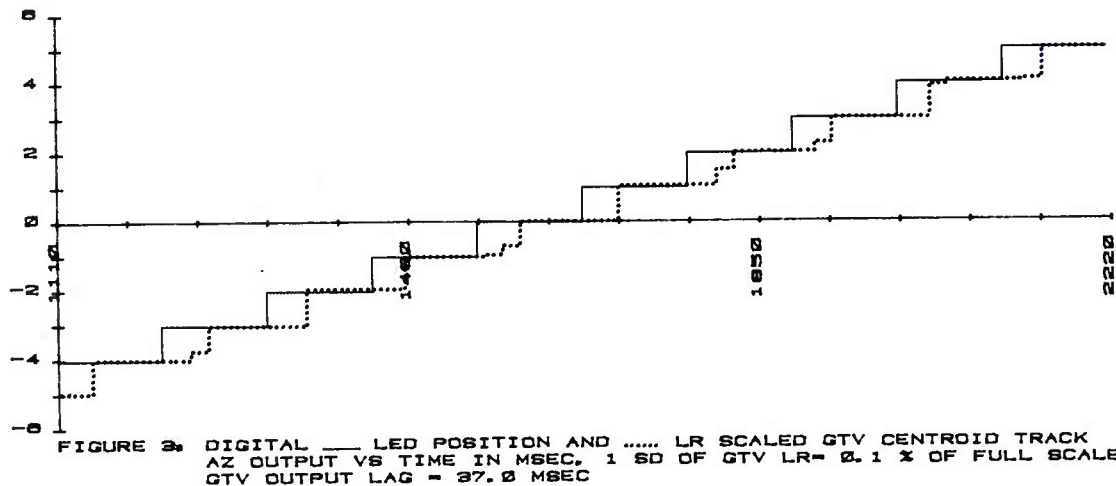
In this experiment the LED driving function was a step function, so the GTV outputs could be considered as a step response having an average rise time of 34.7 milliseconds. Applying the commonly used formula:

$$3\text{-dB frequency} = 0.35 \div \text{rise time}$$

yields a frequency of 10 Hz. Although 3 dB is meaningless in terms of the digital processing within the tracker, this frequency could be considered to be the point above which breaking lock would occur.

It was anticipated that the GTV system would be used extensively for an evaluation of vehicular-mounted systems, an environment that would subject the video camera and lens to considerable vibration. To evaluate the performance of the system in this environment, the camera, lens, and mount were firmly mounted on a package tester and then shaken in such a way that the test items were subjected to accelerations very similar to (if not more severe than) those recorded on the camera position of a moving vehicle.

*BROWN



DISCUSSION OF LABORATORY TESTS

Because most field recording-and-measuring devices are calibrated to within only 1% of full scale, this value was chosen as a criterion by which to judge the overall accuracy of the GTV system. The laboratory static-linearity tests demonstrate that the GTV system (tracker and camera) is linear within 1% of full scale, because all differences between the observed outputs and the linear regression predicted outputs are less than 0.05 volt. Dynamic linearity was calculated using a 2.5 S (standard deviation) estimate level, a strict linearity criterion, predicting that 99% of the time the output will be within the calculated linearity. For centroid track azimuth and elevation, this value was within 1%; for edge track azimuth and elevation, this value was slightly greater than 1% but was within 2%. The choice of a more lenient 1 S or 1.5 S level would have brought the edge-track data within 1% linearity. It is felt that the slightly worse linearity performance of the tracker in the edge mode is because the LEDs were of varying brightness. Any concentrated light source blooms somewhat on a video sensor, depending on the brightness, and sources of varying brightness appear to be different sizes; therefore, the target edge position, in relation to the center, differs with respect to source brightness, greatly affecting edge track while not appreciably affecting centroid track.

System frequency response was verified to the predicted 10 Hz, with the gate being stationary in manual track and all target movement being contained within the gate. The average phase lag was experimentally derived as approximately 35 milliseconds instead of the

expected 17. Slight persistence was perceptible with the CCD camera used during the phase-lag experiment. This slight persistence is noticeable and was experimentally verified as approximately the same for all the video image devices used with the GTV system. Because any persistence or lag in the image sensor carries over into the next field scan, this lag tends to be a quantized time equal to one field scan (approximately 17 milliseconds). This plus the 17 millisecond tracker lag equals 34 milliseconds, within one computer sample time (3.7 milliseconds) of the experimentally derived lag of approximately 35 milliseconds, making the predicted and experimental numbers equal within the sampling accuracy.

FIELD TESTS

The chief purpose of the field tests was to examine the practicality of the GTV system for field instrumentation. During system evaluation, the GTV system was successfully installed on and removed from several different types of vehicles; this required no more time than was involved with photometric techniques. The GTV did not affect vehicle operation, nor did the vehicle affect operation or accuracy of the GTV. Packaging was rugged and convenient enough for field usage. The most practical installation seemed to be with the video camera mounted on the gun tube with its video output sent by microwave link to a data van containing the tracker, monitor, and recorder. The camera and microwave transmitter are both ruggedized units and did function without a problem during all testing. This setup required no operator on board the vehicle, as tracker monitoring, data logging, and recording were all accomplished by one person in the data van. A summary of manpower savings and associated monetary savings during several tracking and stabilizer programs conducted during FY76 is included as Figure 4.

	<u>Savings in \$1000</u>
Data Acquisition (per test)	
Typical tracking test	
previously required 6 weeks, 6 men, \$15,000	
GTV requires 2 weeks, 6 men, \$5,000	
Savings of	10
Typical stabilizer test (same as above)	10
<u>Data Reduction (tracking or stabilizer) (per test)</u>	<u>10</u>
Major Programs (data acquisition and reduction combined)	
Quick Fox (combined tracking/stabilizer/target acquisition series)	200
XM-1 Tank (4 tracking, 4 stabilizer tests)	160
MICV (3 tracking, 10 stabilizer tests)	120
M-60 Tank (4 tracking, 10 stabilizer tests)	280
Graph Angle (air defense tracking series)	200
Total Major Program Savings	<u>960</u>

Figure 4 Summary of GTV Monetary Savings FY76

*BROWN

The parameters evaluated during laboratory testing were examined at various times during field testing. An analysis of this data would indicate system degradation, mechanical or electrical drift, or an operator's inability to make the system function properly in the field. It was apparent from field data and results that there was no drift or linearity degradation beyond the 1.0% tolerance specified earlier. This was true during linearity and boresight (zero) checks during all phases of field testing.

An inspection of the tracking waveforms on a playback oscillograph of analog field data showed few discontinuities that indicated loss of lock. The situations with discontinuities were remedied by a change in the focal length of the lens to increase the FOV. These were the only frequency-response problems; an examination of the waveforms indicated that the majority of the frequencies were well below the 10 Hz response determined during laboratory testing. Since the frequency responses of the turret systems of the vehicles under test were well below 10 Hz, the frequency components at 10 Hz or above in the error waveforms were generally not meaningful tracking data.

CONCLUSIONS

The following conclusions were reached as a result of the evaluation of the GTV instrumentation system:

a. The GTV system is linear and stable in both static and vibration or dynamic environments to within 1.0% of the FOV. The accuracy and precision of the GTV system is, therefore, greater than the maximum 1.0% tolerances of the data recording and processing systems and the 1.0% tolerances required for tracking and stabilizer testing.

b. Dynamic response is a function of the FOV of the video camera. Changes in focal length affect dynamic response and measurement precision inversely, so the focal length of the lens must be carefully tailored to the particular testing situation.

c. Dynamic response is also affected by gate size; increasing the gate size increases the dynamic response, but also decreases optical signal-to-noise performance. Size also must be tailored to the particular testing situation, but can easily be changed by the operator during the test for best results.

d. The GTV system presented no unreasonable problems in installation or operation during field testing. The video-monitoring capability was most valuable and made immediate test-progress analysis possible. The GTV system therefore appears to be a viable field instrumentation system producing good data suitable for direct input

*BROWN

to the ADPE or any data-recording system.

e. Because camera linearity greatly affects data and can vary with time and environment, linearity and other camera specifications must be verified at regular intervals. This requirement is important to vidicon cameras where linearity and aspect ratio are dependent on the electronics that sweep the beam across the vidicon faceplate, but should also be applied to solid-state cameras at less frequent intervals.

FURTHER APPLICATIONS OF VIDEO TO INSTRUMENTATION

The GTV system represented one of the first applications of video techniques for instrumentation instead of merely viewing. Evaluation of the system demonstrated that the video camera is a viable transducer for dynamic and static position data acquisition. The GTV's success has led to many unique applications of video in the instrumentation field. These applications include the: 1) Automated Video Target Scoring System, 2) Video Muzzle Position System, 3) Video Surf Measurement and Analysis System, and 4) Automatic Video Shell Fragment Area and Velocity Measuring Systems.

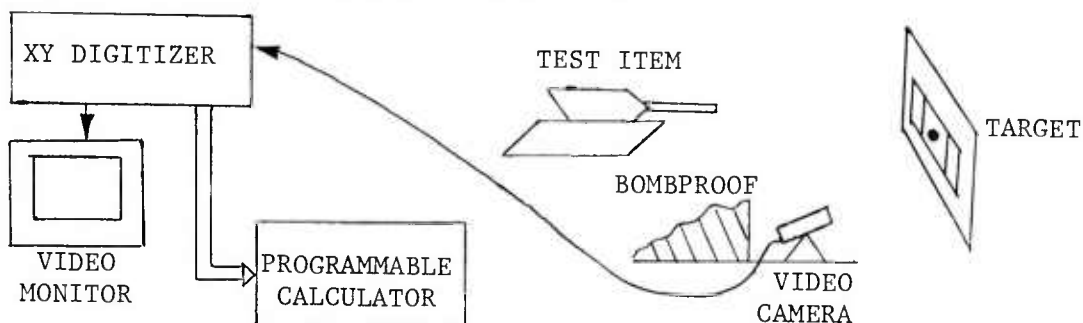


Figure 5 Automated Video Target Scoring System

The Automated Video Target Scoring System (AVTSS) utilizes a video camera, a programmable calculator, and a video XY position digitizer (Figure 5). The XY digitizer is similar to the TEP in the GTV system except that, instead of tracking a target point, an operator positions a cursor in the video FOV and the X and Y positions of the cursor are output as digital words. The video camera is aimed at a cloth target used during testing large caliber weapons. An interactive calculator program steps the operator through appropriate calibration such that distances in the video FOV can be related to actual distances at the target. The gunner's aim point is digitized and then the cursor is placed over the images of holes made in the target by the rounds fired and thus the target is scored remotely. At the completion of firing the calculator outputs the mean and standard deviation of the shot group. The AVTSS eliminates the need for measuring the cloth

*BROWN

target directly thereby saving time and manpower to drive downrange, lower, score and raise the target, and return.

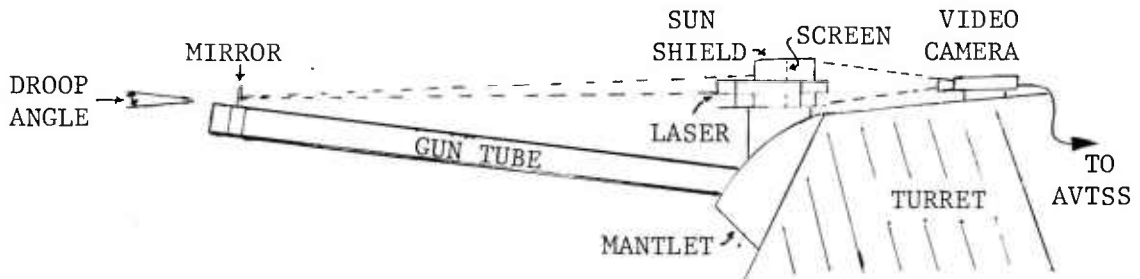


Figure 6 Video Muzzle Position System

The Video Muzzle Position System (VMPS) incorporates the AVTSS to measure gun muzzle movement with respect to the mantlet over a period of time. A small continuous wave laser and translucent screen are mounted on the gun mantlet (Figure 6). A mirror is fastened to the muzzle so that the laser beam is reflected back to the translucent screen. The video camera of the AVTSS is aimed at the back side of the translucent screen and readings are taken by positioning the cursor over the image formed by the reflected laser beam on the screen. The calculator output is angular displacement of the muzzle in mils or other desired scientific units referenced to a position taken at the beginning of testing. Once again, application of video techniques allows this data to be acquired remotely and logged more accurately and rapidly than with other techniques.

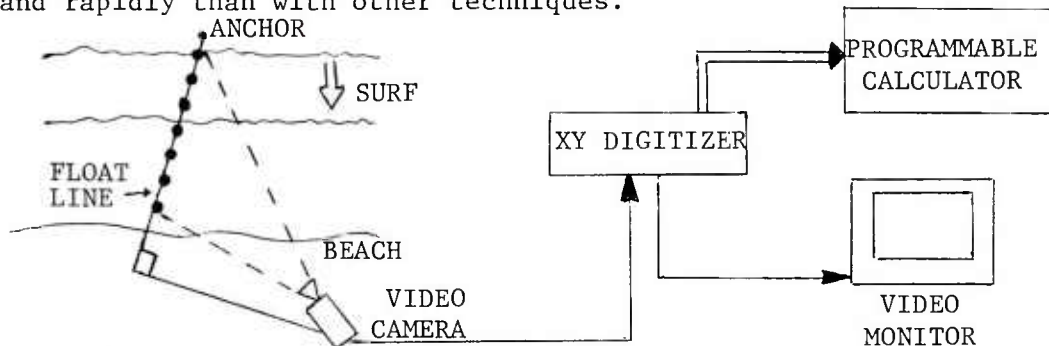


Figure 7 Video Surf Measurement and Analysis System

The Video Surf Measurement and Analysis System was developed to measure the height, velocity, and period of plunging surf throughout an entire surf zone. Testing of a military landing craft required that this data be obtained, however all wave measuring systems in existence prior to the development of the video based system were incapable of measuring surf and could only obtain data for a single location.

A float line is anchored on the beach and beyond the surf zone (Figure 7). The same components of the AVTSS are utilized with the camera positioned on the beach such that its FOV covers the entire portion of the float line in the surf zone. Previously determined angles and distances are entered into the calculator during calibration. An operator positions the cursor at the peak of a breaker and the float line and digitizes this position. After the breaker passes the cursor is then moved vertically to the float line as it rests at the low water level. The calculator program then accomplishes the necessary trigonometry to calculate the height of the breaker. The addition of a digital time base allows the system to yield breaker velocity by digitizing the same wave crest at two positions, and period, by digitizing two succeeding wave crests at a single position. The calculator also logs the data and performs any analysis as required by the test program.

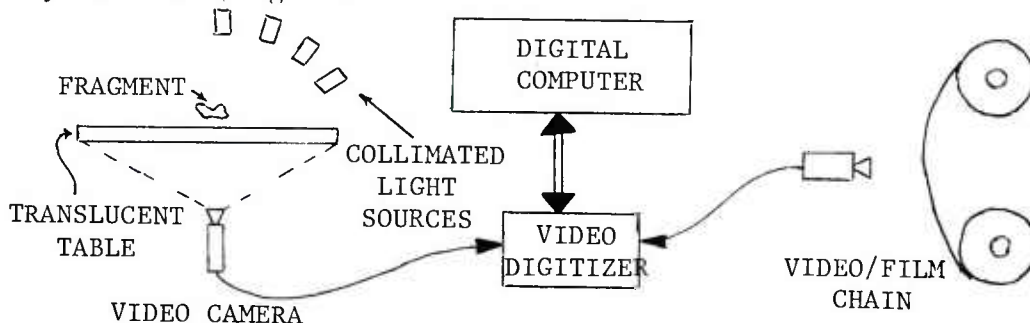


Figure 8 Automatic Video Shell Fragment Area and Velocity Measuring Systems

The Automatic Video Shell Fragment Area and Velocity Measuring Systems are presently under development. Both systems utilize a video camera, a programmable calculator, and a video digitizer (Figure 8). The digitizer allows a single video field to be converted to a matrix of digital words corresponding to the gray scale value of each picture element (pixel) comprising the entire FOV.

To measure shell fragment velocity, a shell is exploded in an arena with dark vinyl sheeting along one side. As the fragments pierce the sheeting a high speed film camera records the holes as light passes through the sheeting from a bank of flash bulbs on the inside. Reading the film frame by frame to determine fragment time and position for velocity calculation is presently a long and arduous task. The video based system will present each film frame to the video camera, digitize the picture, and compare the new picture matrix with that stored in memory from the previous frame thus determining new fragment holes and their location. Data analysis that presently takes days would be accomplished in minutes with the video based system.

*BROWN

The same system can also measure fragment area. Collimated light sources at several angles project silhouettes of the fragment onto a translucent table above which the fragment is suspended. A video camera is placed below the table so that the silhouettes become video images. These video images are digitized and the computer can then determine the area of the various projections and thus the size of the fragment.

Presently, applications of video to instrumentation are limited by the speed and resolution of the video image devices. As research improves these parameters, the applications of video to instrumentation will greatly increase. Image correlation trackers will become more feasible along with much more extensive processing of the video image. There is an enormous amount of data yet to be extracted from the processing of a video image. The GTV evaluation proves the effectiveness of a video based instrumentation system, and the further applications discussed here are only the beginning.

REFERENCES

1. Brown, C. David, Final Report, Special Study on the Video Tracking and Processor System, Gated TV, TECOM Project 5-CO-APO-GTV-301, US Army Aberdeen Proving Ground, Report APG-MT-5069, January 1978. (Distribution controlled by US Army Test and Evaluation Command, ATTN: DRSTE-AD-I)
2. Grob, Bernard, Basic Television, McGraw Hill, New York, 1964.
3. Groff, John N., Final Report, Special Studies of Testing Stabilization Systems for Vehicle-Mounted Automatic Weapons and Procedures for Testing Control Systems of Vehicle Automatic Cannon. TECOM Project 9-CO-001-000-016 and -017, US Army Aberdeen Proving Ground, Report APG-MT-3613, August 1970. (Distribution controlled by US Army Test and Evaluation Command, ATTN: DRSTE-TS-M. AD 883 492L.)

IN SITU MEASUREMENTS OF AEROSOL ABSORPTION
WITH A RESONANT CW LASER SPECTROPHONE

*CHARLES W. BRUCE, DR.

RONALD G. PINNICK, DR.

RALPH J. BREWER, DR.

YOUNG P. YEE, MR.

GILBERT FERNANDEZ, MR.

US ARMY ATMOSPHERIC SCIENCES LABORATORY

WHITE SANDS MISSILE RANGE, NM 88002

A major problem area for DOD electro-optical (EO) systems, from concept to operation, is the transmission losses due to atmospheric gaseous and particulate constituents. Atmospheric effects on high energy lasers are usually considered separately from those affecting other EO systems such as target ranging, laser guidance, target designation, forward looking infrared, etc. This is because high energy lasers interact with the atmospheric medium dynamically, changing its nature, while the performance of the remaining EO systems is determined linearly from measurements of the components of the losses, i.e., absorption, scattering, and refractive effects. However, measurements of these quantities plus additional information such as values of meteorological parameters and identification by absorbing constituents form the basis for nonlinear predictions of the transmission of high energy laser energy as well.

In this paper we discuss in situ measurements of the absorption of infrared radiation primarily by atmospheric particulate matter. Absorption itself is of prime importance in determining the high energy laser beam transmission losses, while for the EO systems it is simply additive with the scattering and refractive losses.

Absorption by particulates is currently a major concern in both of the previously mentioned types of systems because this quantity is highly variable with the meteorological and battlefield

conditions and has not been well identified in either composition or magnitude. It is known that particulate absorption can be orders of magnitude below the gaseous absorption. For example, this is typical for the $10\mu\text{m}$ "transmission window" in a quiescent atmosphere. However, the particulate absorption can be considerably more than two orders of magnitude above that due to gases for windy conditions. Fog and battlefield disturbances (including EO countermeasures) also represent a relatively undefined but potentially strongly absorptive environment in terms of particulate absorption.

Absorption and scattering, the two additive components of extinction for a homogeneous medium, have similar functional dependencies as described in the Lorenz-Mie theory. As a result of this similarity, it is highly probable that, if the absorption is well known and correlates well with the result calculated from particle counting and sizing measurements, the scattering component can also be accurately calculated on the same basis. Thus, even when scattering as well as absorption is important (as for EO systems), measurements whose explicit function is to measure or predict the scattering are less vital if the absorption is measured together with the particle densities as a function of size.

Since atmospheric aerosols are known frequently to be fragile with respect to composition, e.g., particles with volatile components, it is important to measure their absorption in situ or, in other words, without removing them from their natural environment. Other reasons for the importance of in situ measurements are that neither the bulk properties of many atmospheric particulates nor, on a real-time basis, the chemical composition of these particulates are known.

Measurement of atmospheric particulate absorption in situ is difficult and to date atmospheric absorption due to this component has only been done definitively by indirect techniques (1-3). The spectrophone holds promise for making these measurements directly, bypassing altogether measurements of particle shape, size, and complex refractive index. This paper reports results obtained using a calibrated in situ spectrophone compared with results calculated from nearly simultaneous measurements of particle size distributions for particulate samples consisting of known atmospheric absorbing species prepared in this laboratory. We believe these to be the first definitive in situ measurements of aerosol absorption with a spectrophone (4).

The spectrophone technique was conceived nearly a century ago by Alexander Graham Bell (5), but applications using laser sources

(with the advantages of high spectral resolution and power densities) first appeared in 1968 (6); absorption measurements on solids, liquids, and gases have all been reported (7,8). Spectrophones of unique design have been developed in this laboratory for gas and/or aerosol measurements (9). Systems of high sensitivity ($\sim 0.001 \text{ km}^{-1}$) have been developed for measurements with pulsed sources (10) as well as with CW sources (11), with or without windows, and for continuous flow-through operation. Basic to these designs has been an acoustically isolated, resonant inner cavity without windows, which generally operates in a longitudinal acoustical mode in response to the absorption signal. The isolation thereby obtained effectively eliminates the window noise, which has been represented as the perennial spectrophone problem.

A schematic of a microphone is shown in Figure 1. Within this cavity, low-Q longitudinal acoustical modes are established in the gaseous medium from the gaseous or particulate absorption signals.

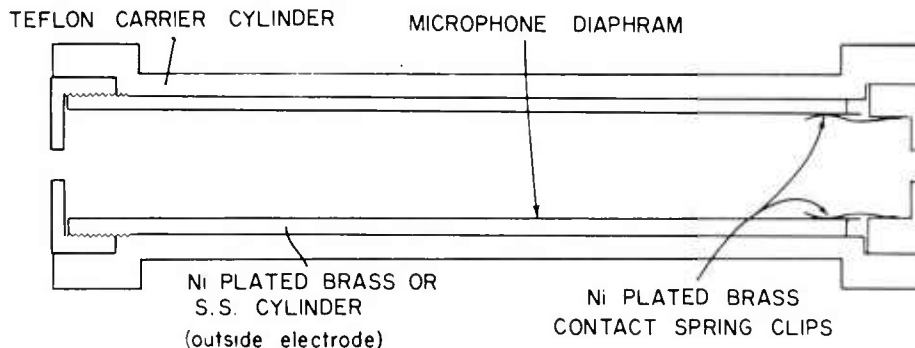


Figure 1. Components of cylindrical microphone.

Spectrophone physical considerations and design parameters have been discussed in a technical report (9); but briefly, the vital points are that the spectrophone system must be designed to produce signals proportional to the absorption coefficient and must be linear with power and be calibratable. Finally, this calibration factor must be independent of the optical wavelength. The accomplishment of these qualities is discussed in the aforementioned report for gaseous absorption, but the same requirements apply to particulate absorption measurements. For the particulates, the above design criteria are manifested in the following concerns: time response of the spectrophone to particulate matter, behavior of volatile components, and variation of the optical constants of the particles as a function of particle temperature. Time response for the particulate material

requires a very complicated calculation for precise results but, as simplified somewhat, provides an upper bound for the frequency of the chopped signal. The spectrophone time response can be shown to vary as the square of the particle diameter. The system parameters were designed such that there is no significant net temperature increase in the particle per on-off cycle of the beam. By keeping the maximum particle temperature rise down ($<30\text{ C}^\circ$) the latter two concerns relative to spectrophone design are successfully treated. Temperature rise in the illuminating beam as a function of particle radius was calculated for three of the four substances studied to assure that the design criteria are not violated.

These particle heating results are highly dependent in both spectral form and magnitude on the complex refractive indices. For example, water particle heating is quite constant as a function of radius below about $2\mu\text{m}$, decreasing slowly above that radius, while particle heating for both quartz and ammonium sulfate were peaked with maximum temperatures much higher than for water droplets. The fourth, calcite, was such a weak absorber that in the $9\mu\text{m}$ - $11\mu\text{m}$ region particle heating was not of concern. Though these heating curves were calculated, this type of information may be inferred from the spectrophone results. When absorption is strong, either beam power or beam power density is reduced to reduce heating.

For these absorption measurements, particles which had been ground with mortar and pestle were dispersed in a 0.5 m^3 environmental chamber. Two methods of measurement were used. In one, two CW-source spectrophones were used in a differential mode (12), with one spectrophone continuously sampling air from the chamber and the other sampling ambient air without the particles. Or, if the absorption coefficients of only a few spectral lines were to be measured, a single spectrophone might be used and the ambient values measured prior to and checked following the measurement of dust absorption.

The predominant ambient absorption at the wavelengths of the CO_2 laser used ($9.2\mu\text{m}$ - $10.8\mu\text{m}$) was by water vapor; and, in fact, a strong absorption (at $\text{R}20$, $10.247\mu\text{m}$) was used to calibrate the spectrophone system for both gaseous and particulate absorption measurements. For this purpose dew point sensors are always in operation during absorption measurements.

The measurement procedure for dry particulates was generally to sample the contents of the holding chamber repeatedly with the spectrophone system, with a filter carousel device to collect particles on Nuclepore filters, and with a light-scattering aerosol counter

developed by Rosen (13). Scanning electron microscope photomicrographs were subsequently taken of particles collected onto the filters; and these, together with the particle counter results, were used to determine the form of the particle size distribution. The question of the calibration of the particle counter may be found elsewhere (14,15). Here the counter, when operated, was used to obtain only the total concentration of particles with radii $\geq 0.17\mu\text{m}$.

A schematic of the spectrophone optical path and particle sampling apparatus is shown in Figure 2. The three sampling flow inlets shown, i.e., to the filter carousel, to the particle counter, and to the vertically mounted spectrophone, are in similar proximity and orientation in the actual system. For measurement of the absorption by liquid droplets a larger chamber was constructed ($\sim 1\text{ m}^3$) and,

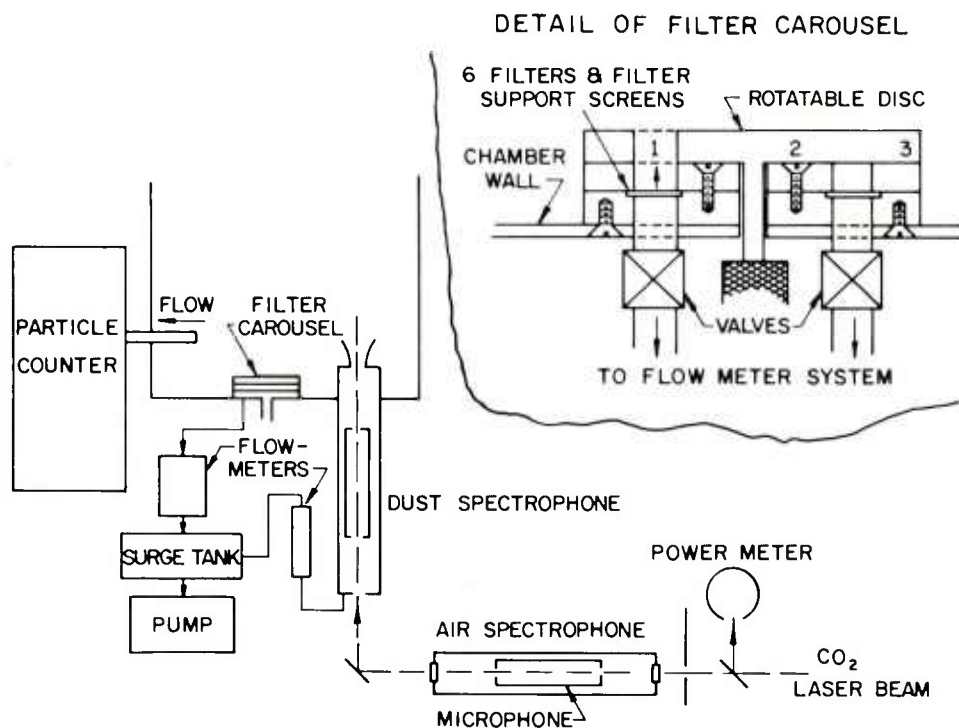


Figure 2. Arrangement of measurement apparatus showing light scattering particle counter, filter particle collectors (6), and spectrophone cutaway. The insert depicts a cross-sectional view of the filter carousel. To obtain a filter sample, the top disk is rotated to expose a filter; and the appropriate valve to the flowmeter is opened.

though the spectrophone arrangement is similar to that described for the dry particulates, a commercial "Knollenberg" particle counter is used in place of the Rosen counter and a liquid content measurement takes the place of the filter sample carousel. Intake orifices for each of these plus that of a dew point sensor are in close proximity to each other. The droplets are generated by a commercial mist generator.

Individual droplets cannot be examined in this case but, since they can be assumed to be quite spherical, the particle counter is expected to be relatively accurate since calibration can be performed using spherical particles having various refractive indices (16).

Several particulate substances were selected for absorption studies, the prime consideration being relevance to the atmospheric absorption problem. Absorption studies for particulate quartz, calcite, and ammonium sulfate and preliminary results on liquid water will be presented in this paper.

Quartz was chosen because of the strong spectral dependency of the absorption in the wavelength region for the laser source used, the ready availability of measurements on the bulk properties of the material, and its stable (and minimally hygroscopic) nature, in addition to its frequency of occurrence and relatively strong absorption within the $9\mu\text{m}$ to $11\mu\text{m}$ region.

Measurement of the absorption of dust as a function of settling time in the environmental chamber effectively provides results for a continually changing size distribution. The strong test of correlation between calculated and measured results was an objective of this procedure. Figure 3 shows spectrophone measurements and Lorenz-Mie predictions of the absolute value of the aerosol absorption coefficient (in units of km^{-1}) as a function of time after the quartz samples are dispersed in the holding chamber. To illustrate some of the difficulties and shortcomings of such a comparison, photomicrographs of the quartz particle filter samples are shown in Figure 4. Each of the samples was collected over a 2-min interval at the times t_1 , t_2 , and t_3 identified in Figure 3. The Lorenz-Mie calculations of the particle absorption coefficients were made for polydispersions of quartz spheres of cross-sectional area equivalent to those measured on enlargements of the photomicrographs. The high degree of irregularity of the particles is immediately apparent; thus, the supposition that particles are spherical is not accurate. To further complicate the comparisons, crystalline quartz is birefringent, with distinctly

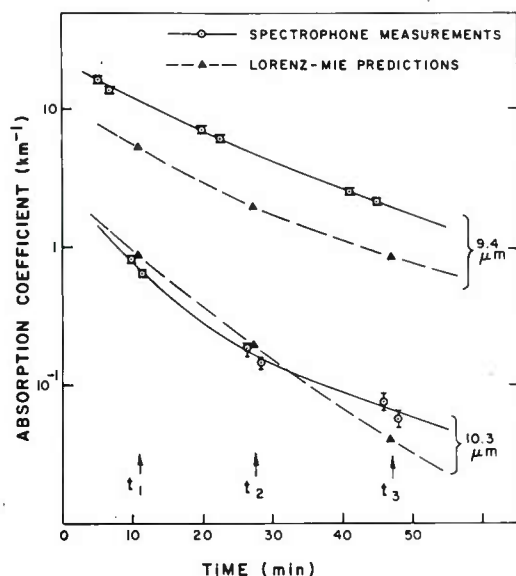


Figure 3. Absorption coefficients of quartz dust as a function of the time after the dust was dispersed in the holding chamber, as predicted by Lorenz-Mie theory and measured with a spectrophone, for two wavelengths. The predictions are based on particle size distributions measured at times t_1 , t_2 , and t_3 .

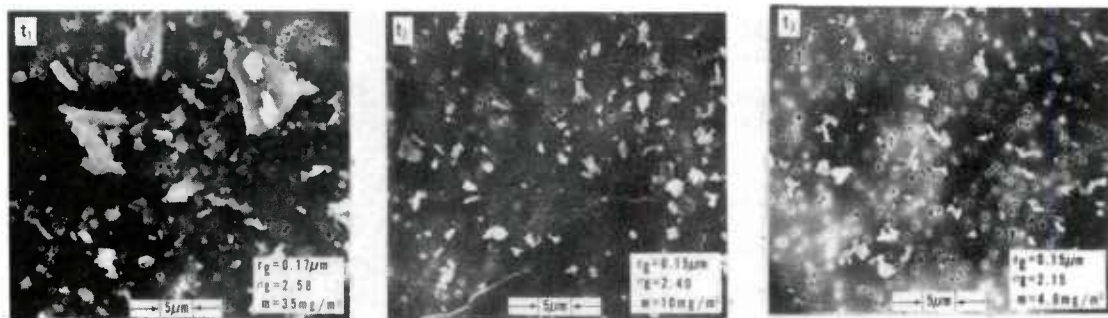


Figure 4. Scanning electron microscope photomicrographs of quartz particles collected onto Nuclepore filters. The corresponding samples were taken at times t_1 , t_2 , and t_3 (see Fig. 3). Each photograph contains the geometric mean radius r_g and geometric standard deviation σ_g for log normal size distributions obtained by determining the radii of spheres of equivalent cross sections from the SEM photomicrographs. Also shown are the corresponding aerosol mass loadings m , as determined from the log normal size distribution parameters by using a quartz density of 2.6 g cm^{-3} . The dark circles are holes in the Nuclepore filter substrate.

different indices of refraction for the ordinary and extraordinary rays. In accordance with the methods of Peterson and Weinman (17), the calculations for the ordinary and extraordinary rays were weighted on a 2:1 basis. These calculations were based on the observed particle size distributions and the complex refractive index for bulk crystalline quartz as reported by Spitzer and Kleinman (18). For the calculational approximations made, the agreement, which is roughly within a factor of 2, should be considered relatively close. It might have been more appropriate to use spheres of equal volume in the calculations of particulate absorption, since the particles were generally small in comparison with the wavelength; however, quantitative estimates of particle volumes from the photomicrographs were not possible.

That the absolute agreement (in magnitude but not in form) of the spectrophone results and the predictions in Figure 3 is better for the $10.3\mu\text{m}$ than for the $9.4\mu\text{m}$ results is probably fortuitous. The measured and predicted time dependencies, however, are quite close for the results at both wavelengths. The slightly different decay times for the two wavelengths are a result of the changing particle size distribution; the agreement between the spectrophone measurements and the predictions of the temporal changes in the particle absorption at both wavelengths is encouraging.

A comparison of spectral dependencies of measured and predicted absorption is shown in Figure 5. The upper three traces are predictions based on samples taken at times t_1 , t_2 , and t_3 , which were previously identified in Figure 3 for different times in the settling process after the ground quartz was dispersed into the holding chamber.

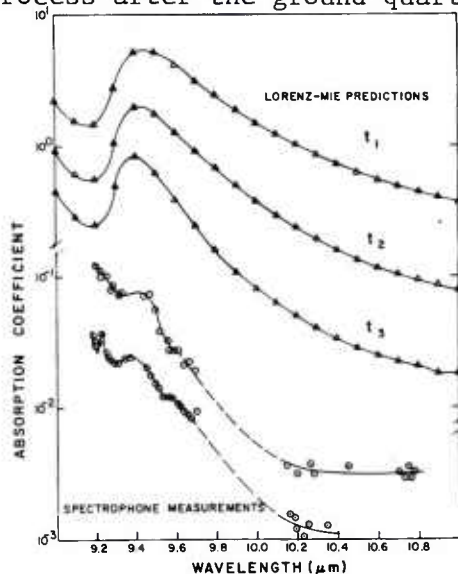


Figure 5. Spectral dependence of the absorption coefficient of quartz dust as predicted from particle size distribution measurements and bulk quartz optical properties by using Lorenz-Mie theory (absolute scale units of km^{-1} , upper three curves) and as measured by using a spectrophone (relative scale, lower two curves). The lower two curves represent data using dust ground from two different bulk quartz samples. The three upper traces correspond to sampling times $t_1 = 11$ min, $t_2 = 27$ min, and $t_3 = 47$ min, as explained in the text.

The lowest two traces refer to spectrophone measurements made at times comparable with the approximate time interval from t_2 to t_3 and should, therefore, be compared with the spectral variations shown in the predictions for those times. The spectrophone results represent a number of different measurements that were all normalized to a fixed time. The spectrophone signal time decay was removed by periodically repeating measurement of the absorption signal at particular laser lines. At the longest wavelengths, the absorption fell well below the ambient air absorption (primarily water), and this stressed the subtraction process, though not severely. The resulting errors in measurements of absorption (which are indicated by the spread in the data points) are larger than practical sensitivity limits because, despite the considerable noise in the laboratory during these experiments, no spectrophone acoustic isolation additional to that designed primarily for closed system gaseous absorption measurements was used.

In any case, there is general agreement of the measured and predicted absorption profiles over the $9.2\mu\text{m}$ - $10.8\mu\text{m}$ spectral range, though significantly different behavior is obtained within the $9.2\mu\text{m}$ - $9.4\mu\text{m}$ wavelength region. The upper absorption spectrum shown in Figure 5 for the spectrophone measurements is representative of a number of samples prepared from a block of high purity crystalline quartz. Some significantly different behavior (e.g., a small, sharp peak at $9.22\mu\text{m}$ shown in the lowest curve) has been observed for samples associated with another source of quartz.

Spectrophone measurements of a number of different quartz aerosol samples prepared by the procedures described above showed repeatability of the time and spectral dependencies evident in Figures 3 and 5.

By contrast with quartz the spectral dependence of calcite is very flat and the absorption quite weak in the $9\mu\text{m}$ - $11\mu\text{m}$ wavelength region. Spectrophone measurements of the spectral dependence of calcite agree well with that of spectrophotometer results obtained by the potassium bromide technique. Because of the weak absorption, this common atmospheric ingredient was not studied further.

Ammonium sulfate is also generally believed to be a frequently occurring atmospheric absorber and, like quartz, is known to have a strong absorption between $9\mu\text{m}$ and $10\mu\text{m}$. Once again the measurement process was initiated by grinding and sieving the coarse-grained reagent-grade material and dispersing it within the environmental chamber. The very large particles always settle out quickly and the mean particle radius (and breadth of the size distribution

with respect to particle radius) becomes smaller with time, though some deviations from a simple settling process can result due to small amounts of turbulence in the chamber. This can be seen in the sample size distribution data of Figure 6 in which the T-values represent successive times in the settling process. Measurements of the absorption were again compared with results calculated from the particle density as a function of size showing the degree of correlation over a continuously changing distribution. Since ammonium sulfate, like quartz, has a peak absorption at about $9.2\mu\text{m}$ and a relatively weak absorption at $10\mu\text{m}$, lines from each region were used in the spectrophone measurements of absorption as a function of settling time.

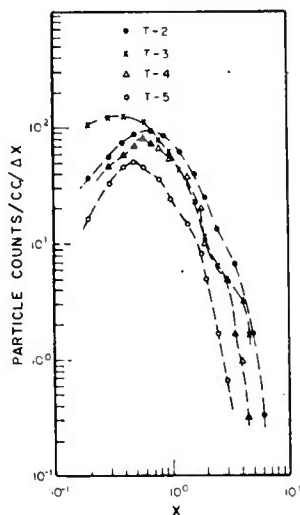


Figure 6. Particle size distribution measurements (shown in units of number/cc/x-interval) vs particle size parameter, $x = 2\pi r/\lambda$. Curves T-2 through T-5 refer to 2-min samples centered at 9.4, 14.4, 20.0, and 33.0 min after dispersal of the dust in the environmental chamber. The first sample, T-1 at 4.2 min showed a heavy concentration of particle clusters and was not analyzed.

Figure 7 shows absorption results for these two frequencies, both as measured and as calculated in the same manner as for quartz. Although ammonium sulfate is trirefringent (quartz was birefringent) the indices do not differ markedly. The size distribution data used in the calculations are shown in the previous figure (6). Here, the $9.25\mu\text{m}$ results are in comparatively close agreement, considering the calculational difficulties, though the time dependencies are different. The $10.22\mu\text{m}$ results for ammonium sulfate show a difference which increases with time--to a point. A consistency which has been noted for the two substances, i.e., quartz and ammonium sulfate, is that the measured spectral dependencies are greater than the calculated quantities and, further, the measured spectral dependence increases with time faster than predicted and then appears to level off. The theory indicates that, for any particulate substance, the spectral dependence becomes independent of particle radius for $r \ll \lambda$. The size distributions of these data approach but do not really satisfy this criterion even at the largest settling time.

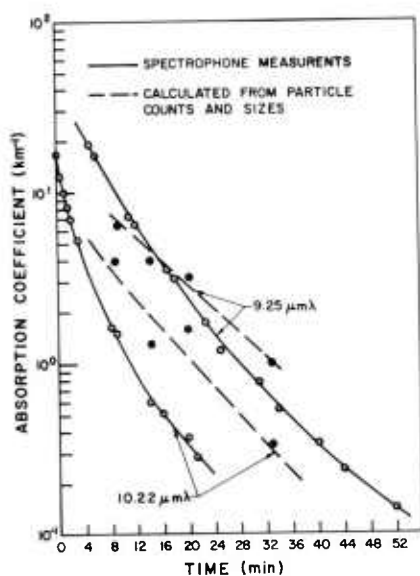


Figure 7. Absorption coefficients of ammonium sulfate dust as a function of time after dispersal, as predicted and measured with a spectrophone, for two wavelengths.

Absorption as a function of particle size parameter is plotted in Figure 8 for $9.25\mu\text{m}$ and $10.22\mu\text{m}$ to show from what size range the significant contributions to the absorption come. For example, at $9.25\mu\text{m}$ approximately 80% of the absorption for the first time period is due to particles between $1.2\mu\text{m}$ and $6.3\mu\text{m}$ and for the last time period particles between $1.0\mu\text{m}$ and $2.9\mu\text{m}$ contribute about the same percentage. Again, at $10.22\mu\text{m}$, 80% of the absorption occurs between $1.2\mu\text{m}$ and $6.3\mu\text{m}$

for the first time period but the radius range is now broader at $1.5\mu\text{m}$ to $3.7\mu\text{m}$ for the last period.

It can be seen from Figure 7 that, particularly for the $10.22\mu\text{m}$ calculated absorption, there is considerable fluctuation with settling time. This apparent sensitivity to details of the size distribution and wavelength was not encountered for quartz. Evidence of this sensitivity is also shown in Figure 9.

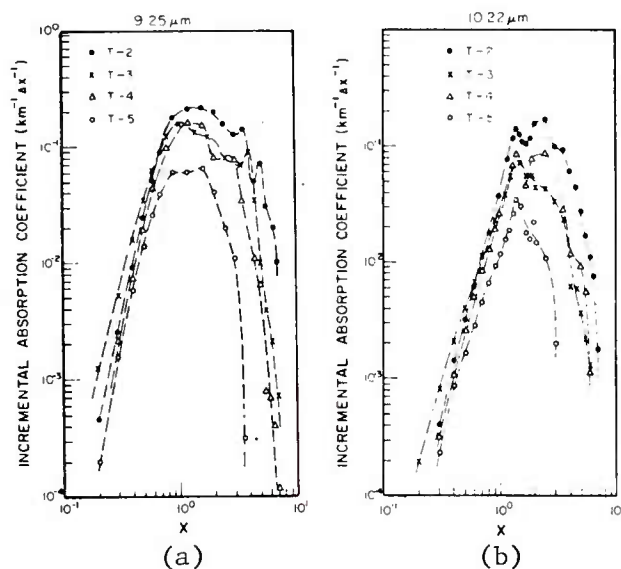


Figure 8. Absorption coefficient per unit size parameter increment as the size parameter for $9.25\mu\text{m}$ (a) and $10.22\mu\text{m}$ (b). Differences in form between the two wavelengths are primarily due to the optical constants.

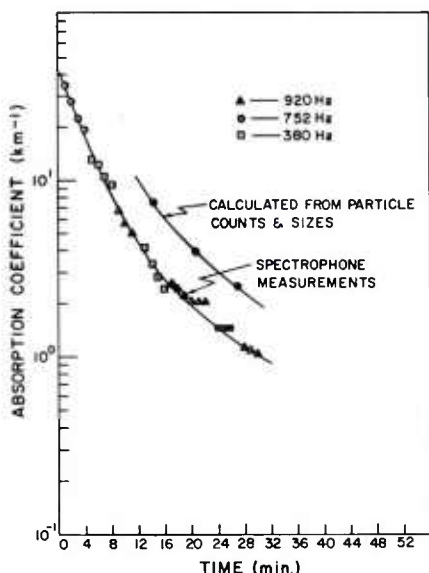


Figure 9. Absorption coefficients of ammonium sulfate dust as a function of settling time after dispersal (similar to Fig. 7) for $9.25\mu\text{m}$ but using three spectrophone chopping frequencies.

This represents another $9.25\mu\text{m}$ series of measurements in the environmental chamber. The spectrophone results are close to those of Figure 7. However, the predicted absorption is now higher though the corresponding "differential" absorption plot does not appear significantly different in form from that of Figure 8.

Another aspect of this particulate data that is of interest relates to the fact that the data were taken for three different chopping rates. The spectrophones were calibrated for each of these rates using the ambient water vapor absorption. Time response calculations indicate that the higher frequencies should be adequate for the particle size distribution encountered, and these data constitute an experimental check on this and should indicate whether the largest particles contributing significantly to the absorption are within the spectrophone response time. The smooth curve shown, in fact, gives no evidence of absorption dependence on the chopping frequencies used.

The overall agreement between spectrophone measurements and calculated values for distributions of particles that are quite irregular, e.g., quartz and ammonium sulfate, is perhaps surprising considering the uncertainty associated with the estimation of the effective cross-sectional areas of the particles. However, measured spectral dependencies do appear to be greater than predicted. The reason(s) for this may be associated with the area estimation just mentioned. If the areas are being overestimated (as is probably the case), the spectral dependence would be reduced. Of course this would also cause a general decrease in magnitude of calculated absorption.

These uncertainties unique to irregular particles should not be of concern in measurement of water droplets found in foggy and hazy environments. The spectral dependence of water absorption in the $9.2\mu\text{m}$ to $10.8\mu\text{m}$ wavelength range is quite flat so the principal question is whether the magnitudes by the two methods agree. The

instrumentation, as discussed, gives several points of comparison in close spatial proximity, i.e., the chamber water vapor partial pressure, the total particulate water mass density, the number density as a function of size using a commercial particle spectrometer and the spectrophone spectral absorption values. The water vapor pressure in the chamber was used to calibrate the spectrophone and is the predominant gaseous absorber for the $10\mu\text{m}$ wavelength region--hence dew point and temperature measurements. A measurement of the liquid water content (LWC) was obtained using the particle spectrometer as well as with separate instruments. The check is an important one because of correlations sought between meteorological parameters and absorption values. Though it is not a very stringent test of the particle spectrometer performance, the LWC results have been found to agree very well with the same quantity as measured with the particle spectrometer.

Since water droplets in fogs tend toward large geometric mean radii, spectrophone time response is once again being investigated as a function of chopping frequency for droplet distributions from the largest to be encountered in the chamber to the smallest. Calibration of the particle spectrometer with respect to response as a function of radius and for density is currently uncertain and under investigation.

A number of spectrophone measurements of the water droplet absorption have been made and compared with calculated results based on the particle spectrometer data. These results show agreement within about a factor of two. Comparative results are currently being obtained for a variety of droplet distributions in an attempt to correlate the absorption with simpler measurements in a simple and direct way.

By the time this paper is presented specially designed spectrophone/particle spectrometer systems will have been operated in a field measurement program to identify atmospheric effects on high energy lasers and EO systems at WSMR. The adaptation of the laboratory in situ absorption measurement systems involves a number of design requirements. As an example, low visibility conditions are of prime interest and, for dusty conditions, this frequently implies high wind velocities. The problem arising from this is that the spectrophone must sample the medium without modifying the size distribution or the densities. To cope with the problem, an intake system based on calculations of aero and particle dynamics has been designed to minimize distortion of the atmospheric quantities.

Another example is found in the requirement to perform sensitive acoustically based measurements in a high noise level environment. A relatively high degree of acoustical isolation is inherent in the spectrophone design (9), but further acoustic damping for the intake system has been designed and tested. Sample results for spectrophone systems operating between 350 and 750 Hz chopping rates are shown in Figure 10. These mechanical dampers are simply inserted in series with the microphone in the spectrophone system.

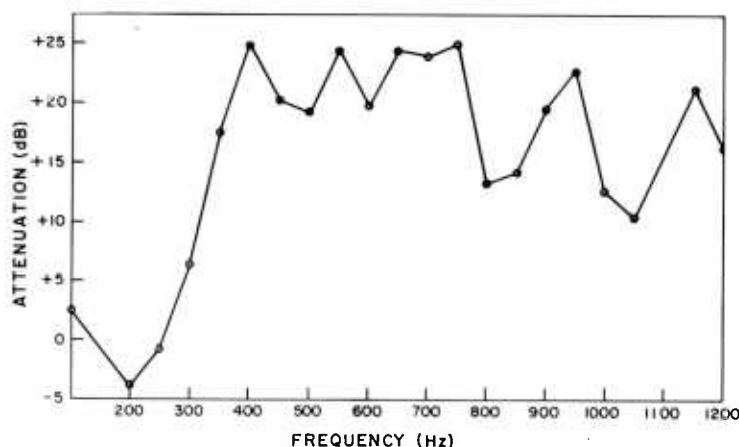


Figure 10. Attenuation of extraneous acoustical noise by spectrophone mechanical filters. Design range from 350 to 750 Hz.

The approach for this measurement program is and has been to study particulate and gaseous absorption under relatively controlled conditions using prototype spectrophone systems and correlative predicted values. On the basis of the experience gained making measurements under laboratory conditions, systems for field measurement programs are then designed, tested and placed in service.

REFERENCES

1. Grams, G. W., I. H. Blifford, Jr., D. A. Gillette, and P. B. Russell, 1974, J. Appl. Meteorol., 13:459.
2. DeLuisi, J. J., P. M. Furukawa, D. A. Gillette, B. G. Schuster, R. J. Charlson, W. M. Porch, R. W. Fegley, B. M. Herman, R. A. Rabinoff, J. T. Twitty, and J. A. Weinman, 1976, J. Appl. Meteorol., 15:455.

*BRUCE, PINNICK, BREWER, YEE, and FERNANDEZ

3. Herman, B. M., R. S. Browning, and J. J. DeLuigi, 1975, J. Atmos. Sci., 32:918.
4. Preliminary results of aerosol absorption measurements (at one wavelength near 10 μ m) made using another type of spectrophone and a modified Royco aerosol counter were published by D. Depatie and J. Lentz, 1974, Air Force Weapons Laboratory Laser Digest Report AFWL-TR-74-344, Air Force Weapons Laboratory, Kirtland Air Force Base, New Mexico. These authors hope to complete their analysis and report their results in the open literature.
5. Bell, A. G., 1880, Proc. Am. Assoc. Advancement Sci., 29:115.
6. Kerr, E. L., and J. G. Attwood, 1968, Appl. Opt., 7:915.
7. Rosencwaig A., 1973, Science, 181:657.
8. Harshbarger, W. R., and M. B. Robin, 1973, Acc. Chem. Res., 6:329.
9. Bruce, C. W., 1976, "Development of Spectrophones for CW and Pulsed Radiation Sources," R&D Report, ECOM-5802, Atmospheric Sciences Laboratory, US Army Electronics Command, White Sands Missile Range, New Mexico.
10. Bruce, C. W., B. Z. Sojka, B. G. Hurd, R. Watkins, K. O. White, and Z. Derzko, 1976, Appl. Opt., 15:2970.
11. Bruce, C. W., 1975, J. Opt. Soc. Am., 65:1163A.
12. Bruce, C. W., 1976, J. Opt. Soc. Am., 66:1071A.
13. Rosen, J. M., 1964, J. Geophys. Res., 64:4673.
14. Pinnick, R. G., J. M. Rosen, and D. J. Hofmann, 1973, Appl. Opt., 12:37.
15. Pinnick, R. G., and D. J. Hofmann, 1973, Appl. Opt., 12:2593.
16. Pinnick, R. G., "Response Characteristics of Knollenberg Light Scattering Aerosol Counters" (to be published).
17. Peterson, J. T., and J. A. Weinman, 1969, J. Geophys. Res., 74:6497.
18. Spitzer, W. G., and D. A. Kleinman, 1961, Phys. Rev., 121:1324.

EXPLOSION ON A SINGLE MOLECULE LEVEL: A CONCEPTUAL
MODEL BASED ON IONIZATION AND FRAGMENTATION
OF TNT UNDER ELECTRON IMPACT

DR. SURYANARAYANA BULUSU
ENERGETIC MATERIALS DIVISION, LARGE CALIBER WEAPON
SYSTEMS LABORATORY, ARMAMENT RESEARCH AND DEVELOPMENT COMMAND
DOVER, NEW JERSEY 07801

INTRODUCTION: Several studies of the mass spectra of nitroaromatic compounds have appeared in recent literature (1) with emphasis on either analytical applications (2-4) or on the modes of electron impact fragmentation (5-7). The chemistry of the nitroaromatic compounds is of considerable interest to the field of explosives because of the central role played by TNT (2:4:6-trinitrotoluene) in a wide variety of armaments. Electron impact studies of explosives and related substances were initiated in this laboratory in order to provide a comparison with other modes of decomposition such as by heat (8,9), light (10) and ionizing radiation. The knowledge thus gained should provide an insight into the nature of the chemical species and processes preceding detonation. Under this program an investigation of nitramines under electron impact was previously published (11) and the present paper deals with 2:4:6-TNT which is the most important of the six isomeric trinitrotoluenes. In addition, to assess the effect of ortho orientation of the CH_3 and NO_2 groups on the fragmentation, comparative data will be presented on 3:4:5-trinitrotoluene which is the only isomer with no nitro groups ortho to the CH_3 group. The purpose of this paper is to indicate the major primary processes in TNT isomers under electron impact and draw attention to the similarities of these to the decompositions initiated by light and heat and to extend the analogies to the processes that precede detonation.

A comprehensive and highly noteworthy study of the electron impact decomposition of TNB (1:3:5-trinitrobenzene) was recently published by Meyerson, et al. (6), which utilized metastable scanning technique and identified a large number of ion transitions. This

study delineates, in a remarkable detail, most of the competing and consecutive reactions which follow the ionization and dissociation of TNB and led to the important conclusion that the loss of NO_2 group was probably the rate determining step in the breakdown of the TNB molecule under electron impact. Our current studies, still in the early stages, are modeled along similar lines to provide a comparison of TNB and TNT in general and to elucidate the key bond breaking steps in the degradation of TNT when ionized by electron impact.

EXPERIMENTAL: Samples of 2:4:6- and 3:4:5- trinitrotoluene were obtained from Stanford Research Institute and recrystallized from isopropyl alcohol before use. The m.p.s were 80.5° and 133.5° , respectively.

The mass spectra reported in this paper were obtained on a DuPont Model 21-492 double focusing mass spectrometer by introducing the samples via the solids probe. An ionization voltage of 70 ev and a source temperature of 200° were used. For metastable scanning the source temperature was maintained at 100° and the sample in the 50 - 70° range to prevent rapid loss of the sample by evaporation and allow time for the metastable scanning.

The metastable scanning technique and its theory have been described in the literature (12, 13). Briefly, it consisted of locating a peak of interest for which precursors were to be found, followed by decoupling the ion accelerating voltage from the electrostatic sector voltage and gradually increasing it. The amplifier output and the variable accelerating voltage were applied to the Y and X coordinates, respectively, of an X-Y recorder. To observe the precursors, it is essential to raise the amplifier gain by a factor of 100 from that employed to detect normal peaks. At the top of each peak the precise accelerating voltage was measured to within ± 0.1 volt using a digital volt meter.

RESULTS AND DISCUSSION: The 70 ev mass spectra of 2:4:6-TNT and the isomeric 3:4:5-TNT are summarized in Tables -1 and -2, respectively. Table -1 for 2:4:6-TNT includes, in addition to the relative abundance, the percentage contribution of ions of each mass to the total ion formation.

Table -3 summarizes the metastable transitions observed by the scanning technique in the high mass end of the TNT spectrum. Current work is aimed to determine as many transitions as possible in the entire mass spectrum. The normal mass spectrum of TNT is rich in metastable peaks and those at the high mass end are shown in Table -4 along with the probable assignments.

TABLE -1: Mass Spectrum of 2:4:6- Trinitrotoluene*

m/e	% Total Ion	RA	m/e	% Total Ion	RA
229	0.13	1.1	92	0.62	5.2
228	0.12	1.0	91	0.37	3.1
227	0.44	3.7	90	1.02	8.5
212	0.31	2.6	89	8.06	67.2
211	1.94	16.2	88	1.76	14.7
210	12.00	100.0	87	1.30	10.8
209	0.72	6.0	86	0.83	6.9
194	0.39	3.3	85	0.31	2.6
193	2.31	19.3	80	0.24	2.0
181	0.25	2.1	79	0.28	2.3
180	2.41	20.1	78	0.56	4.7
179	1.11	9.3	77	1.67	13.9
166	0.80	6.7	76	3.06	25.5
165	0.13	1.1	75	1.57	13.1
164	0.82	6.8	74	1.48	12.3
163	0.37	3.1	73	0.21	1.8
152	0.27	2.3	71	0.62	5.2
151	0.27	2.3	70	0.62	5.2
150	0.19	1.6	69	0.83	6.9
149	1.76	14.7	68	0.28	2.3
136	0.29	2.4	67	0.37	3.1
135	0.41	3.4	66	0.37	3.1
134	2.41	20.1	65	1.02	8.5
133	0.27	2.3	64	1.11	9.3
121	0.14	1.2	63	6.20	51.7
120	0.68	5.7	62	3.52	29.3
119	0.14	1.2	61	0.65	5.4
118	0.22	1.9	60	0.23	1.9
117	0.23	1.9	55	1.95	16.3
116	0.16	1.3	53	0.56	4.7
107	0.22	1.8	52	0.69	5.8
106	0.66	5.5	51	2.50	20.8
105	0.84	7.0	50	2.13	17.8
104	0.72	6.0	49	0.19	1.6
103	0.53	4.4	46	1.2	10.0
102	0.31	2.6	44	0.30	2.5
94	0.27	2.3	43	1.76	14.7
93	0.29	2.4	40	0.17	1.4

TABLE -1: Mass Spectrum of 2:4:6- Trinitrotoluene*

m/e	% Total Ion	RA	m/e	% Total Ion	RA
39	3.61	30.1	29	0.58	4.8
38	0.65	5.4	28	0.65	5.4
37	0.36	3.0	27	0.43	3.6
31	0.12	1.0	26	0.22	1.8
30	8.61	71.8			

*Only peaks of 1% or higher relative abundance are included; however, the % total ion is based on true total ion current.

Several observations can be made from an inspection of the spectra (Tables -1 and -2) of the two trinitrotoluene isomers. In the following discussion only the main features of the spectra will be considered.

The Molecular Ions: In general, the chemical stability of the molecule parallels the stability of M^+ and is therefore reflected in the abundance of M^+ . The stability of M^+ depends on the energy needed to ionize the molecule. If less energy is required to ionize the molecule, more molecular ions will be formed with lower internal energy. Aromatic compounds are well known to yield abundant molecular ions. In the mass spectrum of 2:4:6-TNT the molecular ion at m/e 227 was quite small (3.7%) for an aromatic compound. 3:4:5-TNT, on the other hand, gave rise to a molecular ion of large abundance (70%). This contrast indicates that the instability of the molecular ion of 2:4:6-TNT stems from the position of the nitro group rather than an intrinsic instability associated with the NO_2 group in explosive molecules. This point will be further elaborated later in the following discussion.

The Base Peak in the Spectrum of 2:4:6-TNT: The base peak in the spectrum of 2:4:6-TNT was at m/e 210 (M-17). This represents the loss of an OH group since the elimination of the only other contender, NH_3 , is not a common process (15) and in any case, is unlikely to occur in TNT. A precise mass determination of the peak at m/e 210 assigned the formula, $C_7H_4N_3O_5^+$ to it and confirmed the above conclusion.

Table 2Mass Spectrum of 3:4:5-Trinitrotoluene*

<u>m/e</u>	<u>RA</u>	<u>m/e</u>	<u>RA</u>	<u>m/e</u>	<u>RA</u>
229	1.3	90	7.3	62	25.3
228	6.6	89	60.0	61	6.6
227	69.9	88	4.1	60	1.5
211	3.4	87	10.6	57	3.4
197	4.0	86	8.0	56	1.4
182	1.1	85	2.7	55	2.7
181	1.1	83	1.5	54	1.4
179	1.0	81	3.1	53	10.9
149	1.5	80	1.8	52	23.3
137	1.3	79	2.3	51	55.2
136	1.7	78	8.7	50	16.0
135	10.7	77	51.2	49	1.6
134	3.4	76	9.3	46	34.6
121	1.8	75	4.4	44	10.0
108	1.7	74	7.8	43	4.0
107	20.0	73	2.6	42	1.0
106	2.0	71	2.2	41	5.7
105	6.0	69	3.4	40	3.0
104	1.0	68	1.7	39	41.3
97	1.4	67	4.9	38	4.7
95	1.3	66	2.6	37	4.0
93	1.7	65	12.7	31	1.3
92	1.2	64	9.3	30	100
91	3.4	63	77.3	29	8.7
				27	8.0

*Only peaks of 1% or more of relative abundance are included in this table.

Table 3
Transitions Observed by Metastable
Scanning Technique

<u>Fragment ion</u>	<u>Precursor ion</u>	<u>Approximate Peak Height*</u>
210	227	
193	210	
180	227	29
180	210	6
179	227	35.5
179	209	1.6

*Comparison is valid only when the same fragment was obtained from more than one precursor and gives the relative contribution of each transition to yield the same daughter ion.

Table 4

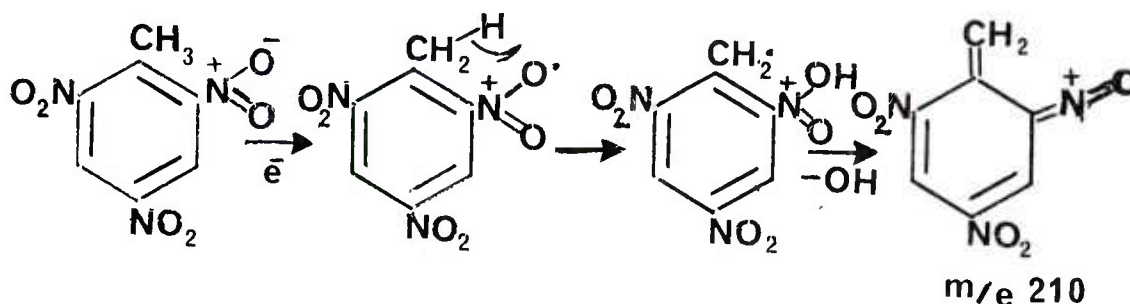
Well Defined Metastable Peaks Observed
in the Normal Mass Spectrum of 2:4:6-TNT
and their Probable Assignment to
Various Likely Transitions

<u>Found</u>	<u>Metastable Peak, m/e</u>		<u>Probable Assignment</u>
		<u>Calculated*</u>	
194.2		194.3	227 $\xrightarrow{-OH}$ 210
192.5		192.4	227 $\xrightarrow{-H_2O}$ 209
177.5		177.4	210 $\xrightarrow{-OH}$ 193
-		164.1	227 $\xrightarrow{-2OH}$ 193
154.5		154.3	210 $\xrightarrow{-NO}$ 180
-		153.3	209 $\xrightarrow{-NO}$ 179
141.3		141.1	227 $\xrightarrow{-H_2O, NO}$ 179
			193 $\xrightarrow{-28}$ 165
128.2		128.1	210 $\xrightarrow{-NO_2}$ 164
124.1		124.0	179 $\xrightarrow{-NO}$ 149

Based on the well known relation, $m^ = m_2^2 / m_1$ for the transition $m_1 \rightarrow m_2 + \text{neutral}$, where m^* represents the apparent mass of m_2 (14).

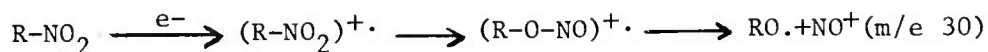
The elimination of a hydroxyl group by TNT obviously requires a prior hydrogen rearrangement (Scheme -I). Such rearrangements by hydrogen atoms and even larger functional groups (e.g., NO_2 migration in HMX noted in Ref. 11) have often been observed (16) in electron impact fragmentations and are facilitated by the ortho orientation of the groups concerned. A very low energy

SCHEME -I

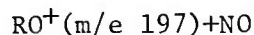


of the transition state involved in this reaction was probably responsible for the low relative abundance of the molecular ion (M^+) in the case of 2:4:6-TNT. When this pathway was preempted in the 3:4:5-TNT the relative abundance of the molecular ion (M^+) increased markedly (70%). Two close analogies are the mass spectra of *o*-nitroaniline and *o*-nitrotoluene both of which gave significant peaks due to the loss of a hydroxyl radical (1) arising from the "ortho" effect.

The Base Peak in the Spectrum of 3:4:5-TNT: The base peak in the spectrum of 3:4:5-TNT was at m/e 30, apparently due to the ion NO^+ . Elimination of a charged or neutral NO is one of the general features of the mass spectra of aromatic nitro compounds (14). If a neutral NO was eliminated, the complementary fragment would retain the charge as was the case here shown by m/e 197. However, the low abundance (4%) of $(M-\text{NO}^+)$ (m/e 197) indicates its relative instability. Nevertheless, the presence of these ions, $(M-\text{NO}^+)$ and NO^+ , suggests yet another interesting bond-forming rearrangement of the nitro group resulting in nitrite formation (Scheme -II).

SCHEME -II

or



where $R = (NO_2)_2 \cdot CH_3 \cdot C_6H_2 -$

The Cleavage of the NO₂ Groups: The loss of NO₂ to give an (M-46)⁺ ion is a common process among nitroaliphatics (17), polynitro compounds (17c), nitrates (18), nitrites (19) and aromatic compounds (1). However, in the spectrum of 2:4:6-TNT the low abundances of the ions, (M-46)⁺ (m/e 181, 2.1%) and NO₂⁺ (m/e 46, 10%) indicated that the loss of NO₂ was either a relatively minor process or that the ions resulting from the NO₂ loss were highly unstable. On the other hand, in the spectrum of 3:4:5-isomer the ion NO₂⁺ was found to have a relative abundance of 35% while the ion (M-46)⁺ had only 1.1% relative abundance.

In the earlier investigation (6) of the electron impact decomposition of 1:3:5-trinitrobenzene (TNB), metastable scanning revealed that the ion (M-46)⁺ was a precursor to almost all the lower mass ions found in its spectrum and not surprisingly, its own relative abundance was of the order of only 5%. This was interpreted to mean that the cleavage of NO₂ group released sufficient internal energy to sustain its decomposition to small fragments. In the two isomeric TNTs under study a similar situation could account for a low abundance of (M-NO₂)⁺ ions. A close comparison of the spectra of the two isomeric trinitrotoluenes shows that the 2:4:6-TNT gives rise to approximately 40% of the total ions in the high mass region (> m/e 89) against about 20% for the 3:4:5-TNT in the same region. This difference is in accord with the ortho effect in the 2:4:6-TNT which should have a stabilizing influence on the molecule by providing a low activation energy pathway.

Metastable Scanning Experiments: Metastable peaks in mass spectra arise from ion decompositions in the field-free region of the analyzer following acceleration (12). With new instrumental techniques such decompositions can be identified and used as a source of information to reconstruct the breakdown paths of molecular ions. Table -3 summarizes the decomposition steps determined. Aside from confirming the one step elimination of OH group by the molecular ion

to give m/e 210, the data indicate several primary steps involving the loss of 2 OH groups, HNO_2 , H_2O , and $\text{H}_2\text{O} + \text{NO}$. Table -4 illustrates the fact that the mass spectrum contains many more metastable transitions which can be potentially correlated with the structure and efforts are being made to accomplish this.

The Relationship of Electron Impact Fragmentation to Photochemistry and Thermal Decomposition of Organic Molecules: The mass spectrum of an organic compound represent the products of ionization and dissociation of isolated molecules in the gas phase which are detected within a few microseconds after electron impact. The lowest energy ionization of a nitro compound in the electron beam will in general remove a non-bonding electron from the nitro group. Much of the photochemistry of the nitro group can be attributed (20) to excitation in the non-bonding orbital. Reactions which emanate from a half-vacant non-bonding orbital should be similar whether the means of removing the non-bonding electron is electronic excitation with light or ionization by electron impact. This is borne out by the parallel between the two important processes observed in the mass spectra of TNTs and the photolysis of nitroaromatics including 2:4:6-TNT. Examples are the photochemical isomerization of *o*-nitrotoluene (21), 2:4-dinitrotoluene (22) and the nitro-nitrite rearrangement of nitroolefines (23) and 9-nitroanthracene (20). In this laboratory *o*-nitrotoluene was found (24) to convert to *o*-nitrosobenzyl alcohol representing an intramolecular hydroxyl migration after the initial hydrogen rearrangement. In fact, most recent ESCA studies (25) of 2:4:6-TNT photolysis showed that this reaction occurs both in the solid and gaseous phases. The loss of an oxygen atom under electron impact by the cleavage of the N-O bond in TNT also has a photochemical equivalent in the formation of nitrosodimethylamine by UV irradiation (10). Several other instances of similarities between mass spectrometric and photochemical reactions were described in the literature (26, 27).

Close parallels between mass spectral and pyrolytic decompositions have often been rationalized as reflecting parallelism of bond energies and vibrational modes in the vibrationally excited neutral molecules and their ionized counterparts. Recent mass spectral and thermal decomposition studies of nitroaromatics (8), polynitropolyphenyls (9) uncovered several reactions common to the two modes. A detailed review (Ref. 28) of these has appeared recently.

The above comparisons between mass spectral reactions and reactions induced by light and heat will serve to support the suggestions often made (23, 26, 27) that the mechanistic mass spectral data helps to rationalize reactions in the other cases and guide the exploratory work.

CONCLUSION: This investigation of the mass spectral breakdown patterns of 2:4:6-TNT and its isomers, has uncovered several interesting reactions which are quite analogous to those inferred in photochemical and thermal decompositions and opened the potential of the metastable scanning approach to enable construction of a multistep decomposition scheme accounting for the disintegration of TNT molecule. Such a correlation was possible in the case of earlier TNB study (6) which showed that the cleavage of the C-NO₂ bond required a high energy transition state and probably constituted the rate-determining step in the disintegration of the TNB molecule. Because of the importance of such information to the understanding of the explosive behavior the authors (6) suggested that the breakdown of TNB in the mass spectrometer source may profitably be viewed as an "explosion" in a system whose dimensions are those of a single molecule. The present work supports this view as a helpful conceptual model.

ACKNOWLEDGEMENT: The author is highly thankful to Dr. Tung-Ho Chen of this laboratory for help with the metastable scanning experiments, and to Mr. J.R. Autera for assistance with instrumentation.

REFERENCES

1. H. Budzikiewicz, C. Djerassi and D.H. Williams, Mass Spectrometry of Organic Compounds, (Holden-Day, San Francisco), p. 512 (1967).
2. R.P. Murrmann, T.F. Jenkins and D.C. Leggett, Special Report #158, Cold Regions Research and Engineering Laboratory, Hanover, NH (1971).
3. R. Boschan and S.R. Smith, US Government Department of Commerce, Office of Technical Services, PB 151120 (1957).
4. J. Yinon, Biomed. Mass Spectrom., 1, 393 (1974).
5. S. Meyerson, I. Puskas and E.K. Fields, J. Ame. Chem. Soc., 88, 4974 (1966).
6. S. Meyerson, R.W. VanderHaar and E.K. Fields, J. Org. Chem., 37, 4114 (1972).
7. J.H. Beynon, Mass Spectrometry and its Applications to Organic Chemistry, Elsevier, Amsterdam, p. 268 (1960).
8. B. Suryanarayana and R.J. Graybush, Proceedings of the 39th Congress on Industrial Chemistry, Brussels, Belgium (1966).
9. S. Bulusu, J.R. Autera and R.J. Graybush, Proceedings of the Army Science Conference, West Point, NY, Vol. 2, p. 423 (1968), Chem. and Ind., 2177 (1967).
10. K. Suryanarayan and S. Bulusu, J. Phys, Chem., 76, 496 (1972).
11. S. Bulusu, T. Axenrod and G.W.A. Milne, Org. Mass Spectrom, 3, 13 (1970).
12. Cooks, Beynon, Caprioli and Lester, Metastable Ions, Elsevier (1973).
13. J.T. Watson, Introduction to Mass Spectrometry, Raven Press (1976).

14. R.W. McLafferty, Interpretation of Mass Spectra, 2nd Edn., W.A. Benjamin, Inc., p. 189 (1973).
15. Ref. 14, p. 159.
16. Ref. 14, p. 56.
17. J. Collin, Bull. Soc. Roy. Sci. Liege, 33, 194 (1954); R.T. Aplin, M. Fisher, D. Becker, H. Budzikiewicz and C. Djerassi, J. Am. Chem. Soc. 87, 4888 (1965); J.T. Larkin, F.E. Saalfeld and L. Kaplan, Org. Mass Spectrom. 2, 213 (1969).
18. R.T.M. Fraser and N.C. Paul, J. Chem. Soc. (B) 659 (1968); R.T.M. Fraser and N.C. Paul, J. Chem. Soc. (B) 1407 (1968).
19. L. D'Or and J. Collin, Bull. Soc. Roy. Sci. Liege 22, 285 (1953).
20. H.A. Morrison, The Photochemistry of the Nitro and Nitroso Groups, in Ch. 4; The Chemistry of the Nitro and Nitroso Groups, Part I, by H. Feuer, Interscience, p. 168 (1969).
21. G. Wettermark, J. Phys. Chem., 66, 2560 (1962).
22. G. Wettermark and R. Ricci, J. of Chem. Phys., 39, #5, pp. 1218-1223 (1963).
23. O.L. Chapman, P.G. Cleveland and E.D. Hoganson, Chem. Comm., 101 (1966).
24. O. Sandus and N. Slagg, Reactions of Aromatic Nitro Compounds, PATR 4385, ARRADCOM, Dover, NJ (1972).
25. J. Sharma, Unpublished results from this laboratory.
26. N.J. Turro, D.C. Neckers, P.A. Leermakers, D. Seldner and P. D'Angelo, J. Ame. Chem. Soc., 87, 4097 (1965).
27. F.W. McLafferty, Anal. Chem., 31, 84 (1959).
28. E.K. Fields and S. Meyerson, Accounts of Chem. Res., 2, 273 (1969).

29. J.C. Hoffsommer, Thermal Stability of Polynitropolyphenyl Compounds at Elevated Temperatures, Report NOLTR 67-118, US Naval Ordnance Laboratory, Silver Spring, MD (1967); J.C. Hoffsommer and J.F. McCullough, Thermal Stabilities of Tetranitrobiphenyl Isomers at Elevated Temperatures, Report NOLTR 68-159, US Naval Ordnance Laboratory, Silver Spring, MD (1968).

HIGH-FREQUENCY HEARING LOSS INCURRED BY
EXPOSURE TO LOW-FREQUENCY NOISE (U)

*CHARLES K. BURDICK, CPT, MSC
JAMES H. PATTERSON, PH.D., BEN T. MOZO, MR.
ROBERT T. CAMP, JR., MR.

U. S. ARMY AEROMEDICAL RESEARCH LABORATORY, FORT RUCKER, AL 36362

The current damage-risk criteria (10) which define the limitations for exposure to continuous noise are specified in terms of the A-weighted levels of noise rather than the unweighted, absolute sound pressure levels of noise. A-weighted levels are derived via an electrical network found in sound measurement equipment. The effect of A-weighting is the de-emphasis or the measurement reduction of the levels of the low-frequency components of a noise. For example, absolute sound pressure levels or octave-band levels appear systemically lower by as much as 70 dB at 10 Hz, 26 dB at 63 Hz, and .8 dB at 800 Hz when measured through an A-weighting network (1). By specifying damage-risk criteria in terms of A-weighted levels, the implicit assumption is that high-intensity, low-frequency sounds are not as harmful to hearing as are high-intensity, high-frequency sounds.

This is of particular concern to the Army because of the large number of vehicles within the Army inventory which generate low-frequency noise at very intense levels. For example, analysis of the running noise generated by the Infantry Fighting Vehicle (IFV) or Cavalry Fighting Vehicle (CFV; formerly designated as the Mechanized Infantry Combat Vehicle, MICV) indicate the presence of high-intensity, low-frequency noise. In general, the greatest amount of acoustic energy occurs below 250 Hz with many low-frequency components exceeding a sound pressure level of 100 dB. The unweighted intensity levels found at 63 Hz are frequently in excess of 120 dB sound pressure level (9).

Another factor which must be considered is that most hearing protectors characteristically provide poor attenuation at the low frequencies (3). Consequently, it is more difficult to protect hearing from high-intensity, low-frequency noise than from high-intensity, high-frequency noise.

Therefore, given the factors that (1) many Army vehicles generate high-intensity, low-frequency noise, (2) most hearing protectors are relatively ineffective for low-frequency noise, and (3) the current damage-risk criteria are based on a measuring scale that de-emphasizes the intensity of low-frequency noise, a research program was developed to determine the potential of high-intensity, low-frequency noise to be a hazard to hearing. The data presented are from three experiments, two with animals and one with humans. An overview and summary of these data are presented.

METHODS AND PROCEDURES

Subjects

Experiment I

The subjects were eight, male binaural chinchillas which were randomly assigned to two groups of four subjects each. Chinchillas were selected as the animal model because their hearing is similar to man's and they are easily trained for behavioral audiometry (7).

Experiment II

Sixteen male, binaural chinchillas served as subjects. They were randomly assigned to four groups of four subjects each.

Experiment III

Five male, young adults served as subjects. Three subjects were tested under one procedure and two subjects were tested with a slightly different procedure.

Procedures

Experiment I

The procedure and apparatus have been discussed in detail elsewhere (2). Briefly, the chinchillas were initially trained

to respond to pure-tone signals. They were trained to move from one side of a shuttlebox to the other when a tone was presented to avoid a mild electric shock. Once trained with clearly audible tones (60-75 dB SPL), hearing thresholds at a number of frequencies were determined by systematically reducing the intensity level of the tones to the point where the subjects failed to emit a response. The level at which the subjects no longer responded was taken as the absolute threshold of hearing for each frequency.

After the baseline hearing thresholds were determined, the subjects were exposed to octave bands of noise in a sound room especially treated to produce a diffuse uniform sound field. The groups exposed to low-frequency noise were exposed to an octave band of noise with a center frequency of 63 Hz and the groups exposed to high-frequency noise were exposed to an octave band of noise with a center frequency of 1000 Hz. These noise bands were selected because they provide an example of two bands of noise that have a 26 dB difference in octave-band level while being equal with regard to A-weighted level. The intensity level of an octave band at 63 Hz is reduced by 26 dB in the conversion from octave-band level to A-weighted level, while the intensity level of an octave band at 1000 Hz is unchanged in the conversion (1). Also, 63 Hz was selected because it is a very high-intensity component of the noise generated by the IFV or CFV (9). The distribution of acoustic energy with frequency of the two noise bands used in the experiments are shown in Figure 1.

Each group of four subjects was exposed on separate occasions to their respective noise band at three intensity levels. The low-frequency group was exposed at levels of 100 dB SPL (74 dBA), 110 dB SPL (84 dBA), and 120 dB SPL (94 dBA). The high-frequency group was exposed at levels of 75 dB SPL (75 dBA), 85 dB SPL (85 dBA), and 95 dB SPL (95 dBA). Adequate time was allowed between exposures for hearing to completely recover and stabilize. The duration of each exposure was three days. Following exposure, the subjects' hearing was monitored until it completely recovered or for 30 days in those cases of permanent hearing loss.

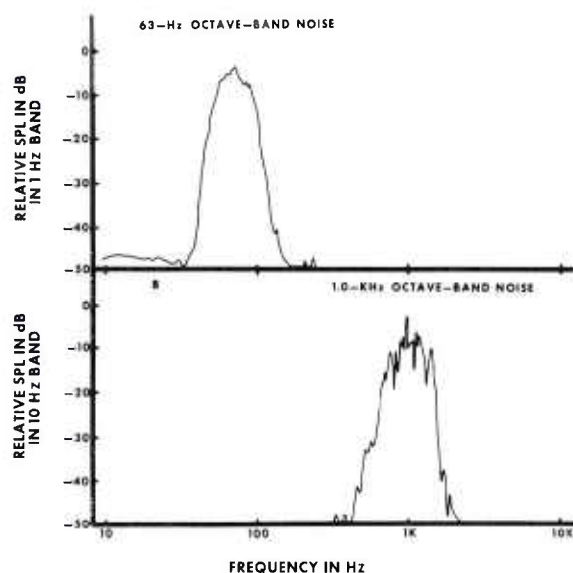


Figure 1. Power spectrum characteristics of the octave bands of noise used for exposure as measured in the sound field. Panel A depicts the octave band with a center frequency of 63 Hz and Panel B depicts the octave band with a center frequency of 1000 Hz.

Experiment II

The training, testing, and exposure conditions were as described above with the following exceptions: (1) of the four groups, two were exposed to the low-frequency noise and two were exposed to the high-frequency noise; (2) each group was exposed only once at a particular intensity level. One low-frequency group was exposed to 110 dB SPL (84 dBA) and the other was exposed to 120 dB SPL (94 dBA). One of the high-frequency groups was exposed to 85 dB SPL (85 dBA), and the other was exposed to 95 dB SPL (95 dBA); (3) the duration of the exposures was nine days.

Experiment III

The baseline hearing thresholds of the five human subjects were determined using the conventional audiometric procedure of tracking. They were then exposed individually to the same noise bands in the same sound field as the chinchillas. The subjects were exposed to the low-frequency noise at levels of 110 dB SPL (84 dBA) and 120 dB SPL (94 dBA), and to the high-frequency noise at levels of 85 dB

SPL (85 dBA) and 95 dB SPL (95 dBA). The duration of the exposures was four hours. The procedural difference between the two groups of subjects primarily involved the use of a different sequence of testing the frequencies during recovery. The hearing thresholds of all subjects were monitored until there was a complete recovery to baseline levels.

RESULTS

Experiment I

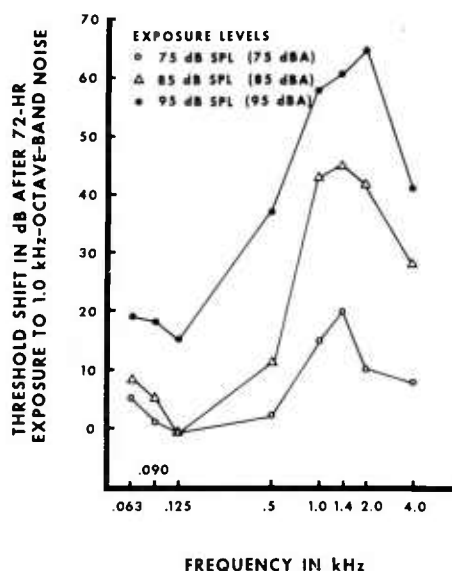


Figure 2. Threshold shifts between .063 and 4.0 kHz after a three-day exposure to octave band noise with a center frequency of 1.0 kHz at three exposure levels.

The threshold shifts for the 1000-Hz octave-band exposures are depicted in Figure 2. These data reflect the traditional pattern of threshold shift found with octave bands of noise with center frequencies of 500 Hz and above (4, 5, 6, 8). For example, as the level of the exposure band was increased systematically, an orderly increase in the amount of threshold shift occurred. For every 10 dB increase in exposure level, a concomitant increase in threshold shift of about 20 dB occurred at the frequencies maximally affected. In addition, the frequency region of greatest sensitivity to the 1000-Hz octave

band of noise occurred one-half to one octave above the center frequency of the exposure band, i.e., 1400 and 2000 Hz. This, too, is a classic finding for high-frequency octave bands of noise (4, 5, 6, 8). At the highest exposure level, the maximum threshold shifts were 61 dB at 1.4 kHz and 65 dB at 2.0 kHz.

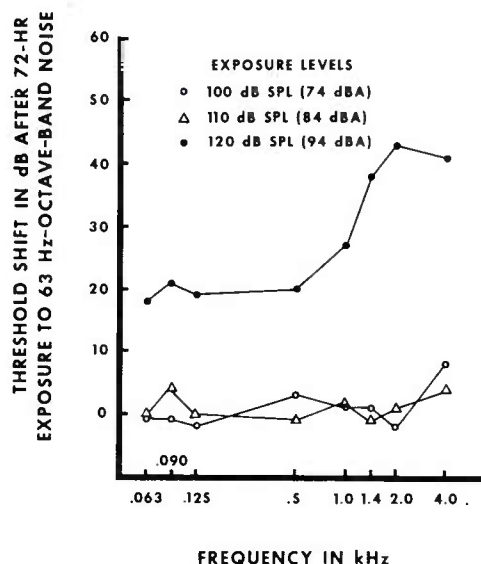


Figure 3. Threshold shifts between .063 and 4.0 kHz after a three-day exposure to octave-band noise with a center frequency of 63 Hz at three exposure levels.

The pattern of threshold shift for the low-frequency band of noise is quite different, however. This is shown in Figure 3. The two lowest exposure levels produced very little, if any, threshold shift. However, the highest level exposure, 120 dB SPL (94 dBA), produced dramatically different results. There was an abrupt appearance of a moderate to substantial threshold shift across the frequency range tested. This was a departure from the traditional results with high-frequency bands of noise. More interesting, however, was the frequency region in which the maximum shift occurred. Rather than occurring one-half to one octave (90-125 Hz) above the center frequency of the exposure band (63 Hz), it occurred five octaves above the center frequency at 2000 Hz. There was a slight elevation at the half-octave frequency of 90 Hz, but this was insignificant by comparison to the shift found at 2000 Hz. The amount of threshold shift found at 1.4 kHz was 39 dB and at 2.0 kHz was 43 dB.

The highest level exposure for the low-frequency noise, 120 dB SPL (94 dBA), produced permanent hearing losses in the subjects, and the highest level exposure for the high-frequency noise, 95 dB

TABLE I. Permanent threshold shifts in dB following three days of exposure to octave-band noise centered at 63 Hz at 120 dB SPL (94 dBA) and to octave-band noise centered at 1000 Hz at 95 dB SPL (95 dBA).

Frequency in kHz	1.4	2.0
63 Hz Exposure Band	11	16
1000 Hz Exposure Band	6	9

SPL (95 dBA), also produced permanent hearing losses. These are given in Table I for 1.4 and 2.0 kHz. All other frequencies completely recovered. Not only did the exposure to the 63-Hz octave band produce a high-frequency hearing loss, a finding not previously known, but also, the low-frequency noise produced nearly twice as much permanent loss as the high-frequency noise. This has particularly important implications for damage-risk criteria since both of the exposure bands were within 1 dBA of each other in level and consequently should be considered equally hazardous. Another interesting result was that the threshold shift from the low-frequency exposures recovered considerably less than the shift from the high-frequency exposures.

Experiment II

The threshold shifts of subjects exposed to the two levels

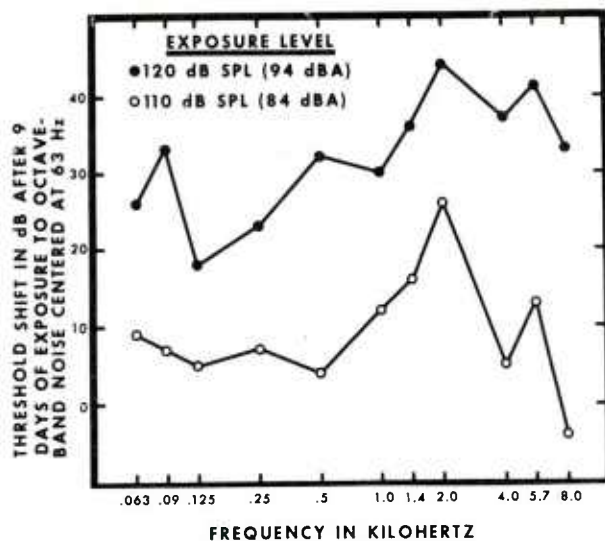


Figure 4. Threshold shifts between .063 and 8.0 kHz after a nine-day exposure to octave-band noise with a center frequency of 63 Hz at two exposure levels.

of high-frequency noise were virtually the same as those found for the three day exposure. Consequently these data are not presented. Figure 4 shows the threshold shifts for the subjects exposed to the low-frequency noise. Unlike the previous experiment, those subjects exposed to the level of 110 dB SPL (84 dBA) showed a small to moderate threshold shift with the maximum shift at 2.0 kHz. This difference between the studies is likely a function of the longer exposure duration. The high-level exposure group, 120 dB SPL (94 dBA), showed considerable shift at all frequencies tested. Again, this is probably due to the longer exposure. This exposure produced the largest shifts in the high frequencies. Threshold shifts in excess of 30 dB and 40 dB occurred at all frequencies between 500 and 8000 Hz as well as at the half-octave frequency of 90 Hz. The maximum shift of 44 dB again occurred at 2.0 kHz. Both groups clearly incurred a high-frequency shift to the low-frequency noise.

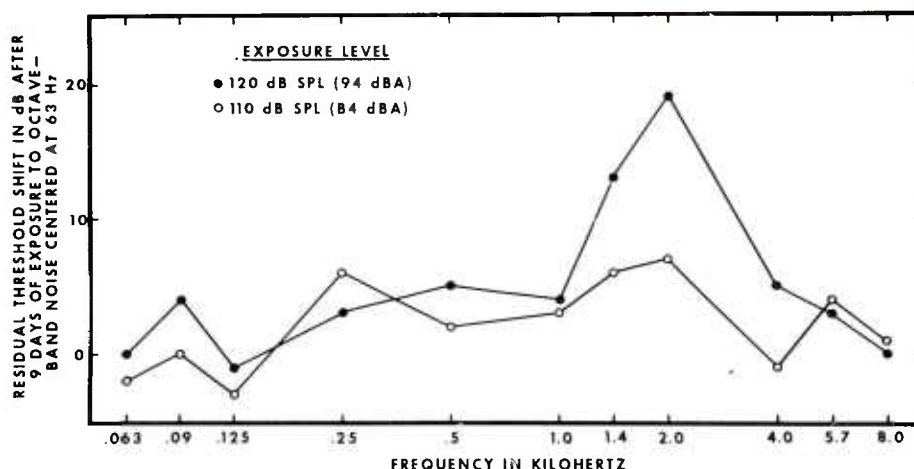


Figure 5. Permanent threshold shifts between .063 and 8.0 kHz obtained 30 days after the termination of a nine-day exposure to octave-band noise with a center frequency of 63 Hz at two exposure levels.

The permanent hearing losses which resulted from the low-frequency exposures are shown in Figure 5. Small permanent losses occurred at 250, 1400, 2000, and 5700 Hz from the 110 dB SPL (84 dBA) exposure, reflecting high-frequency hearing loss to the low-frequency noise. The 120 dB SPL (94 dBA) exposure resulted in permanent losses of 13 dB at 1400 Hz and 19 dB at 2000 Hz which are in excellent agreement with those found for the three-day exposure.

The comparison of permanent hearing losses incurred by the high-level exposure to the low- and high-frequency noise bands is given in Table II. The differences between the two losses are relatively small. Unlike the previous experiment the permanent losses were more in line with the predictions of the damage-risk criteria. That is, the high-frequency noise produced slightly more hearing loss than the low-frequency noise which is compatible with the high-frequency band being 1 dBA more intense. Again, the shifts for the high-frequency noise recovered substantially more than did the shifts for the low-frequency noise.

Table II. Permanent threshold shifts following nine days of exposure to octave-band noise centered at 63 Hz at 120 dB SPL (94 dBA) and to octave-band noise centered at 1000 Hz at 95 dB SPL (95 dBA).

Frequency in kHz	1.4	2.0
63 Hz Exposure Band	13	19
1000 Hz Exposure Band	17	28

Experiment III

The temporary threshold shifts found with the human subjects exposed to low-frequency noise are shown in Figure 6. All subjects recovered to their baseline levels. Although the shape of the threshold shift curves are not identical to those of the chinchillas, a good deal of shift occurred at 1.4 and 2.0 kHz, indicating that low-frequency noise also affects the high-frequency hearing of humans. Some of the difference between the two groups of humans may result from using a different order of testing the various frequencies and an interaction of this with the rapid recovery of hearing that occurs to the short duration exposure. These results indicate that the high-frequency hearing losses found with chinchillas are not unique to the chinchilla and that the human ear tends to respond in a similar way.

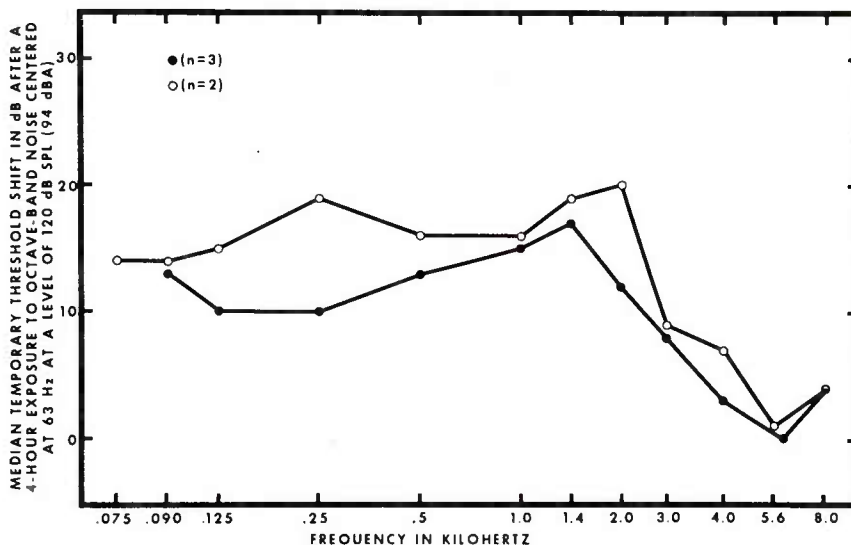


Figure 6. Temporary threshold shifts found in two groups of human subjects between .075 and 8.0 kHz after a four-hour exposure to octave-band noise with a center frequency of 63 Hz at 120 dB SPL (94 dBA).

DISCUSSION

The low-frequency exposures with chinchillas have consistently had their maximum effect on high-frequency hearing. The temporary threshold shifts in humans indicate that the human ear responds to low-frequency noise in a like manner. Also, the high-frequency threshold shifts produced in the animals by low-frequency noise showed considerably less recovery than high-frequency shifts produced by high-frequency noise. This indicates that low-frequency noise may have a more "potent" effect on hearing than high-frequency noise. These results together indicate that high-intensity, low-frequency noise may be a hazard to hearing which was previously unrecognized. Because A-weighting de-emphasizes the effects of low-frequency noise, the current damage-risk criteria may be inadequate and allow exposure of personnel to hazardous acoustic environments. Although the evidence is not clear-cut at this time, the findings of these experiments emphasize the need for much more research and raise serious questions concerning the adequacy of the current damage-risk criteria with regard to low-frequency noise.

Although the implications of the results for the damage-risk criteria are equivocal at this time, the new finding of a high-frequency hearing loss from low-frequency noise is significant. All previous research on noise-induced hearing loss has found that the maximum effect on hearing of a band of noise occurs one-half to one octave above the frequency of the noise band. A persistent problem in attempting to account for noise-induced hearing loss has been the consistent finding of primarily high-frequency hearing losses regardless of the noise source. These previous findings account for the emphasis currently placed on high-frequency noise as the primary hazard to hearing. It is now known, from the results presented in this paper, that the effect of low-frequency noise is far removed in frequency from the frequency of the exposure band. The present findings show, for the first time, that high frequency hearing losses can be induced by low-frequency noise. This may account for the phenomenon of high-frequency hearing loss in individuals exposed to low-frequency noise. These results also clarify the enigma concerning the consistent failure to find permanent low-frequency hearing losses in individuals exposed to low-frequency noise.

CONCLUSIONS

The following conclusions are made: (1) low-frequency noise produces permanent high-frequency hearing loss in chinchillas; (2) the human ear shows a similar pattern of temporary threshold shift; (3) the current damage-risk criteria may be inadequate and may require revision to deal with high-intensity, low-frequency noise.

ACKNOWLEDGEMENTS

In conducting the research in Experiments I and II, the investigators adhered to the "Guide for the Care and Use of Laboratory Animals," as promulgated by the Committee on Revision of the Guide for Laboratory Animal Facilities and Care of the Institute of Laboratory Animal Resources, National Research Council.

In conducting the research in Experiment III, the human subjects participated in this study after giving their free and informed voluntary consent. Investigators adhered to AR 70-25 and USAMRDC Reg 70-25 on USE of VOLUNTEERS in RESEARCH.

REFERENCES

1. ANSI (1971) S1.4, Specification for Sound Level Meters. American National Standards Institute, New York.
2. Burdick, C. K., Patterson, J. H., Mozo, B. T., and Camp, R. T., Jr. Threshold Shifts in Chinchillas Exposed to Octave-Bands of Noise Centered at 63 and 1000 Hz for Three Days. Journal of the Acoustical Society of America, 1978, In Press.
3. Camp, R. T., Jr., Mozo, B. T., Kuc, L. F., and Schott, G. A. Real-Ear Sound Attenuation Characteristics of Hearing Protective Devices Available Through Federal Supply Channels. U. S. Army Aeromedical Research Laboratory, Report LR-72-12-2-6, 1972.
4. Carder, H. M. and Miller, J. D. Temporary Threshold Shifts Produced by Noise-Exposure of Long Duration. Transactions of the American Academy of Ophthalmology and Otolaryngology, 1971, 75, 1346-1354.
5. Carder, H. M. and Miller, J. D. Temporary Threshold Shifts from Prolonged Exposure to Noise. Journal of Speech and Hearing Research, 1972, 15, 603-623.
6. Henderson, D., Hamernik, R. P., Dosanjh, D. S., and Mills, J. H. Effects of Noise on Hearing. Raven Press: New York, 1976.
7. Miller, J. D. Audibility Curve of the Chinchilla. Journal of the Acoustical Society of America, 1970, 48, 513-523.
8. Mills, J. H. Temporary and Permanent Threshold Shifts Produced by Nine-Day Exposures to Noise. Journal of Speech and Hearing Research, 1973, 16, 425-438.
9. Patterson, J. H., Mozo, B. T., and Camp, R. T. Jr. Preliminary Medical Assessment of the Acoustic Hazard of the Prototype Mechanized Infantry Combat Vehicle. U. S. Army Aeromedical Research Laboratory, Report 76-16, 1976.
10. TB MED 251. Noise and Conservation of Hearing. HQDA, 7 March 1972.

INTENSIVE TROPIC FUNCTION TESTING

ELDON M. CADY, MR.
US ARMY TROPIC TEST CENTER
APO NEW YORK 09827

INTRODUCTION:

Tropic testing of US Army materiel includes a storage phase designed to surface the adverse effects of the humid tropics. Failures are sometimes catastrophic, but are usually time dependent. Regulations such as AR 1000-1 (1) require that efforts be made to reduce Development Test time. Project Managers and DARCOM commodity commands have curtailed or foregone Development Tests because of excessive time/cost considerations. It was hypothesized that reducing test calendar time while increasing test functioning time, i.e., increasing the ratio of operational hours to calendar days may yield quicker and still valid test results for some categories of equipment. Large quantities of Reliability, Availability and Maintainability (RAM) data could be generated quickly for immediate analysis using standard RAM data analysis. A methodology investigation was conducted at the US Army Tropic Test Center to validate the intensified testing concept and reassess storage testing.

The tropic storage period results in a relatively long period of dormancy between data generation and the final production of test reports. Because the storage period represents a significant amount of calendar time, it had been repeatedly proposed that the storage phase be shortened or eliminated and that each test item be tested at an intensified functioning rate. Advocates of intensified functioning considered that the same quality of RAM data would be generated in a shorter calendar period. This is based on the assumption that number of operational hours is the only factor required for developing valid RAM data. The Tropic Test Center has observed materiel failures which occurred in the tropic storage phase of Development Tests. Examples of

such failures recently occurred during tests of the Forward Area Alert Radar (2), the OH-58A Helicopter (3), and the Modular Collective Protection Equipment (4). Although these failures were not identified with a single unique aspect of tropic exposure, they occurred during the tropic storage phase.

The US Army's standard 1.5 KW AC generator was selected as the test vehicle because it incorporated both electrical and mechanical components, it could be easily operated and had readily available maintenance support. The generators are high density items in the US Army supply system, i.e., available in large quantities, thus reducing project expenditures. The main advantage of using generators for this investigation was the capability to generate a large quantity of RAM data.

PROCEDURES:

The test site selected was an abandoned concrete pad in the Fort Clayton General Purpose test area in the humid tropics of the Canal Zone. Fifteen 1.5 KW AC generators were separated into three groups of five each (Figure 1), so that one group would be intensively functioned, the second would undergo a storage phase, and the third would simulate usage in the field. The generators were operated over a period of one year in the same test area using a single fuel source and equivalent variable power loads.

One group (Intensive Function Mode) was functioned at a rate of 16 hours per day for one year. The second group (Storage Mode) was functioned at a rate of four hours per day for 100 operational hours, after which the generators were placed in limited field storage as specified in the unit maintenance manuals (5). The storage period lasted six months, with an inspection at the end of the first three months. After the six-month storage, this group was returned to operation with the same usage schedule. The third group (Simulated Tactical Use Mode) was functioned at a rate of four hours per day for one year.

Power was drawn from each unit by a series of five 300-watt light bulbs which were so wired that the power drawn could be varied from 0 to 1500 watts in 300-watt stages. The load bank system is illustrated in Figure 2. Wired into the load banks and connected to a central control panel were meters which measured operational time, voltage output and voltage frequency. Figure 3 shows the control console with an oscilloscope attached for measuring the transients involved in load level changes. Load level changes took place every hour for a duration of 15 seconds, with the power draw change being from 900 to 1500 watts. Figure 4 shows the shunt resistor across which the stable



Figure 1. Test Site with Generators Mounted on a Concrete Pad



Figure 2. Load Bank System for Power Dissipation

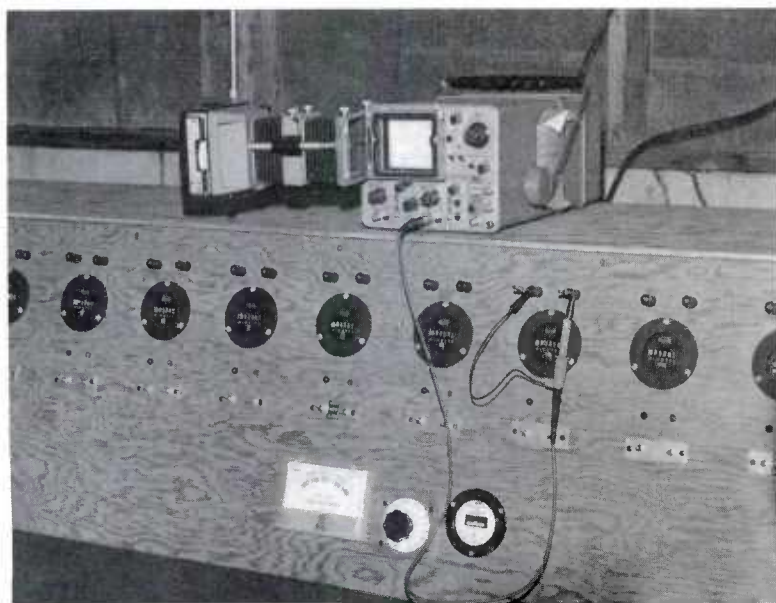


Figure 3. Control Console with Oscilloscope Attached

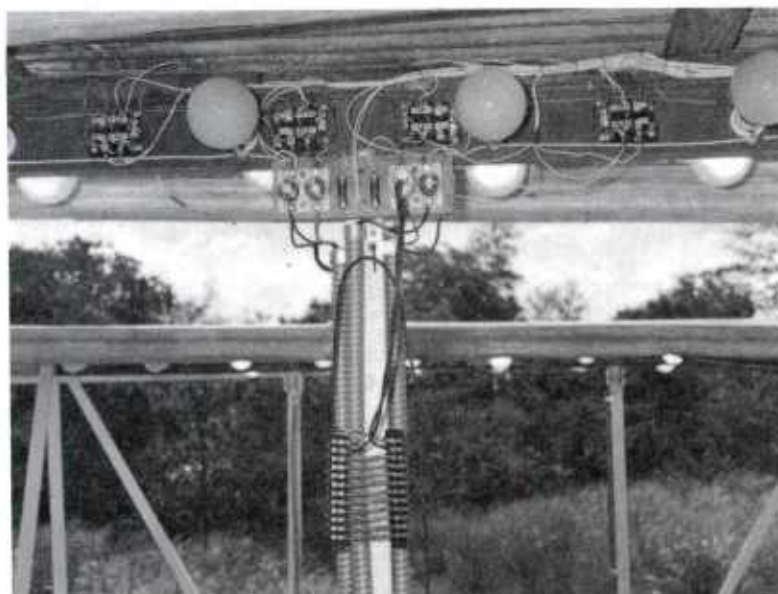


Figure 4. Shunt Resistor for Measuring Current Levels

and transient current levels were measured. Voltage and frequency levels for each generator were monitored and recorded on an hourly basis by an operator who reported malfunctions. Close monitoring was intended to identify the onset of a malfunction in order to initiate prompt maintenance action.

DISCUSSION OF RESULTS:

A summary of the data used in the analysis is presented in Tables I, II, and III. These tables present the basic data and computed RAM parameters for each of the functional modes and individual generators. Of these parameters, Mean Time Between Failures (MTBF) and Mean Time Between Maintenance Actions (MTBMA) are the most useful in assessing RAM performance. The discussion is confined to functional modes rather than individual generator performance. To establish a commonality, the data have been normalized to a per 1000 hours of operation basis.

In Tables I, II, and III there are five items of data per functional mode that lend themselves for RAM performance comparisons. These are the MTBF, MTBMA, Number of Chargeable System Failures per 1000 Hours (CSF/1000), Maintenance Actions per 1000 Hours (MA/1000) and Unscheduled Maintenance Time per 1000 Hours (UMT/1000). The MTBF is inversely proportional to CSF/1000 and the MTBMA is inversely proportional to MA/1000; therefore, only the values for the MTBF, MTBMA, and UMT/1000 are used as comparison parameters.

The MTBF was 243.0 hours for the Intensive Function Mode, 176.6 hours for the Storage Mode and 182.4 hours for the Simulated Tactical Use Mode. This indicates that the Intensive Function Mode had the longest operation period between failures, and that the Storage and Simulated Tactical Use Modes were similar, with the Storage Mode being slightly less. The second parameter, the MTBMA for the Intensive Function Mode, was 134.2 hours; the Storage Mode was 56.2 hours; and the Simulated Tactical Use Mode was 94.8 hours. The ranking for MTBMA is the same as for MTBF, except that the difference between the Storage Mode and the Simulated Tactical Use Mode is larger.

The UMT/1000 in Tables I, II, and III for the Intensive Function Mode was 2.8 hours, whereas the Storage and Simulated Tactical Use Modes were both 7.2 hours. Again, the Intensive Function Mode differs considerably from the other modes.

Additional computed RAM data are provided in Table IV for the functional modes. The MTBF and MTBMA data are also presented for convenience. Another useful RAM parameter is the Mean Time to Repair

TABLE I. SUMMARY OF INTENSIVE FUNCTION MODE DATA

Generator Number	Total Test Hrs	Number* Chargeable System Failures	Number* Chargeable Maint Act per 1000 Hrs	Number Chargeable System Failure Time (Man-Hrs)	MTBF (Hrs)	MTEMA (Man-Hrs)	Active Maintenance Time*			
							Unscd (Man-Hrs)	Scd (Man-Hrs)	Total (Man-Hrs)	Unscd Maint Time per 1000 Hrs
1	878.6	6	4	6.8	4.6	3.6	219.7	146.4	3.0	6.6
2	2033.0	18	10	8.9	4.9	5.9	203.3	112.9	6.1	12.0
3	1867.5	17	10	9.1	5.4	5.7	186.8	109.9	6.1	11.8
4	2077.9	10	4	4.8	1.9	3.4	519.5	207.8	5.1	8.5
5	2133.1	16	9	7.5	4.2	6.7	237.0	133.3	4.6	11.3
Total	8990.1	67	37	—	—	25.3	—	—	24.9	50.2
Mean	1798.0	13.4	7.4	7.5	4.1	5.1	243.0	134.2	5.0	10.0
Standard Deviation	523.4	5.2	3.1	—	—	1.5	—	—	1.3	2.4

TABLE II. SUMMARY OF STORAGE MODE DATA

6	502.4	10	4	19.9	8.0	3.0	125.6	50.2	3.0	11.0	6.0
7	501.3	7	1	14.0	2.0	1.5	501.3	71.6	1.5	8.0	3.0
8	464.4	11	5	23.7	10.8	10.5	92.9	46.4	10.5	8.0	18.5
9	502.7	8	2	15.9	4.0	1.4	251.4	62.8	1.4	8.1	9.5
10	501.5	9	2	17.9	4.0	1.4	250.8	55.7	1.4	7.6	9.0
Total	2472.3	44	14	—	—	17.8	—	—	39.7	57.5	—
Mean	494.5	9.0	2.8	17.8	5.7	3.6	176.6	56.2	7.9	11.5	7.2
Standard Deviation	16.8	1.6	1.6	—	—	3.9	—	—	0.2	4.0	—

TABLE III. SUMMARY OF SIMULATED TACTICAL USE DATA

11	1024.1	5	1	4.9	1.0	1.0	1024.1	204.8	1.0	2.1	3.1	1.0
12	978.2	13	8	13.3	9.2	13.2	108.7	75.3	13.2	3.0	16.2	13.5
13	732.3	15	10	20.5	13.7	13.5	73.2	48.8	13.5	3.5	17.0	18.4
14	1004.0	11	5	11.0	5.0	4.0	200.8	91.3	4.0	4.0	8.0	4.0
15	1002.8	6	2	6.0	2.0	2.5	501.4	167.1	2.5	3.5	6.0	2.5
Total	4741.4	50	27	—	—	34.2	—	—	34.2	16.1	50.3	—
Mean	948.3	10	5.4	10.5	5.7	6.8	182.4	94.8	6.8	3.2	10.1	7.2
Standard Deviation	121.8	4.4	4.0	—	—	6.0	—	—	6.0	0.7	6.2	—

* Excludes oil changes

TABLE IV. SUMMARY OF COMPUTED RAM DATA

1	Mean Time Between Failures	
	Test Modes	
	a. Intensive Function	243.0
	b. Storage	176.6
	c. Simulated Tactical Use	182.4
2	Mean Time Between Maintenance Actions	
	Test Modes	
	a. Intensive Function	134.2
	b. Storage	56.2
	c. Simulated Tactical Use	94.8
3	Mean Time To Repair	
	Test Modes	
	a. Intensive Function	0.0
	b. Storage	1.3
	c. Simulated Tactical Use	1.0
4	Mean Time To Repair Unscheduled Maintenance Actions	
	Test Modes	
	a. Intensive Function	0.7
	b. Storage	1.3
	c. Simulated Tactical Use	1.3
5	Maintenance Ratio	
	Test Modes	
	a. Intensive Function	0.006
	b. Storage	0.023
	c. Simulated Tactical Use	0.011
6	Maintenance Ratio For Unscheduled Maintenance Actions	
	Test Modes	
	a. Intensive Function	0.003
	b. Storage	0.007
	c. Simulated Tactical Use	0.007

(MTTR). It is a measure of the degree of difficulty in performing maintenance actions, and was obtained by dividing the total active maintenance time for that functional mode by the total number of maintenance actions. The MTTR involves all maintenance actions, scheduled and unscheduled. For the MTTR, the Intensive Function Mode is 0.7 hours, the Storage Mode is 1.3 hours, and the Simulated Tactical Use Mode is 1.0 hour. Again, the Intensive Function Mode has required less time for repair than either of the other modes. Another MTTR value is for Unscheduled Maintenance Actions which is concerned primarily with actions involving generator malfunctions. In this case the Intensive Function value is 0.7 while the Storage and the Simulated Tactical Use values are both 1.3. The Intensive Function Mode required about half the time of the other two functional modes. The two remaining parameters in Table IV are the Maintenance Ratio and the Maintenance Ratio for Unscheduled Maintenance Actions. These are computed by dividing the active maintenance time, either total or unscheduled, by the total operational hours of each functional mode. The Maintenance Ratio for the Intensive Function Mode was 0.006, the Storage Mode was 0.023, and the Simulated Tactical Use Mode was 0.011. In the Maintenance Ratio for Unscheduled Maintenance Actions, the results were Intensive Function Mode 0.003, Storage Mode 0.007, and Simulated Tactical Use Mode 0.007. Both types of Maintenance Ratios show that the Intensive Function Mode produces lower values than the other two functional modes, and again the Storage and Simulated Tactical Use Modes had equivalent values.

It was assumed that the computed data (Tables I, II, III, and IV) are measures of the degree of operational severity for each functional mode to permit the construction of an arbitrary matrix for comparative purposes. The matrix (Table V) was based on six parameters (MTBF, MTBMA, UMT/1000, MTTR, MTTR Unscheduled Maintenance Actions, and Maintenance Ratio for Unscheduled Maintenance Actions) selected from the previous tables. For each parameter, a numerical value of one was given to the functional mode that was considered least severe; the value of two was given for the median severity; and the value of three for the most severe. In the case of two modes being equal for a given parameter, a value of 1.5 or 2.5 was assigned to both depending on degree of severity. The data are summarized in Table V. For all six parameters the Intensive Function Mode was given the value of one for being least severe. For three (MTBF, MTBMA, and MTTR) of the six parameters, the Storage Mode was the most severe and for the other three parameters it was equal to the Simulated Tactical Use Mode in severity. The ranking from least to most severe was as follows: Intensive Function Mode, Simulated Tactical Use Mode and Storage Mode, with a proportionality of 1:2.3:2.8, respectively.

TABLE V. SEVERITY OF OCCURRENCE

SOURCE							
TABLES I, II AND III				TABLE IV			
TEST MODE	MTBF	MTBMA	Unscd Maint Time per 1000 Hrs	Mean Time to Repair Unscd Maint Actions	Mean Time to Repair Unscd Maint Actions	Maint Ratio for Unscd Maint Actions	Summary Of Rows
Intensive Function	1	1	1	1	1	1	6
Storage	3	3	2.5	3	2.5	2.5	16.5
Simulated Tactical Use	2	2	2.5	2	2.5	2.5	13.5
1 Least Severe Test							
2 Median Severe Test							
3 Most Severe Test							

NOTE: If equal in value, then assigned value will be 1.5 or 2.5 based on severity rank.

Tables I, II, and III also provide information on individual generators so that questions may be asked concerning the types of malfunctions that occurred and major differences in the types of malfunctions. To examine these questions, Tables VI and VII were prepared to summarize the data in appropriate form. Table VI is a Summary of Unscheduled Maintenance Actions by type of maintenance action performed and time involved for each functional mode. The number of maintenance actions, in decreasing order, were carburetor, ignition and voltage regulator, regardless of functional mode. To establish a baseline for comparative purposes, Table VII presents Normalized Unscheduled Maintenance Actions. The scheduled maintenance actions were not considered because some involved work unique to each functional mode. This table also includes the average time to perform each action. For carburetors, the Storage Mode had more actions per 1000 hours, thus more time charged per 1000 hours, but the actual time per action for all three modes was about the same. This indicates that similar types of maintenance actions were performed on the carburetors, regardless of the functional mode. However, voltage regulators in the Storage Mode, required the least actions per 1000 hours. The Time Charged per 1000 Hours for the voltage regulator, the Storage Mode is half way between the Intensive Function and the Simulated Tactical Use Modes. However, the Time per Action for the storage mode is the highest. This means that, although the frequency of maintenance actions was less on voltage regulators of the Storage Mode, the amount of time required for that action was greater. For the ignition category, it is found that the Storage Mode had the highest number of maintenance actions performed per 1000 hours but the Simulated Tactical Use Mode had the most time per action. Although the number of actions per 1000 hours for Intensive Function Mode ignition maintenance actions is similar to the other two modes, there is a major difference in its value for the Time Charged per 1000 Hours and also for the Time per Action. This indicates that there is a difference in the type and degree of difficulty involved in the ignition maintenance actions performed during the Intensive Function Mode when compared with the other two modes.

This discussion has dealt with examining the three functional modes by operational hours instead of by calendar time. Because operation hours are rapidly accumulated in the Intensive Function Mode and the Storage Mode involves a long dormant period, it was determined that biased data would result from comparing the two modes by calendar time. Although it is recognized that the Intensive Function Mode produces malfunctions rapidly in a much shorter calendar period and produces RAM data in a shorter time, the rate and types of failure produced differ in comparison with the other two test modes.

TABLE VI. SUMMARY OF UNSCHEDULED MAINTENANCE ACTIONS

Type Of Maintenance Action	Intensive Function			Storage Mode			Sim Tactical Use		
	Number Of Actions	Time Charged	Number Of Actions	Time Charged	Number Of Actions	Time Charged	Number Of Actions	Time Charged	Time Charged
Carburetor	14	11.5	7	6.0	6	7.1			
Voltage Regulator	5	3.0	1	1.5	5	4.5			
Ignition	11	7.6	4	4.7	6	10.1			
Fuse Breaker	3	0.4	—	—	—	—			
Valve	1	0.1	—	—	2	2.5			
Frequency Meter and Converter	—	—	—	—	3	2.5			
Miscellaneous	3	2.7	2	5.6	4	7.5			
Total	37	25.3	14	17.8	26	34.2			

TABLE VII. SUMMARY OF NORMALIZED UNSCHEDULED MAINTENANCE ACTIONS

TYPE OF MAINTENANCE ACTION	INTENSIVE FUNCTION				STORAGE MODE				SIMULATED TACTICAL USE			
	Unscd Maint Act per 1000 Hrs	Unscd Maint per 1000 Hrs	Time per Action (Hr)	Unscd Maint Act per 1000 Hrs	Unscd Maint per 1000 Hrs	Time per Action (Hr)	Unscd Maint Act per 1000 Hrs	Unscd Maint per 1000 Hrs	Unscd Maint Act per 1000 Hrs	Time per Action (Hr)	Unscd Maint per 1000 Hrs	Time per Action (Hr)
Carburetor	1.6	1.3	0.8	2.8	2.4	0.9	1.5	1.5	1.5	1.0		
Voltage Regulator	0.6	0.3	0.6	0.4	0.6	1.5	1.1	0.9	0.9	0.9		
Ignition	1.2	0.8	0.7	1.6	1.9	1.2	1.3	2.1	2.1	1.7		
Miscellaneous	0.3	0.3	0.9	0.8	2.3	2.8	0.8	1.6	1.6	1.9		

CADY

CONCLUSIONS:

On the basis of the results of the investigation discussed herein, the following conclusions are offered:

1. In terms of test severity, as depicted by the RAM parameters analyzed, the Intensive Function Mode ranked the least severe, the Simulated Tactical Use Mode next, and the Storage Mode was the most severe.

2. Materiel items having electro-mechanical characteristics similar to the 1.5 KW AC generators and tested in an Intensified Function Mode will probably not produce higher failure rates than Storage or Simulated Tactical Use Modes.

3. Malfunctions which occurred were of similar nature for all three modes (i.e., carburetor, voltage regulator, and ignition); however, the time required to correct the malfunctions was less for the Intensive Function Mode.

4. Pertinence of the findings of this investigation should be explored for other materiel systems such as electronic and various weapon systems.

5. The need for a tropic storage phase in development testing is essential and is supported by results of this investigation. The optimum period for the storage phase requires further investigation.

REFERENCES:

1. AR 1000-1 Basic Policies for System Acquisition by the Department of the Army, 5 Nov 75.

2. US Army Air Defense Board Final Report, DT II, Service Phase, Forward Area Alerting Radar (FAAR) (Tropic Environment) (U), TECOM Project No. 3 EE MPQ 049 034, Mar 74.

3. US Army Aviation Test Board Final Report, Tropic Service Test of OH-58A Helicopter System, TECOM Project No. 4-A1-130-58A-010, 19 Jan 71.

4. US Army Tropic Test Center Final Report, DT II, Tropic Integrated Engineering and Service Phase of Modular Collective Protection Equipment (MCPE), TECOM Project No 2 ES 825 MCP 005, Sep 76.

CADY

5. DA Technical Manual, TM-5-6115-323-15, Generator Set, Gasoline Engine Driven, Skid Mounted, Tubular Frame, 1.5 KW, Single Phase, AC, 120/240V, 9 September 1970.

ENERGETIC TRANSIENT SPECIES FORMED
VIA ELECTRONIC EXCITATION OF S-TNB AND S-TNT (U)CHRISTOS CAPELLOS, PhD AND *SURY IYER, PhD¹
ENERGETIC MATERIALS DIVISION, LCWSL
ARRADCOM, DOVER, NJ 07801

INTRODUCTION: The initiation of exothermic explosive decompositions in materials, such as, s-TNT² (2,4,6-trinitrotoluene) is probably due to the initial formation of energetic transients, like electronically or vibrationally excited states, ionic species or free radicals. The stimuli which can cause such an event can be an electronic, shock wave or thermal pulse. So far, there has been no information in the literature pertaining to the formation and reactivity of these energetic transients in secondary explosives.

In the present paper data are presented on the spectroscopy and chemical reactivity of ionic species, electronically excited states and free radicals of nitro aromatics formed via electronic excitation of the parent molecules. Furthermore, it was discovered in the present work, that formation of the TNT anion, $(C_6H_2)(NO_2)_3CH_2^-$, (B^-) , in molten TNT by additives such as $(CH_3)_4NB_3H_8$, known as QMB₃, NaOH, and NaI lead to deflagration of TNT. These additives sensitized TNT to impact and differential scanning calorimetry yielded data which are consistent with the sensitization of TNT by these additives.

It was also discovered in this work that TNT can be sensitized by additives such as benzoquinone and tetraethyl lead. It is believed that these additives abstract a hydrogen atom from the methyl group of TNT during thermal excitation to form the radical $(C_6H_2)(NO_2)_3CH_2\cdot$, $(B\cdot)$. Work is in progress for the formation and detection of this radical in electronically excited TNT in the gas phase at 215°C.

¹Formerly known as: K. Suryanarayanan. ²Throughout this report the notations TNB and TNT refer to the symmetric form.

EXPERIMENTAL

All solvents used in this work were spectrograde quality. TNB, TNT and the nitronaphthalenes reported here were purified by successive recrystallizations from ethanol or n-hexane followed by sublimation under vacuum until the gas chromatogram of the sample showed that the impurity level was below 0.001%. Deaeration of the solutions was done by argon or helium flushing for 30 minutes in a flask attached to the optical cell. Electronic excitation of liquid and gaseous samples of TNB and TNT was achieved with a flash photolysis system which has been described earlier (1). Spectroscopic and kinetic studies of nitronaphthalenes in the nanosecond range were performed with a ruby laser nanosecond kinetic spectrophotometer as described in detail previously (2). Most of the kinetic data presented here were analyzed with a Nova 840 Computer. The kinetic oscilloscope traces were analyzed for first or second order plots, and the result was displayed on a 4010 Tektronix terminal with the use of graphics subroutines. After determining the order of the reaction, the least square line was plotted from the slope of which the rate constants were calculated.

RESULTS AND DISCUSSION

I. Energetic Transients Formed From Electronically Excited TNB

Electronic excitation of TNB in aerated polar solvents with appreciable proton affinity, such as, i-PrOH (A) and MeOH (B) results in the formation of a transient species with absorption maxima at 430 and 510 nm (Fig 1).

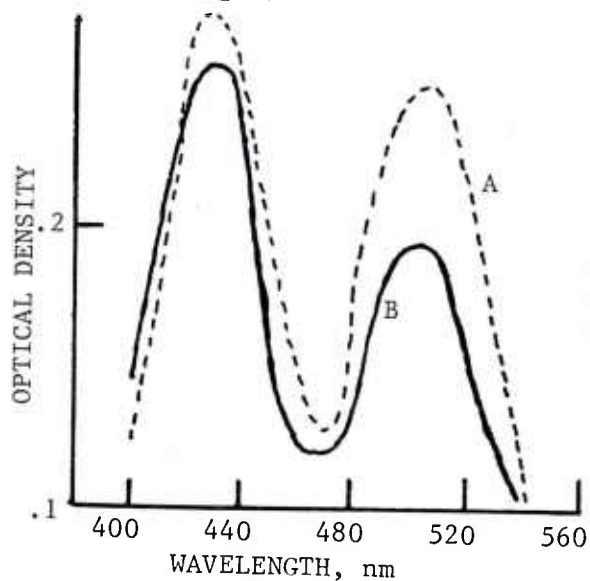


Fig. 1

Absorption spectrum of the transient species formed in flash-photolyzed aerated TNB solutions in (A) i-PrOH medium (B) methanol medium

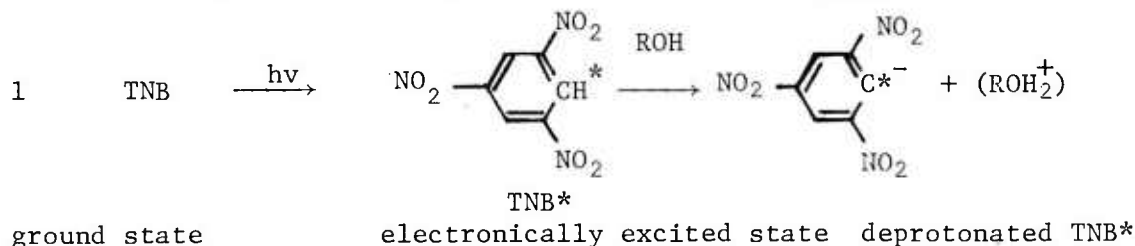
The yield of this species is a function of oxygen concentration, and its rate of formation and decay is viscosity and temperature dependent. In methanol it grows with a half-life of 25 μ sec, while in i-PrOH its half-growth time is 60 μ sec. Its decay obeys second order kinetics and the k/ϵ values in i-PrOH were found as follows:

$$k/\epsilon_{430} = 1.26 \times 10^{-4} \text{ and } k/\epsilon_{515} = 1.22 \times 10^{-4} \text{ sec}^{-1} \text{ cm}$$

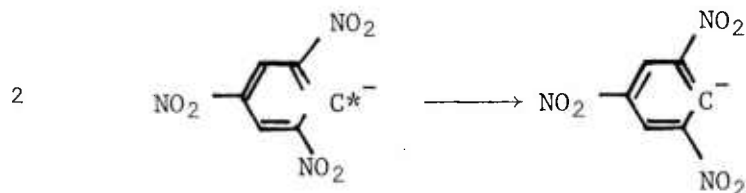
indicating that both absorption maxima are attributed to only one transient species. It is important to emphasize that this transient is not formed in CH_3CN which is a polar solvent without appreciable proton affinity.

Assignment of the Transient and Mechanism

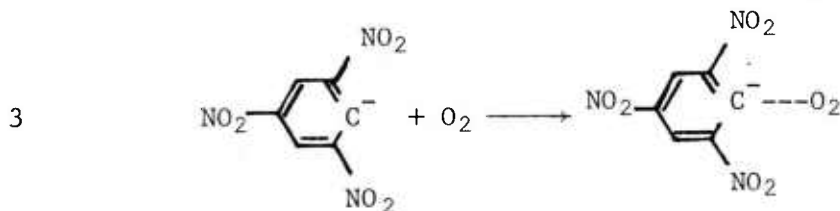
Methanol is a molecule with large proton affinity. A proton ejection mechanism as shown in reaction 1 is very likely to be involved in the photochemical generation of the transient:



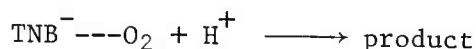
The deprotonated form of TNB^* decays to generate the negative ion of TNB:



It is further proposed that TNB^- interacts with molecular oxygen to form the transient charge transfer complex $\text{TNB}^- \cdots \text{O}_2$, as shown below:



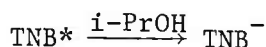
In order to probe further into the polar nature of the proposed mechanism which leads to the formation of $\text{TNB}^{--}\text{O}_2$, the effect of sulfuric acid on the formation of this species was studied. The presence of acid affects not only the formation of $\text{TNB}^{--}\text{O}_2$, but also its decay, by changing the order of decay, namely, from second order when there is no acid, to pseudofirst order in presence of relatively large concentrations of H_2SO_4 ($1.3 \times 10^{-4} \text{ M}$) as compared to the estimated small concentration of $\text{TNB}^{--}\text{O}_2$ (about 10^{-6} M). The effect of the acid strongly supports reactions 1 and 2 and the following neutralization reaction:



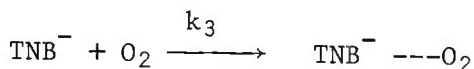
The proton ejection mechanism, reaction 1, assisted by polar solvents with relatively high proton affinity, is further supported by the fact that a similar type of dissociation has been proposed (3) for ground-state TNB in polar media with high proton affinity. It is conceivable that the protons in TNB are acidic because of the large electronegative character of the nitro groups. Furthermore, in the excited state, the protons would be expected to be even more acidic due to intramolecular charge transfer from the aromatic moiety to the nitro group during photoexcitation. However, even in the excited state the removal of proton requires extra energy which is furnished by the proton affinity of the methanol molecule. The acidic behavior of excited TNB seems to be analogous to that of the triplets of ketones and quinones having hydroxy and amino substituents (4).

In order to evaluate the kinetic data showing the growth and decay of $\text{TNB}^{--}\text{O}_2$ in acidic solutions ($2.1 \times 10^{-5} \text{ M H}_2\text{SO}_4$) of TNB in *i*-PrOH, the following reaction scheme was considered:

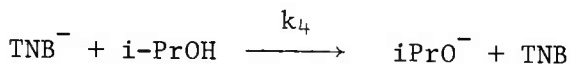
1 and 2



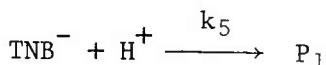
3



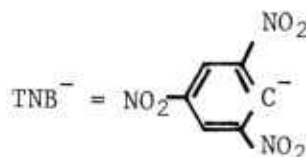
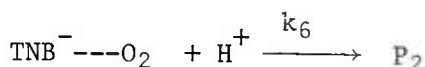
4



5



6



where P_1 and P_2 are products. The first half-life for the decay of $\text{TNB}^{--}\text{O}_2$ in aerated i-PrOH in the absence of added acid is 2.5 milliseconds. However, in solutions of TNB in aerated i-PrOH containing $2.1 \times 10^{-5} \text{ M H}_2\text{SO}_4$, the first half-life for the decay of this transient species is 500 microseconds. Hence, in acidified solutions, the fate of $\text{TNB}^{--}\text{O}_2$ is solely controlled via reaction 6.

Experimental evidence indicates that the rate of formation of TNB^- through reactions 1 and 2 exceeds by far the rate of its disappearance due to reactions 3, 4, and 5. Consequently, the rate determining step for the formation of $\text{TNB}^{--}\text{O}_2$ would depend on the relative reactivity and concentration of oxygen as compared to those for i-PrOH and H^+ . Setting $(\text{TNB}^-) = (X)$ and $(\text{TNB}^{--}\text{O}_2) = (Y)$, the following integrated equation was obtained:

$$7 \quad (Y) = (X'_0) [1 - e^{-(k_3' + k_4')t}] e^{-k_6't}$$

where $(X'_0) = (X_0) k_3'/(k_3' + k_4')$, and primes refer to pseudofirst-order rate constants ($X_0 = X$ at the end of the flash). In deriving Eq. 7, the simplifying assumption $k_5' = k_6'$ was made. Substitution of (Y) by $(\text{O.D.})_Y/\epsilon_Y l$ into Eq. 7 leads to the expression

$$8 \quad (\text{O.D.})_Y = \epsilon_Y l (X'_0) e^{-k_6't}$$

since the term $1 - \exp \{-(k_3' + k_4')t\}$, describing the growth of $\text{TNB}^{--}\text{O}_2$, becomes practically unity after 300 μsec as the data indicate. Equation 8, describing the decay of $\text{TNB}^{--}\text{O}_2$, yielded a linear plot. From the plot, the following values were obtained:

$$k_6 = 2.1 \times 10^7 \text{ M}^{-1} \text{ sec}^{-1}$$

$$\text{intercept} = \epsilon_Y l (X'_0) = 0.26$$

The term $1 - ((\text{O.D.})_Y e^{k_6't} / \text{intercept})$ describes the kinetic behavior of $\text{TNB}^{--}\text{O}_2$, during the growth period. This term plotted on semilog paper versus t gave a linear plot from the slope of which the value of $k' = k_3' + k_4' = 1.2 \times 10^4 \text{ sec}^{-1}$ was obtained for the disappearance of TNB^- by reaction with both oxygen and i-PrOH.

Efforts to observe TNB^- directly by flash photolysis in aerated or deaerated solutions of TNB in alcohols, in the wavelength region 260–700 nm, were unsuccessful because of the low extinction coefficient (about $2000 \text{ M}^{-1} \text{ cm}^{-1}$) of the species (5).

II. Spectroscopy and Chemical Reactivity of Electronically Excited States of Nitroaromatics

A series of mono and dinitronaphthalenes (NO_2N and DNO_2N) was studied, in polar and nonpolar solvents, with a laser nanosecond kinetic spectrophotometer which was designed and assembled in our laboratory. Absorption spectra of the lowest triplet excited states of 1-, and 2-nitronaphthalenes as well as 1,2-, 1,4- and 1,8-dinitronaphthalenes were obtained in nonpolar and polar solvents. The spectra were red shifted in polar solvents. From the spectral shifts, dipole moments of the lowest and upper triplet states of all the nitronaphthalenes studied in this work were calculated according to the method described earlier (6) and the values are reported in Table 1.

Table 1

Values of Dipole Moments of T_1 , T_n and Ground States
of Nitronaphthalenes

	μ_g	T_1	T_n	$\Delta\mu_{T_1 \rightarrow T_n}$
1- NO_2N	4.04	5.24	9.74	4.5
2- NO_2N	4.4	5.7	11.9	6.2
1,2- DNO_2N	6.5	7.1	10.4	3.3
1,8- DNO_2N	7.7	9.6	11.0	1.4

The extent of spectral shifts, lifetimes, quenching by oxygen and other data pertinent to the chemical reactivity of these triplet excited states are summarized in Table 2. These triplet states behave like $n-\pi^*$ states in nonpolar solvents, while in polar media the $n-\pi^*$ character of these states is reduced with a simultaneous increase in their intramolecular charge transfer character (2,7). In the case of 1,4-dinitronaphthalene (7b), however, owing to the symmetry of nitrosubstitution, there is lack of spectral shift and the triplet excited state retains its $n-\pi^*$ character even in polar media.

Table 2
Spectroscopic and Kinetic Data of Nitronaphthalene Triplets

Com- pound	T ₁ -T _n Abs. Max. in Hex- ane (nm)	Red Shift in Etha- nol (nm)	Reaction Rate Constant (M ⁻¹ sec ⁻¹) of the Triplet With									
			Triplet Mean- life (μs) in		Pure Solvent		Oxygen in		H ⁺	NaOH	Et ₃ N	Bu ₃ SnH
			Hex- ane	Etha- nol	Hex- ane (x10 ⁵)	Etha- nol (x10 ⁴)	Hex- ane (x10 ⁹)	Etha- nol (x10 ⁹)	(x10 ⁷)	(x10 ⁸)	(x10 ⁹)	(x10 ⁸)
1-NO ₂ N	525	55	0.93	4.9	1.4	1.2	1.3	3.3	1.8	1.7	2.8	-
2-NO ₂ N	425	45	0.53	1.7	2.5	3.5	1.7	1.6	88	12	6.4	0.4
1,2- DNO ₂ N	490	60	1.0	4.3	1.3	1.3	1.9	3.3	0.17	7.9	5.2	3.4
1,4- DNO ₂ N	545	No Shift	4.6	4.8	0.29	1.2	1.7	2.0	0.003	15	3.0	3.8
1,8- DNO ₂ N	550	40	2.5	5.3	0.52	1.1	0.81	2.1	2.5	5.0	4.6	1.6

III. Energetic Transients Formed From Electronically Excited TNT

Electronic excitation of TNT in aerated nonpolar solvents like benzene or cyclohexane results in the formation of a transient species, designated here as species BH, with absorption maximum at 460 nm as shown in Figure 2,A. This species decays in cyclohexane with a first-order rate constant $k = 1.04 \times 10^3 \text{ sec}^{-1}$.

In aerated polar solvents like methanol or acetonitrile, flash photolysis of TNT leads to a different transient with absorption maxima at 500, 540, and 630 nm, as illustrated in Figure 2, B.

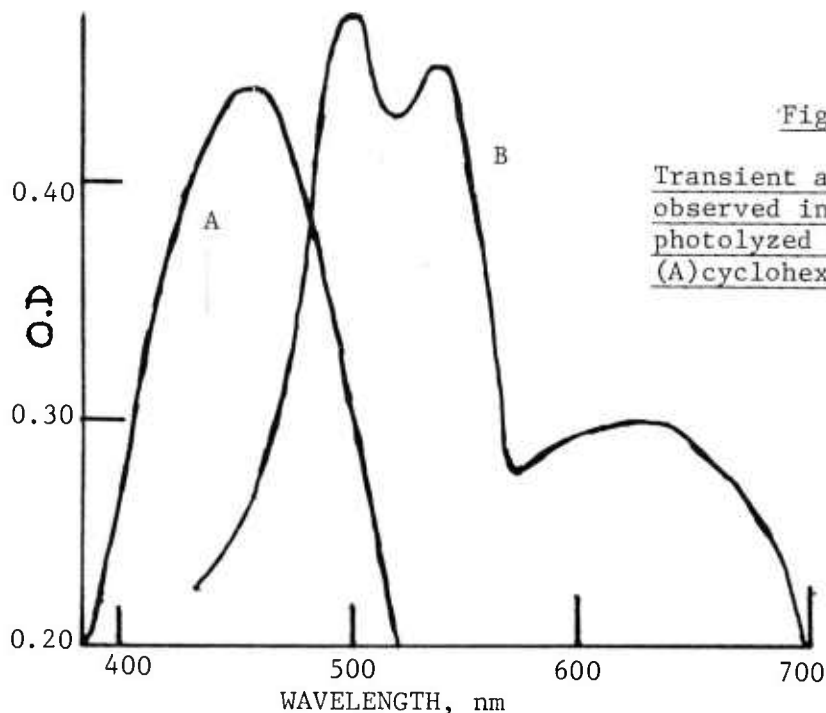


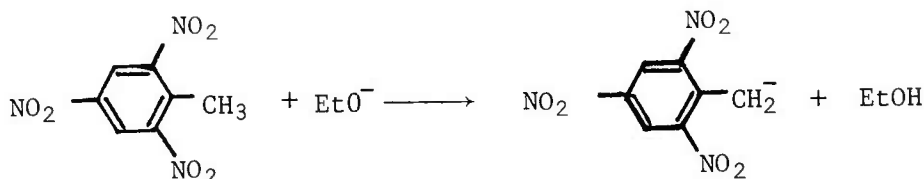
Fig. 2

Transient absorption spectra
observed in aerated flash-
photolyzed solutions of TNT
(A)cyclohexane (B) methanol

The rate constants for the decay of the transient absorption in methanol at these wavelengths were found to be equal within experimental error;

$$k_{500} = 1.61, k_{540} = 1.86 \text{ and } k_{630} = 1.74 \text{ sec}^{-1}$$

which indicates that these transitions originate from the same transient species. This species has an absorption spectrum which is closely in agreement with the spectrum of 2,4,6-trinitrobenzyl anion (8) formed by the reaction



and will be designated as B^- .

In order to substantiate further the possibility, based on spectral similarities, that the transient species B^- formed photo-lytically in polar solutions of TNT is negatively charged, salt effect experiments were carried out.

A. Salt Effect Studies

The Bronsted-Bjerrum theory of ionic reactions combined with the extended Debye-Huckel theory (9) leads to the following relationship

$$9 \quad \log k = \log k_0 + 2AZ_A Z_B [\mu^{1/2}/(1 + \mu^{1/2})]$$

where k is the rate constant at ionic strength μ and k_0 is the rate constant at infinite dilution. The constant A is given by Eq. 10

$$10 \quad A = 1.825 \times 10^6 (DT)^{-3/2}$$

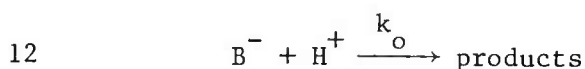
where Z_A , Z_B represent the charges on the reacting ions, and D is the static dielectric constant.

For most cases reported in the literature Eq. 9 was applied to ionic reactions in aqueous solutions. Dainton and coworkers demonstrated, however, that by using LiCl to vary the ionic strength of the solution, Eq. 9 could also be applicable to ionic reactions in methanol (10).

At 298°K the static dielectric constant of methanol is $D = 32.63$ and the constant A calculated from Eq. 10 becomes 1.9. Then Eq. 9 assumes the following form:

$$11 \quad \log k = \log k_0 + 3.8Z_A Z_B [\mu^{1/2}/(1 + \mu^{1/2})]$$

Because of the limiting solubility of TNT in neutral water (11), reaction 12 was studied in methanol:



where H^+ represents the acid added before the flash ($4.4 \times 10^{-5} \text{ M H}_2\text{SO}_4$).

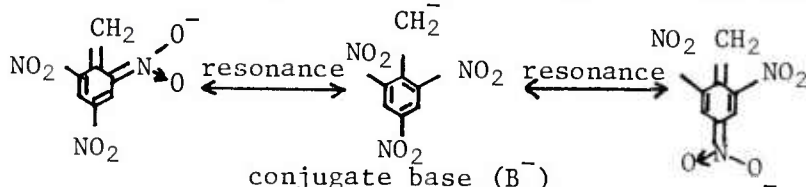
The neutralization reaction 12 was followed by time-resolved absorption spectroscopy at 540 nm where B^- absorbs strongly. The rate constant measured for this reaction was equal to $k_o = 5.0 \times 10^6 \text{ M}^{-1} \text{ sec}^{-1}$. Subsequently the rate constant k of reaction 12 was determined as a function of ionic strength, varied with LiCl, and the data were plotted according to Eq. 11. From the slope of the resulting linear plot, the negative charge on B^- was found to be equal to 1.

B. Assignment of the Transient Species BH and B^-

The species BH and B^- cannot be attributed to a free radical type of species because their lifetimes (in cyclohexane, $\tau_{BH} = 1/k_1 = 0.96 \times 10^{-3} \text{ sec}$; in methanol, $\tau_{B^-} = 1/k_2 = 0.575 \text{ sec}$) were found to be unaffected by the presence of oxygen.

Their assignment to electronically excited states of TNT can also be excluded because their lifetimes are extremely large as compared with the values reported (12) for the electronically excited states of nitroaromatics, and furthermore, molecular oxygen would be expected to quench excited states (12).

The transient species B^- having an absorption spectrum closely in agreement with that of the 2,4,6-trinitrobenzyl anion (8) and possessing a charge equal to -1, is assigned to the following species:

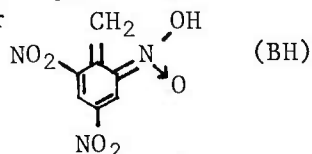


This structure is consistent with the conjugate base participating in the photochromic mechanism of ortho nitro and dinitrotoluenes (13). According to this structure the conjugate base B^- is not paramagnetic and oxygen would not be expected to affect its lifetime as shown to be the case experimentally.

The spectroscopic evidence which indicates that increasing polarity of the solvent converts BH to B^- seems to be in agreement with the aci-quinoid and the conjugate base mechanism suggested by Wettermark et al. (13) to explain the photochromic activity of nitrotoluenes.

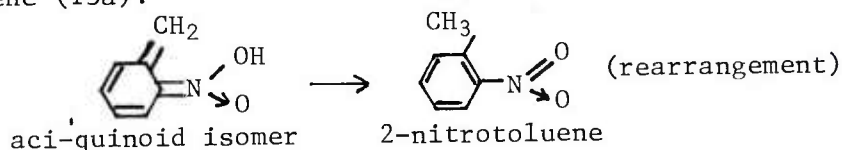
It should be realized, of course, that while the polarity of the solvent is necessary for the stabilization by solvation of the species B^- , the basicity of the solvent is also expected to facilitate the dissociation of BH (13c).

Additional evidence supporting the assignment of the species BH to the aci-quinoid isomer



is obtained by comparing the absorption spectrum of the species BH (Fig 2, A) with those reported earlier (13) for the aci-quinoid isomer of 2-nitro and 2,4- and 2,6-dinitrotoluenes. The comparison shows spectral similarities and a red shift with increasing number of nitro substitutions, which is consistent with the absorption maximum of BH at 460nm.

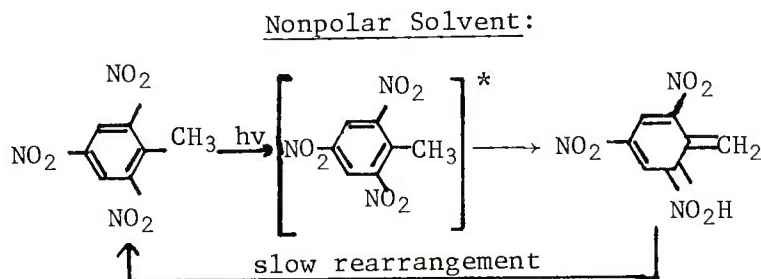
Furthermore, the half-life of BH in cyclohexane $t_{1/2} = 0.69 \times 10^{-3}$ sec is in fair agreement with the half-life $t_{1/2} = 1.1 \times 10^{-3}$ sec for the rearrangement or fading reaction of the aci-quinoid isomer of 2-nitrotoluene (13a):



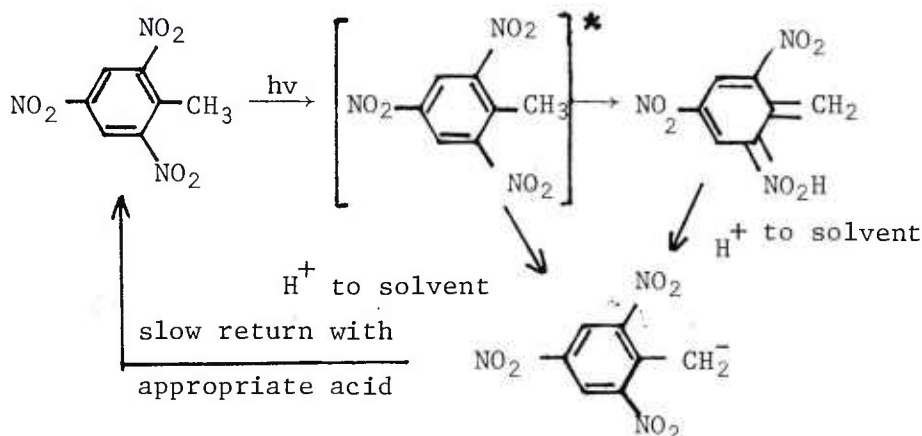
The similarity of the half-lives is not surprising, since the fading reaction represents an intramolecular proton transfer reaction which is not expected to be strongly influenced by the solvent medium.

C. Mechanism of the Photochromic Activity of TNT

Applying closely the mechanism suggested by Wettermark et al. (13) for the photochromism of mono and dinitrotoluenes to the case of TNT, the following mechanism can be proposed:



Polar and/or basic solvent:



SCHEME I

According to Scheme I, BH is formed through intramolecular proton transfer from the methyl group to one of the ortho nitro groups during the lifetime of the electronically excited TNT. In nonpolar solvents, BH rearranges back to ground state TNT. In polar solvents, however, the only observable transient species is B^- , and it is not known whether B^- is formed directly from electronically excited TNT or by rapid dissociation of species BH. Species B^- presumably regenerates TNT through proton abstraction from the solvent. Caldin et al. (14) and Blake et al. (15) have shown that 2,4,6-trinitrobenzyl anion decays in alkaline alcoholic solutions via proton abstraction reaction.

The biphotonic mechanism for the formation of B^- is excluded on the basis of experimental evidence showing that the optical density of B^- at the end of the flash is linearly proportional to the square of the charging voltage applied across the terminals of the flash lamps. This implies that the yield of B^- is linearly proportional to the intensity of the excitation light.

D. Modification of Initiation Thresholds of TNT by Chemical Additives

During the course of the present studies it was discovered that TNT can be sensitized to impact and thermal decomposition by additives which form exothermically a charge transfer complex or salt in molten TNT through the initial formation of TNT anion, B^- . Such additives are NaOH, QMB₃ and NaI. These salts were isolated and

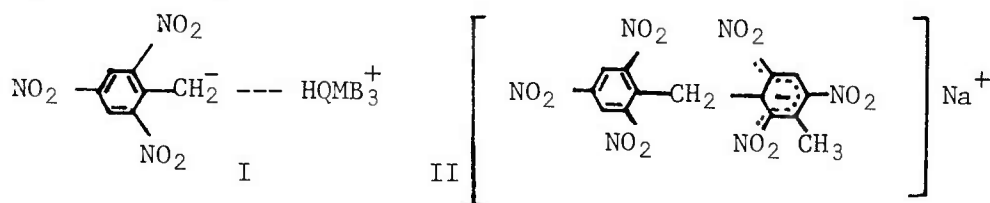
their sensitivity to thermal decomposition and impact were determined and listed in Table 3. Also listed in this table are the thermal and impact sensitivity of the physical mixtures of TNT with 5% W/W NaOH and QMB₃.


Table 3
Thermal and Impact Sensitivity Data

<u>Compound</u>	<u>DSC</u>	<u>* 50% Impact Value (cm)</u>
	<u>Exotherm</u> <u>Max. (° C)</u>	
TNT	330	62 ± 2.59
TNT-QMB ₃ (5%) Phys. Mixt.	152	14.5 ± 1.04
TNT-NaOH	100 (I peak)	-
Heterog. Mixture	151 (II peak)	-
TNT - 7% NaI Phys. Mixt.	257	-
TNT - NaOH Complex (1:1)	272	50.9 ± 2.70
TNT - QMB ₃ Complex (1:1)	No Exotherm	No initiation up to 200 cm

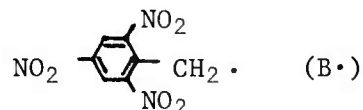
* Picatinny Arsenal unit with Type-12 Tool, sand paper and 2½ kg wt.

Absorption spectroscopic and ESCA studies resulted in the following assignment of the charge transfer complexes of TNT/QMB₃ (I) and TNT/NaOH (II).



Furthermore, it was also discovered that TNT can be sensitized by additives, such as, benzoquinone, , and tetraethyl lead, Pb(C₂H₅)₄.

DSC studies indicate that the exotherm appears at 274°C and 175°C for the mixtures TNT/Benzoquinone and TNT/Tetraethyl lead, respectively. The mechanism of sensitization, in this case, is attributed to the initial formation of the radical,



through hydrogen abstraction. Hydrogen abstraction by the benzoquinone is expected to lead to the semiquinone radical, $\text{HO} \text{---} \text{C}_6\text{H}_4 \text{---} \text{O}\cdot [\text{C}\cdot]$

while in the case of tetraethyl lead, ethane is expected to form since the hydrogen abstraction occurs by the ethyl radicals which are produced thermally from $\text{Pb}(\text{C}_2\text{H}_5)_4$.

Electronic excitation of TNT in the gas phase at 215°C (in the presence of 600mm argon) leads to the formation of a transient species with absorption maxima at 265 and 340 nm and $t_{1/2} = 69$ sec. Oxygen and NO used (at 400mm Hg) instead of argon were found to increase the yield and reduce the lifetime of the transient, $t_{1/2}(\text{NO}) = 14$ and $t_{1/2}(\text{O}_2) = 34$ sec. The increased yield and shorter lifetime of this transient species in the presence of NO and O_2 are consistent with the tentative assignment of this species to the free radical, $\text{B}\cdot$. Both NO and O_2 are expected to assist initially, the formation of this radical through hydrogen abstraction from the methyl group of the electronically or vibrationally excited TNT. Furthermore, when the radical is formed it is also expected to be scavenged by molecules, such as, NO and O_2 .

The results concerning the modification of initiation thresholds appear to be very promising, since eventual control of the sensitivity of explosives should contribute towards reducing pre-matures and vulnerability of munitions.

Work is in progress for the formation of the radical, $\text{B}\cdot$, from electronically excited HNBB (hexanitrobibenzyl, B-B) in the gas phase or solid matrices at 77°K. Furthermore, benzoquinone will be electronically excited in the gas phase in presence of TNT. Electronically excited benzoquinone is expected to abstract a hydrogen atom from the methyl group of TNT and yield the radical, $\text{B}\cdot$, and the semiquinone radical, $\text{C}\cdot$, which has absorption maximum at 415 nm.

REFERENCES

1. C. Capellos and K. Suryanarayanan, Int. J. Chem. Kinet., 5, 305 (1973).
2. C. Capellos and K. Suryanarayanan, Int. J. Chem. Kinet., 8, 529 (1976).
3. V. Baliah and V. Ramakrishnan, Rec. Trav. Chim., 78, 783 (1959), ibid., 79, 1150 (1960); G. Lambert and R. Schaal, Compt.Rend., 255, 1939 (1962).
4. T. S. Godfrey, G. Porter and P. Suppan, Disc. Faraday Soc., 39, 194 (1965).
5. D. Radulescu and V. Alexa, Z. Phys. Chem., B8, 395 (1930).
6. P. Suppan, J. Mol. Spectr., 30, 17 (1969).
- 7.(a) C. Capellos and G. Porter, Faraday Trans. II, 70, 1159 (1974).
(b) C. Capellos and K. Suryanarayanan, Int. J. Chem. Kinet., 8, 541 (1976), (c) ibid., 9, 399 (1977).
8. E. F. Caldin and G. Long, Proc. Roy Soc., A, 228, 263 (1955).
9. C. Capellos and B. H. J. Bielski, Kinetic Systems, Wiley, New York, 1972, P. 120.
10. G. V. Buxton, F. S. Dainton and H. Hammerli, Trans. Faraday Soc., 63, 1191 (1967).
11. T. Urbanski, Chemistry and Technology of Explosives, Vol. 1, MacMillan, N.Y. 1964, P. 292.
12. C. Capellos, Picatinny Arsenal, Dover, NJ Tech. Rep. 4341, 1972.
- 13.(a) G. Wettermark, J. Phys. Chem., 66, 2560 (1962) (b) G. Wettermark and R. Ricci, J. Chem. Phys., 39, 1218 (1963) (c) G. Wettermark et al, J. Amer. Chem. Soc., 91, 2204 (1969).
14. E. F. Caldin and R. A. Jackson, J. Chem. Soc., 2413 (1960).
15. J. A. Blake, et al, Can. J. Chem., 44, 119 (1966).

MILITARY USE OF BULK EXPLOSIVES

*HENDRIK D. CARLETON
U. S. ARMY ENGINEER WATERWAYS EXPERIMENT STATION
VICKSBURG, MS 39180

IntroductionBackground

"Battle in Central Europe against forces of the Warsaw Pact is the most demanding mission the U. S. Army could be assigned....Warsaw Pact doctrine anticipates use of nuclear weapons in...future war, but teaches preparedness to fight without them. For both conditions, it *emphasizes heavy concentrations of armor*.... Forces opposing Soviet equipped and trained troops must expect intense, highly mobile combat. [If initiated,] battle will be fought on a scale and at a tempo rarely seen in all history" (from Field Manual 100-5 (1)).

The U. S. Army is presently developing a bulk explosive system intended to make possible the rapid excavation of obstacles and defensive positions, and to be used against large, prechambered targets. It is also intended that this system will be useful for quarry work, and as a substitute for standard military explosives as may become necessary under emergency conditions. It is not generally recognized that the Army has no truly suitable explosive available for the size of demolition mission between those normally undertaken with military high explosives and those considered suitable for the employment of atomic demolition munitions (ADM's). Bulk explosives will also be desirable in certain cases to substitute for low-yield ADM's where field commanders must have the ability to respond to tactical situations unencumbered by nuclear release procedures and employment constraints.

It is apparent that large quantities of explosives will be routinely used on bulk explosive system missions. It is also readily apparent that the greatest need for obstacle creation will occur at

early times during the battle, and that, to be effective, obstacles will have to be prepared with great speed. The conclusion is inescapable: during the most critical engagements, there will not be time to use a bulk explosive system that requires transport from ammunition supply points to obstacle sites. The system must be suitable for safe storage and use in forward areas, so that it may be rapidly employed as threats are perceived. Also, bulk explosives will have to be used in conjunction with a truly rapid emplacement capability, since timely availability of these explosives will do little good if emplacement means are wanting.

Purpose and scope

This paper summarizes the characteristics of available bulk explosive systems, and evaluates them in terms of military requirements. It also discusses the continuing development of techniques intended to place effective obstacles and defensive positions on time using bulk explosives under anticipated combat conditions.

An Evaluation of Existing Bulk Explosives Systems

Bulk explosives

Bulk explosives are a class of explosives that may be handled by bulk loading techniques, i.e., that are pourable or pumpable during emplacement operations. They are characteristically used in large quantities, and their costs per unit weight are normally very low compared with those of high explosives. Bulk explosives may be used in cartridge form, as are high explosives. However, during the excavation and quarrying applications, for which they are most suited, they are more typically placed directly into holes in the ground without packaging.

In this report, eight bulk explosive types will be discussed:

- a. ANFO is a dry blasting agent composed of ammonium nitrate and fuel oil, usually in a 94 percent to 6 percent ratio by weight. Neither ANFO nor any other blasting agent contains any chemical classified as an explosive, and all require high-explosive primers to induce detonation. ANFO may be purchased from explosives manufacturers or field-mixed.
- b. Aluminized ANFO is ANFO containing up to 28 percent particulate aluminum by weight. It may be purchased from explosives manufacturers or field-mixed.

- c. Basic field-mix slurry blasting agents are ammonium nitrate-based formulations that may be mixed in the field from commonly available chemicals using pioneer handtools or transit-mix concrete delivery trucks. These and other blasting agents are essentially mixtures of oxidizers (such as ammonium nitrate) and fuels (such as fuel oil or aluminum) in a liquid medium thickened with a gum and gelled with a chemical called a cross-linker. Slurry blasting agents are strongly water-resistant, and may be used in wet soil charge emplacements without protective packaging.
- d. Developmental field-mix aluminized slurry blasting agents are advanced two- or three-component slurries that may be mixed in large quantities in the field without heating, using equipment no more complex than transit-mix concrete delivery trucks. A few explosives manufacturers have done research to develop such products, and are presently considering their sale on the open market.
- e. Commercial slurry blasting agents are, for the purposes of this evaluation, commercially available slurry blasting agents that contain less than 25 percent aluminum by weight. These products are factory-mixed by explosives manufacturers.
- f. Highly aluminized commercial slurry blasting agents are commercially available slurry blasting agents that contain from 25 to 35 percent aluminum by weight. These products are factory-mixed by explosives manufacturers.
- g. Commercial slurry explosives are similar to commercial slurry blasting agents except that they contain explosive compounds as sensitizers in place of (or in addition to) nonexplosive fuels such as fuel oil or aluminum. They may or may not be cap-sensitive, but all are classified as explosives for shipping purposes, whereas blasting agents are not. These products are factory-mixed by explosives manufacturers.
- h. Gelled nitromethane is a mix of nitromethane (a chemical used in the pharmaceutical, dye, insecticide, and textile industries), a modified guar gum, and, sometimes, a cross-linker that increases the thickness and water resistance of the mix. It is field-mixed with specialized equipment, and is not normally cap-sensitive.

The fact that blasting agents contain no chemicals that are classified as explosives can be misleading under certain circumstances. ANFO can be produced in cap-sensitive form by using finely pulverized ammonium nitrate, and slurry blasting agents incorrectly made with finely flaked aluminum become similarly cap-sensitive. Because of the extreme variability possible with these products, care must be taken to ensure that the particular product being considered for use in the field will be properly handled. A more complete background on blasting agents and slurry explosives (2-4), a description of a recent experiment in the field mixing of a slurry blasting agent (5), and a discussion of gelled nitromethane (6) are available.

Candidate systems versus
required characteristics

Table 1 lists required characteristics and mission statements provided by the U. S. Army Engineer School (USAES), and the author's judgment of the acceptability of each of the eight defined bulk explosive systems for each listed requirement or mission. Though the order of the listed items has been changed somewhat, the number associated with each item is that from the original USAES listing. This section and the following section qualify the evaluations found in Table 1.

ANFO and aluminized ANFO. These products are rated unsatisfactory for required characteristics 4, 5a, and 9 because they become ineffective when exposed to water. ANFO is considered unsatisfactory for required characteristic 11 because it is bulky and the least efficient cratering explosive (in terms of charge weight) of the eight bulk explosives being considered herein. Its great popularity in industry is explained by its extremely low cost and ease of mixing. Watertight packaging is frequently used to protect ANFO products from ground moisture, but such packaging is not always effective and may lead to other difficulties such as static electricity hazards or toxic gas formation upon detonation.

Slurry blasting agents. These products are rated marginal for required characteristic 5b because many slurries become rigid at freezing temperatures, in addition to becoming less sensitive. However, special formulations are available that maintain fluidity at freezing temperatures. Factory-produced slurries have a shelf life of from 6 months to 1 year, depending on storage conditions. Thus, they have been rated unsatisfactory for required characteristic 2. As with the dry blasting agents, slurry blasting agents become more efficient cratering explosives with the addition of significant amounts of aluminum. The basic field-mix slurry blasting agent has been rated as unsatisfactory for required characteristic 4 because several of its six components are damaged if exposed to water before

Table 1

Author's Evaluation of Bulk Explosive Systems for Military Purposes

	Dry Blasting Agents			Slurry Blasting Agents				Commercial Slurry Explosives	Gelled Nitro-methane
	ANFO	Aluminized ANFO	Basic Field Mix	Develop-mental Field-Mix Aluminized	Commercial Aluminized Commercial	Highly Aluminized Commercial			
3. Number of components required in the field.	2	3	6	2 or 3	1	1	1	1	2 or 3
1. Meets all safety requirements in storage, mixing, transportation, and handling.	+	+	+	+	+	+	+	0	0
2. Storage as stable, nonexplosive components.	+	+	0	0	-	-	-	-	0
4. Capable of mixing in all weather conditions with simple equipment.	-	-	-	0	N/A	N/A	N/A	N/A	-
5a. Capable of wet weather emplacement and firing.*	-	-	+	+	+	+	+	+	0
5b. Capable of emplacement and firing through wide temperature variation.*	+	+	0	0	0	0	0	0	+
6. Mixed explosive not bullet- nor cap-sensitive.	+	+	+	+	+	+	+	0	0
7. Mixed explosive has low sensitivity to variations in mix proportions and mixing times.	+	+	0	+	N/A	N/A	N/A	N/A	-
8. Components and mixed explosive should be nontoxic and nonirritating.	+	+	+	+	+	+	+	+	0
9. Mixed explosive should remain potent for 5 days minimum in all types of ground emplacement.	-	-	+	+	+	+	+	+	0
11. Maximum power to weight ratio.	-	+	0	+	0	+	+	0	0
10. Note: Manufacturers show great flexibility in packaging;	all systems can be made to meet the single-man-lift packaging requirement.								
1. Rapid, efficient, economic explosive excavation for creation of large, effective antiarmor obstacles.	-	0	0	+	0	0	0	0	-
2. Nonnuclear alternative to small ADM's.	-	0	0	+	0	0	0	0	-
3. Safe, economical storage near planned obstacles for minimum logistics impact.	+	+	+	0	-	-	-	-	-
4. Able to replace military dynamite and ammonium nitrate cratering charges for a wide variety of military explosive missions.	0	0	0	+	-	-	-	-	-

Key to rating symbols: + = satisfactory; 0 = marginal; - = unsatisfactory; N/A = not applicable. Ratings are qualified in the accompanying text.
 * Using standard caps and TMI or C-4 primers.

*CARLETON

mixing is completed. It is rated marginal for required characteristic 7 because the addition of the minor constituents (guar gum and cross-linker) must be carefully handled to get the right consistency in the final mix. The developmental field-mix aluminized slurry blasting agents are rated marginal for required characteristics 2 and 4 because these are foreseen as potential problem areas for which satisfactory performance is yet to be demonstrated.

Commercial slurry explosives. These products have many characteristics in common with commercial slurry blasting agents. Thus, they are rated unsatisfactory for required characteristic 2, and marginal for required characteristic 5b for the same reasons as the commercial slurry blasting agents. Because they contain components that are classified as explosive for shipping purposes, they have been placed in the marginal category for required characteristics 1 and 6.

Gelled nitromethane. The guar gum used in this formulation must be quickly dispersed throughout the liquid nitromethane during mixing to prevent the formation of lumps that are difficult to break up. Thus, this product is rated unsatisfactory for required characteristic 7. Liquid nitromethane is soluble in water, and in fact can be sensitized somewhat by the addition of small amounts of water. It is therefore rated as unsatisfactory for all-weather mixing (required characteristic 4). It is rated marginal for required characteristics 1 and 6 because it is not compatible with several materials, some of which are used in containers. Its storage is also subject to special rules. High-explosive primers need protection when used in nitromethane, since this chemical can partially dissolve most military high explosives. For this reason, and because gelled nitromethane is more susceptible to charge deterioration under severe groundwater conditions than are slurries, it is rated marginal for required characteristics 5a and 9. It is rated marginal for required characteristic 8 because liquid nitromethane vapors are considered a moderate health hazard.

Candidate systems versus military missions

Rapid, efficient, economic explosive excavation for creation of large, effective, antiarmor obstacles. Because of their cratering efficiency and water resistance, highly aluminized slurry blasting agents are the products of choice for this mission. However, the factory-mixed product does not meet storage requirements, and frequent stock turnover would be uneconomical. The ANFO products do not have the required water resistance to guarantee mission accomplishment. Safety is considered problematical with slurry explosives and gelled nitromethane.

Nonnuclear alternative to small ADM's. The comments of paragraph 13 apply also to this mission. However, charge compactness has added importance in this application because the ADM's to be replaced are very small devices in proportion to their yields. Even a very compact, efficient bulk explosive might not be suitable for certain small ADM missions. Because confinement is especially beneficial to the performance of an aluminized product, some surface missions might be accomplished at reduced efficiency.

Safe, economical storage near planned obstacles for minimum logistics impact. The ANFO products and the field-mix slurry blasting agents are the products of choice for this mission. Long-term storage is not practical with state-of-the-art factory-mixed slurries. Safety is considered problematical with slurry explosives and gelled nitromethane.

Able to replace military dynamite and ammonium nitrate cratering charges for a wide variety of military explosive missions. A readily mixed, two-part aluminized slurry blasting agent would be ideal for these missions because of its efficiency and the elimination of the explosive storage requirement. All of the factory-mixed slurries would require some sort of special storage, as would gelled nitromethane. The ANFO products and the basic field-mix slurry blasting agents would be susceptible to water damage during mixing, and the ANFO products could be rendered ineffective by groundwater as well. However, it should be kept in mind that the ANFO products, because of the universal availability of their components and their ease of mixing, could be used as emergency supplements to the Army demolitions system.

Techniques for the Use of a Bulk Explosive System

A concept for explosive barrier ditch creation

Figure 1 presents an explosive barrier ditching plan that enables troops in the field to use single-, double-, or triple-ditch designs, depending upon the perceived threat and available preparation time. It is anticipated that the single-ditch design would be most appropriate for use where the barrier trace is to be mined and adequately covered by the fires of defending units. However, where the defeat of armored vehicle launched bridges (AVLB) is a requirement, as might be the case where defending fires are overcommitted, the double- or triple-ditch designs could be used (time permitting) to increase the obstacles' effectiveness.

Figure 2 shows schematic cross sections of the ditches produced by each of the three designs shown in Figure 1. The double-ditch option is the double-ditch tank trap described in an earlier

*CARLETON

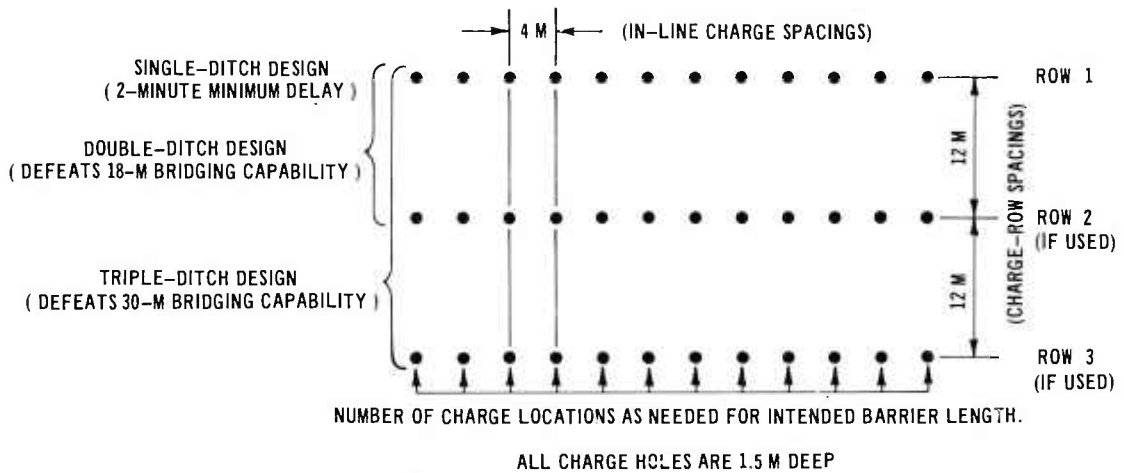


Figure 1. Explosive barrier ditching designs (plan view)

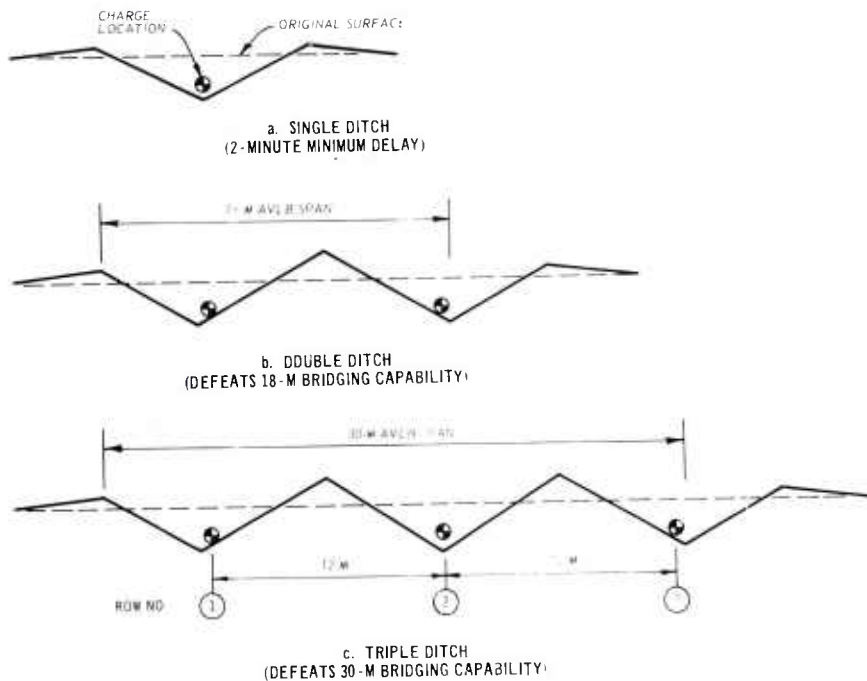


Figure 2. Explosive barrier ditch cross sections

report (7). Note that the charge hole depths (1.5 m*), the in-line spacings between charges (4 m), and the spacings between charge rows (12 m, if used) are always the same, regardless of the nature of the medium being cratered. All three ditching designs are excavated by the simultaneous detonation of all of the buried charges after the charge holes have been stemmed (backfilled above the emplaced charges) to the original surface with native soil from the digging operation.

The 1.5-m charge hole depth has been chosen because it is normally practical with shaped charges and hand tools, rapidly excavated with backhoe or auger, and close to the average of optimal charge burial depths for cratering purposes in a variety of soils. The availability of the JD 410 tractor within the Army materiel system allows troops to use either the backhoe or an auger to make charge holes, depending upon the specific conditions with which they are faced. The backhoe will generally be best for the digging of charge holes for stacked charges such as TNT, and for the emplacement of large charges of any type explosive. Preliminary field test results also indicate that for smaller charges in clay soils, backhoe-dug charge emplacements produce steeper crater side slopes than do auger-dug charge emplacements. Thus, the use of the backhoe may be advantageous for the production of obstacles and defensive positions in certain situations. However, any such advantage may be offset by the greater digging speeds possible with the auger in many situations. Further field tests have been scheduled to quantify the differences between backhoe- and auger-dug charge emplacements in a variety of soils, and to investigate new techniques for charge emplacement.

Table 2 shows the explosive charge weights required in each hole for any of the plans in Figure 1 for obstacle ditching in a variety of earth materials. As can be seen from the wide variation in the weights of the required charges, crater size is greatly influenced by the composition of the cratered medium. The effects of soils on ditch size may vary greatly, even within a very small area where composition differences are not readily apparent. For this reason the recommended charge sizes have been chosen to produce at least the minimum required effect in typical earth materials of the types named. However, in cases where the soil type has not been determined, the largest individual charge weight for the explosive considered (shown in boldfaced type in Table 2) may be used as a rule of thumb.

By projection from earlier estimates (7), a 10-man crew with a JD 410 tractor should be able to complete 150 m of the triple-ditch design in a 12-hour day. Alternately, the same crew

* To convert metres to feet, multiply by 3.280839.

Table 2
Design Data for Explosive Barrier Ditch Production

<u>Cratered Medium</u>	<u>Individual Charge Weight of TNT*</u>	
	<u>kg</u>	<u>lb</u>
Saturated silty clay	20	44
Dry sand, weak sandstones, and shales	40	88
Wet clay	60	132
Dry gravelly sand	100	221
Dry sandy clay	160	353

Note: For all designs in this table: charge hole depths = 1.5 m; in-line charge spacings = 4 m; spacings between rows = 12 m. (Design basis is Reference 8.)

* If used, highly aluminized slurry blasting agent charges would be about 60 percent as large as equally effective TNT charges. Exact charge weights for any bulk explosive to be adopted by the Army will be available before the scheduled date for type classification. Note: To convert kilograms to pounds (mass), multiply by 2.205.

should be able to complete 225 m of the double-ditch design, or 450 m of the single-ditch design, in a 12-hour day.

An explosively excavated tank position

Figures 3-5 show front, side, and rear views, respectively, of an explosively excavated tank position created during recent field tests at Fort Polk, Louisiana. Requirements for the design were as follows:

- a. The tank hull must have full protection from the front and both sides of the position.
- b. The tank's main gun must be able to make a full-circle traverse at normal firing elevations.
- c. The tank must be able to enter and leave the position from the rear without difficulty, and without any additional preparation after the initial excavation.



Figure 3. Front view of an M60 tank in an explosively excavated defensive position



Figure 4. Side view of an M60 tank in an explosively excavated defensive position



Figure 5. Rear view of an M60 tank in an explosively excavated defensive position

- d. The explosive excavation must be done without the use of delay caps.

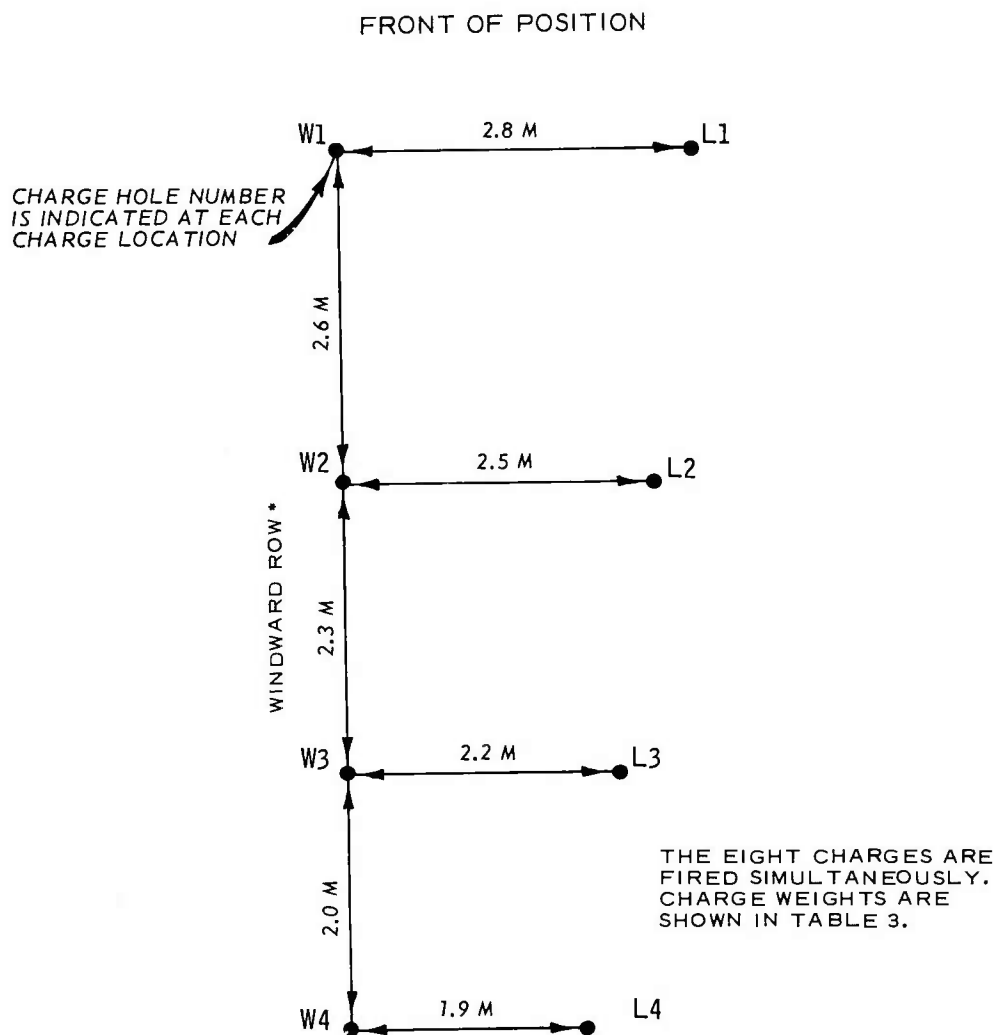
All design requirements were met by the experiment, which was completed in 1 hour.

Figure 6 and Table 3 give specifications for the creation of the hull-down tank position using TNT in a wet clay soil. It should be remembered, however, that this design is still under development. Tests have been scheduled to simplify the charge array, and to examine the effects of different soils on its successful execution.

Conclusions and Recommendations

When the primary bulk explosive system requirements for cratering efficiency, water resistance, and long-term nonexplosive storage are considered together, only one candidate system survives: the easily field mixable, highly aluminized, slurry blasting agent. However, such an advanced system is still at a developmental stage within the industry. The present Army effort to accelerate completion of this development appears very likely to succeed; it is recommended that the project be given appropriate priorities to

*CARLETON



* NO WIND IS NECESSARY. THIS CONFIGURATION ENSURES THAT STRONG WINDS, IF PRESENT, WILL NOT UNDO THE EFFECT OF THE EXPLOSIVE DESIGN, WHICH PREVENTS FALLBACK FROM THE SHOT.

Figure 6. Explosive hull-down tank position design (plan view)

Table 3

Design Data for Explosive Hull-Down Tank Position Production

Charge Hole No.	Hole Depth		Charge Weight of TNT	
	m	ft	kg	lb
W1	0.60	2.0	41	90
W2	0.52	1.7	27	60
W3	0.46	1.5	18	40
W4	0.40	1.3	12	27
L1	1.15	3.8	20	45
L2	1.01	3.3	14	30
L3	0.88	2.9	9	20
L4	0.77	2.5	6	13
Total			147	325

Note: Design basis is wet clay curve (8). Hole depth refers to the depth of the empty charge hole before the charge is added. Holes must be backfilled after charges are placed to get full excavation. If used, highly aluminized slurry blasting agent charges would be about 60 percent as large as the TNT charges listed above. Slurry blasting agents are also much more readily emplaced than block munitions. A final design for this excavation using the bulk explosive to be developed for the Army will be available before its scheduled type classification date.

enable the DARCOM Project Manager for Selected Ammunition to achieve on-time completion.

The constant charge depth and spacing technique for single-, double-, and triple-ditch obstacle production (Figure 1) is ideal for use with a bulk explosive system. Use of the DARCOM slurry blasting agent with this simple employment technique should give the Army an improved capability for swift and effective ditch production. Future emphasis in the Military Engineering Applications of Commercial Explosives (MEACE) program will be on increasing charge emplacement speed, determining the exact cratering characteristics of the DARCOM selected slurry blasting agent, and completing designs for explosively excavated defensive positions. It is recommended that close coordination be maintained between USAES and the U. S. Army Engineer Waterways Experiment Station to ensure that newly developed bulk explosive system employment techniques will be ready for incorporation into doctrinal literature prior to the system's availability to troop units.

References

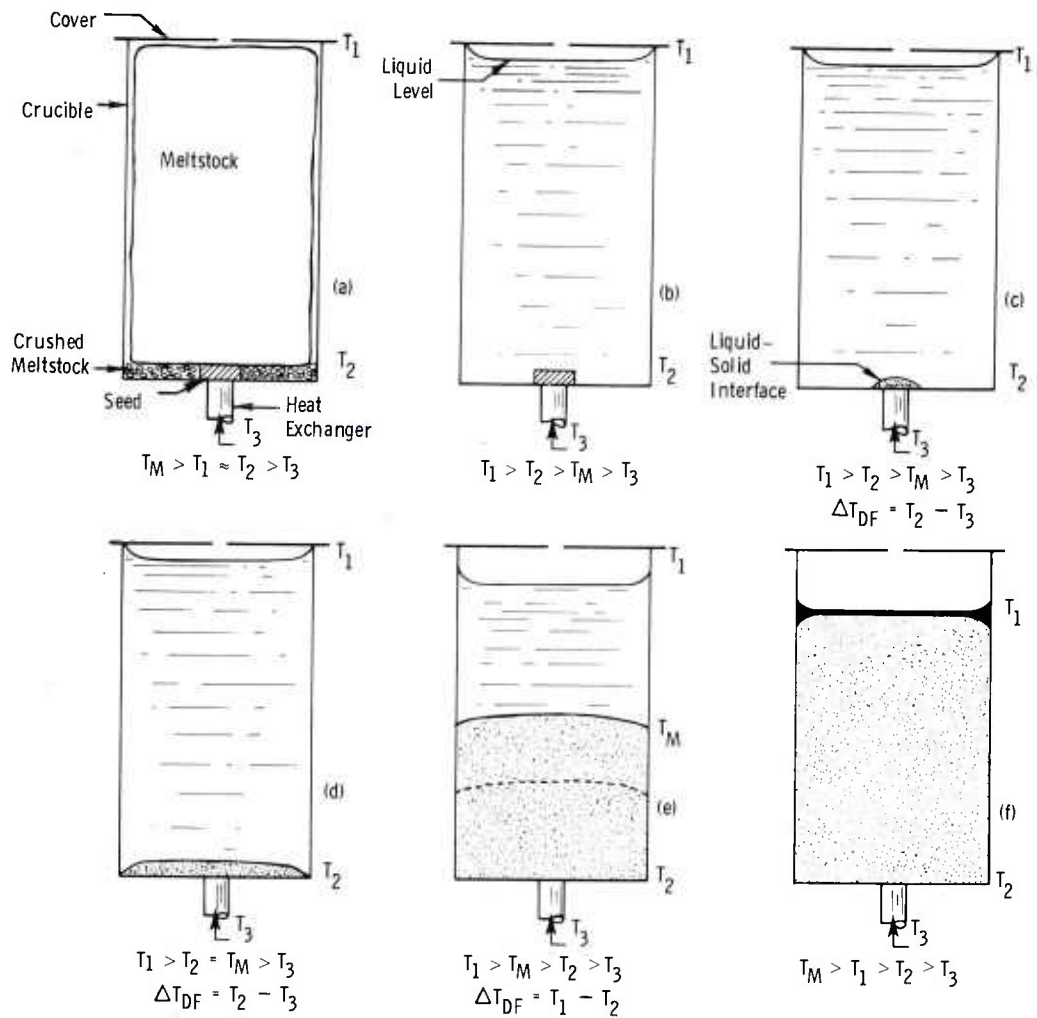
1. Headquarters, Department of the Army, "Operations," Field Manual FM 100-5, Jul 1976, Washington, D. C.
2. Dick, R. A., "The Impact of Blasting Agents and Slurries on Explosives Technology," Information Circular No. 8560, 1972, U. S. Department of the Interior, Bureau of Mines, Washington, D. C.
3. Penn, L. et al., "Determination of Equation-of-State Parameters for Four Types of Explosive," UCRL-51892, Aug 1975, Lawrence Livermore Laboratory, University of California, Livermore, Calif.
4. Reed, H. H., "A Review of Explosives Used in Explosive Excavation Research Laboratory Projects Since 1969," Miscellaneous Paper E-74-6, Dec 1974, U. S. Army Engineer Waterways Experiment Station, Explosive Excavation Research Laboratory, CE, Livermore, Calif.
5. Lane, W. B. et al., "An Experiment in the Field Mixing of Bulk Explosives," Miscellaneous Paper N-77-5, May 1977, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss.
6. Reed, H. H., "Explosive Evaluation: Gelled Nitromethane and Slurry as Military Bulk Explosives Systems," Miscellaneous Paper N-76-9, Jun 1976, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss.
7. Carleton, H. D., "Antitank Ditching with Explosives," Technology Transfer Report E 76-3, Dec 1976, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss.
8. Müller, A. M. and Carleton, H. D., "Explosive Ditching with TNT," Miscellaneous Paper N-77-7, Jul 1977, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss.

LASER RODS FABRICATED FROM AMMRC GROWN Nd:YAG

JAROSLAV L. CASLAVSKY, PhD
*DENNIS J. VIECHNICKI, PhD
ARMY MATERIALS AND MECHANICS RESEARCH CENTER
WATERTOWN, MASSACHUSETTS 02172

INTRODUCTION

The laser will be used extensively on the modern battlefield to greatly improve the accuracy of new Army weapon systems. Some examples of applications are laser range finders in the XM-1 tank and other armored vehicles and laser target designators for infrared guided missiles and projectiles such as HELLFIRE and COPPERHEAD. While lasing ability is found in gases, liquids, and solids (both crystalline and noncrystalline), the laser preferred in the above Army application, because of its high efficiency and narrow spectral linewidth, is a single crystal of a compound of yttrium oxide and aluminum oxide with small additions of neodymium oxide as the active element. The compound $\text{Y}_3\text{Al}_5\text{O}_{12}$ has a garnet crystal structure similar to the naturally occurring gemstone $\text{Mg}_3\text{Al}_2\text{Si}_3\text{O}_{12}$, hence the name yttrium aluminum garnet, acronym YAG or Nd:YAG when doped with neodymium. Currently Nd:YAG laser rods 6.4 mm diameter by 75 mm in length cost the Army more than \$600 each. The objective of this program was to lower the cost of these rods by employing a different growth technique, the heat exchanger method (HEM) (1) shown schematically in Figure 1. This method was originally developed for the growth of large sapphire single crystals for transparent armor applications (2) and has several cost saving features when compared to the state-of-the-art Czochralski (CZ) technique (3). Single crystal boules 3.0 to 3.7 cm diameter by 18 cm in length can be routinely obtained by the CZ technique. A cross section of a boule of this material is seen on the right in Figure 2 and Nd:YAG grown at AMMRC by the HEM technique is seen on the left. In addition to the size difference between the two crystals, another obvious difference is the occurrence of facets, i.e., the flat faces of low index planes, on



- (a) Crucible, cover, meltstock, and seed prior to melting.
 T_M = melting point of Nd:YAG
 T_1, T_2, T_3 = temperatures at various points in crucible.
 (b) Meltstock melted.
 (c) Seed partially melted to insure good nucleation.
 ΔT_{DF} = the temperature difference which is the driving force for crystal growth.
 (d) Growth of crystal commences and covers crucible bottom.
 (e) Unidirectional crystal growth. Note nearly planar liquid-solid interface.
 (f) Crystal growth completed.

Figure 1. Growth of Nd:YAG by the heat exchanger method (HEM).

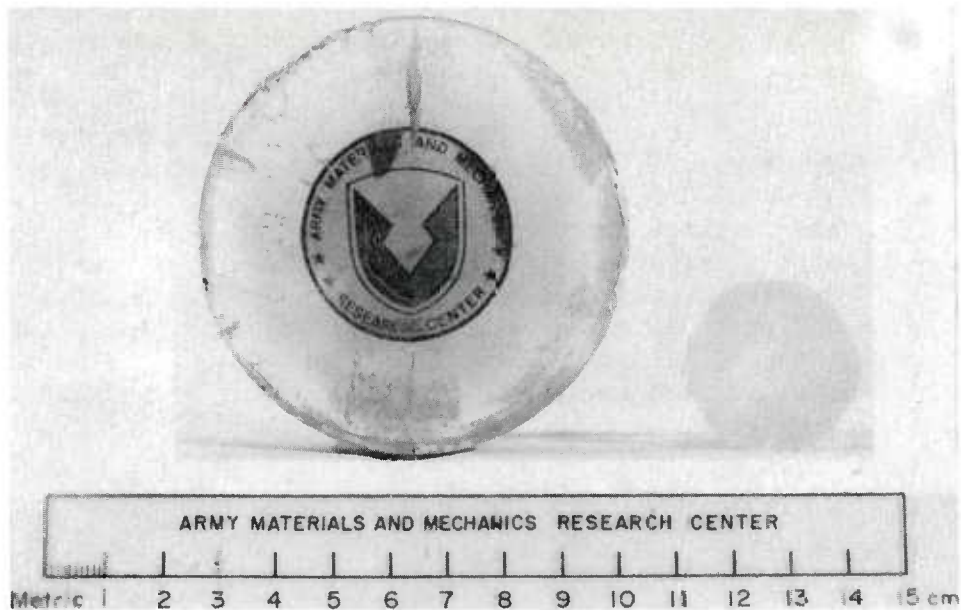


Figure 2. Cross sections of Nd:YAG single crystals grown by HEM (left) and Czochralski technique (right).

the CZ material. Faceting usually occurs during the growth of garnets, and is very difficult to prevent when Nd:YAG is grown by the CZ technique (4). Aside from the flat faces of the crystal, the faceting also manifests itself as a central strained core of material immediately surrounding the axis of the boule. This limits the amount of the boule that can actually be used for laser rods to no more than 26% of a 3-cm diameter boule and 36% of a 3.7-cm diameter boule when 6.4-mm diameter laser rods are considered. The efficiency of material utilization increases with increasing boule diameter, and would increase more if the material in the center of the boule could be used. For example, if 6.4-mm rods could be cut from a 7.5-cm diameter boule, including the center, material utilization efficiency would increase to 83%.

The CZ growth technique for Nd:YAG is also inefficient from the point of material utilization in that only about 50% of the melt can be solidified in a controlled manner to produce a single crystal. The rest of the melt is frozen quickly in the crucible and remelted later after its stoichiometry is adjusted, or discarded when impurity levels become too high. All the material grown by the HEM technique is solidified in a controlled manner and 80% to 90% is available for laser rod fabrication.

Expensive iridium crucibles (>\$10,000 each) are used in the CZ growth of Nd:YAG while low-cost molybdenum crucibles (\$200) are used in

HEM growth of Nd:YAG. Even though the iridium crucibles are reusable and the molybdenum crucibles are not, the latter are still less expensive. Also, since large crystals can be grown only from large crucibles, scale-up of the CZ technique is limited by the cost and availability of large iridium crucibles.

This paper will describe the transfer of the technology gained during the sapphire growth program for transparent armor to the growth of large Nd:YAG single crystals. It will describe starting materials preparation, modifications to the crystal growth process, and growth of 7.5-cm-diameter single crystals 9.7 cm high weighing 2000 g (the largest man-made Nd:YAG crystals in the world), fabrication of 3-mm-diameter by 30-mm-long laser rods, and the lasing ability of these test rods.

EXPERIMENTAL

Starting Materials

Starting materials did not present a problem during the sapphire growth program because of the availability of high purity sapphire crackle, i.e., cracked Vernueil grown boules with impurity levels of less than 40 ppm. These could be placed into crucibles with packing efficiencies of 70%. Any higher packing efficiency was unnecessary because more material in the crucible would "boil over" due to the 22% volume change of sapphire upon melting. No comparable Nd:YAG crackle was available, so meltstock had to be made in-house by sintering isostatically pressed cylindrical billets of mixed oxide powders to 65% to 75% of theoretical density. After firing, the billets would be small enough to slip inside the crucibles. Also, unlike sapphire, very accurate control of composition is necessary so that the melt stock is of the correct stoichiometry, $\text{Nd}_{0.03}\text{Y}_{2.97}\text{Al}_5\text{O}_{12}$. Incongruent vaporization of a constituent during sintering had to be minimized to maintain proper stoichiometry. To determine the process parameters for the Nd:YAG meltstock, a matrix experiment was conducted.

Starting materials were high purity oxides, 99.992% pure alumina,* 99.999% pure yttria,† and 99.0% pure neodymia.‡ These were weighed in correct proportions in 300-g batches to form Nd:YAG, i.e., 37.1 mol % Y_2O_3 /62.5 mol % Al_2O_3 /0.4 mol % Nd_2O_3 . The powders were either mixed by allowing ethanol to boil away in a 2-gallon mill made

*0.3 μm Extra Pure, Adolf Meller Company, Providence, Rhode Island 02904

†563 Y_2O_3 , Molycorp, White Plains, New York 10604

‡ Nd_2O_3 Code 629.9, American Potash and Chemical Corp., Rare Earths Division, West Chicago, Illinois 60185

of high alumina, or milled in the same mill under several different conditions. The time of milling, the number of 1/2"-high alumina balls, and the milling fluid, ethanol or tertiary butanol, were systematically varied. After drying, the powders were placed in a rubber bag and isostatically pressed at 30,000 psi to form rods 15 mm in diameter by 100 mm long. These were cut into 15-mm lengths with a dry diamond saw, and their bulk densities were taken by weighing and then measuring with a micrometer. Volumes were calculated from the physical dimensions. The rods were heated at 1650 C in air in a zirconia felt-lined MoSi₂ furnace for 1, 3, 10, or 10-1/4 hours. After firing determinations of weight loss, shrinkage, and bulk density were made. Phases present were determined by X-ray powder diffraction from powdered samples. The morphology of selected powders before and after milling were studied using a scanning electron microscope (SEM).

Crystal Growth

A hollow, helium-cooled tungsten/molybdenum heat exchanger was used for growth of sapphire to protect the seed and to extract heat from the crucible. With the furnace temperature held constant, the helium flow was increased and crystal growth commenced. This mode of growth resulted in a highly curved solid-liquid interface which was adequate for sapphire but would produce faceting in Nd:YAG (4). This mode of growth also produced large temperature gradients within the crystal and it was feared a serious cracking problem would result. Lastly, these hollow heat exchangers were not very durable and had to be replaced after five or ten crystal growth runs. For these reasons the mode of crystal growth was changed to a "gradient freeze" mode where a temperature gradient was imposed on the melt by redesign of the heating element, and the furnace temperature varied with time according to a predetermined program.

A solid heat exchanger consisting of a tungsten cap, a graphite body, and a water-cooled copper base was substituted for the hollow helium-cooled heat exchanger. Although this heat exchanger did not have the flexibility for varying the seed temperature during the run independently of furnace temperature, it was more durable and still would prevent the seed from melting.

The crystal growth furnace contains a cylindrical graphite heating element. The element is naturally hotter at the bottom than at the top; this is opposite the way it should be for this mode of crystal growth. An engineering redesign study was undertaken to overcome this situation. A series of graphite muffles and a new type of graphite element were machined to try to overcome this difficulty.

An actual crystal growth run takes 12 to 14 days from the time the furnace cover is closed until it is opened again. The first day is used for heat up and melting; the next half day is seeding; the next 8 to 9 days are growth; and the remaining 2-1/2 to 4-1/2 days are cool down. Growth rates vary between 0.3 and 0.8 mm/h. Temperature gradients within the furnace vary between 1 to 10 C/cm.

Fabrication

Slabs of Nd:YAG were rough cut with a diamond saw from selected regions of single crystals. The slabs were then cut into rectangular prisms, and subsequently centerless ground into 3-mm-diameter by 30-mm-long rods. The ends of the rods were ground and polished flat to within a tenth of a wavelength and parallel to each other to within 10.0 seconds. The ends were free of pits, scratches, and streaks. They were then coated with an anti-reflective coating of MgF_2 a quarter wavelength thick or $\sim 4700 \text{ \AA}$.

Characterization

As-grown boules were studied visually in normal and polarized light for gross flaws such as cracks and voids. The regions containing second-phase inclusions were determined with a microscope in oblique illumination. When necessary, plates were cut from crystals and polished. From these plates the progress of the solid-liquid interface during the growth process could be determined and correlated with changes in furnace conditions, especially the thermal gradient imposed by the furnace. The polished plates were observed with a transmitted light microscope and details of second phase and other scattering centers determined.

The scattering centers in the laser rods were determined by examining them visually under a strong light. The flatness of the ends was determined using an optical flat and then counting the fringes from a photograph. Fluorescent lifetimes were measured and compared against calibration curves to determine the actual concentration of Nd in the rods. Long pulse lasing efficiency was determined using an AN/GVS-5 cavity with a 20- μf capacitor.

RESULTS AND DISCUSSIONS

Starting Materials

Pressed meltstock billets had to be fired at 1650 C for one hour or more to attain densities above 65% of the theoretical density of Nd:YAG (4.56 g/cm^3). Figure 3 shows the results of several mixing

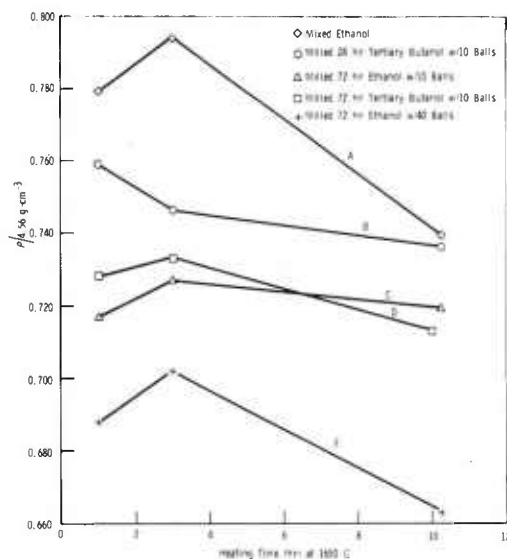


Figure 3. Densification of sintered billets of Nd:YAG expressed as bulk density of billets divided by the theoretical density of Nd:YAG as a function of heating time in air at 1650 C and different mixing and milling conditions.

that the meltstock must be single-phase YAG to get a controlled melt and the desired stoichiometry in the Nd:YAG crystal.

Furnace Design

The furnace element used for the sapphire growth program, on the right in Figure 5, is a "ribbon" type where power is brought in from the top. The current follows on a ribbon or sinusoidal path through the element as evidenced by the pitted regions on its surface. This type of heating element contains alternate hot and cold spots along the ribbon which produce circumferential thermal gradients and is hotter at the bottom than at the top. This was not a problem for growth of sapphire because the cooling efficiency of the helium-cooled heat exchanger could compensate for the reversed gradient and the hot spots and still produce a sufficient thermal gradient for growth. With Nd:YAG where a flat solid-liquid interface is needed, the helium-cooled heat exchanger is too powerful. It was necessary to build the upward temperature gradient into the heating element. It was difficult to machine the desired taper in a heating element such as the one on the right of

and milling procedures on the density of sintered billets of Nd:YAG as a function of sintering time at 1650 C. In all cases a maximum is reached after three hours. This would be enough if the $Y_2O_3-Al_2O_3-Nd_2O_3$ system were a simple ternary. However, the $Y_2O_3-Al_2O_3$ binary contains two other compounds besides $Y_3Al_5O_{12}$: $YAlO_3$ ($\rho = 5.35 \text{ g/cm}^3$) and $Y_4Al_2O_9$ ($\rho = 4.41 \text{ g/cm}^3$) (5). When X-ray diffraction phase analysis was done on the pressed billets prepared under the conditions specified in Figure 3, it was found that they were usually not single-phase, but contained mixtures of all three of the above compounds plus traces of the starting materials. Figure 4 shows that heat treatment alone will not produce single-phase meltstock of only $Y_3Al_5O_{12}$. It is important to mill the material to insure good mixing and to break up agglomerates which would react together and form unwanted compounds. Previous studies (6) had indicated

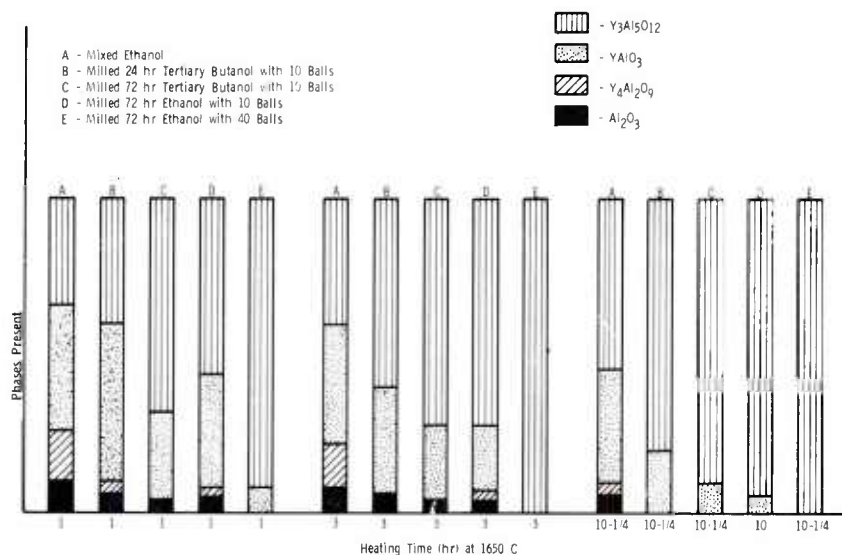


Figure 4. Phases present after heating powder mixtures of composition 37.1 m/o Y_2O_3 /62.5 m/o Al_2O_3 /0.4 m/o Nd_2O_3 .

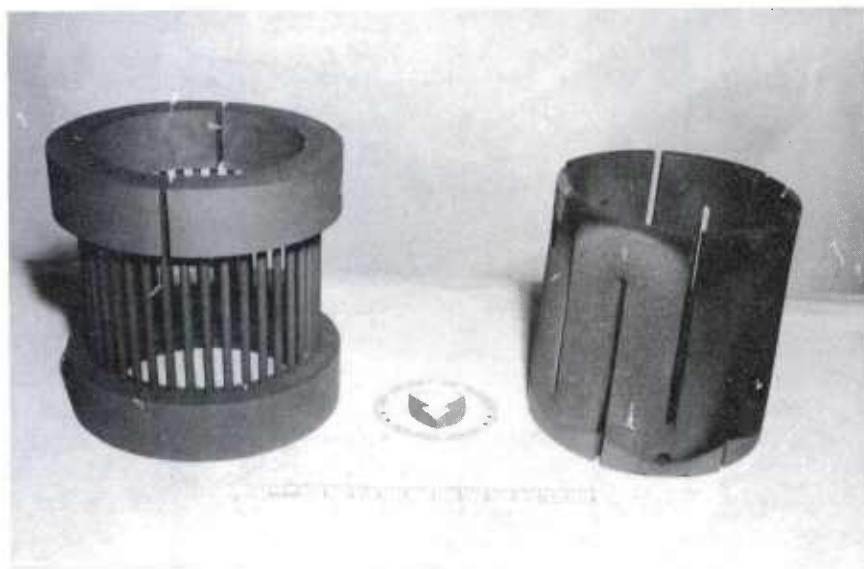


Figure 5. Graphite heating elements used in HEM furnace: "Birdcage" type is on the left and "ribbon" type is on the right. Note pitted regions on "ribbon" type indicative of electric current path.

the figure. Interim solutions, such as contoured muffles which fit inside the element, were not totally satisfactory because not enough of a gradient would be introduced this way. Therefore, the furnace element design at the left was selected. This "birdcage" design is all graphite and consists of a complete bottom ring, a split upper ring, and many tapered rods connecting the upper and lower portions. The individual rods can be tapered as needed to introduce the desired temperature gradient. Initially, the power feeds came through the top ring. It was found that because the power feeds act as powerful heat sinks the gradient in the element was still not satisfactory. This problem was corrected by inverting the element, which brought the power feeds to the bottom of the element. With the heat sink at the bottom, a sufficient gradient, i.e., 10 C/cm was built into the element over the 12-cm height of the crucible.

Crystal Growth

Unfaceted Nd:YAG single crystals have been grown by HEM. Scale-up efforts over the last several years have increased crystal sizes from 5.32 cm diameter by 3.50 cm high weighing 355 g to 7.62 cm diameter by 10.5 cm high weighing 2180 g, as seen in Figure 6. It is believed that the 2180-g crystals are the largest ever made in the world. The crystals shown were doped at levels between 0.25 and 0.75 at. % Nd which is below the 1.1 at. % required. These crystals are free of scattering centers throughout 90% of their volume. At dopant levels of 1.1 at. %, crystals are free of scattering centers for no more than 50% of their

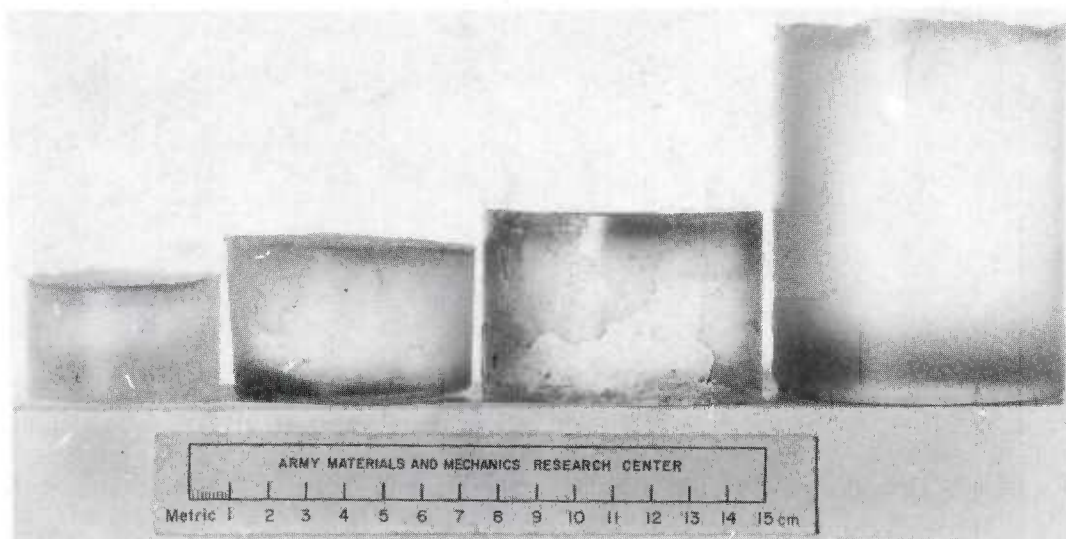


Figure 6. Scale-up of Nd:YAG single crystals.

volume. This is because growth becomes more difficult with greater dopant concentrations. More careful attention must be paid to attaining and maintaining stoichiometry because of the complex phase equilibria (5), and to achieving the required temperature gradients and growth rates. Nd:YAG must be grown slower than undoped YAG (7). While an average growth rate can be calculated from crystal size and time of growth, the instantaneous growth cannot yet be measured directly. Thus, although the change of growth rate with time should equal zero, it does not. This means that different parts of the crystal grow at different rates.

Scattering centers can be seen with the unaided eye. Detailed structure (Figure 7) can be seen with a transmitting light microscope. The scattering centers are channels which are mostly empty except for some second-phase α -Al₂O₃ or YAlO₃, depending upon whether the liquid is Al₂O₃-rich or Y₂O₃-rich. This structure is characteristic of one formed when the planar solid-liquid interface breaks down because of constitutional supercooling (8). The channels are originally liquid regions which freeze well below the melting point of the rest of the crystal because they contain higher concentrations of impurities and/or dopant. Thus the crystal freezes around them. When they finally freeze, the material contracts leaving mostly empty space in the channels.

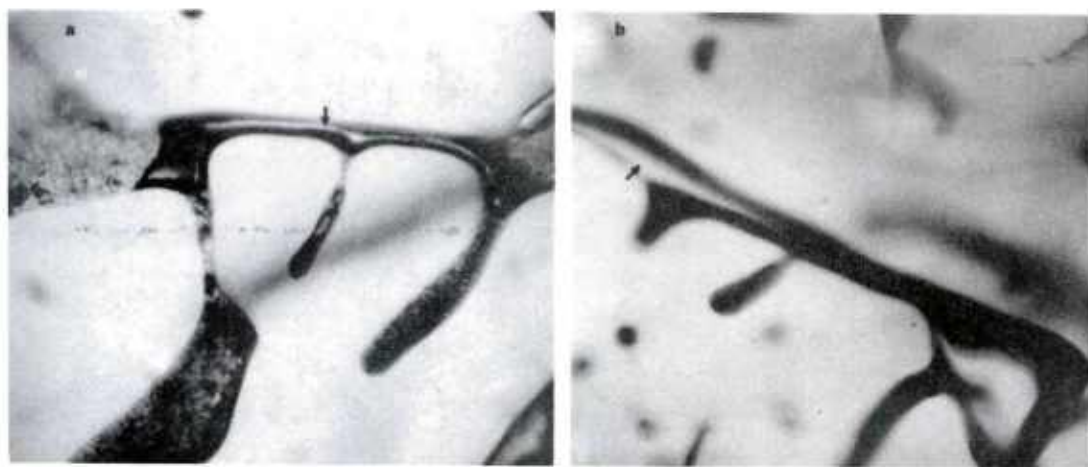


Figure 7. Optical photomicrographs of second-phase inclusion in YAG single crystals. Plate a - inclusion of perovskite; plate b - inclusion of α -alumina. The arrows denote voids in channels containing inclusions. These voids were created due to the liquid-solid contraction. Therefore it could be concluded that the second-phase liquid solidified in already solid YAG matrix.

Laser Rod Fabrication

Laser rods were fabricated from the lower half of a nominally 1.1 at. % Nd:YAG. As seen in Figure 8, these were 3 mm diameter by 30 mm long and were cut along a radius of the single crystal. The size was dictated by the lasing chamber which was used for characterization.

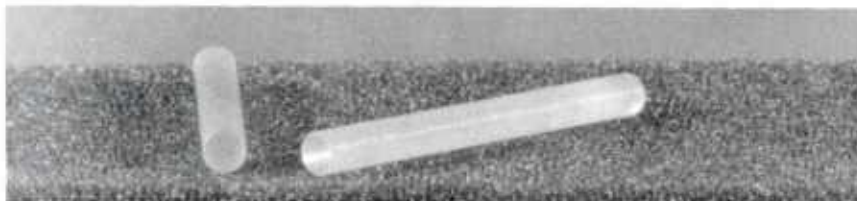


Figure 8. 3 x 30 mm laser rods made from AMMRC-grown Nd:YAG single crystal - scale 1:2.

Lasing Properties

The fluorescent lifetime of the rods was determined to be 254 ± 2.5 μsec . When compared with predetermined calibration curves relating fluorescent lifetime to Nd concentration in the rods, AMMRC rods were found to have a concentration of 0.6 to 0.8 at. % Nd.* Lasing efficiency for long pulse operations is seen in Figure 9. The slope of the curve $\eta = 0.2\%$. This is less than obtained from CZ Nd:YAG where

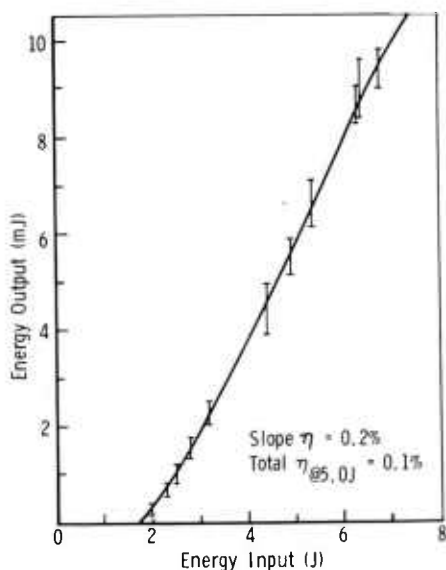


Figure 9. Energy output versus energy input for HEM-grown laser rods. $R = 75\%$.

*STROZYK, J. ERADCOM, Private Communication, 28 February 1978.

$\eta \approx 1.0\%$. This value is also dependent upon Nd concentration. The purpose of these tests was to demonstrate that material made by HEM could lase, which it did. No attempt was made to optimize laser performance. It did appear, however, that performance could have been better if the material contained higher concentration of Nd. Studies are now underway to increase Nd concentration in AMMRC laser rods, including a detailed investigation of the ternary phase equilibrium diagram in the region near $Y_3Al_5O_{12}$.

SUMMARY

This paper has described HEM and its application to the growth of Nd:YAG. It has detailed the advantages of growing Nd:YAG by HEM as compared to the state-of-the-art CZ technique. The technology transfer from a technique originally developed for the growth of large sapphire ingots has been covered. The particular problem areas have been melt-stock preparation, furnace redesign to insure a planar solid-liquid growth interface and hence facet-free crystals, and definition of process parameters. HEM-grown Nd:YAG does lase, and studies are now underway to optimize its lasing properties by increasing Nd concentration in the laser rods.

ACKNOWLEDGMENTS

The authors wish to thank J. Smith, Crystal Optics, Inc., for laser rod fabrication, R. F. Belt, Litton Industries, Inc., for the anti-reflective coatings, J. Strozyk, ERADCOM, for measurement of lasing properties, and W. H. Earle, AMMRC, for valuable technical assistance throughout the program.

REFERENCES

1. VIECHNICKI, D. J., and SCHMID, F. *Crystal Growth Using the Heat Exchanger Method (HEM)*. J. Crystal Growth, v. 26, no. 1, 1974, p. 162-164.
2. VIECHNICKI, D. J., and SCHMID, F. *Growth of Large Single Crystals of Al_2O_3 by the Gradient Furnace Technique for Transparent Armor Applications (U)*. Army Science Conference Proceedings, v. 4, 20-23 June 1972, p. 362-374, AD 522726.
3. BELT, R. F., PUTTBACH, R. C., LATORE, J. R., and DENTZ, D. *Production Engineering of Nd:YAG Laser Rods for Laser Illuminator Transmitters*. First Report October 1970. December 1972, Contract DAAB-25-71-6-2611, Department of the Army, U.S. ERADCOM, 225 S. 18th Street, Philadelphia, Pennsylvania 19103, AD 908626.
4. BASTERFIELD, J., PRESCOTT, M. J., and COCKAYNE, B. *An X-ray Diffraction Topographic Study of Single Crystals of Melt-Grown Yttrium Aluminum Garnet*. J. Matls. Sci., v. 3, 1968, p. 33-40.
5. TORPOV, N. A., BONDAR, I. A., GALAHBOV, F. Ya., NIKOGOSYAN, Kh. S., and VINOGRADOVA, N. V. Izv. Akad. Nauk SSR, Ser. Khim, no. 7, 1964, p. 1158-1164.
6. VIECHNICKI, D. J., and CASLAVSKY, J. L. *Solid State Formation of $Nd:Y_3Al_5O_{12}$ (Nd:YAG)*. Army Materials and Mechanics Research Center, AMMRC TR 78-7, February 1978.
7. BELT, R. F., PUTTBACH, R. C., and LEPORE, D. A. *Crystal Growth and Perfection of Large Nd:YAG Single Crystals*. J. Crystal Growth, v. 13/14, 1972, p. 268-271.
8. CHALMERS, B. *Principals of Solidification*. John Wiley & Sons, Inc., New York, 1964, p. 150-154.

FLIR IMAGE ENHANCEMENT BY AUTOMATIC LOW
FREQUENCY GAIN LIMITING (U)

*SEN-TE CHOW, MR.
JOHN J. PUPICH, MR.
NIGHT VISION & ELECTRO-OPTICS LABORATORIES
FORT BELVOIR, VA. 22060

I. INTRODUCTION

An inherent problem with existing Military FLIR Common Module systems⁽¹⁾ is that the dynamic range of the thermal scenes far exceeds that of the system display. Current displays used in FLIR systems have a dynamic range of approximately 30 db (32:1) and typical infrared scenes can be several orders of magnitude greater. A common example being the terrain-sky horizon which an Airborne FLIR encounters when the helicopter or aircraft makes a steep bank. In this scenario some of the channels in the vertical array of detectors will be viewing both the hot terrain and cold sky. This scene dynamic range can be as high as 500 to 1, or 54 db, with a temperature difference, ΔT of 100°C and the noise equivalent differential temperature (NE ΔT) of 0.2°C . The display of the channels will be such that the terrain saturates to white and the sky is suppressed to black, assuming that the system polarity switch is in the white/hot mode. Now if the operator attempts to reduce the gain of the system, such that both areas are within the dynamic range of the display, he is unable to perceive detail in both regions due to a lack of contrast. At this moment the operator may adjust the brightness level and maintain the contrast high enough to perceive either the terrain or the sky detail but not both simultaneously. In addition, the operator has to constantly adjust both the brightness and the contrast controls for the various types of scenarios encountered in flight and if the operator is the pilot, this reduces his ability to navigate. To accommodate approximately 60 db of scene information onto a 30 db display a large degree of signal compression is required. Classical methods of automatic gain control and logarithmic compression techniques are unable to preserve the detail signal over such a large

dynamic range and provide the fast response time required in missile tracking applications. This paper describes an electronic approach to solve this problem called "Automatic Low Frequency Gain Limiting"⁽²⁾. This circuit limits the amplitude of the signal pedestal and enhances the high frequency scene detail. Thus it expands the perceivable scene dynamic range and provides the fast response time required by many military applications.

II. PRINCIPLE OF AUTOMATIC LOW FREQUENCY GAIN LIMITING

A. Principle of Operation

The Automatic Low Frequency Gain Limiting circuit operates in two modes depending on the input signal level. When the signal amplitude is within the prescribed detection levels, it is fed directly to the output. However, as the signal amplitude either exceeds the white detection level or is below the black detection level, it compresses the pedestal, the average value of the local background, and preserves detail, the temperature variation of interest in all parts of the scene. Since this approach essentially preserves most of the high frequency detail and automatically limits the average value of the local background, which mainly consists of low frequency components, it is called "Automatic Low Frequency Gain Limiting (ALFGL)."

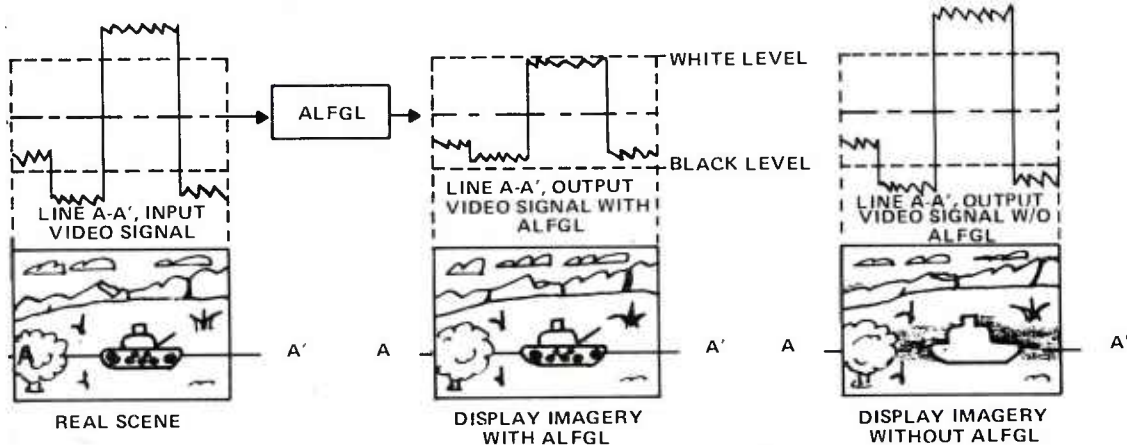


Fig. 1A.
System with ALFGL

Fig. 1B.
System without ALFGL

Figure 1A exhibits the display imagery and the video signals at line A-A' of the system with ALFGL. In this case: (i) the tree signal is within the detection levels, thus, it is fed directly to the

output, (ii) the signal level of the hot tank exceeds the white detection level, therefore, the pedestal is compressed, yet the tank detail is preserved and (iii) the terrain signal levels on both sides of the tank are below the black detection level, where the terrain pedestals are clipped and the detail signals are preserved. In Figure 1B the system does not have ALFGL. In this case, the tank signal saturates to the display white level and the cold terrain signal is suppressed below the display black level. Therefore, the scene detail is lost in the terrain area on both sides of the tank and is also lost in the local horizon.

B. Signal Transfer Characteristic

The ALFGL signal transfer characteristic is shown in Figure 2. In the nonlimited condition, i.e., the input signal is within $-e_1$ and e_1 , both pedestal and detail signal operate in the linear region A' O A. When the input pedestal exceeds e_1 , the output

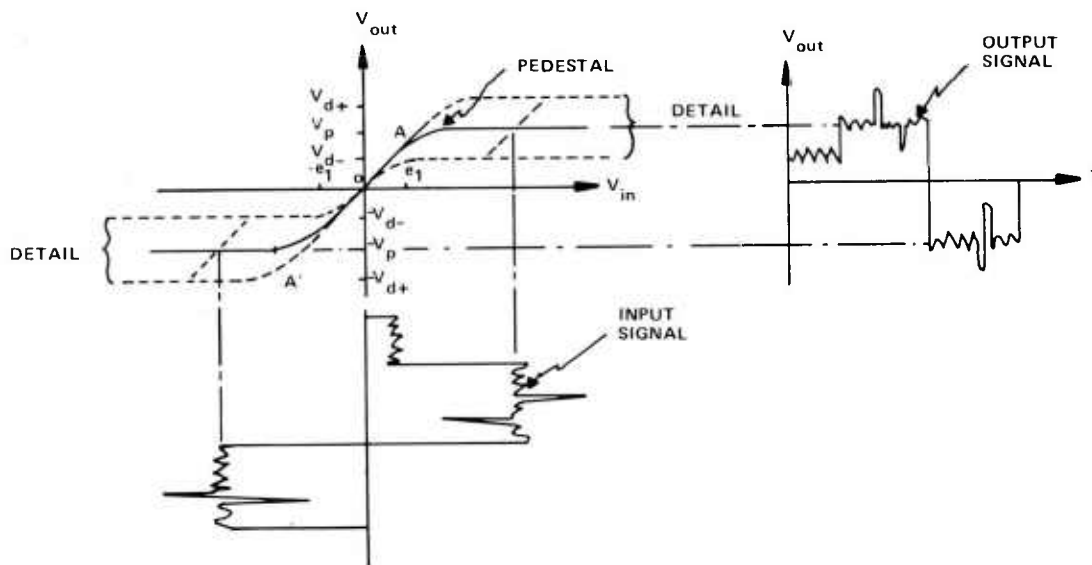


Fig. 2. Signal transfer characteristics

of the pedestal is limited to V_p , the detail signal is superimposed on this limited pedestal and operates within the limited range of V_{d+} to V_{d-} . Similarly, when the input pedestal is more negative than $-e_1$, it is limited to $-V_p$ and the detail signals operate within the limited range of $-V_{d+}$ to $-V_{d-}$.

C. Frequency Response

The frequency response of the ALFGL is shown in Figure 3.

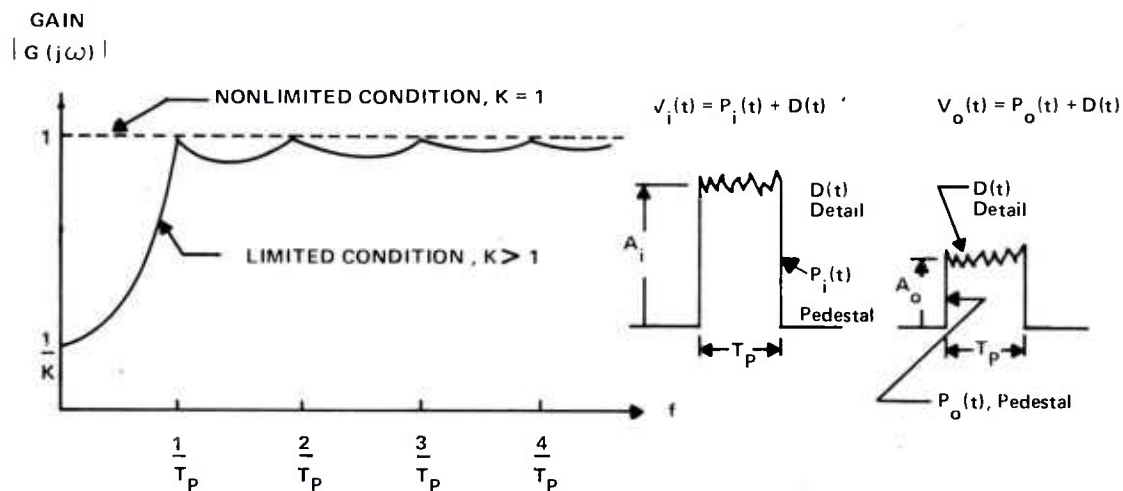


Fig. 3. Frequency response of ALFGL

Here we assume that the input signal, $V_i(t)$ is the sum of the pedestal, $P_i(t)$ and its detail, $D(t)$. The width of the pedestal is T_p . Similarly, the output signal, $V_o(t)$, is the sum of the limited pedestal, $P_o(t)$ and its detail, $D(t)$, the pedestal width is T_p .

If we take the Fourier transform of both input and output signals, then the transfer function is:

$$G(j\omega) = \frac{V_o(j\omega)}{V_i(j\omega)} = \frac{P_o(j\omega) + D(j\omega)}{P_i(j\omega) + D(j\omega)} = \frac{1 + \frac{D(j\omega)}{P_o(j\omega)}}{K + \frac{D(j\omega)}{P_o(j\omega)}} \quad (1)$$

$$\text{where: } K = \frac{P_i(j\omega)}{P_o(j\omega)} = \frac{A_i}{A_o} \begin{cases} > 1, \text{ limited condition} \\ = 1, \text{ nonlimited condition} \end{cases} \quad (2)$$

$$\text{and } P_o(j\omega) = A_o T_p \frac{\sin \omega T_p / 2}{\omega T_p / 2} \quad (3)$$

If we assume that the detail is much smaller than the output pedestal, then by inspection of equations (1) through (3) we have:

- (i) $|G(j\omega)| \approx \frac{1}{K}$ at $f = 0$
- (ii) $|G(j\omega)| = 1$ at $f = \frac{n}{T}$ where: $n = 1, 2, 3, \dots$
- (iii) Attenuation is more dominant in the low frequency.

The approximate frequency response of the ALFGL in the limited condition is shown in the solid curve of Figure 3. In the nonlimited condition, where $K = 1$ and the $|G(j\omega)| = 1$ for all values of f . This is shown in the dotted line of the same figure.

III. ALFGL IMPLEMENTATION

Figure 4 depicts the block diagram of the ALFGL and its position in the common module video chain. The advantages of inserting

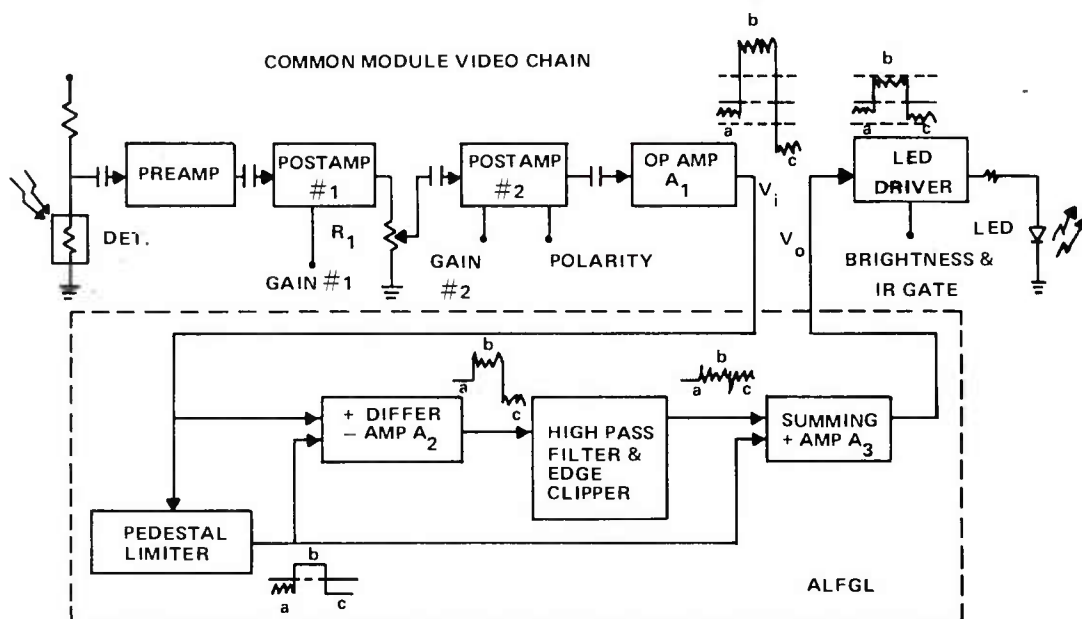


Fig. 4. ALFGL block diagram

ALFGL before the LED driver are: (i) it reduces the DC level shifting since video signals are directly coupled from the ALFGL to the display, (ii) semiconductor diodes can be selected for both black and white levels of detection since the signal amplitude is 0.5 to 1.5 volts, (iii) the system gain control function precedes the ALFGL circuit and thereby a change in the gain control setting will not effect the pedestal limiting level. This assures that the output signals are always within the dynamic range of the display and (iv) for remote view FLIRs, dynamic range expansion is not limited by the dark current of either the CCD image buffer or vidicon tube.

To analyze the operation of the ALFGL circuit consider input signals at three different signal levels: "a", "b" and "c". Assume that: level "a" is within the display dynamic range, level "b" exceeds the black detection level (it is inverted because the LED driver has a negative gain), and level "c" exceeds the white detection level.

(i) When an input signal of level "a" is fed into the ALFGL circuit, it passes through the pedestal limiter without attenuation and is directly fed to the lower input side of the summing amplifier, A_3 . The signals at both inputs of the differential amplifier, A_2 , are identical. Therefore, the differential amplifier output voltage is zero as is the upper side input of the summing amplifier, A_3 . Consequently, the output of the summing amplifier is the lower side input signal, which is the same signal fed into the ALFGL circuit.

(ii) When the input signal level "b" is fed into the ALFGL circuit, the pedestal limiter clips the signal amplitude to the black limiting level. Then, it is directly fed into the lower input side of the summing amplifier, A_3 . The output of the differential amplifier, A_2 , contains both the detail and a reduced amplitude pedestal; the latter is then removed using a high pass filter and edge clipper to limit leading and trailing edge of the spikes. Thus, the output of the summing amplifier, A_3 , is the sum of the input detail and the black level limited pedestal.

(iii) When the input signal level "c" is fed into the ALFGL circuit, the circuit is similar to case (ii) except that the pedestal limiter clips the signal to the white limiting level and the output of the summing amplifier, A_3 , is then the sum of the detail and the white level limited pedestal.

Figure 5 shows the ALFGL schematic diagram where the operational amplifier A_1 is provided for dynamic range expansion. This avoids the presaturation of the post amplifier #2, which has an output

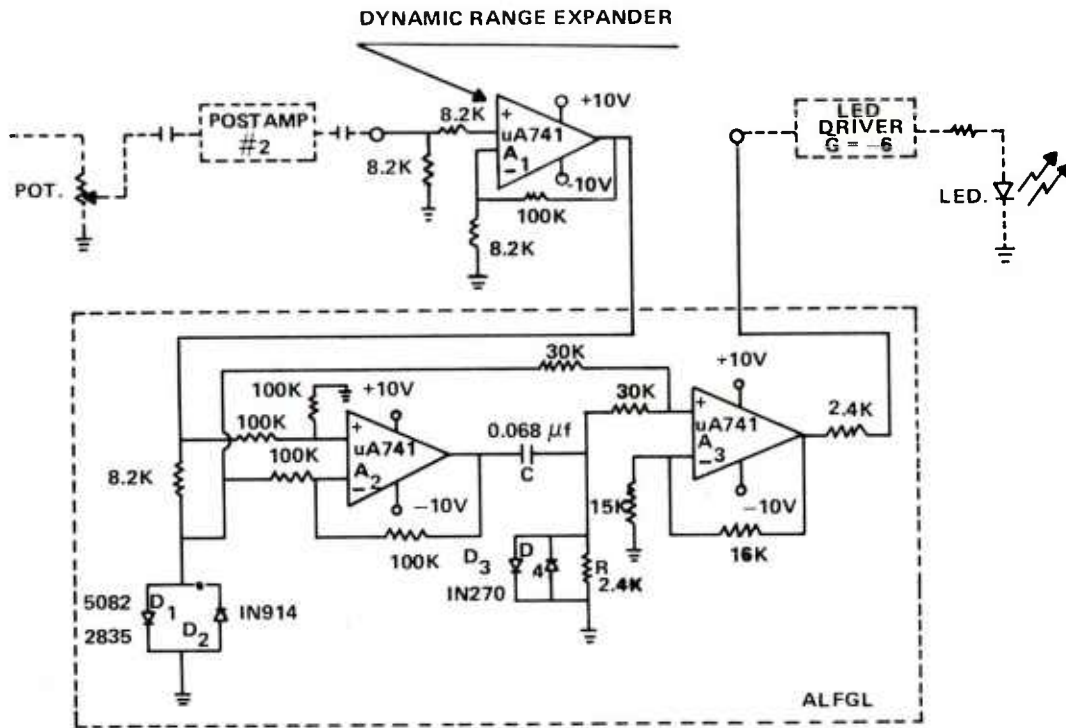


Fig. 5. ALFGL schematic diagram

voltage swing of 1.5 volts peak-to-peak. This is achieved by scaling down the resistor pot and compensates the attenuated gain by the operational amplifier A_1 which has a large output voltage swing. Diode D_1 is a black level detector which has a threshold voltage of 0.34 volt and diode D_2 is a white level detector with a threshold voltage of 0.5 volt. Amplifier A_2 is a differential amplifier, and the capacitor C and resistor R constitute a high pass filter for the detail signal path. Diodes D_3 and D_4 clip the high peak leading and trailing edges. Amplifier A_3 is a summing amplifier, which sums the detail signal across the resistor R and the pedestal across the diodes D_1 and D_2 . The output of the amplifier A_3 is fed to the LED driver, which has a gain of -6.

IV. PERFORMANCE DATA

A. Voltage Waveforms

Typical voltage waveforms at the corresponding nodes of the ALFGL circuit are depicted in Figure 6 where the input signal is comprised of a 2 volts, 1.5 milliseconds wide pedestal and a 0.3 volts

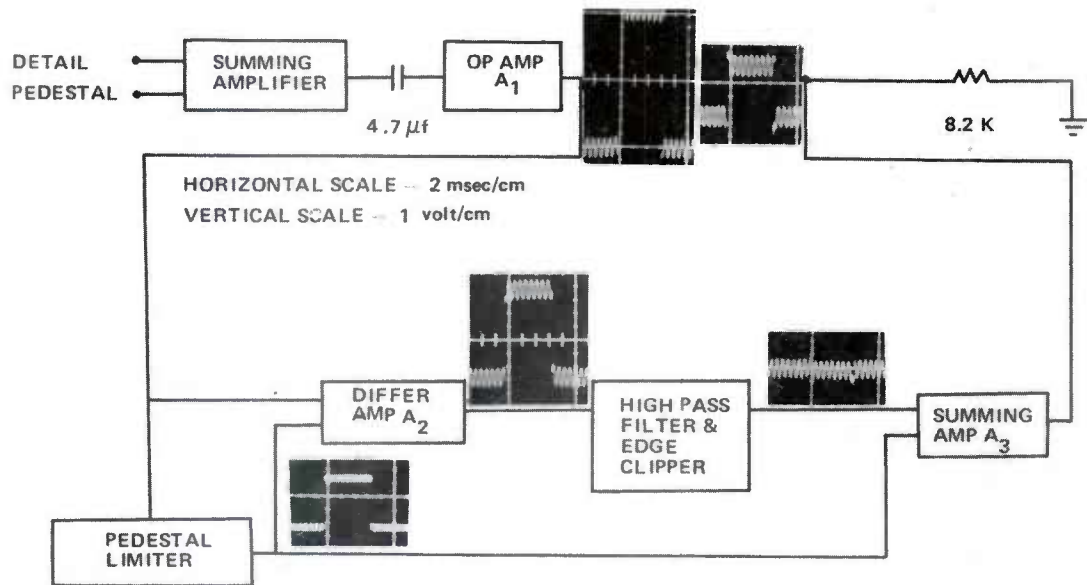


Fig. 6. Voltage waveforms

peak-to-peak 6 KHz sine wave. The signals which exceed either +0.35 or -0.5 volts are clipped off by the pedestal limiter and the signal which is lost at the pedestal limiter appears at the output of the differential amplifier, A_2 . A high pass filter and edge clipper remove the remainder of the pedestal and clip the leading and trailing edge of the spikes. Thus the output of the high pass filter and clipper is the detail signal which was lost at the pedestal limiter. Consequently, the summing amplifier, A_3 , simultaneously adds the same input detail signal and limited pedestal.

B. Dynamic Range Expansion

Amplifier A_1 in Figure 5 has a large output voltage swing of 17 volts peak-to-peak when biased with ± 10 volts. Since one volt peak-to-peak to the input of the LED driver will saturate the display the dynamic range expansion contributed by the ALFGL is 17/1 volts/volt or approximately 25 db. This allows the systems operator to maximize the gain of the system and enhance scene detail without saturating the display.

C. Minimum Resolvable Temperature (MRT)

Figure 7 depicts the system MRT, with and without ALFGL. Test results indicate that no degradation of MRT occurs with ALFGL circuitry in the system. This is logical since MRT is measured with a low signal level below the detection levels of the ALFGL circuit and is therefore fed directly to the output. Therefore, MRT should not be effected by ALFGL.

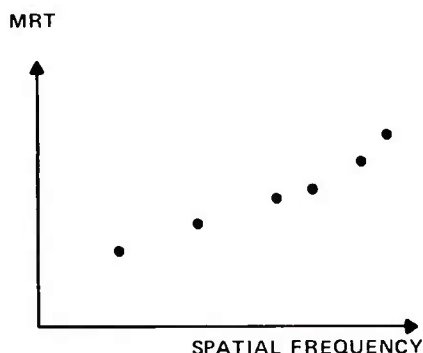


Fig. 7. Minimum Resolvable Temperature with and without ALFGL

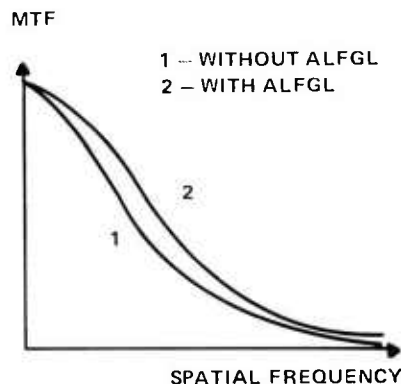


Fig. 8. Modulation Transfer Functions

D. Modulation Transfer Function (MTF)

Figure 8 shows how the system MTF is boosted when ALFGL is employed. This occurs because during the MTF measurement the signal amplitude generally exceeds the detection level of the ALFGL circuit. This results in boosting since the low frequency components are attenuated more than the high frequency components.

E. Narcissus Effect and $1/f$ Noise

The Narcissus effect, i.e., a dark image that appears in the center area of a display due to the detectors sensing their own cold surfaces relative to their warm surrounding⁽³⁾ and $1/f$ noise are substantially reduced by the ALFGL circuits. This occurs since both the narcissus effect and $1/f$ noise are essentially low frequency components which are limited by the ALFGL circuitry.

V. FIELD TEST RESULTS

A. Combat Vehicle Applications

In a combat vehicle battlefield environment, targets are frequently encountered in the vicinity of large fires and burning vehicles. For a FLIR to perceive target detail in this scenario it must have a large dynamic range. Fire tests were conducted in the field to determine the capability of the ALFGL circuitry in this extreme thermal scenario. The tests indicate that the 25 db of dynamic range expansion of the ALFGL FLIR is barely adequate.

B. Airborne Applications

Helicopter flight tests were conducted with a FLIR system in which half the FLIR channels contained ALFGL circuitry and the other half contained standard common module electronics. The results of these tests clearly indicate that the serious image streaking problem (loss of scene detail) in the vicinity of the horizon is eliminated. This is shown in Figure 9 where the top half of the scene illustrates the imagery of the ALFGL circuits. Figure 10 shows the IR imagery of a tower and illustrates how the upper tower details are enhanced by the ALFGL circuitry.

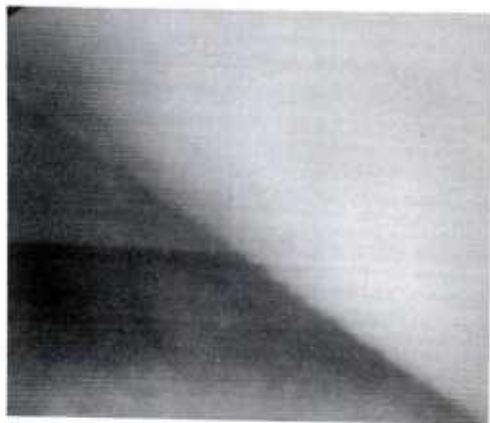


Fig. 9. Horizon.
Upper half with ALFGL
Lower half without ALFGL



Fig. 10. Tower.
Upper half with ALFGL
Lower half without ALFGL

C. Missile Tracking Applications

Another problem with Common Module FLIR systems was their inability to perceive targets in the vicinity of the rocket plume. A static firing test was conducted with an ALFGL modified FLIR, where simulated targets were placed in the vicinity of the rocket plume. Figure 11 shows the results of this test and indicates that the simulated targets can be observed during the rocket firing.



Fig. 11. Rocket plume test. Square targets are not obscured with ALFGL circuitry

D. Ground Applications

In ground applications targets are frequently obscured by hot backgrounds such as buildings, roads and runways. Figure 12A depicts FLIR imagery of a man target that is obscured while standing on a hot runway where the IR channels in the upper half of the scene contain ALFGL circuitry and those in the lower half of the scene contain standard common module circuitry. Figure 12B is the identical scene except that the man is standing in the upper portion containing the ALFGL circuits and is clearly recognized. Figure 13A is the FLIR imagery of a tree-sky scene without ALFGL where the trees are saturated to white and the sky is suppressed to black. Figure 13B is the identical scene with ALFGL showing detail enhancement.

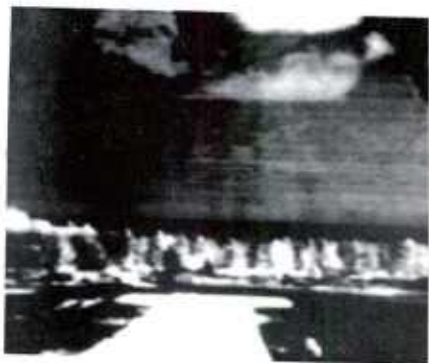


Fig. 12A. Man on a runway
is obscured in the lower
half of the scene without
ALFGL



Fig. 12B. Man on a runway
is seen in the upper half
of the scene with ALFGL



Fig. 13A. Trees and sky
without ALFGL



Fig. 13B. Trees and sky
with ALFGL

VI. ALFGL DEVELOPMENT

During 1975 an internal development program was initiated at the Night Vision Laboratory to design and fabricate the ALFGL circuitry. Demonstration of the first ALFGL circuit was held in June 1975 with 80 channels of circuitry designed into the common module M.O.D. FLIR system. Once circuit feasibility was demonstrated a contract was awarded to Circuit Technology Incorporated⁽⁴⁾ to package the circuitry using thick film technology for airborne system flight testing. This circuitry provided 25 db of dynamic range expansion and operated over the temperature range of -54°C to $+71^{\circ}\text{C}$. Common Module FLIR post amplifier boards are currently being redesigned to incorporate this ALFGL circuitry. These ALFGL post amplifier boards will be used in the Pilot Night Vision Sight and Target Acquisition Designator FLIR systems being designed for the Advanced Attack Helicopter. In 1977 a development contract was awarded to the Hughes Aircraft Corporation⁽⁵⁾ to develop a low cost, monolithic integrated circuit version of ALFGL with 34 db of dynamic range expansion and automatic gain control for hands-off operation of the FLIR. This improved circuitry will be installed in all Army FLIR systems beginning FY81.

VII. CONCLUSIONS

The principal of Automatic Low Frequency Gain Limiting and its application to parallel scan FLIR systems has been described. ALFGL allows the system gain to be maximized without saturating the display. This enhances scene detail and improves the systems recognition capability. Field tests have demonstrated that 25 db of ALFGL dynamic range expansion is sufficient for airborne applications to eliminate horizon streaking and enhance terrain features. Tests have also demonstrated that the response time is adequate for missile tracking applications. Development efforts are currently being directed towards designing an ALFGL Circuit with 34 db of expansion which is required in some missile tracking applications and in scenarios containing fires. Automatic Gain Control will also be incorporated into the circuitry to provide hands-off capability to the FLIR.

VIII. REFERENCES

1. P.J. Daly et al., United States Patent Number 3760181, "Universal Viewer for Far Infrared", 18 September 1973.
2. Sen-Te Chow, "Automatic Low Frequency Gain Limiting for A-C Coupled Signal Processing Systems", Proceedings of IRIS, I.R. Imaging Special Group, February 1976.
3. J.M. Lloyd, "Thermal Imaging Systems", Plenum Press, 1973, pg 275.
4. Contract Number DAAG53-76-C-0192.
5. Contract Number DAAK70-77-C-0204.

COHERENT OPTICAL CORRELATION IN REAL TIME
FOR MISSILE TERMINAL GUIDANCE (U)

*CHARLES R. CHRISTENSEN, PhD, JURIS UPATNIEKS, M.S.E.E.
AND BOBBY D. GUENTHER, PhD
US ARMY MISSILE RESEARCH AND DEVELOPMENT COMMAND
REDSTONE ARSENAL, ALABAMA 35809

INTRODUCTION

The use of area correlation in terminal guidance requires that the system cross correlate a stored reference with the observed scene and have the capacity for handling variations in aspect angle, rotation, scale and intensity. This correlation must be made in real time at a low false alarm rate.

Digital techniques can accomplish the preceding requirements but have several limiting characteristics. The number of resolution elements that can be processed is limited by the available core memory. Even with well-chosen algorithms, a large number of multiplications and additions are required and these increase with the number of resolution elements. For the hypothetical situation of cross-correlation of a reference pattern having 100×100 resolution elements against a scene with 200×200 elements, 10^4 multiplications and additions are required for each possible location of the patterns. Because there are 10^4 possible locations with 100% overlap, 10^8 operations are required to perform a complete cross-correlation; scale and orientation compensations increase this number further. Parallel processing can reduce the time required to perform this very large number of operations but requires increased complexity and cost.

Permanent Address: Environmental Research Institute of Michigan,
PO Box 8618, Ann Arbor, Michigan 48107.

A recent study (1) has been performed to establish the hardware requirements for using a digital correlation technique. It was concluded in this study that approximately 500 integrated circuits (IC) requiring 350 W of power are required to cross-correlate two 128×128 pixel pictures using 16-bit arithmetic precision. It was also assumed that eight reference maps would be carried in the processor. Six correlations could be performed on the 128×128 reference and 128×128 input scene in 0.5 sec. It was assumed that some parallel processing was used to obtain this speed.

Optical techniques can be used to perform cross-correlation and have the following advantages. An optical processor has an inherently large information capacity. A relatively modest optical system can handle scenes having over 10^7 resolution elements. Such a system handles two-dimensional data in a parallel and isotropic manner with a response time dictated by the time it takes light to travel the length of the processor, plus the time required for data input and output. An increase in the number of required resolution elements does not increase the response time or size of the optical system.

Optical data processing techniques can be divided into two general categories, incoherent and coherent. Incoherent optical processing operates on the intensity of the images to be correlated, that is, it handles only positive functions. Coherent processing makes use of the phase and amplitude of the images and can therefore handle complex functions. A study has been performed that compares the two optical processing techniques (2). This study demonstrated that typically a larger output signal-to-noise ratio and a greater precision can be obtained using coherent rather than incoherent processing.

Optical processing is only one of several analog techniques that can be used to perform correlation. A comparison of these analog techniques with digital processing has been made (3) and the results show that optical processors now out perform digital systems in speed and cost.

In all correlation systems, variations in the input scene when compared to the on-board reference scene can cause a reduction or loss of the correlation signal. The ability of a processor to handle variations in the input scene will determine if a particular correlation technique is successful. The most common scene deviations are scale, rotational orientation, intensity, aspect angle, and overlap. A typical processor can handle errors of $\pm 5\%$ in scale. Larger errors can be handled by using additional reference images or by change in magnification of the input image. Variation in rotational orientation can be reduced by providing attitude control to the missile. A typical

optical processor can handle $\pm 2^\circ$ rotational errors although this is a severe error for a digital system. Other compensation techniques for rotational variations include using additional reference images or rotating the input optically, electronically, or digitally. A change of intensity or shading can be a serious problem for a noncoherent processor or a digital processor that matches scenes in real space by an overlay process or by identification of key features. It is not a problem for those systems that first obtain the Fourier transform of the scene (such as a coherent processor) for they can bandpass filter the spatial frequencies of the scene before correlation. A small change in aspect angle is a distortion of the scene and can be handled by a nonuniform magnification change across the scene area. Large aspect angle changes require that additional reference scenes be stored on board. Overlap of the input scene beyond the boundary of the reference scene causes a reduction in the signal-to-noise ratio of the correlation. This problem can be handled by making the reference scene larger than the input scene.

It should be remembered when considering all of these errors that, because of energy requirements, the missile should not be designed to correct to a predetermined trajectory but should be designed to home on the target from a point on the actual trajectory.

In the design of optical processors very little attention is paid to the number of resolution elements in the input because this parameter does not affect the speed of the correlations. However, in the design of digital systems the reduction of this parameter is a major consideration. Thus, the digital and optical systems which have been designed are not equivalent devices and cannot be directly compared.

A sensor on board a missile will typically provide a low resolution scene for the terminal guidance system. A previous study (4) has demonstrated that the use of low resolution imagery reduces the sensitivity of the system to scale and rotation errors in the input scene while still providing an adequate correlation signal (signal-to-noise ratio greater than 15 dB). Additional advantages are also obtained by the use of low resolution imagery. The size of the optical elements required in the processor is reduced and the coherence requirements on the light source for the coherent optical processor are reduced, allowing laser diodes to be used.

In this paper, the design of a real time coherent optical processor will be described that will operate using realistic, low resolution input imagery. The design incorporates a bank of reference images

to provide the capacity for handling variations in aspect, rotation, and scale. This bank is scanned in time so that it can be determined which reference images are providing correlation signals.

THE EXPERIMENTAL SYSTEM

A coherent optical correlator operating at TV frame rates and utilizing a liquid crystal optical modulator was constructed for these experiments. Figure 1 shows a schematic of the system. Details of its operation are given in a previous report (5).

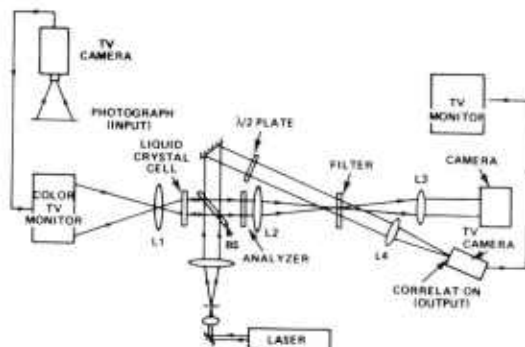


Figure 1. Schematic of coherent optical filter.

CROSS CORRELATION EXPERIMENTS

A series of experiments was performed using aerial photographs to demonstrate real time performance of the optical processor. Three classes of scenes were chosen which are representative of potential military targets. Photographs made in 1962 were used to record matched filters. These photographs were cross correlated with photographs of the same areas made in 1970. No attempt was made to optimize the filter parameters for each input image although a K-ratio was chosen to record spatial frequencies between approximately 0.5 and 4.0 ℓ/mm . This yielded good correlation signal-to-noise ratios for input images of urban and rural areas and structures such as airports and bridges.

Figure 2A is a photograph of a 1962 airport scene on the correlator input monitor. Figure 2B is the output from the liquid crystal modulator. The matched filter made from this input is shown in Figure 2C. The 0-50 scale corresponds to 1 mm. The matched filter was illuminated with the reference beam to reconstruct the image of Figure 2D. The features on which the filter will correlate are readily seen. The bandpass filtering property of the matched filter is apparent from the edge enhancement in the reconstructed image.

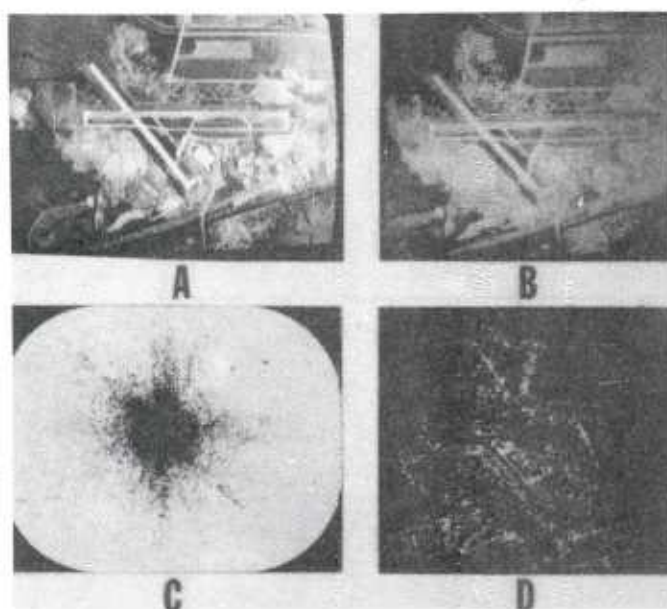


Figure 2. (A) TV image of airport in 1962, (B) modulator output; (C) matched filter made from this scene, and (D) image reconstructed from filter.

A 1970 aerial photograph of the same scene was correlated with the filter made from the 1962 photograph. Cross-correlation and tracking are demonstrated in Figure 3. Figure 3D is a triple exposure showing the correlation spot for each of the three positions of TV input image. The displacement from Figure 3A to Figure 3C corresponds to 680 m on the ground. An oscilloscope trace of a horizontal TV camera scan line through the center correlation spot is shown in Figure 3E. The noise in these traces is electronic noise from the TV camera. The halfwidth at half height for the correlation peak represents approximately 15-m displacement in the scene. Similar results were obtained for other classes of scenes.

EXPERIMENTS WITH RADAR IMAGERY

Tests were performed with imagery used in another terminal guidance program. The purpose of these experiments was to determine if the simulated radar imagery would correlate with actual radar imagery and to obtain performance data on which to base an improved correlator design. The simulated radar imagery is shown in Figure 4A as it appeared on the TV monitor input to the correlator. This photograph shows an area estimated to be 5.2 km in diameter. Figure 4B shows the image formed by the liquid crystal cell. A matched filter (Figure 4C)

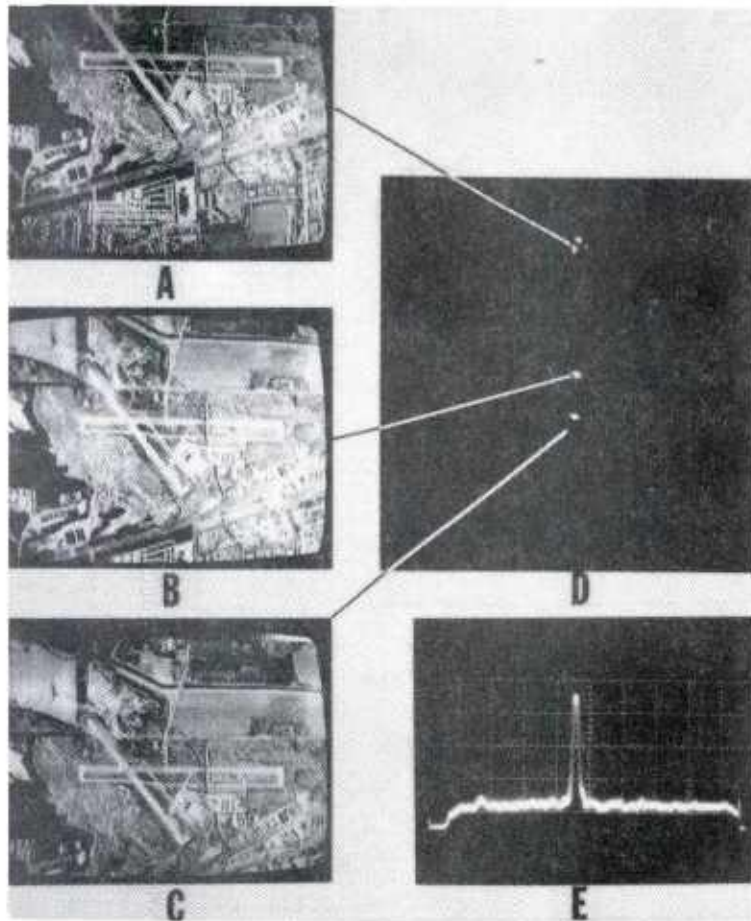


Figure 3. (A, B, C) positions of 1970 airport image corresponding to (D) peaks in correlation plane, (E) oscilloscope display of TV line scan through the central cross-correlation peak.

was constructed for this image with a spatial frequency bandpass of 0.2 to 0.8 l/mm , chosen to match the frequency content of the radar imagery. The image reconstructed from the filter is shown in Figure 4C. Its low quality is due in part to problems of extraneous scattered light and low light level.

An actual radar image that was not grossly distorted relative to the simulated image could not be obtained. A scheduling problem prevented the recording of high quality images. The remaining tests were performed with the original simulated radar image and therefore were auto-correlations.

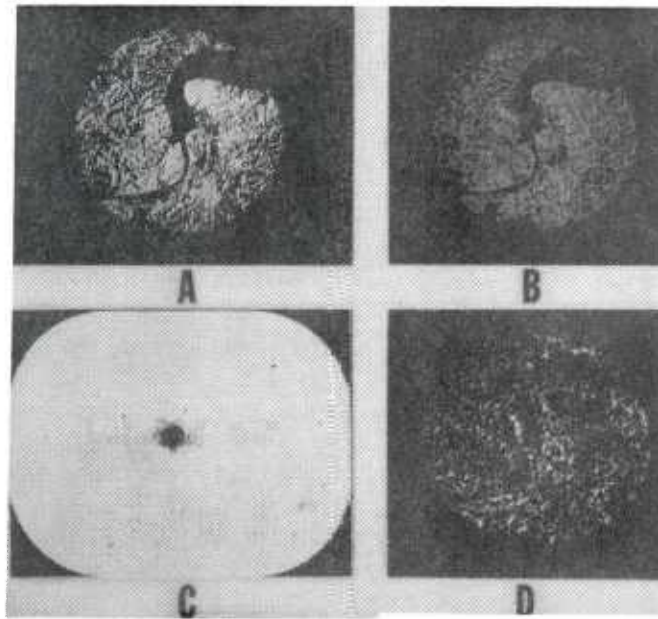


Figure 4. (A) TV display of simulated radar image, (B) modulator output image, (C) matched filter, and (D) image reconstructed from matched filter.

Figure 5A shows the correlation plane and Figure 5B is a trace through the center of the auto-correlation peak as detected by the TV camera at the output of the correlator. The correlation is distinct and well above noise level. The half-power width is $1/60$ of the width of the input image. It should be possible to estimate the center of this peak to at least $1/5$ of its width, thus giving an estimated accuracy of $1/300$ or 0.3% of image width. The correlation peak was also detected and displayed with a photodiode array to confirm the feasibility of using this detector array in a compact breadboard system.

The required filter alignment and input scale and rotational alignment accuracies were determined by adjusting the system until 50% decrease in the height of the correlation peak was observed. The tolerances measured are $\pm 12 \mu\text{m}$ for lateral filter position, $\pm 2^\circ$ for image or filter rotation and $\pm 4.3\%$ for image scale change.

A measurement was made of the light energy incident on the input image and the energy in the correlation peak. Energy input over the 12.5-mm diameter image was 1.5 mW and the energy in the correlation peak was $5 \times 10^{-3} \mu\text{W}$. No attempt was made to improve or optimize the system efficiency.

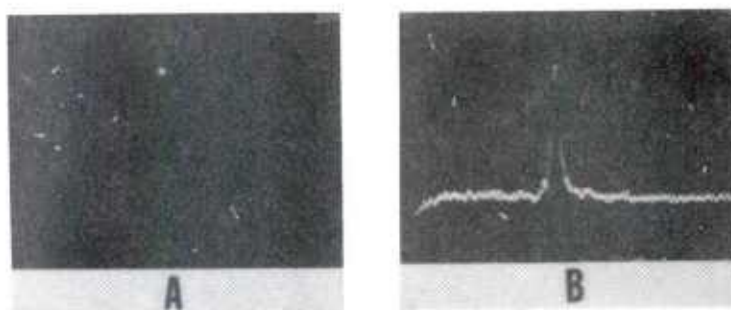


Figure 5. (A) correlation plane and (B) oscilloscope display of TV line scan through the simulated radar image autocorrelation peak.

In a tactical correlator it would probably be necessary to generate the filter on a separate optical system and load it into the correlator. A matched filter was recorded using one liquid crystal modulator and the correlation was performed in the same system using a different modulator. The correlation spot amplitude and signal-to-noise ratio were the same as those measured using one modulator to record the filter and to input the image to be correlated.

PROPOSED CORRELATOR CONFIGURATION AND PERFORMANCE ESTIMATES

A. Image and Filter Format

Correlator configurations are dependent upon the input image size and resolution. It will be assumed that this correlator is to operate on low resolution images with relatively few pixels. As the input data, the image will be assumed to consist of 128×128 or 1.64×10^4 pixels and the reference image from which the matched filter is made to consist of 256×256 or 6.55×10^4 pixels. Having the reference larger than the input insures that the correlation peak amplitude will not vary due to relative lateral displacement of the images.

For a reasonable balance between input image and its Fourier transform size, an input image format of 22 pixels/mm which gives an image size of 6×6 mm and a reference image twice as large will be chosen. If the Fourier transform lens has a 200-mm focal length and laser diodes are used as light sources with $\lambda = 820$ nm, the maximum diameter of the Fourier transform is 3.6 mm for data and 7.2 mm for sampling frequency. For a correlation peak displayed at a distance of 200 mm from the Fourier transform plane, the minimum size of the correlation spot should be approximately 200 μ m. The location of this spot should be within an area 6×6 mm in size if the input image is to overlap the reference

completely. A detector having a resolution of $100\text{ }\mu\text{m}$ should be sufficient and the area covered need not exceed $6 \times 6\text{ mm}$. As a minimum, $100\text{-}\mu\text{m}$ detector resolution is needed to detect correlation peaks while $20\text{-}\mu\text{m}$ resolution would provide a much better estimate of the location of the correlation peak, to approximately $1/5$ the width of the correlation peak.

Using a cathode ray tube (CRT) or an equivalent input device (Figure 1), scale search can be performed by changing the CRT deflection amplifier gain to change the image size. By changing the horizontal gain as a function of vertical position, small aspect angle changes or distortions can be searched.

B. Filter Multiplexing

Multiplexing can be performed by the use of several input image illuminating beams and numerous parallel filters at the Fourier transform plane. Figure 6 shows the basic arrangement. At the left are several light sources which can be turned on either one at a time or simultaneously. These sources might be laser diodes, for example. Light from each source passes through the input image and forms a Fourier transform that is separate from those of adjacent light sources. A different matched filter can be located at each transform location. The correlation from each source can be made to coincide at the output plane or appear at separate locations. If they coincide, then one detector can be used for all filters and the filters would be used in time sequence. If the correlations appear at separate locations at the output plane, then each correlation would have its own detector and the correlations could be performed simultaneously; the latter arrangement is faster but requires multiple detector arrays. All of the light from each source is used to perform correlations with one filter. To keep the complexity to a reasonable level, an array of up to 5×5 light sources for a total of 25 parallel processors could be used.

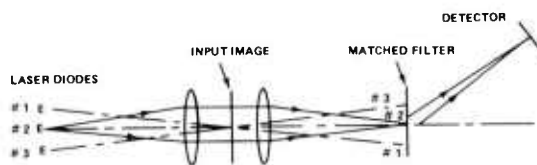


Figure 6. Matched filter multiplexing with multiple light sources.

Another method of multiplexing was described by Vander Lugt (8). Numerous filters are superimposed at the same location in the Fourier transform plane so that correlation peaks from each are located separately at the output. This arrangement requires multiple detectors in the output plane. The number of such superimpositions is limited by space available at the output plane and by the fact that light is equally divided between all correlations and thus decreases as $1/N$, where N is the number of superimposed filters. Vander Lugt demonstrated the use of nine superimposed filters which seems to be a realistic maximum number.

Using both multiplexing techniques simultaneously, a total of 9×25 or 225 different filters could be recorded at the Fourier transform plane. If the scale search for ten different image sizes for each of the 225 filters is included, this correlator could perform a total of 2250 different correlations. These 225 filters might include different images, angular orientations, or aspect angles.

C. Estimate of Correlation Time

An optical correlator takes a Fourier transform and performs correlations almost instantaneously. The readout of data is limited by the rate of scanning the output device and by the light energy used in the processor to charge light detector cells. The time to load the image into the processor is determined both by the scan rate of the sensor or sensor display and by the response time of the light modulator.

Output Detector - As an example, a commercial 100×100 element detector array with elements spaced on $60\text{-}\mu\text{m}$ centers and a 6×6 mm active area will be considered. The usable range of sensitivity extends from a minimum of 0.16 ergs/cm^2 to a saturation exposure of 8 ergs/cm^2 . The maximum scan rate of 10 MHz permits one complete output plane scan in 1 msec. The power consumption for a detector array and its associated electronics is approximately 10 W.

Light Sources - A 10-mW laser diode with output at 820 nm and having 4 nm spectral bandwidth can be used in the correlator. A typical diode has an emitter area of $2 \times 13 \mu\text{m}$ and an overall package diameter of 10 mm. Input power is less than 1 W. Its switching time is less than 1 nsec and therefore can be considered instantaneous. Its wavelength matches the peak response of the detector array. It is estimated that approximately 10% of its output energy will enter the correlator.

Correlator Input Devices - Several types of input devices using liquid crystal or photoconductor-thermoplastic materials could be employed in an optical correlator. Because previously described experiments were performed with the Hughes Aircraft noncoherent-to-coherent image converter and complete specifications are available, the computational estimates used here will be based on this device using a CRT as the source of the image and a lens to image the CRT picture onto the image converter.

Correlation Time - In generation of the input image, it will be assumed that the image data are collected and stored in a digital memory and read onto a CRT which is imaged onto the liquid crystal cell. Thus for a 128×128 point array there are 1.6×10^4 points. These can be scanned at a 1.6-MHz rate so that the image is read onto the CRT in 10 msec. Because the image converter response time is 15 msec and turn-off time is 25 msec, it will be assumed that a usable image exists during a 10-msec period from 20 to 30 msec after the start of the scan, and that an additional 20 msec are needed for a complete image turn-off. During these 10 msec, five sequential sets of correlations, each with nine parallel correlations, can be performed for a total of 45 correlations. Thus, in 50 msec, 45 correlations can be performed at an average rate of 900 correlations/sec.

The data arrangement for the correlator would depend upon factors such as the angular search or scale search required, the storage of multiple targets, and the total operating time for the correlator. Table 1 summarizes the performance capability of the correlator for a few examples of input and filter combinations.

TABLE 1. SUMMARY OF SPECIFICATIONS FOR PROPOSED OPTICAL CORRELATOR

Reference data format (filters)	256 · 256 points
Input image format (real time)	128 · 128 points
Total number of reference filters	225
Maximum number of scale changes	10
Filter format	25 sets of filters in time sequence, nine filters per set at a time
Correlation time	2 msec per set
Image enter and erase time	50 msec (includes 10 msec computation time)
Correlator power consumption	· 2 W
Correlator size	1500 cm ³ (100 in. ³ or 0.05 ft ³)
Correlator weight	5 kg (12 lb)

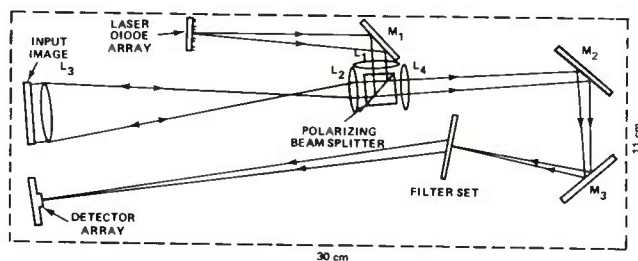
Computational Capability per Half-Second Period for Various Data Arrangements			
New Images	Scale Increments per Image	Filters Accessed	Correlations
1	1	225	225
1	10	45	450
1	10	9	90
2	5	45	450
5	1	90	450

Since the laser diodes require 1 W of power and the liquid crystal input device and detector array a fraction of a watt, the total power consumption should be less than 2 W. This does not include the power requirements of the CRT and computer components expected to be used with the correlator. The charge-couple-device (CCD) addressed liquid crystal modulator under development by Hughes could be used directly in place of the incoherent-to-coherent image converter (9). With the latter arrangement, the computer could be packaged into 1500-cm^3 (0.05-ft^3) volume and would probably weigh less than 5 kg (11 lb).

A BRASSBOARD CORRELATOR

Figure 7 shows a proposed design for a correlator to demonstrate the concepts discussed in the previous section. Current state-of-the-art components are used. While the general arrangement is the same as that discussed in the previous sections, this correlator has four superimposed filters instead of nine. This change was considered necessary to reduce complexity and to increase light level in each correlation peak. The image input device is a Hughes Aircraft liquid crystal TV display having 4 pixel/mm resolution. This low resolution makes the input light modulator approximately 33×33 mm in size and gives a very small Fourier transform. To increase its size, the Fourier transform is magnified by lenses L_2 and L_4 by the ratio of f_4/f_2 , where f is the lens focal length. This provides a Fourier transform focal length of 200 mm. The use of the combination L_2, L_3 allows for the use of a small polarizing beamsplitter and reduces the overall correlator size. The lens combination L_2 and L_3 also reduces the image size to 6×6 mm to the right of lens L_2 . The preceding lens focal lengths were chosen primarily to achieve a convenient scale and do not represent the minimum possible (10).

The matched filter consists of an array of 5×5 filters each occupying 7.2×7.2 mm of space. At each filter location, four different filters are superimposed. The correlation peaks from each fall on a detector array such as a Reticon RA 100×100 which has a spectral sensitivity matching the laser diode output. One-fourth of this detector could be allocated for each filter giving 50×50 elements for each correlation. For the parameters shown in Figure 8, the width of the correlation peak can be expected to be approximately $100\text{ }\mu\text{m}$ while the detector elements are on $60\text{-}\mu\text{m}$ centers. This allows for some improvement in estimating the location of the peak.



LENSES	FOCAL LENGTH (mm)	DIAMETER (mm)
L ₁	100	10
L ₂	25	12.5
L ₃	125	50
L ₄	200	10

Figure 7. Brassboard optical correlator.

The light sources are laser diodes such as RCA Type C30127. Of the 10-mW output, approximately 10% can be utilized and should give sufficient light output at the detector array.

The matched filter array would be constructed on an optical system separate from the correlator. The filters could either be recorded on a high resolution photographic emulsion, on dielectric materials such as dichromated gelatin for higher efficiency, or on thermoplastic photoconductive recording materials. The latter would be most suitable for operational systems because it is nearly real time and the recording is permanent until erased.

SUMMARY AND CONCLUSIONS

Coherent optical correlators are well known to give distinct auto-correlation and cross-correlation peaks between data having precise scale, orientation, and contrast match. These peaks are generally quite narrow and have a low background level because correlations are performed on the high-frequency content of the input image, such as edges and other details. Correlation time is independent of the number of data points on the reference filter and the input image, although in practice the time required to obtain a correlation is determined by the data read-in time and the correlation read-out time.

*CHRISTENSEN, UPATNIEKS, and GUENTHER

Experiments were performed with an existing optical correlator to determine typical correlator operating parameters. Special emphasis was placed on correlation of low resolution images because these would be typical for terminal guidance situations. Distinct correlations were obtained with simulated radar images and data on alignment accuracy of the optical system were obtained. Additional data have been compiled in a separate report (5).

An optical correlator suitable for guidance applications was designed and its correlation time estimated. The reference filter was assumed to contain 256×256 pixels; the input data had 128×128 pixels. It was estimated that 50 msec is required to complete one read-in and erase cycle and that during this time, up to five sequential correlations and up to nine parallel correlations can be performed and read out. The filter library could store up to 225 different reference functions; in addition, scale can be changed by rewriting input data on the input light modulator.

This proposed optical correlator would have very low power demands. The power consumption of the optical correlator itself, including laser diode light sources and photodiode detector arrays, would be less than 2 W. Additional power would also be required for electronic circuitry to handle data input and output and to control the operation of the correlator.

The optical correlator can be packaged into 1500 cm^3 (0.05 ft^3) of space and is expected to weigh less than 5 kg (12 lb). Additional size and weight reductions could be achieved by miniaturization of components and the use of holographic optical components in place of lenses.

A brassboard correlator was designed with presently available components. This correlator can be made small and lightweight, with a capability of storing up to 100 reference filters. A variety of tests could be performed with this correlator to demonstrate the feasibility of compact coherent optical correlators.

REFERENCES

1. Clary, J. B., "All-Digital Correlation for Missile Guidance," Final Report for Battelle Task D. O. No. 0144, 17 December 1976.
2. Guenther, B. D., Christensen, C. R., Greer, M. O., and Rotz, F. B., "Optical Cross-Correlation for Terminal Guidance," Proceedings of the AIAA Guidance and Control Conference, Stanford, California, 14-16 August 1972.
3. Preston, K., Jr., "A Comparison of Analog and Digital Techniques for Pattern Recognition," Proceedings IEEE, Vol. 60, 1972, p. 1216.
4. Shareck, M. W. and Castle, J. G., Jr., Area Correlation by Fourier Transform Holography, Final Report, USA MICOM Contract DAAH01-72-C-0916, University of Alabama in Huntsville, November 1973.
5. Upatnieks, J., Guenther, B. D., and Christensen, C. R., Real-Time Optical Correlation for Missile Terminal Guidance, US Army Missile Research and Development Command, Redstone Arsenal, Alabama, 1978, Report No. H-78-5.
6. Grinberg, J., Jacobson, A. J., Bleha, W., Miller, L., Fraas, L., Boswell, D., and Myer, G., "A New Real-Time Non-Coherent to Coherent Light Image Converter The Hybrid Field Effect Liquid Crystal Light Valve," Opt. Eng., Vol. 14, 1975, p. 217.
7. Gara, A. D., "Real-Time Optical Correlation of 3-D Scenes," Appl. Opt., Vol. 16, 1977, p. 149.
8. Vander Lugt, A. and Leith, E. N., "Techniques in Optical Data Processing and Coherent Optics," Ann N. Y. Acad. Sci., Vol. 157, 1969, p. 99.
9. Grinberg, J., Bleha, W. P., Braatz, P. O., Chow, K., Close, D. H., Jacobson, A. D., Little, M. J., Massetti, N., Murphy, R. J., Nash, J. G., and Waldner, M., "Liquid-Crystal Electro-Optical Modulators for Optical Processing of Two-Dimensional Data," Proceedings SPIE, Vol. 128, 1977, p. 253.
10. Fienup, J. R., Colburn, W. S., Chang, B. J., Leonard, C. D., "Compact Real-Time Matched Filter Optical Processor," Proceedings SPIE, Vol. 188-04, 1977.

PHYSICAL AND MECHANICAL RELATIONSHIPS IN
ELECTRO SLAG (ESR) STEEL

*VITO J. COLANGELO PhD
GARY P. LESSEN, MR.
BENET WEAPONS LABORATORY

Electroslag remelting of gun steel has been proposed as an alternate manufacturing process because of reported improvements in mechanical properties and potential economic benefits. These improvements are a result of the effective control of the many variables inherent in the process.

In order to obtain a uniform microstructure while avoiding piping, porosity and macrosegregation, a unidirectional solidification mode is desired⁽¹⁾. This mode of solidification, together with the relatively short holding times of the molten metal, also affects the distribution of second phase particles, for example, it results in a more uniform carbide distribution because the carbides are not allowed to grow in the transverse or longitudinal directions.

The slag layer plays many roles. The composition of the slag is a vital factor in determining the net composition of the final product. A slag that is too reactive removes desired alloy elements, while one that is not reactive enough results in inadequate cleanliness⁽²⁾.

Another advantage of the slag is its presence between the mold and the solidified metal. This results in a smooth ingot surface requiring minimal finishing and conditioning. However, in order to avoid excess slag along the sides of the mold, the distance between the electrode tip and the molten metal pool should be kept at a constant value; the exact distance depending on the specific alloy being produced.

The amount of non-metallic inclusions can be controlled within certain limits with the slag layer⁽³⁾. The non-metallic matter becomes dissolved in the slag, lowering the amount of inclusions in

the finished product and distributing those that do result more evenly. One of the most deleterious elements removed in this manner is sulfur^{(4),(5)}. Desulfurization of the metal usually results in better mechanical properties and weldability than can be obtained with conventionally cast materials.

Controlling the hydrogen level in the ingot also enhances the weldability, ductility, fracture toughness and fatigue life of the alloy produced⁽⁶⁾. Vacuum degassing of the electrodes prior to melting reduces their hydrogen level thereby reducing the amount of hydrogen in the ingot⁽²⁾.

High ingot yield is another reported advantage of the electroslag remelt process. It has been shown that the depth of the molten metal pool is critical in achieving this. The optimum condition is a shallow pool, which can be maintained by correct control of the power source. Preserving a constant electrode to molten metal pool distance is also an aid in attaining high yields.

Electroslag remelting of other alloys is also possible because of the uniform microstructure that results. In experiments on Cr-Fe alloys, changing the ingot diameter from 60mm to 250mm and altering the Cr content had no effect on the uniformity of the microstructure⁽⁷⁾.

Another procedure that can help in obtaining a uniform microstructure is constant agitation of the molten metal pool to ensure mixing of all the alloying elements. This can be affected by stirring of the pool or rotating the base plate of the apparatus. However, this is only supplemental to maintaining a uni-directional solidification mode in obtaining a uniform microstructure.

In the current study, it was decided that in order to evaluate the cited potential benefits of ESR material, several ingots of ESR melted 4335 + V should be obtained from various manufacturers to the same chemical requirement. It should be noted that at the time of the inception of this program, the primary emphasis of ESR melting was in the production of tool steels and superalloys with little going toward the production of alloy steel billet.

Procedure: ESR material corresponding to a nominal 4335 + V alloy steel was ordered from five (5) producers including one European producer. While it would obviously have been desirable to obtain the complete log of melt practice from the suppliers, such information could not be obtained since it was considered proprietary by the producers. However, some information was obtained and is shown in Table I.

The aim composition, together with the actual compositions, are shown in Table II. The original ingots were 12" diameter which were forged to 7-1/4" billets. In one case, a 10" diameter billet was obtained and forged to 7-1/4" diameter in order to evaluate the

effects of forging reduction.

The ingots were then sectioned and mechanical properties, chemical analysis, microprobe analysis obtained. The mechanical properties were derived from transverse bars heat treated as 1" square blanks, thereby effectively eliminating heat treatment variations as a possible cause of variations. Tensile tests were conducted at 72°F (21°C) while impact tests were conducted at -40°F (-40°C).

The material tested was tempered at varying temperatures. However, 975°F and 1125°F are the only tempering temperatures for which data was obtained for all five heats. Other tempering temperatures of 1000, 1050 and 1100°F were applied to selected heats, as shown in Table III.

The hardenability of the various heats was determined using the method proposed by Hollomon and Jaffee. This method, similar to others, uses hardenability factors for each element (Table IV) which are multiplied by the concentration of each element to determine the maximum diameter of an ideal round which can be hardened.

Microhardness tests were run on a Tukon Hardness Tester employing Knoop hardness. This was felt to be superior to Vicker Hardness Tests because of the greater sensitivity of the Knoop hardness to microstructural changes. One hundred random readings were obtained on polished specimens from each heat tempered at 975°F.

Optical metallography was performed on samples from each heat tempered at 975°F in order to observe the microstructure and the dendrite arm spacing.

A microprobe analysis for Cr, Ni, and Mn was done on samples tempered at 975°F from each manufacturer. The method used in this analysis was to perform a continual scan on the specimen over a specific region. The total number of counts for a fixed time period was determined for each element of interest. This count was then corrected for background level and expressed as total counts as a function of distance in microns. In this way, compositions as a function of position could be determined. The specimens were oriented so that the scan would be perpendicular to the axis of the dendrites.

Discussion and Results:

a. Chemical Analysis - The results and the general chemical analysis are shown in Table II. As can be seen, most of the chemical compositions are within the aim chemistry with a few minor variations. However, there are two factors which should be discussed regarding chemical composition. The first is that even though the chemical compositions are within the aim, a range of compositions does exist and this variation has the capability to affect the mechanical

properties. For this reason, a hardenability calculation was made on the actual compositions derived for the various heats. As can be seen from an examination of the data presented in Tables II and IV, there does exist a range of hardenability with the C material exhibiting the greatest hardenability and the D material the least. The ramifications of this variation will be discussed somewhat more fully in a later section.

A second factor involving chemical analysis which should be discussed is the distribution of the compositional elements. In many respects, it is the distribution of the elements, rather than the nominal composition, which controls mechanical properties. For this reason, a microprobe analysis was also conducted on the material as previously described.

d. Microprobe Analysis - The results of the microprobe study for the 5 heats (3 elements) were also examined. These data illustrate the distribution of a particular element in a linear scan over the surface. It was obvious from our examination of the data that some of the heats were more uniform in concentration and displayed less periodicity than do others. For example, in examining the traces of concentration vs distance, the distinct presence of a periodic distribution was quite evident in Heat C.

In order to determine the variation in uniformity of the microprobe data, a statistical evaluation was undertaken using the following technique.

The numerical value of the concentration (counts) was determined at each 10 μ interval in distance. The data was then analyzed statistically to determine the mean concentration (expresses as counts) and the standard deviation and variance from that mean. The data are shown in Table V. The D heat was considerably more uniform than any of the others whereas the C heat showed the greatest variation in concentration exhibiting the greatest variance in all three elements.

c. Hardenability - The hardenability of each heat, expressed as the diameter of an ideal round (DI) is shown in Table II. The values range from 14.459 inches for Heat C to a low of 13.334 inches for Heat D. It is interesting to note that the low hardenability resulted not from a lack of any single element, but rather from the fact that all elements were on the low side of the compositional range. The significance of the hardenability as a factor in determining the obtained properties will be discussed in a later section, (Mechanical Properties).

d. Microhardness - The microhardness data has been summarized and statistically analyzed. The summary of these data are presented in Table VI. The data shows that the D heat exhibits the

least variation in the microhardness, a fact which is consistent with the results observed in the microprobe testing. This, however, was the only heat where a correspondence was observed. In general, the microhardness variance data did not correlate well with the ranking of the heats on the basis of microprobe tests. However, when one examines the mean microhardness level, one finds that the resultant microhardness level is affected both by the alloy level in the heat (hardenability) and by the microdistribution of these elements. For example, if one projects the expected microhardness for each heat on the basis of the calculated hardenability, it would be expected that the C heat would exhibit the highest mean value, whereas in fact, it is ranked lower. Conversely, it would be expected that the D heat would exhibit the lowest mean microhardness on the basis of its calculated hardenability. In reality, the mean microhardness is higher than expected. The significance of the finding is simply that a higher hardenability can compensate for larger dendrite spacing which might occur in melting due to poor control over melt or cooling rates.

Mechanical Properties: The mechanical properties of the 5 heats after forging and tempering at 975°F are shown in Table III. Examination of the data also shows that while there is not a significant variation in the yield strength and tensile strength, there are large variations in the ductility parameters (Heat E being the poorest) and in impact strength. The data also shows the Heats A and B at the top based upon both their impact strength and ductility. This is consistent with their generally high ranking in hardenability, dendrite spacing and microprobe distribution.

Heat E displayed the poorest overall performance exhibiting not only the lowest ductility, but low impact values as well. Because of the abnormal behavior of this heat relative to the expected ductility for this grade material, it was subjected to gas analysis. This gas analysis revealed that the oxygen content was approximately 198 ppm, thereby accounting, in part, for the low ductility and impact strength.

After tempering at 1125°F, Heats A and B are still ranked at the top of order with respect to impact strength and ductility; however, the yield strength level is somewhat lower than Heat C.

Also, Heats C and D have reversed the ranking from the 975°F temper with Heat D now exhibiting superior impact strength at the 1125°F temper.

Summary and Conclusions: A review of the various properties obtained on the 5 heats (summarized in Table VII) reveals that a wide variation of mechanical properties, more specifically the ductility and impact values, is possible even though all heats are within the specified chemistry range and are melted utilizing acceptable melt practices and procedures.

The variables which appear to affect the properties most profoundly are the hardenability and the dendrite arm spacing which presumably is related to the rate of melting.

The mean microhardness level was related both to the hardenability for a particular heat and the microprobe distribution of the elements, a heat with good hardenability being adversely affected by an uneven distribution of the alloying elements, the latter in turn, being affected by the melt rate for the heat. The significance of this information is that a higher hardenability would be expected to permit a wider variation in the melt rate without adversely affecting the mechanical properties. In summary, all the heats, with the exception of Heat E, were within acceptable limits for the alloy under study. Two of the heats, A and B, showed exceptional properties and undoubtedly represent the potential of the process for making high quality, high strength alloy steel economically.

BIBLIOGRAPHY

1. G. Yuasa, T. Yasima, K. Shina
"Quality Improvement of High Alloy Steels by ESR", Fourth
International Symposium on ESR, Tokyo, Japan, 1973, pp 229-239.
2. W. E. Duckworth and G. Hoyle, "Electroslag Refining", Chapman
and Hall, London, 1969.
3. R. A. Swift and J. A. Gulya.
"Property Evaluation of Electroslag Remelted A533B Plate",
Welding Research Supplement, Dec. 1973, p. 537.
4. R. A. Swift, "Electroslag Remelting Improves Properties of
Plate Steels", Metals Progress, May 1973.
5. M. Kepka, "Electroslag Remelting of Steel", Dept. of the
Army Publication.
6. "Plate Steels Score Breakthrough in Quality" Iron Age, October 12,
1972.
7. K. Suzuki, "Production of Cu-Fe Alloy by Electroslag Remelting
Process", ESR Symposium, Tokyo, Japan, p. 183.

TABLE I
MELTING PARAMETERS

HEAT	A	B	C	D	E
Mold Dia.	12"	10"	12"	12"	12"
Current Type	AC	AC	AC	AC	AC
Flux Type	CaF ₂ -60%	CaF ₂ -60%	NA	CaF ₂ -60%	CaF ₂ -70%
	CaO-12	CaO-12	NA	CaO-10	
	Al ₂ O ₃ -28	Al ₂ O ₃ -28	NA	Al ₂ O ₃ -30	Al ₂ O ₃ -30
Flux wt. (lbs)	55.5	40	NA		52
Mold Material	Copper	Copper	Copper	Copper	Steel
Current, Avg.(amp)	7180	6300	NA	6000	9200
Voltage, Avg(volt)	44	47	NA	52	32
Melt Time (min)	136	110	NA	240	172
Ingot Weight (lb)	1130	960	3200	3415	1250

NOTE: NA - Not Available

TABLE II
CHEMICAL ANALYSIS

<u>ALLOY</u>	<u>C</u>	<u>Mn</u>	<u>P</u>	<u>Ni</u>	<u>S</u>
Aim	.30-.35	.38-.50	.010 max	2.10-2.35	.010 max
A	.32	.52*	.009	2.28	.003
B	.33	.49	.009	2.25	.003
C	.30	.60*	.009	2.30	.003
D	.30	.40	.006	2.08	.004
E	.35	.45	.013	2.31	.010

<u>ALLOY</u>	<u>Si</u>	<u>Cr</u>	<u>Mo</u>	<u>V</u>	<u>**Hardenability Ideal Round Diameter (in)</u>
Aim	.25 max	.88-1.12	.45-.55	.06-.12	
A	.21	1.08	.49	.10	14.035
B	.17	1.05	.48	.10	13.708
C	.13	1.02	.60*	.10	14.449
D	.13	1.14*	.48	.07	13.334
E	.18	1.13*	.52	.07	13.684

*Outside specified range

TABLE III
SUMMARY OF MECHANICAL PROPERTIES
(Tempered at 975°F)

PRODUCER	.1% YS (PSI)	UTS (PSI)	EL (%)	RA (%)	CHARPY (-40°F) (Ft-lbs)
A	172,850	191,250	18.1	55.2	27.4
B	173,500	189,675	16.1	50.8	24.9
C	170,500	192,000	15.7	50.4	14.0
D	169,800	186,800	15.7	49.6	12.0
E	175,025	191,550	9.3	24.6	14.0

(Tempered at 1000°F)

C	169,000	192,300	16.8	54.4	14.2
D	167,600	186,800	14.7	40.4	18.5

(Tempered at 1050°F)

C	164,700	187,300	15.7	49.6	14.7
D	164,600	182,800	16.8	54.2	23.6

(Tempered at 1100°F)

C	159,300	183,200	17.1	55.8	17.5
D	157,700	172,900	14.3	42.0	41.7

(Tempered at 1125°F)

A	151,056	164,045	19.3	60.9	59.3
B	150,752	162,476	18.3	57.9	48.1
E	149,548	157,463	6.1	11.4	23.8
C	157,500	181,200	19.1	57.5	19.5
D	156,100	170,900	17.1	43.5	43.1

TABLE IV - EFFECT OF CARBON AND ALLOYING ELEMENTS
ON HARDENABILITY* (IDEAL ROUND)

Element	Pearlitic Harden- ability Factor	Bainitic Harden- ability Factor
Carbon (50% pearlite or bainite (almost no pearlite or bainite))	$0.338 \times \% C \text{ in. } /$ $0.254 \times \% C \text{ in. } /$	$0.494 \times \% C \text{ in.}$ $0.272 \times \% C \text{ in.}$
Manganese	$1 + 4.10 \times \% Mn$	$1 + 4.10 \times \% Mn$
Phosphorus	$1 + 2.83 \times \% P$	$1 + 2.83 \times \% P$
Sulfur	$1 - 0.62 \times \% S$	$1 - 0.62 \times \% S$
Silicon	$1 + 0.64 \times \% Si$	$1 + 0.64 \times \% Si$
Chromium	$1 + 2.33 \times \% Cr$	$1 + 1.16 \times \% Cr$
Nickel	$1 + 0.52 \times \% Ni$	$1 + 0.52 \times \% Ni$
Molybdenum	$1 + 3.14 \times \% Mo$	1
Copper	$1 + 0.27 \times \% Cu$	$1 + 0.27 \times \% Cu$

*After Hollomon and Jaffe (277)

/ For grain size A.S.T.M. 7

TABLE V
STATISTICAL EVALUATION OF MICROPROBE DATA

Heat No.	A	B	COUNTS C	D	E
Mean					
Ni	917.50	848.33	749.44	974.09	860.67
Mn	867.41	403.89	622.78	417.27	303.33
Mo	400.00	300.69	155.00	115.00	121.33
Standard Deviation					
Ni	47.73	53.50	97.19	45.95	59.55
Mn	29.45	32.29	120.86	36.49	57.90
Mo	20.00	17.34	53.68	9.49	36.86
Variance					
Ni	2278.57	2862.50	9446.52	2114.09	3545.95
Mn	867.41	1042.36	14606.95	1331.82	3352.38
Mo	400.00	300.69	2881.25	90.00	1358.81

TABLE VI
SUMMARY OF MICROHARDNESS DATA

<u>HEAT</u>	<u>MEAN</u>	<u>MEDIAN</u>	<u>STANDARD DEVIATION</u>	<u>MODE</u>	<u>VARIANCE</u>
A	522	504	20	509	400
B	494	465	19	475	361
C	505	504	18	498	324
D	521	480	10	480	100
E	459	434	16	438	256

TABLE VII
RANKING OF HEATS BY VARIOUS ATTRIBUTES
TEMPERED AT 975°F

	1st	2nd	3rd	4th	5th
UTS	C	E	A	B	D
.1% Y.S.	E	B	A	C	D
%E1	A	B	C/D		E
%RA	A	B	C	D	E
Impact	A	B	C/E		D
Dendrite Spacing	D	A	B	C	E
Microhardness Mean	A	C	D	B	E
Standard Deviation	D	E	C	B	A
Hardenability	C	A	B	E	D
Microprobe	D	A	B	E	C

COLLETT

A FOUR-CHANNEL POLARIMETER TO MEASURE NANOSECOND LASER PULSES

EDWARD COLLETT, PhD
US ARMY ELECTRONIC WARFARE LABORATORY
US ARMY ELECTRONICS R&D COMMAND
FORT MONMOUTH, NEW JERSEY 07703

INTRODUCTION

With the advent of nanosecond lasers, new opportunities have arisen as well as problems of scientific and military interest. One problem is the measurement of the polarization state of nanosecond laser pulses. We solved this problem by developing a novel four-channel polarimeter which simultaneously measures the four Stokes polarization parameters of an optical pulse. This design is enhanced by proper selection of the polarization state of the transmitted beam, which makes it possible to determine the ABCD polarization matrix of an optical system, using a single optical pulse.

The polarimeter is automated, has no moving parts, and is controlled by a Hewlett Packard (HP) 9825A desk calculator and an HP-IB Interface Bus System with an HP 6940B multiprogrammer.

Two important factors were critical to the adoption of this polarimeter design. The first was the realization that if an optical system was illuminated with a nanosecond laser pulse, the traditional methods for measuring the polarization state, or equivalently the Stokes polarization parameters, are completely inadequate. In order to measure the Stokes parameters, four distinct settings of a waveplate/polarizer combination are needed. It is obviously impossible to do this in such a short time frame when dealing with nanosecond pulses. Since four distinct settings must be made, a polarimeter must contain four separate channels in which the wave plate/polarizer is preset and permanently fixed. With this design it is then possible to determine the Stokes polarization parameters simultaneously.

We wish to emphasize, however, that this arrangement allows us to measure the Stokes parameters of an optical beam irrespective of any time duration of the signal.

The second important factor in the design of the polarimeter was the fact that most optical systems of military interest contain only polarizing and phase shifting elements, as we shall show in the following sections. This being the case, it is possible to show that the 4x4, sixteen-element, Mueller-Stokes polarization matrix has only four unknown matrix elements, A, B, C, and D. This specialized matrix form is called the ABCD polarization matrix. Further analysis shows that if an optical component or system is illuminated with either circularly polarized or linearly polarized light, each of the detected Stokes parameters is equal to one of the A, B, C, or D matrix elements.

The design and operation of the polarimeter are directly related to the polarizing behavior of optical systems. Consequently, we first devote several sections of the paper to a detailed mathematical discussion of the polarization behavior.

DISCUSSION

Mueller-Stokes Matrices for Optical Components

The basic properties of the Stokes polarization parameters have been described in the classic texts by Chandrasekhar¹ and Born and Wolf.² Further properties, as well as their relation to the Mueller-Stokes polarization matrices, have been treated by Shurcliff and other authors.^{3,4,5}

There are three types of optical components that can change the state of polarization of an optical beam. These components are 1) a polarizer, 2) a compensator (phase shifter), and 3) a rotator. In this section we present the Mueller-Stokes matrix for each of these components and refer the reader to the referenced works for their derivation.

Polarizer

$$M_{\text{pol}}(p_x, p_y) = 1/2 \begin{pmatrix} p_x^2 + p_y^2 & p_x^2 - p_y^2 & 0 & 0 \\ p_x^2 - p_y^2 & p_x^2 + p_y^2 & 0 & 0 \\ 0 & 0 & 2p_x p_y & 0 \\ 0 & 0 & 0 & 2p_x p_y \end{pmatrix}, \quad (1)$$

where $0 \leq p(x,y) \leq 1$. For $p(x,y) = 1$ we have perfect transmission, while $p(x,y) = 0$ corresponds to no transmission (total absorption). For example, $p_y = 0$ for a Nicol prism or Polaroid, so

$$M_{\text{POL}}(p_x, p_y = 0) = p_x^2/2 \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (2)$$

Compensator (Phase Shifter)

$$M_{\text{COMP}}(\phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\phi & -\sin\phi \\ 0 & 0 & \sin\phi & \cos\phi \end{pmatrix}, \quad (3)$$

where ϕ is the total phase shift between the orthogonal axes. Specifically, for a quarter-wave plate, $\phi = \pi/2$, so

$$M_{\text{COMP}}(\phi = \pi/2) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (4)$$

Rotator

$$M_{\text{ROT}}(2\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(2\theta) & \sin(2\theta) & 0 \\ 0 & -\sin(2\theta) & \cos(2\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (5)$$

where θ is the angle of rotation.

These expressions are valid only if the axes of the components are aligned with those of the incident beam. If the optical device is rotated through an angle θ then the transformed matrix is given by

$$M(2\theta) = M_{\text{ROT}}(-2\theta) M M_{\text{ROT}}(2\theta) \quad (6)$$

It is possible to construct an eigenvector which provides a simultaneous diagonal basis for the Mueller-Stokes matrices for the polarizer and the phase shifter, but not the rotator. To show this, the Stokes parameters are defined in terms of the optical field by

$$\begin{aligned} S_0 &= E_x E_x^* + E_y E_y^* = I_{xx} + I_{yy} \\ S_1 &= E_x E_x^* - E_y E_y^* = I_{xx} - I_{yy} \\ S_2 &= E_x E_y^* + E_y E_x^* = I_{xy} + I_{yx} \\ S_3 &= i(E_x E_y^* - E_y E_x^*) = i(I_{xy} - I_{yx}). \end{aligned} \quad (7)$$

The eigenvector that allows Eq 1 and 3 to be written in a diagonalized form is

$$\begin{pmatrix} I_{xx} \\ I_{yy} \\ I_{xy} \\ I_{yx} \end{pmatrix}, \quad (8)$$

and the diagonalized polarization matrices for the polarizer and the phase shifter are, respectively,

$$\tilde{M}_{POL}(p_x, p_y) = \begin{pmatrix} p_x^2 & 0 & 0 & 0 \\ 0 & p_y^2 & 0 & 0 \\ 0 & 0 & p_x p_y & 0 \\ 0 & 0 & 0 & p_x p_y \end{pmatrix} \quad (9)$$

and

$$\tilde{M}_{COMP}(\phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e^{-i\phi} & 0 \\ 0 & 0 & 0 & e^{+i\phi} \end{pmatrix} . \quad (10)$$

where the tilde, \sim , stands for diagonalized matrix representation. The relation between the diagonalized forms, Eq 9 and 10, and the nondiagonalized Mueller-Stokes matrices, Eq 1 and 3, is

$$M = M_D^{-1} M M_D , \quad (11)$$

where

$$M_D = (1/2) \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & i \\ 0 & 0 & 1 & -i \end{pmatrix} \quad (12)$$

and

$$M_D^{-1} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -i & i \end{pmatrix}. \quad (13)$$

The use of a diagonalized matrix representation allows us to write the total system matrix for a system with m polarizing surfaces and n phase-shifting elements. This is found to be

$$M_{POL}(m) = \begin{pmatrix} P_{x_1}^2 & 0 & 0 & 0 \\ 0 & P_{y_1}^2 & 0 & 0 \\ 0 & 0 & P_{x_1} P_{y_1} & 0 \\ 0 & 0 & 0 & P_{x_1} P_{y_1} \end{pmatrix} \cdot \begin{pmatrix} P_{x_m}^2 & 0 & 0 & 0 \\ 0 & P_{y_m}^2 & 0 & 0 \\ 0 & 0 & P_{x_m} P_{y_m} & 0 \\ 0 & 0 & 0 & P_{x_m} P_{y_m} \end{pmatrix}. \quad (14)$$

or

$$\tilde{M}_{POL}(m) = \begin{pmatrix} \prod_{i=1}^m P_{x_i}^2 & 0 & 0 & 0 \\ 0 & \prod_{i=1}^m P_{y_i}^2 & 0 & 0 \\ 0 & 0 & \prod_{i=1}^m P_{x_i} P_{y_i} & 0 \\ 0 & 0 & 0 & \prod_{i=1}^m P_{x_i} P_{y_i} \end{pmatrix}. \quad (15)$$

COLLETT

Similarly,

$$\widetilde{M}_{\text{COMP}}^{(n)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \exp -i \sum_{j=1}^n \phi_j & 0 \\ 0 & 0 & 0 & \exp -i \sum_{j=1}^n \phi_j \end{pmatrix} \quad (16)$$

In terms of the non-diagonalized Mueller-Stokes matrices, we can then write

$$\widetilde{M}_{\text{POL}}^{(m)} = \frac{1}{2} \begin{pmatrix} \prod_{i=1}^m P_{x1}^2 + \prod_{i=1}^m P_{y1}^2 & \prod_{i=1}^m P_{x1}^2 - \prod_{i=1}^m P_{y1}^2 & 0 & 0 \\ \prod_{i=1}^m P_{x1}^2 - \prod_{i=1}^m P_{y1}^2 & \prod_{i=1}^m P_{x1}^2 + \prod_{i=1}^m P_{y1}^2 & 0 & 0 \\ 0 & 0 & 2 \prod_{i=1}^m P_{x1} P_{y1} & 0 \\ 0 & 0 & 0 & 2 \prod_{i=1}^m P_{x1} P_{y1} \end{pmatrix} \quad (17)$$

and

$$M_{\text{COMP}}^{(n)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\left(\sum_{j=1}^n \phi_j\right) & -\sin\left(\sum_{j=1}^n \phi_j\right) \\ 0 & 0 & \sin\left(\sum_{j=1}^n \phi_j\right) & \cos\left(\sum_{j=1}^n \phi_j\right) \end{pmatrix} \quad (18)$$

By direct matrix multiplication, the total system matrix is found to be

$$M = \frac{1}{2} \times$$

$$\begin{pmatrix} \prod_{i=1}^m p_{x_i}^2 + \prod_{i=1}^m p_{y_i}^2 & \prod_{i=1}^m p_{x_i}^2 - \prod_{i=1}^m p_{y_i}^2 & 0 & 0 \\ \prod_{i=1}^m p_{x_i}^2 - \prod_{i=1}^m p_{y_i}^2 & \prod_{i=1}^m p_{x_i}^2 + \prod_{i=1}^m p_{y_i}^2 & 0 & 0 \\ 0 & 0 & 2 \prod_{i=1}^m p_{x_i} p_{y_i} \cos\left(\sum_{j=1}^m \phi_j\right) & -2 \prod_{i=1}^m p_{x_i} p_{y_i} \sin\left(\sum_{j=1}^m \phi_j\right) \\ 0 & 0 & 2 \prod_{i=1}^m p_{x_i} p_{y_i} \sin\left(\sum_{j=1}^m \phi_j\right) & 2 \prod_{i=1}^m p_{x_i} p_{y_i} \cos\left(\sum_{j=1}^m \phi_j\right) \end{pmatrix} \quad (19)$$

The reader should note carefully the form of Eq 19.

The ABCD Polarization Matrix

Equations 1, 3, and 19 are seen to have the form

$$M = \begin{pmatrix} A & B & 0 & 0 \\ B & A & 0 & 0 \\ 0 & 0 & C & -D \\ 0 & 0 & D & C \end{pmatrix} \quad (20)$$

Thus, for the polarizing matrix, Eq 1, we see that A, B, C, and D of Eq 20 are

$$\begin{aligned} A &= (1/2) (p_x^2 + p_y^2) & B &= (1/2) (p_x^2 - p_y^2) \\ C &= 2 p_x p_y & D &= 0, \end{aligned} \quad (21)$$

while for the compensator, Eq 3, we have

$$A = 1, B = 0, C = \cos\phi, \text{ and } D = \sin\phi. \quad (22)$$

COLLETT

Henceforth, Eq 20 will be called the ABCD polarization matrix, or simply the ABCD matrix.

We can show that the following equation holds for the Stokes parameters for a perfectly polarized beam:

$$S_0^2 = S_1^2 + S_2^2 + S_3^2 \quad . \quad (23)$$

Using Eq 20 and 23, we readily find that the following relation exists between the A, B, C, and D parameters:

$$A^2 = B^2 + C^2 + D^2. \quad (24)$$

In practice, this relation is very useful since it obviously allows us to find the fourth matrix element if the other three are known. Moreover, the relation can serve as a useful check on all four measured or calculated elements.

Consider that we now illuminate a system described by the ABCD matrix with linearly $+45^\circ$ polarized light. The Stokes vector for the incident light is $\{1, 0, 1, 0\}$. Matrix multiplication with Eq 20 yields

$$\begin{pmatrix} S'_0 \\ S'_1 \\ S'_2 \\ S'_3 \end{pmatrix} = \begin{pmatrix} A & B & 0 & 0 \\ B & A & 0 & 0 \\ 0 & 0 & C & -D \\ 0 & 0 & D & C \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad (25)$$

so

$$\begin{pmatrix} S'_0 \\ S'_1 \\ S'_2 \\ S'_3 \end{pmatrix} = \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix}. \quad (26)$$

Thus we have the very useful result that if an optical system consists only of polarizers and/or phase shifters and if the optics are illuminated with linearly $+45^\circ$ polarized light, each of the measured Stokes parameters is equal to one of the ABCD matrix elements.

The ABCD matrix elements can also be extracted using right circularly polarized light. In this case, the incident Stokes vector is $\{1, 0, 0, 1\}$, so

$$\begin{pmatrix} S'_0 \\ S'_1 \\ S'_2 \\ S'_3 \end{pmatrix} = \begin{pmatrix} A & B & 0 & 0 \\ B & A & 0 & 0 \\ 0 & 0 & C & -D \\ 0 & 0 & D & C \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad (27)$$

and

$$\begin{pmatrix} S'_0 \\ S'_1 \\ S'_2 \\ S'_3 \end{pmatrix} = \begin{pmatrix} A \\ B \\ -D \\ C \end{pmatrix} \quad (28)$$

We note that D and C are interchanged when Eq 28 is compared with the result for linearly $+45^\circ$ polarized light, Eq 27.

The previous analysis assumes that the axes of the transmitted field and the optical system being illuminated are aligned. In general, this is not true. Nevertheless, the A, B, C, D parameters can still be determined. This can be done by transmitting two sequential orthogonally polarized pulses. The optimum choices are either right and left circularly polarized light or linearly $+45^\circ$ and -45° polarized light. The angle of rotation between the axes can then be determined and the A, B, C, D parameters found. In practice, this operation is always done with the polarimeter.

Measurement of the Stokes Polarization Parameters and the ABCD Matrix Elements

We have seen that each of the Stokes parameters is equal to one of the ABCD matrix elements for polarizing and/or phase-shifting systems. In this section we discuss the measurement of the Stokes parameters, or equivalently, the ABCD matrix elements.

The Stokes parameters of an optical beam can be measured by the appropriate positioning of a wave plate followed by a polarizer, set at an angle θ . This arrangement is shown below in Fig. 1.

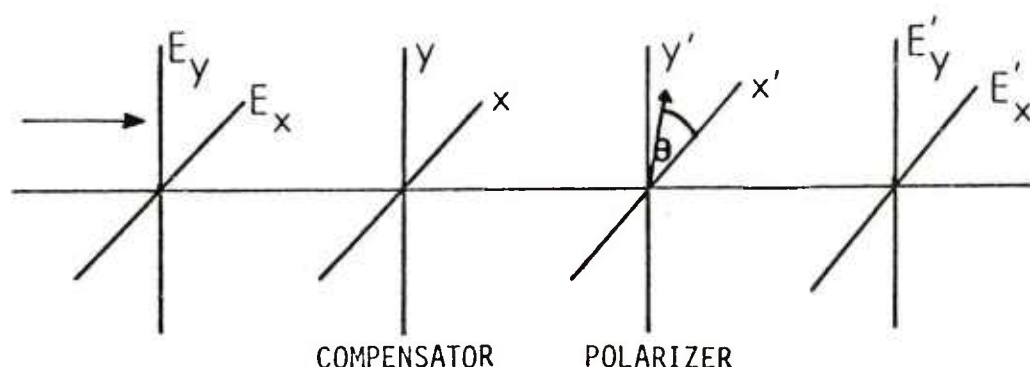


Figure 1. Measurement of the Stokes polarization parameters.

A straightforward analysis shows that the detected intensity, $I(\theta, \phi)$, is

$$I(\theta, \phi) = (1/2) [S_0 + S_1 \cos(2\theta) + S_2 \sin(2\theta) \cos \phi - S_3 \sin(2\theta) \sin \phi], \quad (29)$$

where S_0 , S_1 , S_2 , and S_3 are the detected Stokes parameters, θ is the angle between the horizontal x -axis and the polarizer transmission axis, and ϕ is the phase shift between the orthogonal axes of the wave plate.

The Stokes parameters can be obtained by sequentially setting θ to either 0 , $\pi/2$, or $\pi/4$ and inserting or removing a

quarter-wave plate in the optical train. The respective detected intensities will be

$$I(0, 0) = (1/2)[S_0 + S_1], \quad (30a)$$

$$I(\pi/2, 0) = (1/2)[S_0 - S_1], \quad (30b)$$

$$I(\pi/4, 0) = (1/2)[S_0 + S_2], \text{ and} \quad (30c)$$

$$I(\pi/4, \pi/2) = (1/2)[S_0 - S_3], \quad (30d)$$

from which the Stokes parameters or the equivalent A, B, C, D parameters are found to be

$$S_0 = I(0, 0) + I(\pi/2, 0) \quad (31a)$$

$$S_1 = I(0, 0) - I(\pi/2, 0) \quad (31b)$$

$$S_2 = 2I(\pi/4, 0) - I(0, 0) - I(\pi/2, 0) \quad (31c)$$

$$S_3 = I(0, 0) + I(\pi/2, 0) - I(\pi/4, \pi/2) \quad (31d)$$

In the measurement shown in Fig. 1, the assumption was made that the source of radiation was continuous. This allows time for selective setting of the polarizer/wave plate. We can represent the continuous beam measurement schematically, as shown in Fig. 2, below.

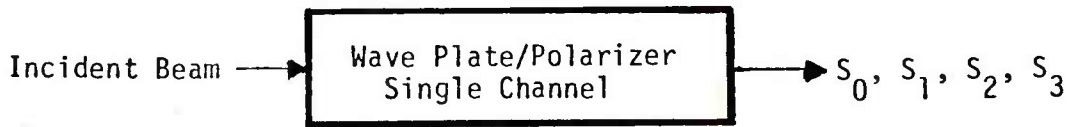


Figure 2. Sequential measurement of the Stokes

Thus an incident beam propagates through a single optical train (referred to as a single channel), and four sequential settings and measurements are then made.

It is clear that the CW method for determining the Stokes parameters for a nanosecond pulse is inappropriate, and, therefore, a different arrangement is required. This can be accomplished by introducing four separate channels, which correspond to each of the equations in Eq 30. This four-channel polarimeter, shown in Fig. 3, permits the simultaneous measurement of the Stokes parameters, A, B, C, and D elements.

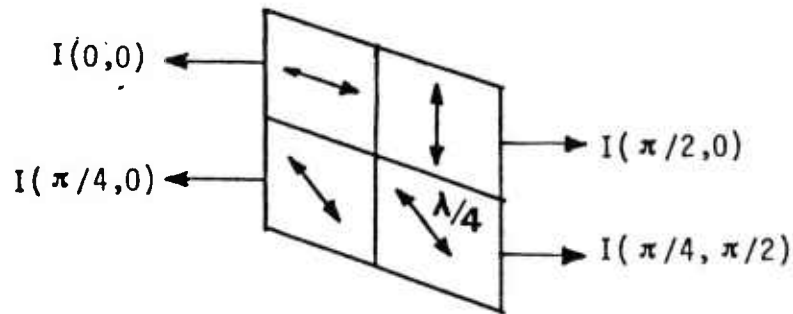


Figure 3. Parallel measurement of the Stokes polarization parameters.

A quarter-wave plate is placed over the fourth channel, $I(\pi/4, \pi/2)$.

The arrangement shown in Fig. 3 can be schematically represented by

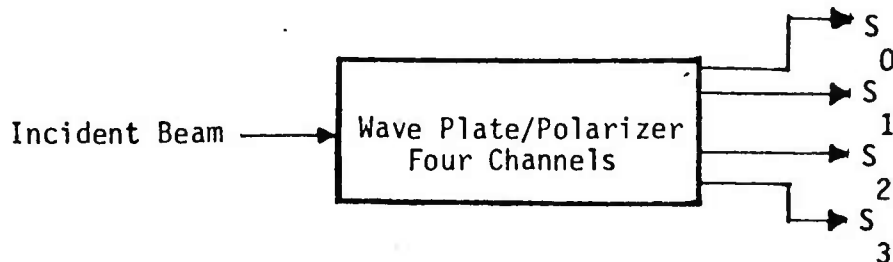


Figure 4. Configuration of the four-channel analyzer.

We emphasize that this arrangement of the polarimeter has no moving parts and, therefore, allows either continuous or pulsed polarized radiation to be measured. Because we actually use the polarimeter to measure the ABCD parameters, the polarimeter is known as either the four-channel polarimeter or the ABCD polarimeter.

In the following and final section we describe the implementation of the polarimeter and its integration into an automatic measurement system. This is done using an HP 9825A desk calculator and a Hewlett-Packard Interface Bus System, in conjunction with the HP 6904B multiprogrammer.

Computer Interfacing in the Four-Channel Polarimeter

During the past decade there has been remarkable progress in the development of minicomputers, desk calculators, and interfacing equipment. In view of this fact and the large amount of data to be collected and analyzed, the Hewlett-Packard 9825A desk calculator was selected to operate the four-channel polarimeter.

The interfacing between the calculator and the polarimeter was accomplished using the Hewlett-Packard Interface Bus System (HP-IB). Briefly, the HP-IB transfers data and commands in parallel between the components of an instrumentation system and the HP 9825A calculator. There are 16 signal lines, of which eight are data input/output (I/O) lines reserved for the transfer of data and other messages in a byte-serial, bit-parallel manner. The remaining eight signal lines are used for data byte transfer control (3) and general interface management (5).

In order to provide even greater flexibility, Hewlett-Packard has developed a general laboratory instrument known as the HP 6940B multiprogrammer. The system is placed in parallel on the HP-IB. This device allows a wide choice of measurements or instrument controls to be made. The HP 6940B multiprogrammer holds up to 15 plug-in cards for various purposes; e.g., control of stepper motors, voltage monitoring, relay control, etc. In our system, an HP 693330A relay card was inserted into the multiprogrammer and used to sequentially open and close relays connecting the detector/amplifiers to a digital voltmeter.

While a voltage monitor card is available for use in the multiprogrammer and permits readings of 150 times per second, it is limited to 12-bit analog to digital conversion. Since resolution is of greater interest than speed, an HP 3490A digital voltmeter, which has a resolution of 1 microvolt on a 100-millivolt scale, was used on the HP-IB. Because of this much greater resolution, the number of readings per second, however, is reduced to 5 per second.

Finally, the HP-IB Interface Bus uses an 8-bit ASCII coded output. This must be converted to the HP 6940B multiprogrammer

16-bit format. In order to make this conversion, an HP 595000A interface unit must be used between the calculator and the multiprogrammer.

We pointed out earlier that the configuration of the four-channel polarimeter can be used with a CW optical source as well as a pulsed optical source. A description of the CW mode will be given first.

To illustrate the basic technique of determining the Stokes parameters, a block diagram of the four-channel polarimeter and the computer/interfacing is shown in Fig. 5. Several versions of the four-channel optical head were built, using Polaroids initially and precisely mounted calcite prisms and quarter-wave plates later. Each Polaroid was cut from the same sheet, and each channel normalized to the first channel. The same was done to the detector/amplifier channels. The alignment of the Polaroids was measured to be within $10'$ of arc, while the calcite prisms polarizers were measured to be within $2'$ of arc.

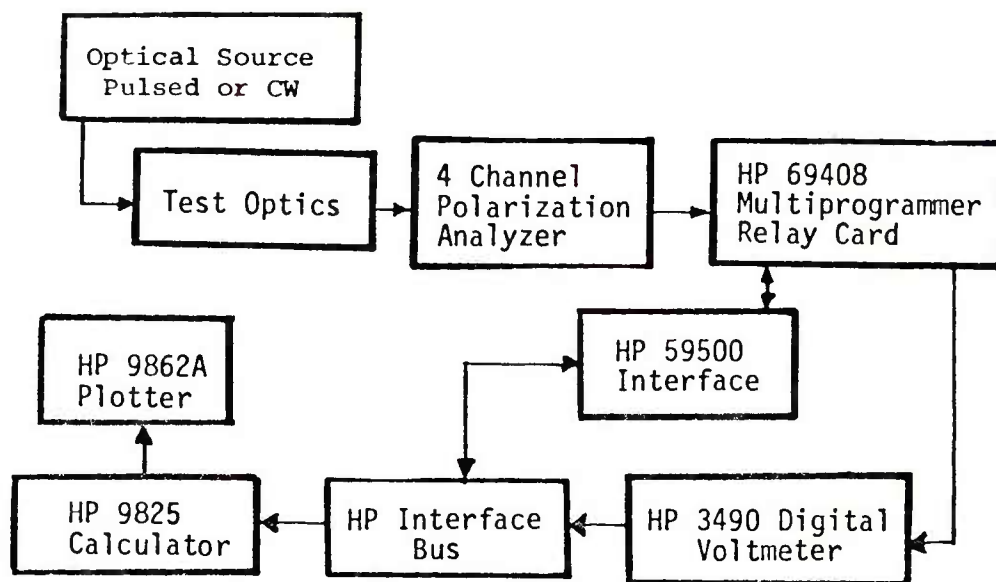


Figure 5. Block diagram of the computer-controlled polarimeter.

The four channels of the optical head were uniformly illuminated, detected, and converted from a current to a voltage signal, using UDT 101A transimpedance amplifiers. By programming the calculator, each relay was closed for N seconds, and a maximum of five readings per second was made on each channel. This was done for

COLLETT

all four channels and the measured data were then stored in the calculator. At the end of the read cycle, the average intensity reading in each channel was determined and converted to the Stokes parameters, or equivalently, to the A, B, C, and D parameters. The results were then printed out on an HP 9871A impact printer. In some instances it was more useful to display the corresponding polarization ellipse, using an HP 9862A plotter. At a future date a CRT will be used to speed up the data presentation.

The system configuration for a pulsed source is similar to that of the CW operation. The difference is the increase in complexity of the detection circuit, Fig. 5, and the addition of an HP 69434A Event Sense Card and an HP 69600B Programmable Timer Card in the multiprogrammer.

The detection circuit converts the nanosecond pulse current to a voltage that is held by the peak detector. The laser is fired once per second, and a trigger is sent to the Event Sense Card. The output of this card initiates the calculator for reading each of the relays. Each relay is programmed to be closed for 100 milliseconds and sequentially read at the beginning of each 200-millisecond interval. All four relays have then been read at the end of 800 milliseconds. During the last 200 milliseconds of the 1-second cycle, a single 1-millisecond pulse is sent to field effect transistors from the HP 6900B Programmable Timer Card. This pulse discharges the capacitors in the detection circuit. Finally, after the laser has fired N times, usually taken at N=20, the data are averaged and the A, B, C, D parameters printed out.

CONCLUSIONS

We have shown that a four-channel polarimeter can be used to determine the polarization characteristics of optical components, systems, and other related phenomena. While the primary purpose of the four-channel configuration is to measure the four Stokes polarization parameters in a nanosecond duration, there is an additional benefit to the design. This is shown by illuminating an optical system with right circularly polarized light or linearly $+45^\circ$ polarized light. Each of the measured Stokes parameters is then equal to one of the unknown A, B, C, D matrix elements. In many situations of military interest this property could be invaluable.

The development of the polarimeter described in this paper evolved from the recognition that all optical beams carry

COLLETT

polarization information. If this information is processed in real time in the manner described, then it may be possible to provide significantly improved performance of military systems.

ACKNOWLEDGEMENT

The author wishes to express his deep appreciation to the US Army Electronic Warfare Laboratory, the Department of the Army and US Air Force for their support in the development of the concepts expressed in this paper.

REFERENCES

1. S. Chandrasekhar, Radiative Transfer, p 28, Oxford University Press, London, 1950.
2. M. Born and E. Wolf, Principles of Optics, Pergamon, New York, 1965, 3rd ed.
3. W. A. Shurcliff, Polarized Light, Harvard U. P., Cambridge, Mass., 1962.
4. E. Collett, "The Description of Polarization in Classical Physics," Amer J Phys 36, 713, 1968.
5. E. Collett, "Mueller-Stokes Matrix Formulation of Fresnel's Equations," Amer J Phys 39, 517, 1971.

ANALYSIS OF A NONLINEAR ELECTROMAGNETIC
FIELD PENETRATION PROBLEMWILLIAM J. CROISANT* and PAUL NIELSEN
U.S. ARMY CONSTRUCTION ENGINEERING RESEARCH LABORATORY
CHAMPAIGN, ILLINOIS 61820

INTRODUCTION

Typical electromagnetic shielding problems involve the calculation of the time-dependent electric or magnetic field penetration of electrically conducting media. Two material properties which enter into such calculations are the electrical conductivity σ and the permeability μ . In most cases the electrical conductivity does not significantly vary with electromagnetic field level and can be regarded as a constant for a given material. In nonmagnetic materials the permeability is also independent of field level and has the value for free space (vacuum) $\mu_0 = 4\pi \times 10^{-7} \text{H/m}$. For media whose material properties are independent of the applied electromagnetic field level, the partial differential equations which govern electromagnetic field propagation are linear. The analytical theory for electromagnetic field calculations in media with constant permeability has been well developed and extensive analyses exist. In fact, quite often the exact analytical solution can be achieved by any one of a number of standard linear mathematical techniques.

On the other hand, ferromagnetic materials, which are commonly used in electromagnetic shield fabrication, not only have a permeability that is dependent on the magnetic field strength H but also exhibit the phenomenon of saturation. Although ferromagnetic materials usually have relatively large (in some cases relatively enormous) permeabilities compared to nonmagnetic materials, the permeability can be greatly reduced if the material is driven into saturation. For these materials the partial differential equations which govern electromagnetic field propagation are nonlinear and

considerably less amenable to analytic solution. Due to the difficulties associated with the nonlinear partial differential equations which arise in this case, relatively few analyses are available so that it is difficult to assess the effect of a variable permeability on the time-dependent electromagnetic field penetration.

This study examines the time-dependent penetration of a step increase in magnetic field strength H into a semi-infinite conducting medium having a permeability which varies with magnetic field strength. The medium is considered to be isotropic, homogeneous, and initially demagnetized. The medium at any point is presumed to follow its initial magnetization curve for which a simple approximation is assumed. The analytical approach consists of mathematical analysis supplemented with numerical calculations. The problem can be simplified from one involving a second order nonlinear partial differential equation to one involving a second order nonlinear ordinary differential equation utilizing a simple transformation of variables. Although a formal parametric representation of the solution of the second order differential equation is considered, a simple closed form solution in terms of elementary functions does not seem to exist. Hence several analytical techniques, as well as numerical calculations, have been employed to deduce properties of the solution.

MATHEMATICAL FORMULATION OF THE PROBLEM

The propagation of electromagnetic fields in a conducting medium (where the displacement current can be neglected) is governed by the equations

$$\nabla \times \vec{E} = - \frac{\partial \vec{B}}{\partial t} \quad (1)$$

and

$$\nabla \times \vec{H} = \sigma \vec{E} \quad (2)$$

where \vec{E} is the electric field, \vec{H} is the magnetic field strength, and \vec{B} is the magnetic flux density. In a homogeneous and isotropic medium, it can be assumed that $\vec{B} = B(\vec{H})$, i.e., that the magnetic flux density is a function of the applied magnetic field strength and is in the same direction. With the differential permeability defined as $\mu_d(H) = \partial B / \partial H$, (1) can be written as:

$$\nabla \times \vec{E} = - \mu_d(H) \frac{\partial \vec{H}}{\partial t} \quad (3)$$

Consider a homogeneous, isotropic, conducting medium of semi-infinite extent ($x \geq 0$). In the present problem it is presumed that the medium is initially demagnetized, and at time $t = 0$, a step increase in magnetic field strength in the \hat{y} direction occurs at the

surface $x = 0$, e.g., as the result of an imposed surface current in the \hat{z} direction.

Since it is presumed that $\vec{H} = H_y(x,t)\hat{y}$, the propagation of the electromagnetic fields following the onset of the step increase in $H_y(x,t)$ is governed by the equations

$$\frac{\partial E_z(x,t)}{\partial x} = \mu_d(H_y) \frac{\partial H_y(x,t)}{\partial t} \quad , \quad (4)$$

$$\frac{\partial H_y(x,t)}{\partial x} = \sigma E_z(x,t) \quad . \quad (5)$$

By eliminating $E_z(x,t)$ from (4) and (5), it is found that the partial differential equation governing the penetration of $H_y(x,t)$ is

$$\frac{\partial^2 H_y(x,t)}{\partial x^2} = \sigma \mu_d(H) \frac{\partial H_y(x,t)}{\partial t} \quad . \quad (6)$$

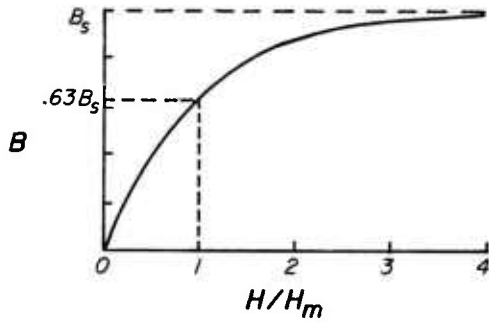
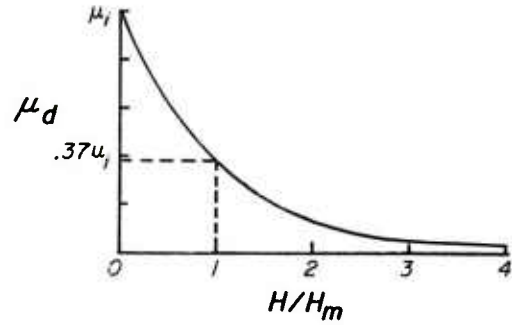
In general, $B(H) = \mu_0[H + M(H)]$, where $M(H)$ is the magnetization. As H is increased from zero, $\mu_d(H)$ usually starts from an initial value μ_i (μ_i is usually larger than μ_0), increases to some maximum value, and then decreases to μ_0 as the material undergoes saturation. In many cases before saturation, the variation of B with H [hence the variation of $\mu_d(H)$] is primarily due to $M(H)$. In this study it is assumed that over the range of interest $B(H)$ can be represented by the simple approximation

$$B(H) = B_s[1 - \exp(-H/H_m)] ,$$

where B_s represents a saturation value for B and H_m is a parameter which indicates the shape of the magnetization curve for the medium. With this representation of the magnetization curve,

$$\mu_d(H) = \mu_i \exp(-H/H_m) \quad ,$$

where $\mu_i \equiv B_s/H_m$ is the initial slope of the B - H curve. The variation of $B(H)$ with H is shown in Figure 1a and the corresponding variation of $\mu_d(H)$ with H is shown in Figure 1b. For large H , $\mu_d(H)$ actually approaches μ_0 as a lower limit; hence this simple representation is appropriate as long as $\mu_i \exp(-H/H_m) > \mu_0$ over the range of H under consideration.

FIGURE 1a $B(H)$ vs. H FIGURE 1b $\mu_d(H)$ vs. H

The present problem, therefore, requires the solution of the nonlinear partial differential equation

$$\frac{\partial^2 H_y(x,t)}{\partial x^2} = \sigma \mu_i \exp(-H/H_m) \frac{\partial H_y(x,t)}{\partial t} \quad (7)$$

subject to the initial condition

$$H_y(x,0) = 0 \quad (8)$$

and the auxiliary conditions

$$H_y(0,t) = H_o \quad (9)$$

and

$$H_y(\infty,t) = 0 \quad (10)$$

The electric field can be determined from the solution to the above problem by noting from (5) that

$$E_z(x,t) = \frac{1}{\sigma} \frac{\partial H_y(x,t)}{\partial x} \quad (11)$$

REDUCTION IN THE NUMBER OF INDEPENDENT VARIABLES

Using an approach similar to that used for nonlinear diffusion phenomena analysis, the number of independent variables appearing in the problem can be reduced by the transformation

$$H_y(x,t) = H(\zeta); \quad \zeta = \frac{\sqrt{\sigma\mu_i}}{2} \frac{x}{\sqrt{t}} .$$

The new variables are introduced by evaluating the partial derivatives in terms of the new variables and substituting these quantities into the original partial differential equation. On calculating the partial derivatives it is found that

$$\frac{\partial H_y(x,t)}{\partial x} = \frac{\partial \zeta}{\partial x} \frac{dH(\zeta)}{d\zeta} = \frac{\sqrt{\sigma\mu_i}}{2\sqrt{t}} \frac{dH(\zeta)}{d\zeta} ,$$

$$\frac{\partial^2 H_y(x,t)}{\partial x^2} = \frac{\partial^2 \zeta}{\partial x^2} \frac{dH(\zeta)}{d\zeta} + \left(\frac{\partial \zeta}{\partial x} \right)^2 \frac{d^2 H(\zeta)}{d\zeta^2} = \frac{\sigma\mu_i}{4t} \frac{d^2 H(\zeta)}{d\zeta^2} ,$$

and

$$\frac{\partial H_y(x,t)}{\partial t} = \frac{\partial \zeta}{\partial t} \frac{dH}{d\zeta} = -\frac{1}{2} \frac{\sqrt{\sigma\mu_i}}{2} \frac{x}{t^{3/2}} \frac{dH(\zeta)}{d\zeta} = -\frac{1}{2} \frac{\zeta}{t} \frac{dH(\zeta)}{d\zeta} .$$

Upon substituting these quantities into (7) and cancelling a factor of $\sigma\mu_i/4t$, it is found that the magnetic field strength satisfies the equation

$$\frac{d^2 H(\zeta)}{d\zeta^2} = -2\zeta \exp(-H/H_m) \frac{dH(\zeta)}{d\zeta} , \quad (12)$$

in which the variables x and t no longer appear explicitly. The electric field is related to $H(\zeta)$ by

$$E_z(x,t) = \frac{1}{2} \sqrt{\frac{\mu_i}{\sigma}} \frac{1}{\sqrt{t}} \frac{dH(\zeta)}{d\zeta} \quad (13)$$

in which a factor of \sqrt{t} appears. Noting that both $x = 0$ and $t = \infty$ imply that $\zeta = 0$ and that both $x = \infty$ and $t = 0$ imply that $\zeta = \infty$, it follows from (9) that

$$H(0) = H_0 \quad (14)$$

and from (8) and (10) that

$$H_y(\infty) = 0 . \quad (15)$$

For computational purposes, it is convenient to normalize the problem in terms of H_0 . With

$$F(\zeta) = H(\zeta)/H_0 = H_y(x,t)/H_0, \quad (16)$$

the problem has been reduced to solving the nonlinear second order ordinary differential equation

$$\frac{d^2 F}{d\zeta^2} = -2\zeta \exp(-\alpha F) \frac{dF}{d\zeta} \quad (17)$$

subject to the conditions

$$F(0) = 1 \quad (18)$$

and

$$F(\infty) = 0, \quad (19)$$

where $\alpha = H_0/H_m$. Having solved for $F(\zeta)$, the electric field can then be evaluated from the relation

$$\frac{dF(\zeta)}{d\zeta} = \frac{2}{H_0} \sqrt{\frac{\sigma}{\mu_i}} \sqrt{t} E_z(x,t). \quad (20)$$

ANALYTICAL ASPECTS OF THE ORDINARY DIFFERENTIAL EQUATION

Classical Linear Solution

In classical analyses of electromagnetic field penetration, it is assumed that the permeability is constant, which corresponds to $\alpha = 0$ in the present investigation. For $\alpha = 0$ the problem reduces to

$$\frac{d^2 F}{d\zeta^2} = -2\zeta \frac{dF}{d\zeta} \quad (21)$$

subject to conditions (18) and (19). This equation can be integrated in a straightforward manner, to obtain the classic solution

$$F(\zeta) = 1 - \text{erf}(\zeta), \quad (22)$$

where the error function is defined as

$$\text{erf}(\zeta) = \frac{2}{\sqrt{\pi}} \int_0^\zeta \exp(-\zeta_1^2) d\zeta_1.$$

It follows immediately that

$$\frac{dF(\zeta)}{d\zeta} = -\frac{2}{\sqrt{\pi}} \exp(-\zeta^2). \quad (23)$$

Formal Solution of the Nonlinear Problem

Most of the standard linear techniques are of relatively little use in achieving an exact solution in nonlinear cases ($\alpha \neq 0$); however, in the course of an investigation of a related problem in nonlinear diffusion phenomena, Fujita [1] found a formal solution which can be directly adapted to the problem presently under investigation. The problem is satisfied with F and ζ being given by the parametric representations

$$F = \frac{2}{\alpha} \int_0^{\xi} \left[\xi_1^2 - \beta \ln(\xi_1^2) \right]^{-1/2} d\xi_1, \quad (24)$$

$$\zeta = \frac{1}{(2\beta)^{1/2}} \left\{ \left[\xi^2 - \beta \ln(\xi^2) \right]^{1/2} - \xi \right\} \times \\ \exp \left\{ \int_0^{\xi} \left[\xi_1^2 - \beta \ln(\xi_1^2) \right]^{-1/2} d\xi_1 \right\}, \quad (25)$$

where the derived quantity β is related to the given quantity α by

$$\alpha = 2 \int_0^1 \left[\xi_1^2 - \beta \ln(\xi_1^2) \right]^{-1/2} d\xi_1. \quad (26)$$

The parametric representation of $dF/d\zeta$ can be evaluated from

$$\frac{dF}{d\zeta} = \left(\frac{dF}{d\xi} \right) / \left(\frac{d\zeta}{d\xi} \right) \\ = - \frac{4}{\alpha \beta^{1/2}} \xi \exp \left\{ - \int_0^{\xi} \left[\xi_1^2 - \beta \ln(\xi_1^2) \right]^{-1/2} d\xi_1 \right\}. \quad (27)$$

The parameter ξ ($0 \leq \xi \leq 1$) relates a value of F to the corresponding value of ζ . (Note that $\xi = 0$ implies $\zeta = \infty$ and $F = 0$, while $\xi = 1$ implies $\zeta = 0$ and $F = 1$.) To evaluate F versus ζ , a value of ξ is selected and the value of F and the corresponding value of ζ are evaluated. The integrals appearing in the formal solution do not appear to be evaluated in closed form, thus numerical integration seems to be required; however, the above representation offers the advantage of allowing the computational effort to be directed to the

very accurate calculation of the solution at only a few points, or one point for that matter, since the knowledge of intermediate points is not required. Of special interest is $dF/d\zeta$ evaluated at $\zeta = 0$ where

$$\frac{dF(0)}{d\zeta} = - \frac{4}{\alpha\beta^{1/2}} \exp(-\alpha/2) . \quad (28)$$

In addition to its relation to the electric field at $x = 0$, the calculated value of $dF(0)/d\zeta$ can be used to transform the problem from a two-point problem (with a condition at $x = \infty$ as well as one at $x = 0$) to an initial value problem (with both boundary conditions specified at $x = 0$). Such a transformation is often useful from a computational standpoint.

Taylor Series Expansion of the Solution

If $F(\zeta)$ was a known function of ζ and if $F(\zeta)$ and its derivatives were continuous, then $F(\zeta)$ could be expanded in a Taylor (Maclaurin) series about $\zeta = 0$

$$F(\zeta) = \sum_{m=0}^{\infty} \frac{d^m F(0)}{d\zeta^m} \frac{\zeta^m}{m!} = F(0) + \frac{dF(0)}{d\zeta} \zeta + \frac{d^2 F(0)}{d\zeta^2} \frac{\zeta^2}{2!} + \dots ,$$

where F and its derivatives are evaluated $\zeta = 0$. For example, the solution (22) to the linear case ($\alpha = 0$) has the series representation

$$F(\zeta) = 1 - \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n \zeta^{2n+1}}{(2n+1)n!} = 1 - \frac{2}{\sqrt{\pi}} \left(\zeta - \frac{\zeta^3}{3} + \frac{\zeta^5}{5 \cdot 2!} - \dots \right) .$$

A series expansion approach can also be used to develop the solution $F(\zeta)$ from the second order differential equation. If $F(0)$ and $dF(0)/d\zeta$ are given as boundary conditions, then the approach is straight-forward since the second and higher order derivatives can, in principle, be determined from successive differentiation of the second order ordinary differential equation. Although $F(0)$ is one of the auxiliary conditions in the problem presently under investigation, it is evident that a difficulty arises because the second condition $F(\infty) = 0$ is specified at $\zeta = \infty$. For this reason we are led to consider the associated initial value problem

$$\frac{d^2 F}{d\zeta^2} = - 2\zeta \exp(-\alpha F) \frac{dF}{d\zeta} \quad (29)$$

subject to

$$F(0) = 1 \quad (30)$$

and

$$\frac{dF(0)}{d\zeta} = \gamma , \quad (31)$$

where γ is to be chosen such that $F(\infty) = 0$. This value for γ can be computed from Eq. (27).

The second derivative can be determined directly from (29) from which it follows that

$$\frac{d^2F(0)}{d\zeta^2} = 0 .$$

The third derivative can be found by differentiation of Eq. (29)

$$\frac{d^3F}{d\zeta^3} = -2 \left\{ \exp(-\alpha F) \frac{dF}{d\zeta} + \zeta \frac{d}{d\zeta} \left[\exp(-\alpha F) \frac{dF}{d\zeta} \right] \right\} , \quad (32)$$

from which it follows that

$$\frac{d^3F(0)}{d\zeta^3} = -2 \gamma \exp(-\alpha) .$$

Differentiating (32) and evaluating the result at $\zeta = 0$ yields

$$\frac{d^4F(0)}{d\zeta^4} = -2\alpha\gamma^2 \exp(-\alpha) .$$

Similarly, it is found that

$$\frac{d^5F(0)}{d\zeta^5} = -6\alpha^2\gamma^3 \exp(-\alpha) + 12\gamma \exp(-2\alpha) .$$

The approach could, in principle, be extended indefinitely; however, in practice it becomes rather tedious after the first few terms.

To fifth order terms the solution $F(\zeta)$ is given by

$$\begin{aligned} F(\zeta) = & 1 + \gamma\zeta - \frac{\gamma \exp(-\alpha)}{3} \zeta^3 + \frac{\alpha\gamma^2 \exp(-\alpha)}{6} \zeta^4 \\ & + \frac{2\gamma \exp(-2\alpha) - \alpha^2\gamma^3 \exp(-\alpha)}{20} \zeta^5 + \dots . \end{aligned} \quad (33)$$

The series representation (33) can be used to calculate $F(\zeta)$ for small values of ζ .

In the study of pulse penetration it is frequently of interest to know the time that it takes for the fields at a certain point to rise to a given value. For this purpose a convenient manipulation is the reversion of series. Given a series for the dependent variable y in terms of the independent variable x ,

$$y = ax + bx^2 + cx^3 + dx^4 + ex^5 + \dots$$

Using a reversion of series one can write x as a function of y :

$$x = Ay + By^2 + Cy^3 + Dy^4 + Ey^5 + \dots,$$

where

$$A = \frac{1}{a}; B = -\frac{b}{a^3}; C = \frac{2b^2 - ac}{a^5}; D = \frac{5abc - a^2d - 5b^3}{a^7};$$

$$E = \frac{6a^2bd + 3a^2c^2 + 14b^4 - 2lab^2c - a^3e}{a^9}; \dots$$

For example, the classic linear solution (22) can be rewritten as

$$(1-F) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)n!} \zeta^{2n+1} = \frac{2}{\sqrt{\pi}} \left(\zeta - \frac{\zeta^3}{3} + \frac{\zeta^5}{10} - \dots \right).$$

A reversion of series yields

$$\zeta = \frac{\sqrt{\pi}}{2} \left[(1-F) + \frac{\pi}{12} (1-F)^3 + \frac{7\pi^2}{480} (1-F)^5 + \dots \right]$$

which is suitable for calculating the value of ζ at which the solution $F(\zeta)$ has a specified value for small values of $(1-F)$ (which implies small ζ).

Equation (33) can be rewritten as

$$(1-F) = -\gamma\zeta + \frac{\gamma \exp(-\alpha)}{3} \zeta^3 - \frac{\alpha\gamma^2}{6} \exp(-\alpha)\zeta^4 \\ - \frac{2\gamma \exp(-2\alpha) - \alpha^2\gamma^3 \exp(-\alpha)}{20} \zeta^5 + \dots$$

A reversion of series yields

$$\zeta = \frac{1}{\gamma} \left[(1 - F) + \frac{\exp(-\alpha)}{3\gamma^2} (1 - F)^3 + \frac{\alpha \exp(-\alpha)}{6\gamma^2} (1 - F)^4 \right. \\ \left. + \left(\frac{1}{20} \frac{\alpha^2 \exp(-\alpha)}{\gamma^2} - \frac{7}{30} \frac{\exp(-2\alpha)}{\gamma^4} \right) (1 - F)^5 + \dots \right]$$

which can be used [for small values of $(1 - F)$] to examine the variation with α of the value of ζ at which $F(\zeta)$ has a specified value.

NUMERICAL CALCULATIONS

For a specified value of β the corresponding value of α can be computed by numerical integration of (26) and the value of $dF(0)/d\zeta = \gamma$ can be calculated from (28). Some numerical calculations were performed and the results are shown in Table 1. An examination of Table 1 reveals that rather large changes in β correspond to relatively small changes in α . This situation can make it difficult to determine the value of β which corresponds to a specified value of α . On the other hand, it is evident that $dF(0)/d\zeta = \gamma$ exhibits a relatively moderate variation with α . While α varies in the range $0 \leq \alpha \leq \infty$, γ varies in the range $-2/\sqrt{\pi} \leq \gamma \leq 0$. This moderate variation makes it possible to accurately represent γ as a function of α using the regression equation

$$\frac{dF(0)}{d\zeta} = \gamma = -1.1284 + 2.0490 \times 10^{-1} \alpha - 3.4426 \times 10^{-2} \alpha^2 \\ + 4.4139 \times 10^{-3} \alpha^3 - 4.1562 \times 10^{-4} \alpha^4 \\ + 2.7025 \times 10^{-5} \alpha^5 - 1.0736 \times 10^{-6} \alpha^6 \\ + 0.9395 \times 10^{-7} \alpha^7 - \dots,$$

which provides a convenient means of evaluating γ for $0 \leq \alpha < 10$.

The variations of F and $dF/d\zeta$ with ζ were calculated for $\alpha = 0$ ($\beta = \infty$), $\alpha = 2$ ($\beta = 3.959 \times 10^{-1}$), and $\alpha = 5$ ($\beta = 5.922 \times 10^{-3}$) and are shown in Figures 2a and 2b. The magnetic field strength H can be determined from (16) and the electric field from (20). In shielding applications, the time variation is often of interest; therefore the variations of F and $dF/d\zeta$ with $1/\zeta^2$ are shown in Figures 3a and 3b.

Table 1
Numerical Calculations of α and $dF(0)/d\zeta$.*

β	α	$\gamma = \frac{dF(0)}{d\zeta}$
∞	0	$-2/\sqrt{\pi}$
5×10^4	0.0112	-1.1260
1×10^4	0.0249	-1.1233
5×10^3	0.0351	-1.1212
1×10^3	0.0773	-1.1127
5×10^2	0.1083	-1.1066
1×10^2	0.2326	-1.0825
5×10^1	0.3197	-1.0662
1×10^1	0.6424	-1.0098
5×10^0	0.8475	-0.9769
1×10^0	1.5058	-0.8847
5×10^{-1}	1.8681	-0.8414
1×10^{-1}	2.8683	-0.7432
5×10^{-2}	3.3552	-0.7043
1×10^{-2}	4.5797	-0.6255
5×10^{-3}	5.1378	-0.5965
1×10^{-3}	6.4838	-0.5392
5×10^{-4}	7.0801	-0.5183
1×10^{-4}	8.4931	-0.4767
5×10^{-5}	9.1115	-0.4612
1×10^{-5}	10.5651	-0.4300

*The calculations were performed on a CDC 6600 computer using simple trapezoidal numerical integration.

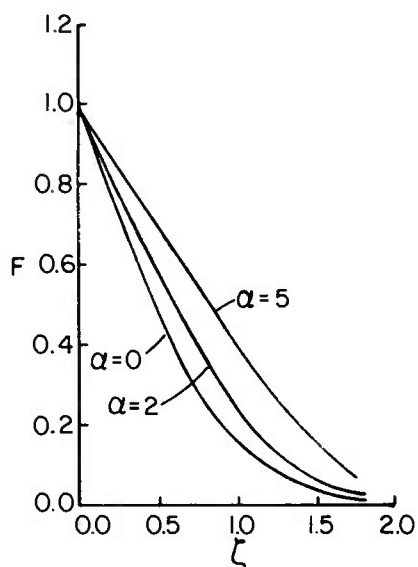


FIGURE 2a. $F(\zeta)$ VERSUS ζ

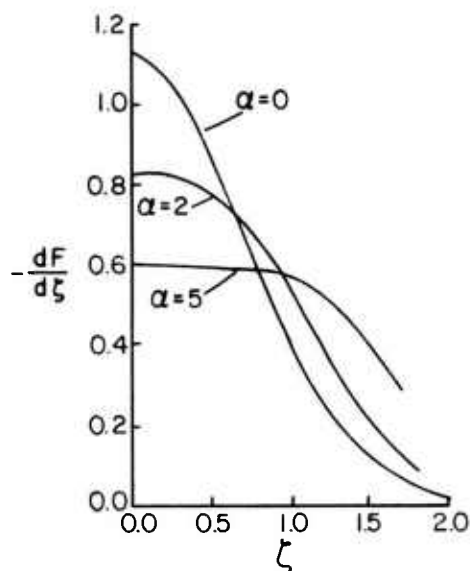


FIGURE 2b. $dF/d\zeta$ VERSUS ζ

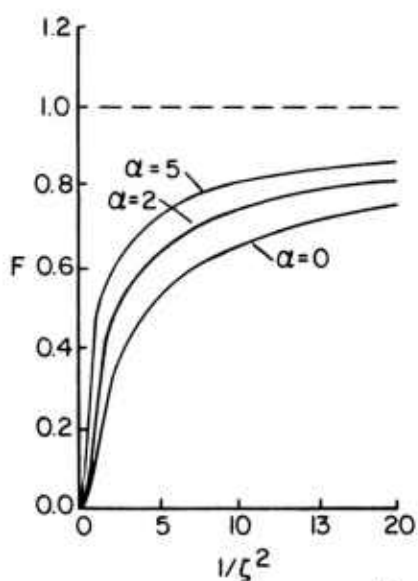


FIGURE 3a. $F(\zeta)$ VERSUS $1/\zeta^2$

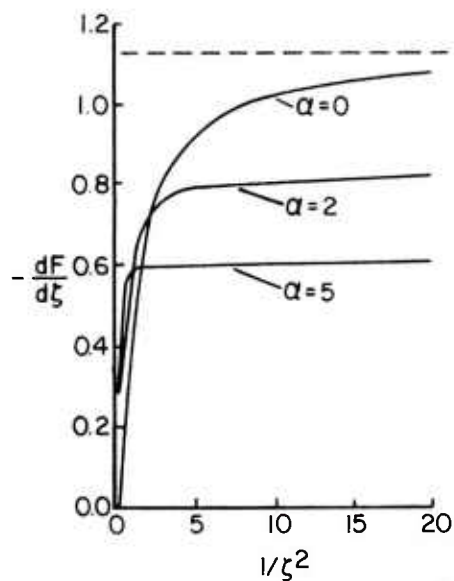


FIGURE 3b. $dF(\zeta)/d\zeta$ VERSUS $1/\zeta^2$

CONCLUSIONS

In the linear case ($\alpha = 0$), the solution $F(\zeta)$ is linearly scaled by H_0 and the shape of the solution $F(\zeta)$ remains unchanged. Examination of the preceding analyses and numerical calculations reveals that the nonlinear case is more complicated. As might be expected, the results in the nonlinear case may not be simply scaled in the usual linear fashion. Although by the nature of the problem the final value H_0 is the same, the manner in which this value is approached depends on the value of H_0 . The variation in permeability with applied magnetic field strength in the present problem results in a more rapid penetration of the medium by the magnetic field. In other words, at a given position in the medium, the magnetic field strength H will reach a given percentage of the applied field H_0 in a shorter time. Alternatively, at a given time the point at which the magnetic field strength has a prescribed value will have penetrated to a greater distance. In addition it is evident that for a given value of the parameter α the solution $F(\zeta)$ at some $\zeta = c$ will have the same value for all combinations of x and t which are related by $\zeta = c$. This indicates that if a certain value for the magnetic field strength H has reached location x at some time t , then the same value of H will occur at some location $x' = ax$ at some time $t' = a^2t$. It might be pointed out that for a given α , the solution $F(\zeta)$ will have reached a prescribed value F_p at some ζ_p . Thus, one can consider a "penetration thickness"

$$\delta_p \equiv \zeta_p (2/\sqrt{\sigma\mu_1})\sqrt{t}$$

beyond which the magnetic field strength has changed by less than 'p' percent of H_0 . For example, the linear solution (22) has a value less than 0.01 when the error function argument is about 2. If it is necessary to calculate the electromagnetic fields in a shield of finite thickness, then the solution $F(\zeta)$ will be a good approximation if the "penetration thickness" is small with respect to the shield thickness.

REFERENCE

- [1] J. Crank, *Mathematics of Diffusion*, Oxford University Press, London, 1964, p. 166-170.

LASER-INDUCED OPTO-ACOUSTIC PULSES IN A FLAME

DAVID R. CROSLEY, DR. .
BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MD 21005

INTRODUCTION

Laser Probes and Combustion Chemistry

The application of lasers to the study of combustion processes offers a wealth of information of diverse kinds. Many of these new techniques - in particular laser-excited fluorescence, ordinary Raman scattering, and coherent anti-stokes Raman spectroscopy - are aimed at the measurement of concentrations of individual molecular species at partial pressures ranging from several atmospheres down to 10^{-10} torr.(1) The use of scanning lasers permits one to map out energy level population distributions within ground electronic states, with an ease and accuracy previously possible only for excited states. From such information can be obtained temperatures corresponding to different degrees of freedom. Measurement of laser absorption line-shapes within flames provides data on energy transfer rates and translational temperatures. The recent successful observation of two-photon excitation of Na in a 1-atm pressure flame(2) suggests numerous possibilities for probing those species now inaccessible by present-day commercial available lasers with one-photon excitation.

The use of the full arsenal of these laser techniques so far demonstrated would provide a nearly complete experimental characterization of a process undergoing combustion. Of course, the use of lasers means, almost necessarily, a high degree of spatial resolution: focusing to 100 μ beam diameter is generally easy. Pulsed lasers of various types yield concomitant temporal resolution down to several nsec. Consequently spatial profiles, over time periods unblurred by turbulence, can be obtained for these quantities of interest. It

should be noted that, in addition, laser probes are non-perturbative. That is, they introduce no physical barriers into the gas flow or provide surfaces which could cause heterogeneous catalysis influencing the chemistry. Although the presence of particulate matter can be a serious problem, extremes of temperature and pressure offer no hostility to the laser beam itself.

The field likely to benefit most from this detailed picture is that of combustion chemistry. The information on transient, reactive species will provide the key to the mechanistic chemical kinetics heretofore usually only inferred from the composition of final products. Data on state distributions will directly address the questions of disequilibrium in combustion systems, widely recognized as existing but seldom accounted for in mechanistic schemes. In fact, this combination of experimental probes using laser-based techniques together with theoretical chemical kinetic modelling on modern large computers has been heralded⁽³⁾ as promising - finally - significant advancement in the enormously complex problem of the chemistry of combustion. An understanding of the chemistry pertinent to ballistic systems will undoubtedly require such models tied in detail to copious laser-based experiments (as well as measurements with other techniques) under controlled laboratory conditions and proven by agreement with the necessarily more limited data available on real systems.

Nearly all of the currently available laser probe methods rely on the scattering or emission of light as the mode of detection. While this can be a highly sensitive method under many conditions, there are other situations in which it is less suitable. For example, high excited state quenching rates reduce the degree of emitted fluorescence, and Raman-scattered photons could be absorbed by some interfering molecular species. One recent alternative method is the optogalvanic effect,⁽⁴⁾ in which the yield of collisionally produced ions is enhanced by promotion, using laser excitation, to excited electronic states lying nearer to the ionization continuum than does the ground state.

Opto-Acoustic Pulses

We discuss here another means of detection of the absorption of laser radiation by individual species in a flame environment: the production of opto-acoustic pulses in a flame at 1-atm pressure.⁽⁵⁾ The flame is seeded with alkali atoms and a pulsed laser is tuned to the appropriate absorption line. While some of the electronically excited atoms fluoresce, the vast majority lose their energy by collision with other gases present in the flame. This energy ultimately

(though rapidly) is converted into translational kinetic energy of the flame gases, producing a pulsed pressure wave which expands from the region illuminated by the laser. This sound wave is readily audible to an observer in the vicinity of the burner.* Quantitative measurements are made using a microphone, and show that the technique is very sensitive and accurate, and perturbs neither the gas dynamics nor chemical kinetics of the flame.

Now the opto-acoustic effect in general, *viz.*, the conversion of optical to acoustical energy through absorption and collisional quenching, has long been known as a means of sensitively detecting the absorption of radiation. What is here newly demonstrated is the nature of its occurrence following pulsed excitation, and the existence of the effect under *in situ* conditions on a 1-atm pressure flame burner (in contrast to the typical mode of operation using specially constructed cells at low pressure).

In addition to exploiting its detection capabilities, we have also utilized the effect to perform measurements of the speed of sound within the flame on a spatially resolved basis. These data provide information on the translational motion of the flame gases. This aspect may prove especially useful for environments too hostile to insert thermocouple probes or processes too rapid to obtain usable thermocouple response.

Finally, we note that the form of the pressure wave appears to contain information concerning density gradients within the flame, and may be influenced by bulk energy transfer within the flame gases. However, we have not pursued these possibilities in a quantitative manner in this study.

* In fact, a typical mode of locating the Na or Li resonance lines within the lab has become tuning the laser while listening for the sound wave produced when on resonance.

CHARACTERIZATION OF THE EFFECT

Experimental Details

The overall experimental arrangement is depicted in Fig. 1. The flame utilized in the experiments is an approximately stoichiometric mixture of C_2H_2 and air, seeded with Na atoms. (However, the generation of opto-acoustic pulses has also been accomplished using Li in the C_2H_2 /air flame, and with Na in a CH_4 /air flame). This is burned on a standard burner used for atomic absorption analytical measurements, so that the Na may be introduced by aspirating NaI solution. The Na atom number density was typically of the order of 10^{10} – 10^{11} cm^{-3} , as determined from both absorption and absolute fluorescence intensity measurements. The primary reaction zone at 1 atm extends only a few mm above the burner surface. The experiments are carried out at heights ranging from 0.5 to 3 cm above the surface, that is, in the region of the secondary reaction zone containing partially combusted gases.

The laser used is a Chromatix CMX-4, a commercially available flashlamp pumped tunable dye laser having a pulse duration of 1 μ sec. Although the laser is capable of delivering 10 mJ/pulse in the region of the Na resonance lines, it was here operated typically at < 1 mJ/pulse. The bandwidth is nominally 3 cm^{-1} and can be narrowed to 0.16 cm^{-1} by insertion of an etalon. (In a separate experiment, not further described here, opto-acoustic pulses have also been observed with a N_2 -laser-pumped-dye laser. The 5 nsec pulse lengths available with this instrument offer useful versatility with the technique).

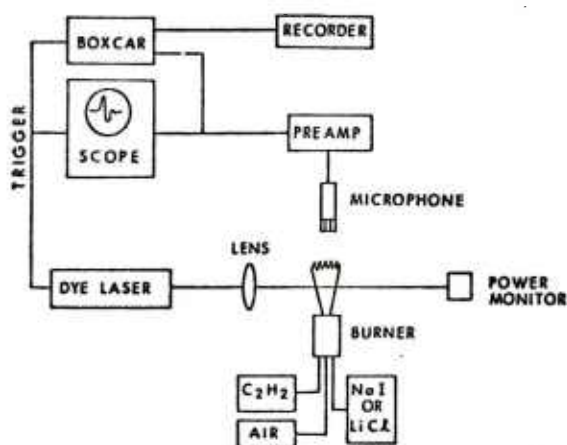


Fig. 1. Experimental arrangement. The laser scanner, and a spectrometer often used for fluorescence observations, are not shown.

The laser is focused to a few mm beam diameter and directed into the flame at a spot ~ 2 cm above the burner surface. When the laser is tuned to either of the two components of the 3^2P-3^2S transition in Na, the energy released in the quenching of the resonantly excited $3P$ state produces a pressure wave which propagates outward from the region illuminated by the laser. This sound wave is detected by a condenser microphone located ~ 4 cm from the flame. After amplification, the microphone signal is fed to an oscilloscope and to a boxcar (gated) integrator, both of which are triggered by a pulse from a photodiode detecting the laser pulse itself. Fig. 2 is a recorder tracing of the amplitude of the pressure wave signal vs. laser frequency. The laser is operated here in its narrowed mode and automatically scanned over the wavelength region which includes both doublet components. (The spikes marked 'etalon reset' are artifacts of the scanning sequence.) It is clear that the pressure waves result from the resonant electronic excitation and not from some other mode of laser energy deposition within the flame gases.

Magnitude of the Effect

We here calculate the amplitude of the pressure wave, given a simple picture of its formation, and compare the result with a measured value. The reasonable agreement achieved lends confidence to the physical picture, and shows that the pressure wave should not perturb the flame to any noticeable degree.

We consider a particular (though typical) experiment in which the Na absorbs 0.22 mJ from a single laser pulse, as the beam transverses a 1 cm path through the flame with a beam diameter of 1.5 cm. Over 99% of this energy is quenched; we assume that it immediately (on the

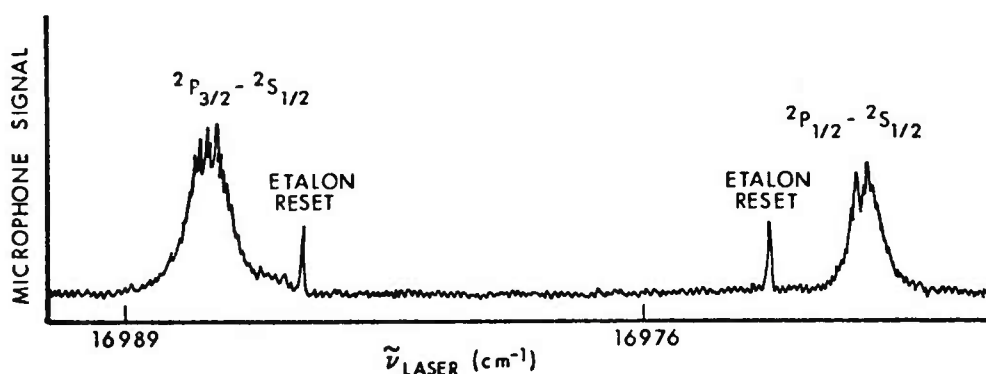


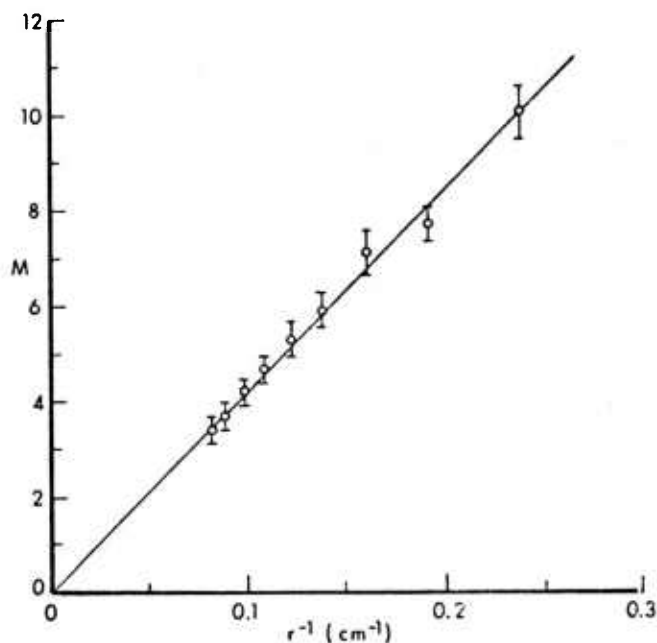
Fig. 2. Opto-acoustic pulse signal as a function of laser frequency.

time scale of the laser pulse duration) is converted to translational energy* so that the illuminated region is effectively locally heated. This energy will then begin to spread into the surrounding regions of the flame. However, the laser pulse length, equivalent to a heating time, is shorter than the time it takes sound to traverse the heated region. Thus it is possible that a shock wave of very small amplitude is initially formed; if so, it soon degrades into a pulsed pressure wave.

The flame gases are assumed diatomic here, with a density appropriate to 1 atm and the flame temperature (from speed of sound measurements described below) of 2300°K. The irradiated volume of 1.75 cm³ thus has a heat capacity of 0.34 mJ/degree and is heated 0.64°K. This is a truly negligible amount of heating compared with the temperature of the flame gases themselves.

A pressure rise of 0.22 torr within the heated volume is calculated using the ideal gas law. This would produce an amplitude of 0.044 torr at the position of the microphone. (The expected inverse dependence on distance is verified by experiment; see Fig. 3.) The actual measured value** is 0.13 torr. In view of some contributions

Fig. 3. Dependence of pressure wave amplitude on flame-to-microphone distance r .



* See below for further comment on this process.

**Using a calibrated microphone.

due to non-uniform irradiation, and especially the lack of knowledge concerning the actual mode of formation of the pressure wave, we consider this to be reasonable agreement.

This model suggests that the amplitude of the effect should be proportional to both the laser energy and the Na concentration. Fig. 4 shows such dependency. (In Fig. 4a, the laser power is well below the values later used for optically saturating the transition.) The high values of Na concentration in Fig. 4b exhibit non-linearity due to the large amount of absorption occurring.

Form of the Pressure Wave

The form of the pressure wave is quite complex, and has been extensively examined although no quantitative information could be extracted. It is quite reproducible on a pulse to pulse basis for a fixed laser beam/flame/microphone geometry, though it varies as that

Fig. 4. a) Opto-acoustic signal as a function of laser power
b) opto-acoustic signal as a function of sodium concentration

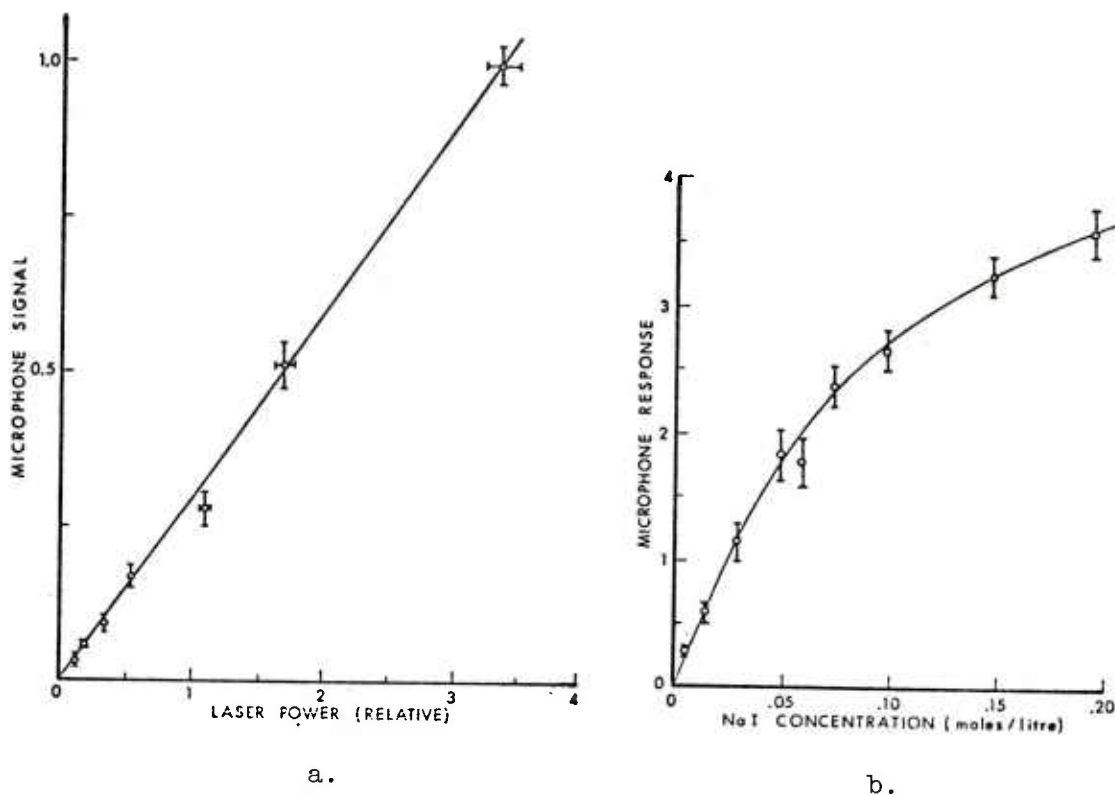
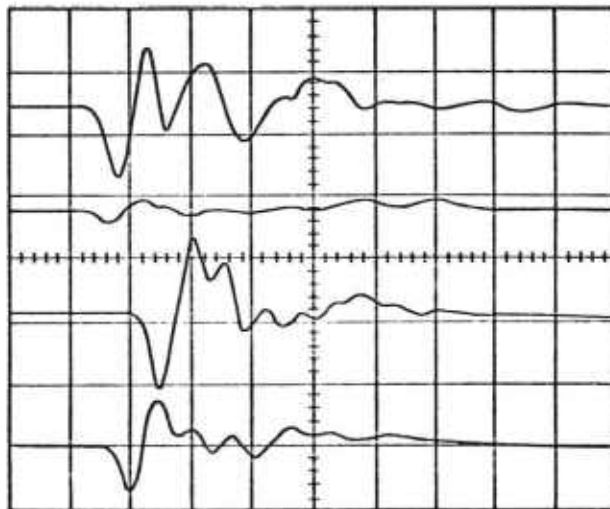


Fig. 5. Oscilloscope traces of opto-acoustic signal waveform, for 4 different geometrical arrangements. Each results from a single laser pulse. Horizontal axis: 20 μ sec per major division.



geometry is changed. Fig. 5 shows four oscilloscope traces for four different geometrical arrangements. The scope is triggered on the left by the photodiode, and the time scale here is 20 μ sec/cm. (The amplitude referred to throughout this paper is measured by setting the two boxcar gates to sample the first positive and negative extrema of the waveforms, and taking the difference.) The characteristics of the waveform, especially the smaller-amplitude oscillations following the first large wave, can be readily varied by introducing disturbances, particularly turbulence, into the flame. This implies that the form is dictated by reflections and/or interference from density gradients within the flame, though no analysis was attempted.

A sound wave having a frequency of the order of the inverse of some relaxation time (e.g., a vibrational energy transfer rate) within the system will undergo attenuation and phase shifts due to that relaxation process. Since vibrational relaxation times at 1 atm are of the order of the laser pulse duration, and since the laser pulse contains a spectrum of frequencies, it was hoped that a Fourier analysis of the waveform might yield information on these relaxation times within the flame. This approach was not successful, due to the dominance of the density gradient effects and the bandwidth (140 kHz) of our fastest microphone. "Air" mixtures of N_2/O_2 , Ar/O_2 , and CO_2/O_2 were tried in this series of experiments, but no differences could be discerned. With a faster microphone it remains possible that this will be a useful application of this technique.

Even though the vibrational relaxation could not be directly observed in this manner, it is safe to conclude that the energy is transferred through the internal levels of the flame gases. N_2 , which constitutes some 60% of the total gas flow, has long been known to have a very high quench rate for electronically excited alkalis, especially Na. The suspicion that such high efficiency is due to the internal levels of the N_2 accepting the Na electronic energy has recently been borne out in a detailed experiment investigating the process.(6)

QUENCH RATE MEASUREMENTS

Descriptive Equations

Under conditions of high laser intensity, one may create an appreciable population in the electronically excited state. This will be the case when the rate, per atom, for absorption of laser light becomes comparable to the rate at which the atoms exit the upper state. Then, the number of excited atoms is no longer linear in the laser intensity. For Na, it is relatively easy to attain the necessary laser power density by focussing the beam, even for the high quench rates found in flames. This phenomenon is here used to measure the quenching rate of the excited state, which is the dominant term in determining its lifetime.

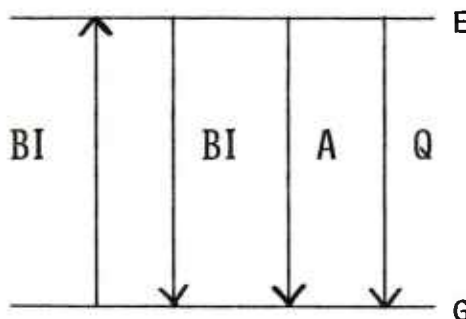


Fig. 6. Schematic illustration of 2-level system near optical saturation.

We consider a two-level system (see Fig. 6) and write a steady-state equation for the excited state number density N_e .

$$\frac{dN_e}{dt} = 0 = BIN_g - (BI + Q + A)N_e \quad (1)$$

The assumption of a steady state is valid since the time scales involved for the absorption and quenching processes are significantly shorter than the laser pulse duration. Here, B and A are the Einstein

coefficients for absorption and spontaneous emission, respectively; the term BIN_e accounts for stimulated emission. I is the laser spectral power density, measured in (erg/sec) per cm^2 per unit frequency interval. Q is the total quench rate, per sec, for the upper level.

Since the total Na number density is a constant N_0 , i.e.

$$N_e + N_g = N_0,$$

Eq. (1) may be solved for the steady-state value of N_e

$$(N_e)_{s.s.} = \frac{BIN_0}{Q+A+2BI} . \quad (2)$$

Now the opto-acoustic pulse signal M is proportional to $(N_e)_{s.s.}$:

$$M = c(N_e)_{s.s.} ,$$

where the constant represents the efficiency of conversion of electronic to translational energy, the microphone and amplifier characteristics and geometrical considerations. Substituting into Eq. (2), one has for the inverse of the microphone signal

$$M^{-1} = c' \left[2 + \left(\frac{Q+A}{B} \right) \frac{1}{I} \right] . \quad (3)$$

Thus a plot of M^{-1} (in arbitrary units) vs. I^{-1} (in absolute units) should yield a straight line with a slope to intercept ratio of

$$\frac{Q+A}{2B} = \frac{Q/A+1}{2B/A} .$$

The relationship between B and A is known from thermodynamic considerations:

$$B = \frac{\lambda^3}{2hc} A ,$$

so that the experimental result serves to determine Q/A . But if A has been separately measured, then an absolute value of Q may be extracted.

Experimental Results

Measurements were made at several heights above the burner surface. At each position, the microphone signal was measured as a function of laser intensity by inserting a series of neutral density filters into the laser beam. The data were taken in a manner alternating reference measurements with no filter and the signals with each given filter, in order to compensate for any drift in the overall laser power. The power measurements are made on an average power basis and reduced to instantaneous spectral power density

($\text{erg cm}^{-2} \text{ sec}^{-1} \text{ Hz}^{-1}$) using the laser bandwidth, an assumed Gaussian gain profile, and observations of the pulse duration.

Fig. 7 shows a plot of the data for one run, in the form of Eq. (3). The error bars represent estimates of the readability of the box-car signals; the line is an unweighted least squares fit to the data. In this instance we obtain $Q/A = 410 \pm 40$. Using the value⁽⁷⁾ $A = 6.3 \times 10^7 \text{ sec}^{-1}$, this results in a $Q = (2.6 \pm 0.3) \times 10^{10} \text{ sec}^{-1}$.

The results for several heights above the burner are shown in Fig. 8. There is a marked falloff with height. This is at first surprising. Since the pressure across this flame is nearly constant, and the temperature decreases over this region, the number density is increasing with height. In addition, quenching rates should show little temperature dependence due to the exoergicity of the process. (In fact, the scanty data that does exist shows⁽⁸⁾ the $\text{N}_2\text{-Na}^*$ cross section dropping slightly in this temperature range.) This means that the observed variation of Fig. 8 must be caused by the presence of some species which itself has a large concentration gradient through this region. On the other hand, N_2 constitutes most of the flame gases. This molecule already has a large cross-section, so the species responsible must have a very large quench rate indeed. It is possible that the decrease reflects contributions from reactive quenching, an interpretation in accord with recent results of Schofield and Steinberg⁽⁹⁾ who find that in some flames, the overall Na concentration

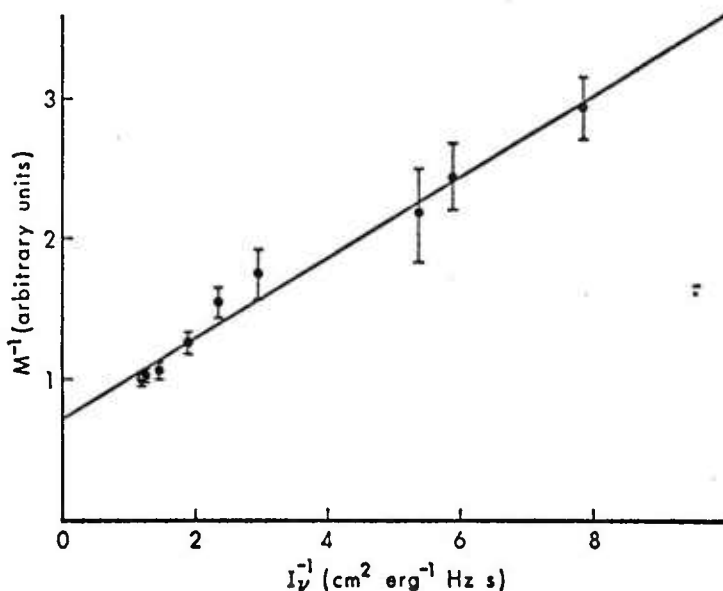
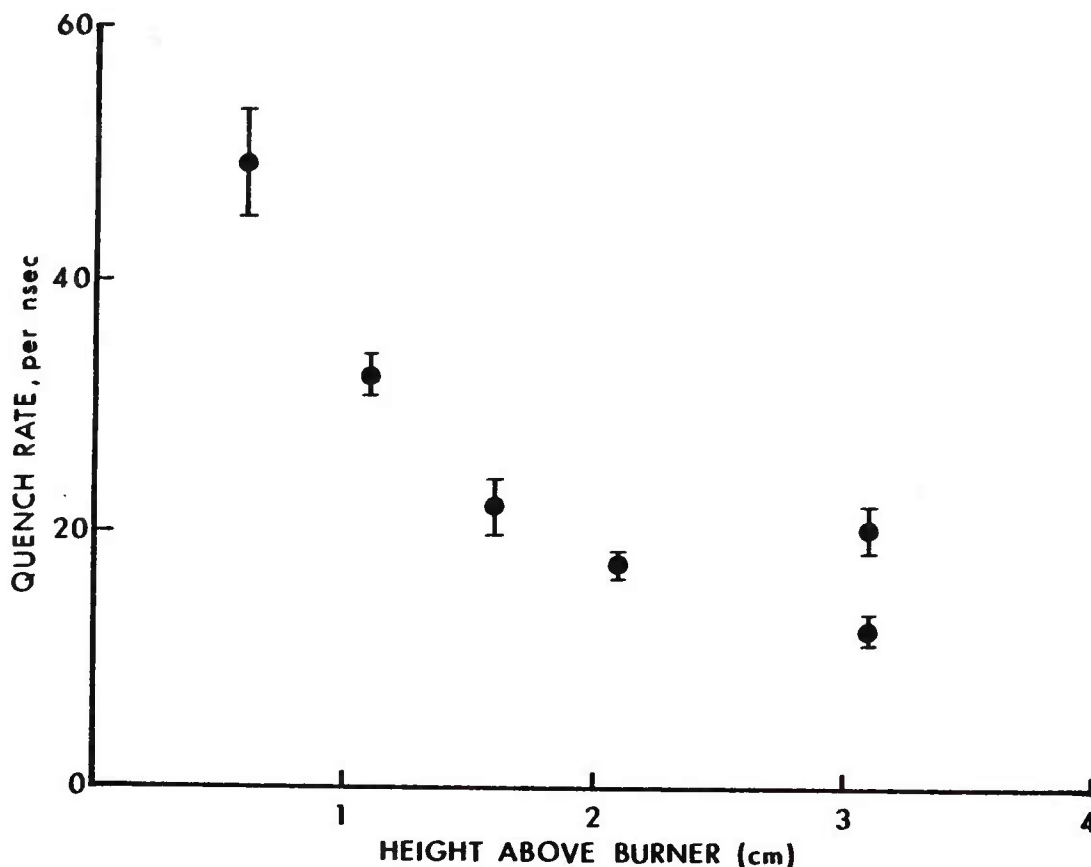


Fig. 7. Plot of inverse of microphone signal vs. inverse of laser spectral power density for quench rate determination.

Fig. 8. Quench rate for Na 3^2P level as a function of height above the burner in a 1 atm acetylene-air flame.



drops with height due to NaOH formation. Also, the measured quench rates (10 to 50 per nsec) are substantially higher than the value (3 per nsec) calculated using the measured(10) bimolecular quench rates in a CH_4-O_2 flame and an assumed stoichiometric composition(11) for the current flame. This further suggests the presence of some intermediate species removing the Na($3P$) state by some means, reactive or non-reactive.

It should be noted that of course the $3P$ level of Na is a doublet, so that the simplified two-level scheme outlined above is inadequate. Since only one component of the doublet is pumped in any given experiment, collisional transfer rates between the two components enter into the steady-state expressions. The necessary accounting for such transfer can be performed only by making measurements

on dispersed fluorescence following initial pumping of one level by the laser. Preliminary results(2) from such experiments indicate that the quench rates shown in Fig. 8 are uniformly higher than actual by about 11%. This of course would not change any of the inferences about the cause of the variation of Q with height.

For this series of measurements, the opto-acoustic pulses have served simply as a sensitive detector, and the fluorescence itself could have been used. However, the acoustic signal suffers from no reabsorption problems as does fluorescence, facilitating comparisons within different regions of the flame. This can be in general (and here is, in particular) an important advantage.

SPEED OF SOUND MEASUREMENTS

Both the oscilloscope and the boxcar are triggered by a pulse from a photodiode which responds to the laser firing. This enables the arrival time of the leading edge of the pressure pulse to be accurately measured. If the laser beam is then moved toward or away from the microphone, remaining in the flame and at the same height, the difference in arrival times together with the distance moved furnishes the speed of sound within the flame. (Although these measurements were made physically moving the beam for successive shots, a better method would be to split the beam into two or three components and obtain the information from one laser pulse, thus adding time resolution).

The measurements were made at three different heights well into the region of partially burned gases. The results were all the same, $(9.7 \pm 0.5) \times 10^4$ cm sec⁻¹, within experimental error. The uncertainty could be reduced by a more careful measurement of distances than was done for this demonstration experiment.

For a gas of average molecular weight \bar{M} and average heat capacity ratio $\bar{\gamma}$, the speed of sound u_s is related to the translational temperature by

$$u_s = [\bar{\gamma}RT/\bar{M}]^{1/2}$$

Again, a stoichiometric composition is used to calculate \bar{M} and $\bar{\gamma}$; the results are insensitive to this choice. The measured u_s corresponds to a temperature of $2280 \pm 230^\circ\text{K}$. For comparison, the ^sadiabatic flame temperature for a stoichiometric acetylene air flame is 2545°K .(11)

The translational temperature is a most important flame characteristic, from the standpoint of both the gas dynamics and the

chemical kinetics. The opto-acoustic pulse method offers the possibility of measuring the closely related speed of sound on a basis of high spatial and temporal resolution. No perturbation such as a thermocouple need be introduced into the flame, and extremely high temperatures (which are hostile to non-optical probes) appear to pose no fundamental difficulties. Such speed of sound measurements could well form the unique contribution of this method as a combustion diagnostic tool.

CONCLUSIONS

The experiments described here demonstrate that the pulsed opto-acoustic effect in a flame forms a sensitive and reproducible detector of the absorption of laser radiation by individual species within the flame. On the basis of the data such as that shown in Fig. 4, the magnitude of the observed effect can be scaled to very low values of overall absorption. Using the CMX4 laser, and this C_2H_2 /air flame, Na should be detectable at concentrations of $< 10^7$ atoms/cm³ with usable signal to noise; this is a partial concentration of < 0.01 ppb.

The method should be suitable for other species as well, as long as the overall quench mechanism is similar to the $E \rightarrow V \rightarrow T$ process here. It should be especially useful at high pressures where little of the absorbed laser energy appears as fluorescence but rather the excited state is predominantly quenched. For molecular systems, it holds promise for the detection of species which absorb but do not fluoresce at all, for example, due to rapid intersystem crossing (singlet to triplet) followed by quenching.

In addition, the method affords the ability to measure, rapidly and with spatial resolution, the speed of sound in a combustng mixture. The form of the pressure pulse produced, if it yields to analysis, holds information on density gradients within the flame and perhaps on relaxation phenomena within the flame gases.

ACKNOWLEDGMENTS

This project owes its success to the participation of, and contributions by, Drs. John E. Allen, Jr., and William R. Anderson, both NAS-NRC Postdoctoral Research Associates. The author gratefully acknowledges a pleasant and fruitful collaboration.

REFERENCES

- (1) See, e.g., A. C. Eckbreth, P. A. Bonczyk and J. F. Verdick, United Technologies Research Center Report R77-952665-6, Hartford, Conn. (1977).
- (2) J. E. Allen, Jr., W. R. Anderson, D. R. Crosley and T. D. Fansler, 17th Symposium (International) on Combustion, Leeds, England, Aug. 1978.
- (3) D. Hartley, M. Lapp and D. Hardesty, *Physics Today*, Dec. 1975, p. 37.
- (4) R. B. Green, R. A. Keller, P. K. Schenck, J. C. Travis and C. G. Luther, *J. Amer. Chem. Soc.* 98, 8517 (1976).
- (5) A preliminary account is given in J. E. Allen, Jr., W. R. Anderson and D. R. Crosley, *Optics Letters* 1, 118 (1977).
- (6) I. V. Hertel, H. Hofmann and K. A. Rost, *Phys. Rev. Lett.* 36, 861 (1976); *Chem. Phys. Lett.* 47, 163 (1977).
- (7) A compilation of Na lifetime measurements is contained in N. Ioli, F. Strumia and A. Moretti, *J. Opt. Soc. Am.* 61, 1251 (1971).
- (8) P. L. Lijnse and R. J. Elsenaar, *J. Quant. Spectr. Radiat. Trans.* 12, 1115 (1972).
- (9) M. Steinberg, Univ. Calif. at Santa Barbara, private communication, Feb. 1978.
- (10) H. P. Hooymayers and C. Th. J. Alkemade, *J. Quant. Spectr. Radiat. Trans.* 6, 847 (1966).
- (11) A. G. Gaydon and H. G. Wolfhard, Flames, 3rd Ed., (Chapman and Hall, London 1970).

THE INFLUENCE OF LARGE RUNWAY SURFACE
ROUGHNESS ON AIRCRAFT RESPONSE

*LANDON K. DAVIS, MR.

WALTER J. HORN, DR.

U. S. ARMY ENGINEER WATERWAYS EXPERIMENT STATION
VICKSBURG, MISSISSIPPI 39180

An accurate and reliable evaluation of the effective damage produced by military attacks against airfield runways cannot be made until suitable criteria for "effective" damage levels are developed. Obviously, a large crater in a runway will be an effective obstacle. However, from 30 to more than 90 percent of the damaged runway areas associated with each impact point of conventional bombs, cluster bombs, or tactical nuclear warheads may actually consist of damaged pavement surrounding the crater. Even the craters that are produced by some attacks, such as strafing, mortar, or artillery bombardments, may not be impassable obstacles.

From a defensive standpoint, it is important to know what levels of pavement damage can be tolerated by military aircraft in combat scenarios or how well such damage must be repaired (in terms of surface smoothness) during rapid runway repair operations. The level of pavement damage or runway roughness that will deny use of a runway is therefore an important parameter for both offensive and defensive aspects of warfare. The objective of this study is to develop such criteria for specific combat aircraft.

STUDY CONCEPT

The pilot and the landing gear were selected as the critical components of the aircraft system in this analysis. The failure criterion selected for the pilot was one of maximum tolerable vertical

acceleration. It was assumed that aircraft control cannot be maintained during take-off when such accelerations exceed a previously established short-term tolerance limit of 5 g's (1). The failure criterion for the landing gear was a dynamic force exceeding the calculated strength of any critical component in the mechanical linkage of the nose or main gear structures.

A basic element of this investigation was the use of an aircraft simulation computer code, TAXI, originally developed at the Air Force Flight Dynamics Laboratory to compute the acceleration response of aircraft to low-level runway roughness (2). The code was modified to (a) accept the structural and aerodynamic characteristics of a selected fighter aircraft and (b) compute the acceleration response at selected locations on the fuselage and forces within the landing gear system as various idealized types and magnitudes of runway roughness (or pavement damage) were encountered during take-off.

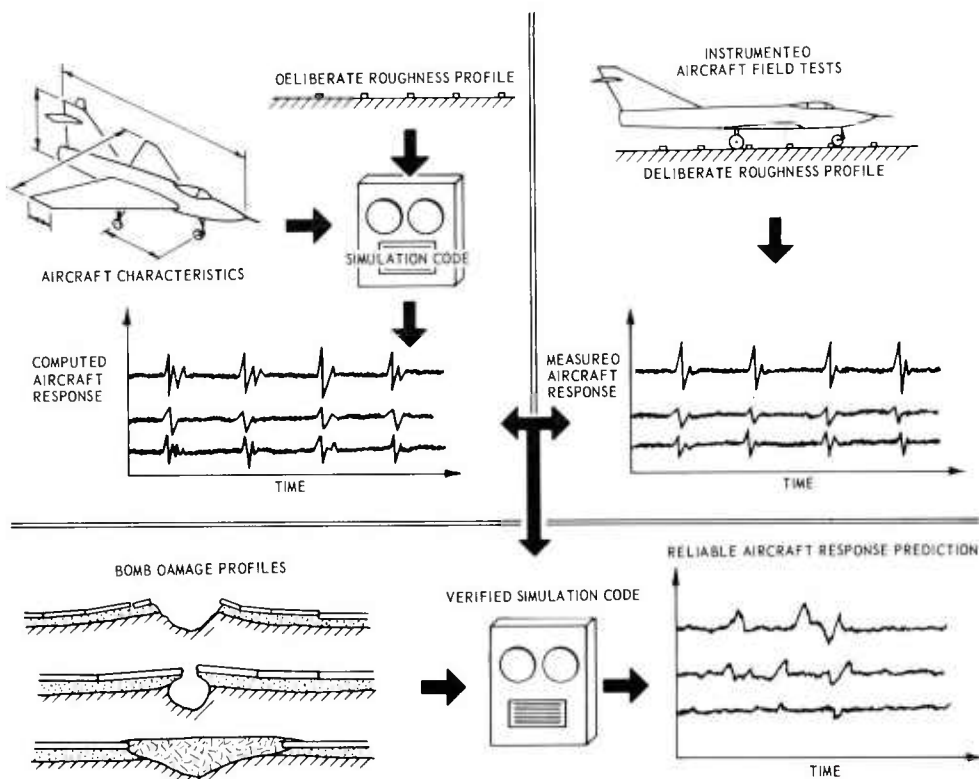


Figure 1. Concept of runway damage study.

It was necessary to verify and "tune" the computer code operation to develop confidence in the results of the TAXI calculations. This was done by comparing the initial output of the code with measurements of aircraft response to low levels of roughness obtained experimentally. Where significant differences existed between the experimental and the calculated results, the computer code was adjusted to match the experimental results. When the code was verified and adjusted, it was then used independently to predict the aircraft response to higher levels of roughness until aircraft failure levels were exceeded. The concept of this approach is illustrated in Figure 1.

RUNWAY DAMAGE DEFINITIONS

An examination of measured profiles of runway damage by various munitions (3-5) revealed four basic types of pavement damage in terms of surface profile distortions. These types, identified as "bumps" and "humps", are shown in Figure 2. A positive bump is a vertical displacement in the pavement profile extending above the normal surface, while a negative bump extends below the normal surface. Similarly, a positive hump is a mound or rise in the pavement, and a negative hump (or slump) is a depression. The boundaries of these basic profile forms are always defined by joints or cracks in concrete pavements. In asphaltic pavements, the boundaries are often less distinct and may be smooth transitions in profile. Since pavement damage intervals can correspond to spacings between impact points of cluster munitions, strafing round impacts, etc., as well as to joint spacings in pavements, the possibility of a resonant response of combat aircraft to a series of bumps was also investigated. Individual and series of bumps or humps were termed "single encounter" or "multi-encounter" problems, respectively.

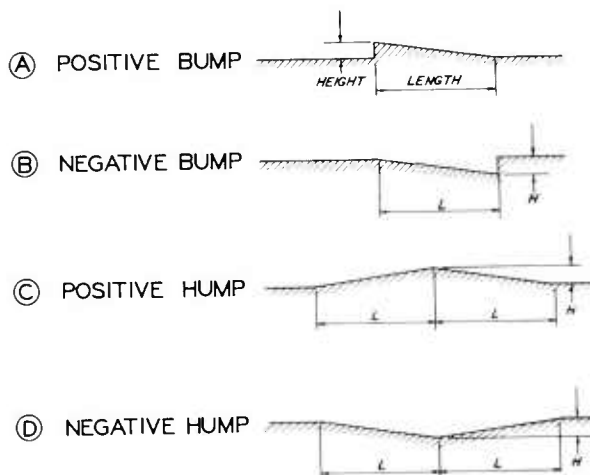


Figure 2. Idealized forms of four major types of runway roughness representing pavement damage.

EXPERIMENTAL PROGRAM

Two phases of field tests were conducted to verify the results of the aircraft response simulation computer code. In Phase 1, response data for a full aircraft/landing gear/pilot system was obtained by taxiing the aircraft across small, artificially-constructed bumps on a runway. Because of the value of the aircraft, it was necessary to restrict the severity of the simulated pavement damage for the Phase 1 tests to low levels of roughness to insure no risk of damage.

Pressure gages, accelerometers, and velocity and displacement transducers were used to monitor the response of the nose and main landing gear components as the aircraft was taxied over the artificial runway bumps. Accelerometers were also mounted on the cockpit floor. Signal conditioning and tape recording equipment was mounted on the wing of the aircraft, with electrical power taken directly from the aircraft power system.

The artificial pavement damage was simulated by plywood strips nailed to the runway surface. The single-encounter bumps ranged in height from one to three centimeters. The multiple-encounter bumps were all one centimeter high and were placed in three groups of 18 bumps each at spacings of 0.35, 1.5 and 3.5 metres. These spacings were determined by initial code calculations to be those most likely to generate a resonant aircraft response. Taxi runs were made over the bumps at speeds of 10, 20, 30 and 40 mph (Figure 3).

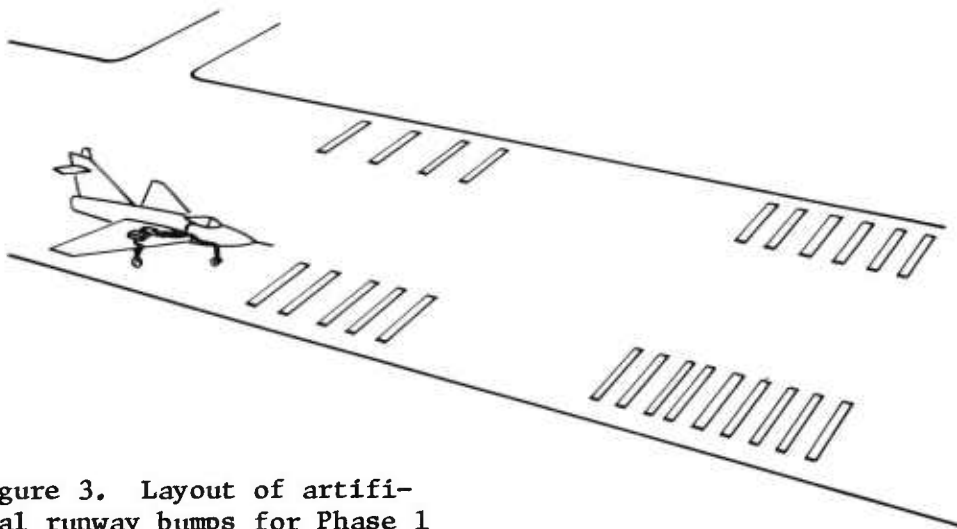


Figure 3. Layout of artificial runway bumps for Phase 1 taxi tests.

Phase 2 of the experimental verification consisted of tests similar to those of Phase 1, except that only the landing gear assembly was tested. A special test carriage was constructed to contain the gear to be loaded with a weight equal to the normal aircraft load and towed behind the truck over a series of artificially-constructed bumps, similar to those used in the Phase 1 tests. The Phase 2 tests were identical in nature to those of the Phase 1, except that the bump heights ranged from 1.0 to 7.5 centimeters, and encounter speeds from 10 to 35 mph.

VERIFICATION OF CALCULATIONS

Figure 4 shows typical comparisons of the aircraft response measured in the test programs and the response calculated by the first run of the TAXI code. Although the initial correlations were poor for very low aircraft speeds, the response measured in the full aircraft tests approached within 10 to 30 percent of the first calculated values for speeds greater than 20 mph. The measurements made in the Phase 2 tests (landing gear alone) were not expected to match the calculated values of those measured on the full aircraft, but to simply show the rate of increase in response up to larger bump heights than could be measured on the full aircraft.

After comparisons were made between all measured and calculated response parameters, small adjustments were made to the TAXI program to produce calculated results that were consistent with the measured data.

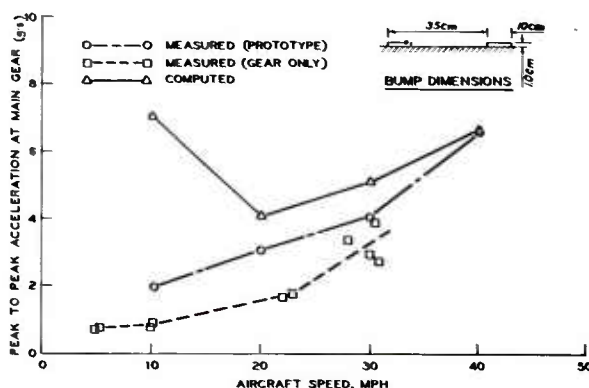


Figure 4. Comparison of measured and calculated main gear accelerations.

RESULTS AND DISCUSSION

When the verification of the TAXI computation results were judged to be satisfactory, the program was run for aircraft speeds ranging from 10 to 160 mph (8.7 to 140 knots) and for bump heights ranging from 1.3 to 15 centimeters (0.5 to 6 inches). Runs were made for each of the four types of characteristic bumps shown in Figure 2. Figure 5 shows the calculated peak acceleration in g's at the pilot station for the positive bump with a pavement slab length of 4 metres (13.12 feet).

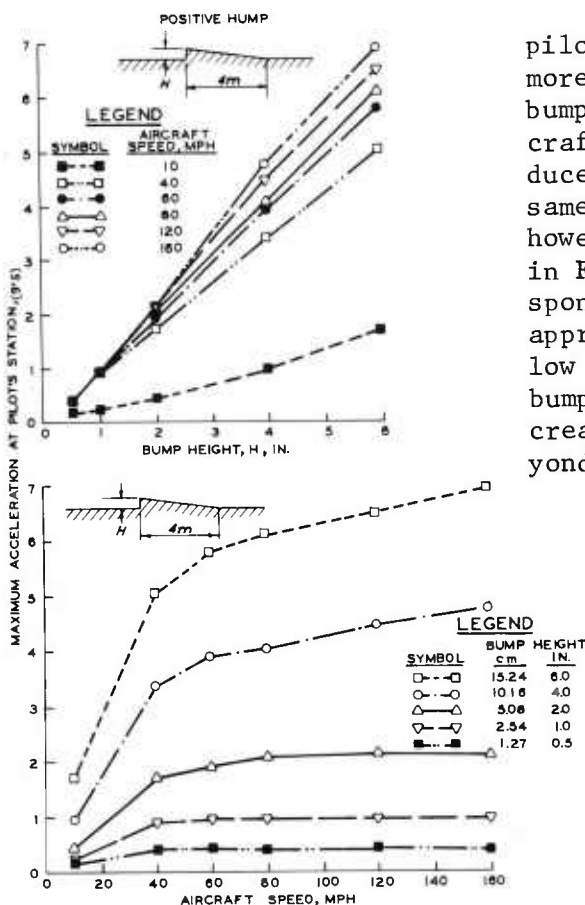


Figure 5. Two types of graphs showing calculated maximum accelerations at pilot station versus speed and bump height.

As might be expected, the pilot station acceleration increases more or less evenly with increasing bump heights, and the higher aircraft speeds generally appear to produce the highest response. When the same data is viewed another way, however, it can clearly be seen in Figure 5 that acceleration response to the smaller bumps approaches its maximum at fairly low speeds. Even for the largest bumps, there is only a small increase in pilot acceleration beyond an encounter speed of 40 mph.

Similar response data for encounters with the other three forms of pavement damage are shown in Figure 6. Figure 6a shows that for a negative bump, the pilot accelerations are similar to those generated by a positive bump up to an encounter speed of about 40 mph. As speed increases, however, the aircraft response drops to much lower levels, with the amount of drop-off increasing with the height (or depth) of the bump. This is due to the

aircraft tendency to "fly" across the bump depression as speed increases.

Pilot accelerations generated by positive and negative humps (Figures 6b and 6c) are much less than those produced by positive bumps of equal height at all encounter speeds. The tendency for response to positive humps to peak at about 80 mph is apparently due to the natural frequency of the aircraft/landing gear spring-mass system. As with a negative bump, the aircraft has a tendency to "fly" over a negative hump at higher speeds.

The forces in selected components of the landing gear for large bump sizes and high encounter speeds were also calculated with the adjusted computer code. Figure 7 illustrates the basic mechanical linkages of the nose and main landing gear assemblies, along with force diagrams showing notations for the shear or tensile forces calculated.

Figure 8 shows a typical force graph, with force plotted as a function of bump height (for a positive bump) at different aircraft encounter speeds. From graphs such as this, a bump height was determined for each speed that represented the point at which the force exceeded the limiting strength of the gear component. These failure limits were calculated for each critical component based on an ultimate shear strength of 8×10^5 psi for normal aircraft steel.

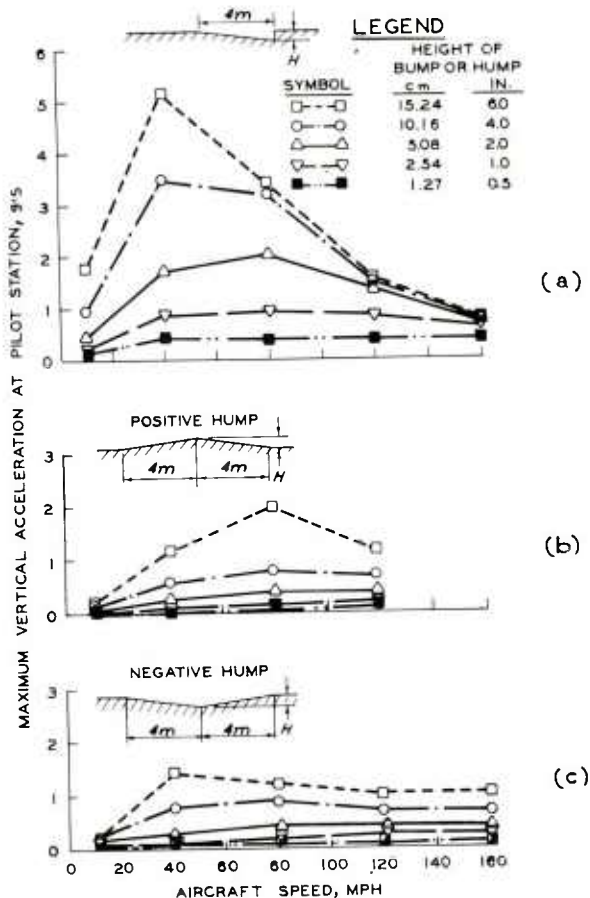


Figure 6. Calculated maximum accelerations at pilot station for different forms of pavement damage.

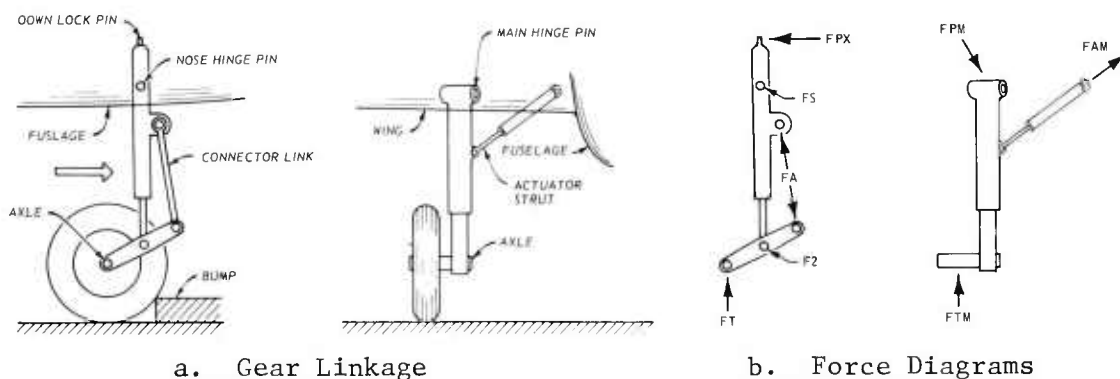


Figure 7. Mechanical linkage and force diagrams for nose (left) and main (right) gear assemblies.

Figure 9 shows failure limits, as a function of bump height and aircraft speed, for the weakest components of the landing gear/pilot system. As mentioned earlier, the criterion for pilot failure was the acceleration level at a frequency of one cycle per second which will begin to produce injury to the pilot, or about 5 g's. Also shown in Figure 9 is the failure limit for the nose gear down-lock pin for a component made of high-strength, chrome-alloy steel. The failure limits for the other gear components would be raised by an equal amount if they are also constructed of such a material.

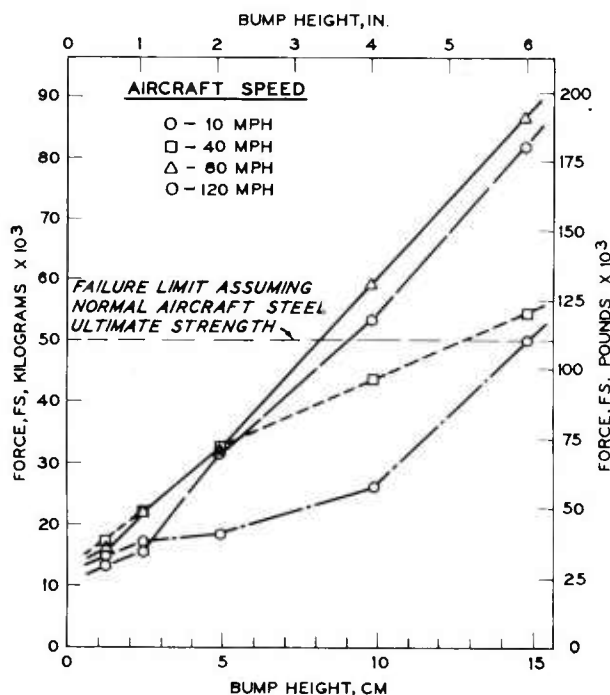


Figure 8. Force on nose gear pivot pin versus bump height and aircraft speed for positive bump.

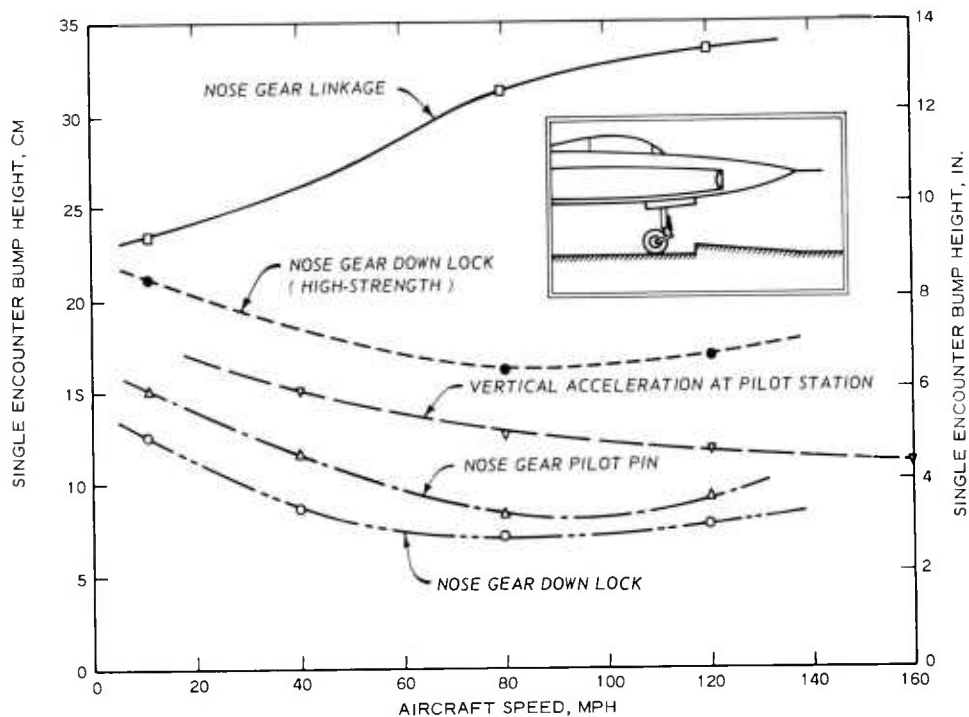


Figure 9. Failure limits of gear components and pilot control as a function of bump height and aircraft speed. Normal aircraft steel strength assumed except as noted.

REFERENCES

1. Goldman, David E. and von Gierke, Henning E.; "Effects of Shock and Vibration on Man"; Shock and Vibration Handbook, Volume 3, 1961, UNCLASSIFIED.
2. Gerardi, A. G., and Lohwasser, A. K.; "Computer Program for the Prediction of Aircraft Response to Runway Roughness"; AFWL-TR-73-109 (two vols.); April 1974; Air Force Weapons Laboratory, Kirtland AFB, NM, UNCLASSIFIED.
3. Davis, L. K.; "Vulnerability of Airfield Runway Pavements to Conventional Munitions (U); TR N-76-2, April 1976, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, MS, CONFIDENTIAL.
4. Davis, L. K.; "ESSEX VI - Effects of Low Yield Nuclear Explosions on Airfields and Railway Systems: CY 1973 Experimental Effort"; TR N-75-4 (DNA PR 0007), September 1975; U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, MS, UNCLASSIFIED.
5. Landa, D. A.; "Demonstration of Runway Cratering by Gun Projectiles, Rocket Warheads, and Rockeye Bombs"; ADTC-TR-77-31, May 1977; Eglin AFB, FL, UNCLASSIFIED.

HYBRID POWER SOURCE FOR VEHICULAR PROPULSION

EDWARD J. DOWGIALLO, MR.
MOBILITY EQUIPMENT RESEARCH & DEVELOPMENT COMMAND
FORT BELVOIR, VIRGINIA 22060

BACKGROUND

The only current off-the-shelf contender for powering electric vehicles, the lead acid battery, needs help! Gasoline contains about 5000 watt-hours of energy per pound (11 KWH/kg) while its lead acid "challenger" produces about 11 watt-hours per pound (24 WH/kg) in powering a typical electric vehicle. This paper presents a means of overcoming this energy disparity - a hybrid lead acid battery-fuel cell power source.

The Army has a requirement for higher energy density power sources for electric vehicles, the most immediate of which is for material handling equipment (MHE) such as fork lift trucks. Future military needs are anticipated in utility trucks and pick ups or vans. The two main paths being explored are higher energy density molten salt batteries, 82 WH/lb (180 WH/kg), and hybrid, or load sharing power sources. The hybrid fuel cell-battery power source technology is a prime candidate for providing a viable electric vehicle power source. Compared to present battery and engine technology, the hybrid technology should increase range, improve performance, reduce air pollution, increase efficiency and expand the options available to meet requirements in cost, size and weight. The need for improvements is particularly evident during emergencies such as the Vietnam War, when increased ammunition movement from narrow door storage igloos, up ramps to trucks was seriously hampered by the poor performance, range and large volume of present batteries in electric fork lift trucks.

INTRODUCTION

Hybrid power sources are composed of two separate power sources, capable of operating at the same time into a common load. They typically use a high specific energy (WH/kg) source such as an engine or fuel cell coupled with a high specific power (W/kg) and storage source such as a battery or flywheel. A battery-flywheel hybrid is currently under development by Garrett Air Research (1). In this case, the battery is the energy source.

The underlying purpose of the hybrid approach has been to allow operation of the major energy source near optimum design by supplying only the base or average load requirements. The storage source then supplies the transient load increases and is recharged when the load drops below average. Thus, the load variation experienced by the energy source is reduced in frequency and magnitude. This allows for design and operation within a band representing the average, rather than design for peak with dominant operation under off-peak conditions. Reduced emissions and increased efficiency are expected benefits from operation of hybrid systems.

This paper will present the design and performance characteristics of a hybrid, fuel cell-battery power source for use in MHE and other military electric vehicles.

HYBRID POWER SOURCE

The hybrid, fuel cell-battery power source is shown in Figure 1.

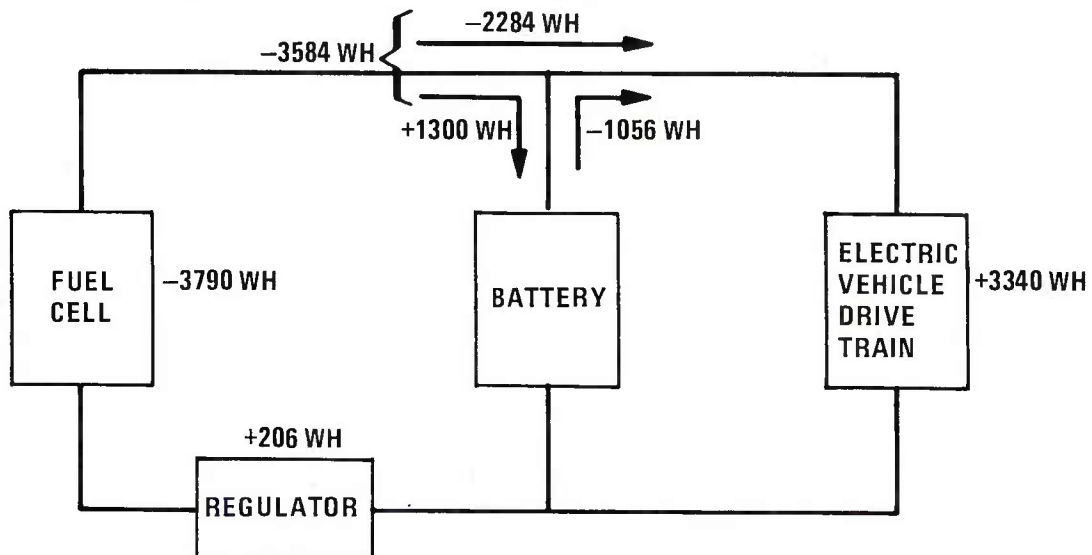


Figure 1 FUEL CELL/BATTERY HYBRID POWER SOURCE

Figure 1 shows a fuel cell energy source with a series regulator for limiting the voltage from the fuel cell to charge the battery and for limiting the current out of the fuel cell to a value that will maximize fuel cell life. The fuel cell and battery supply power to the load, in this case, an electric vehicle drive train consisting of a controller and a differential-motor drive. The interplay between the fuel cell and battery is illustrated in Figure 2. The system terminal voltage common to the load and battery is represented by the thick line in the case of a charged battery and the thick dashed line in the case of a discharged battery. The fuel cell load line extends from OCV_{FC} (the open circuit fuel cell voltage) to the relative current abscissa. The battery discharge follows a load line starting between OCV_{B1} or OCV_{B2} depending on the state of charge of the battery. The regulator can be preset to limit the maximum battery charge voltage to $MAX V_B$, the battery is charged by the fuel cell; some fuel cell current goes to the battery and some to the load depending on the relative currents at the same voltage. Charging occurs in the stripped regions. When the battery drops below minimum voltage, $MIN V_B$, motor current becomes excessive. A continuation of this will eventually cause the motor thermal cut out to be activated. The fuel cell current is held at I_{FC} by the regulator. The regulator's power dissipation is the product of its voltage drop, V , and I_{FC} .

The fuel source in lift truck feasibility studies was cylinders of compressed hydrogen gas. A methanol steam reformer is under development at MERADCOM to supply the hydrogen in the future. The fuel cell power source consisted of four, one kilowatt, stacks that were connected electrically in parallel. The stacks were fabricated by increasing the number of cells in existing 750 watt stack technology, built by Engelhard Industries. These stacks are based on phosphoric acid technology and utilize hydrogen and air. Each stack has its own blower for air and temperature control and is supplied fuel from a common hydrogen manifold. Each stack is rated to produce one kilowatt at 42 volts. The battery consisted of six auto ignition batteries. Two groups of three series connected 12 volt batteries were connected in parallel. The ignition batteries were chosen because of their low internal resistance and high rate capability. The regulator was a modified transistorized load bank. It consisted of 66 transistors connected in parallel which are capable of dissipating 1000 watts at currents up to 110 amperes.

LOAD PROFILE ACQUISITION

To design a hybrid power source, data on the temporal frequency distribution of the load is necessary (2). Information on the power required during travel and lifting with various loads was available but the frequency, duration and magnitude of these events had not been adequately recorded. Thus, the initial phase of designing a lift truck power source required a determination of these parameters for a conventional, battery powered truck.

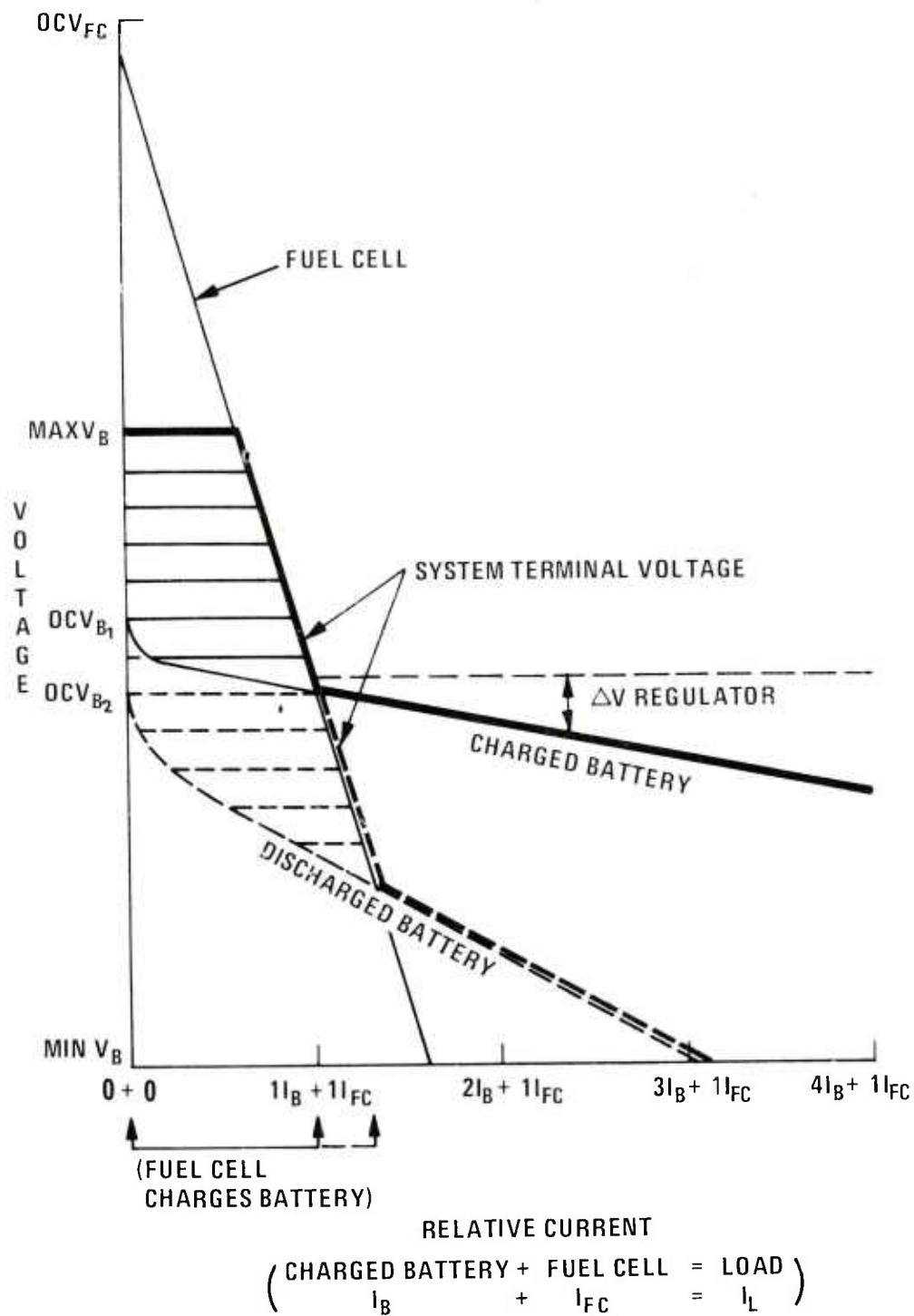


Figure 2 HYBRID CHARACTERISTICS

Instrumentation consisted of signal conditioning circuits and a magnetic tape recorder for recording analog signals of electrical parameters. The tape recorder was operated at 1.875 inches (4.763 cm) per second. The signal conditioning circuitry consisted of a main battery voltage divider, a shunt-voltage amplifier for current monitor, an analog multiplier and circuits for averaging power and current to increase the effective frequency response of the recorder. The parameters measured were battery voltage, current and power and vehicle velocity. Current and power were recorded both raw and electronically averaged. The raw recorded values preserved the rapid switching transients associated with the solid state controller. This instrumentation package was mounted on an electric vehicle during field operation.

The field recorded analog signals were then reproduced and prepared for analysis by conversion to a digital format with a high speed 15 bit A/D converter. From this data, frequency of occurrence distributions for velocity, current and power were derived. This was accomplished by counting how often preset incremental values of these parameters were exceeded.

Programs are available for estimating hybrid behavior using the data and a simple model. The model assumes that the average power is available from some primary (energy) source. When the measured demand exceeds this level it is supplied by a storage device. Conversely when the demand drops below that available the surplus is returned to the storage device. Allowance was made in the programs for some variation in transfer efficiencies in charge and discharge of the storage device. The size of the device is based on the maximum and minimum energy levels determined during processing of the profile data.

LOAD PROFILE ANALYSIS

The results of analysis of field data will be presented for a 4000 pound capacity fork lift truck that was instrumented as described above and operated over the Military Standard Endurance test course (3). This test course is a cyclic route involving three load placements and a ramp. Since load placement is done on one lap and removal is performed on another, it required two laps of the route to yield one cycle. A segment of the strip chart, Figure 3, displays the analog current and velocity from this profile and the identifiable events. The frequency of occurrence distribution for power is shown in Figure 4. The negative power values result from the minor noise about zero and some transients due to tape head vibration. These have subsequently been reduced to insignificance with new Hall effect devices and improved recorder design and mountings. This distribution shows clearly the relationship between peak to average power demand. The average power demand was determined by computer integration of the power levels.

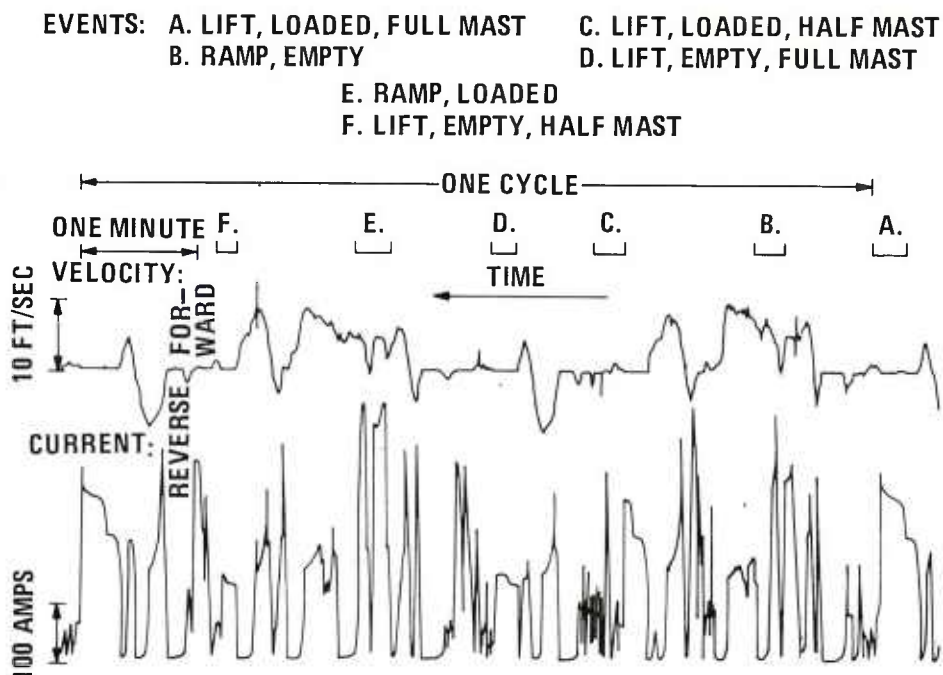


Figure 3 CURRENT-VELOCITY PROFILE FOR MIL. STD. TEST COURSE

The cyclic MIL. STD. route, wherein the same tasks are performed regularly, results in a distribution with local maxima. The full capacity lifts consumed considerable time and energy. The stopped/no work mode is an integral part of the work cycle. No work breaks are included. The data can yield the distance and travel time for discrete periods of motion. An analysis is shown in Figure 5 where the percent-of-total values indicate the percent of the total travel time represented by an interval.

The last seven entries of Table 1 represent characteristics of the hybrid derived from the computer simulation. The time spent in charging the energy storage device was roughly equal to that in discharge. Of particular importance is the energy transferred through the storage device. Forty-four percent of the energy used while performing the task came from, i.e. passed through, the storage device. Because the magnitude of the pulses, discharge and charge are small, the theoretical storage required is quite small, 0.18 Kwh. The small theoretical size indicated and the large fraction transferred implies high transfer rates are required. A term to express this is the capacity cycles per hour; this represents the number of times energy, equivalent to the capacity, is cycled through both discharge and charge. Cycles in the normal sense have no real meaning relative to a hybrid since the device may be fully discharged only once during many capacity cycles.

The above value of storage required was based on a model that assumed 100% transfer efficiency. This is not realistic. The next section describes laboratory operation with a fuel cell, batteries and regulator responding to field recorded load profiles.

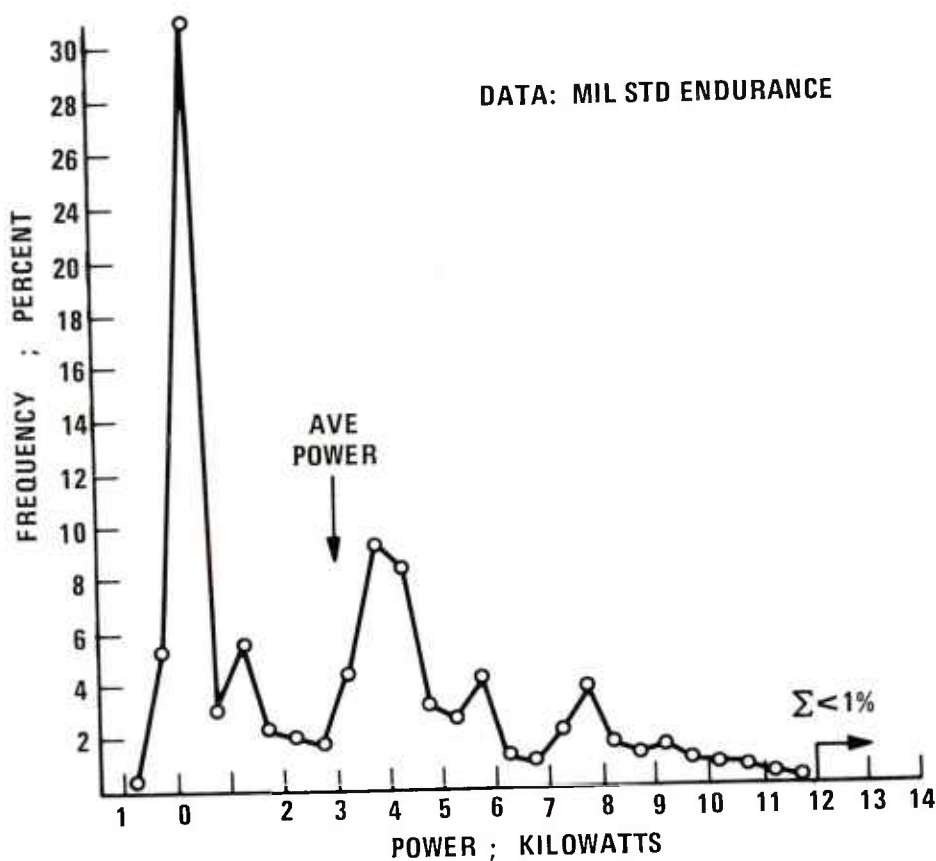
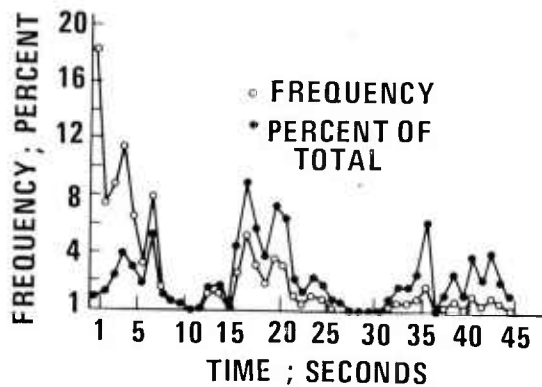


Figure 4

FREQUENCY-OF-OCCURENCE DISTRIBUTION FOR VELOCITY

FORWARD TRAVEL
DATA: MIL STD ENDURANCE



REVERSE TRAVEL
DATA: MIL STD ENDURANCE

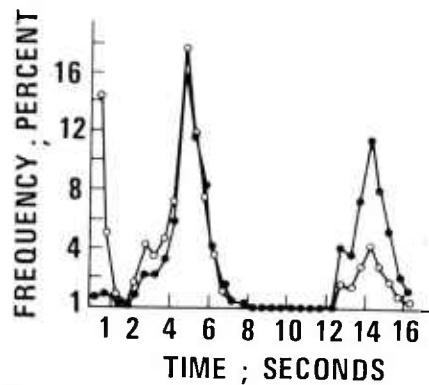


Figure 5

DISTRIBUTIONS FOR TRAVEL PERIODS, TIME, FREQUENCY AND PERCENT-OF-TOTAL

Table 1

SUMMARY OF ANALYSIS OF MILITARY STANDARD ELECTRIC FORK LIFT TRUCK PROFILE

Duration Min.	147.
Average Power, Watts	3040.
Kilowatt-Hr/Mile	1.54
Forward:	
% Time	46.
% Energy	46.
Reverse:	
% Time	17.
% Energy	16.
Stopped/Lift	
% Time	24.
% Energy	38.
Stopped/No Work	
% Time	14.
Charge, % Time	51.
Avg. Power	2580.
Discharge, % Time	49.
Avg. Power	2710.
Energy, % Transferred	44.
Storage Size, KWH	0.18
Capacity Cycles/Hr.	7.4

LABORATORY TESTING WITH FIELD PROFILES

The following field profiles for testing a hybrid were recorded by instrumenting a 4000 pound capacity electric fork lift truck at the Red River Army Depot, Texarkana, Texas. RRD #5 - An electric fork lift truck replaced a gasoline lift truck in the shipping department where a variety of items were picked up and loaded into trucks; RRD #4 - Ammunition was unloaded from a train freight car to a truck, and then removed from the truck, taken down a ramp, and placed in a storage igloo; RRD #3 - Ammunition was removed from an igloo, taken up a long ramp to a truck and the fork lift truck returned, empty, for more ammunition.

The data in Table 2 was acquired by operating with the previously described fuel cell-battery hybrid power source including the regulator against these field profiles. The discharge profiles were controlled in the laboratory by a unique in-house simulator (4). This simulator is ideal for hybrid studies. It can change from a charge to a discharge current of 1000 amperes in 0.2 seconds. To generate the data in Table 2, it was programmed by the field recorded magnetic tape profiles set to reproduce the actual field current levels.

The energies from Table 2 for the RRD #5 profile are shown in Figure 1, next to their associated components for clarity. Adsorbed energy is positive and removed energy is negative. The +1300 WH represents energy into the battery. The energy actually accepted by the battery was not measured. Future tests are planned to measure this as well as the effects on the battery life.

The profiles in Table 2 are presented in the order of increasing current and power. The MIL. STD. profile is included in the table for comparison. The MIL. STD. and RRD #3 profiles have a significant net battery capacity loss. Both profiles represent an unusual use of a fork lift truck. The endurance run is designed as a test of the mechanical components of the truck not the power source. It exceeds all field requirements. No coffee or lunch breaks are included in the profile. These non-drain periods would allow fuel cell recharge of the battery. The RRD #3 ammunition run is mainly used during war emergencies. The RRD #3 operation only lasted about one hour when the truck's motor overheated. The battery voltage had decreased rapidly under the high power ramping conditions. This required increasing currents to maintain the required power. Since generated heat is directly proportional to the square of the current, the motor overheated.

Adding one 1 KW fuel cell module should significantly extend the duration of the MIL. STD. and the RRD #3 profiles. The ramping operation in the RRD #3 profile should not increase in duration due to the stiffer system voltage. the time saved could be used for recharge by the fuel cell. The increased system voltage should prevent the motor over-

DOWGIALLO

heat experienced while determining the field profile and thus operation time should be dependent on the quantity of fuel.

TABLE 2

LABORATORY PERFORMANCE OF HYBRID POWER SOURCE WITH FIELD LOAD PROFILES

Profile	RRD #5	RRD #4	MIL. STD.	RRD #3
Test Time, Hours	2.0	1.4	2.6	1.2
<u>System Load:</u>				
Avg. Current, Amperes	44	80	97	134
Avg. Power, Watts	1670	3030	3435	4710
Energy, Watt Hours	3340	4242	8931	5652
<u>Fuel Cell Subsystem:</u>				
Avg. Current, Amperes	48	83	94	94
Avg. Power, Watts	1895	3240	3490	3420
Energy, Watt-Hours	3790	4536	9074	4104
Regulator Loss, %	5.4	4.9	3.5	4.1
<u>Battery Subsystem:</u>				
Charge				
% Time	76	61	55	42
Avg. Current, Amperes	20	37	70	42
Avg. Power, Watts	855	1515	2745	1630
Energy, Watt-Hours	1300	1294	3925	822
Discharge				
% Time	24	39	45	57
Avg. Current, Amperes	60	62	101	108
Avg. Power, Watts	2200	1223	4101	2538
Energy, Watt-Hours	1056	1223	4101	2538
Net Change, Amp.-Hours	+2.0	-1.4	-20.5	-52.3
Net Change, Watt-Hours	+244	-71	-176	-1716

DOWGIALLO

HYBRID-FUEL CELL OPERATIONAL EXPERIENCE

The hybrid power source was operated in a 4000 pound capacity fork lift truck for three years. During this period the fuel cell experienced about 54 thermal (on-off) cycles and accumulated over 250 hours of operation. The only discernable change has been a slight increase in operating temperature. This temperature normally ranges between 200 and 210°F. The system sealing seemed to be holding well. Hydrogen cross leaks were not detectable and there was no indication of acid leaks.

FUTURE MILITARY APPLICATIONS

Extending the range and performance of electric fork lift trucks with a hybrid power source would make it possible to replace not only current electric trucks but make significant inroads on gasoline powered trucks. Twenty-eight battery-powered cargo carriers of a 2-passenger, 4-wheel type with a flat cargo bed have been purchased for a facilities engineering work force at the Military District of Washington. The low maximum speeds of 10-12 mph and the lack of power to negotiate hills could be overcome with a hybrid power source, with increased range.

Several military truck manufacturers have indicated that they would respond to a documented army requirement for electrics. An assessment of present and future military use of electric vehicles is needed. A preliminary study was published in 1976 (5). The potential for replacing gasoline powered vehicles with electrics was determined for three military bases: McClellan Air Force Base, Fort Ord Army Base and the Long Beach Naval Shipyard. The results are summarized in Table 3. The study was based on lead acid electrics. Hybrid electrics' enhanced performance should assure rapid replacement of such conventional vehicles.

MERADCOM, under an interagency agreement with the Department of Energy, is conducting state-of-the-art assessments of electric and hybrid vehicles. A representative electric passenger vehicle, with a gross weight under 4000 pounds (1814 KG), has performance requirements that could be met by a hybrid power source up to about 35 mph as shown in Table 4. The peak to average power ratios are similar to that of the fork lift truck, namely 4 or 5 to 1. Cruise peak power was not given since this peak only occurs one time during the initial acceleration and is below the schedule B and C powers. For example, the 48 mph acceleration peak power is about 22 KW.

TABLE 3

EXAMPLES OF CONVENTIONAL MILITARY VEHICLES
POTENTIALLY REPLACEABLE BY ELECTRICS

Vehicle Type	Number of Vehicles in Class	Potentially Replaceable by Electrics*	Average Daily Mileage	Vehicle Application and Comments
1/2 Ton Pick-Up	390	197	44	Onbase Maintenance and Repairs, some personnel Transportation and Messenger Service.
3/4 Ton Utility Truck	155	75	22	Onbase Maintenance and Repairs.
29 Passenger Bus	17	11	49	Onbase Bus Service
1 Ton Metrovan	40	40	22	Used for Mail Distribution, Aircraft Maintenance, and delivery of Flight Crews on Base.
3 Wheel, Low Speed Truck	41	41	13	Onbase Messenger and Light Repair Service.

*Criteria: Maximum daily use would be 40 miles and top speed requirement would be 25 miles per hour.

TABLE 4

POWER REQUIRED BY A TYPICAL ELECTRIC PASSENGER VEHICLE

Mode of Operation	Peak Power (KW)	Average Power (KW)	Ratio PK/Avg.	Duration (Hours)
<u>SAE-J227a</u>				
Schedule B	26	5.3	5	1.7
Schedule C	30	7.6	4	1.2
<u>CRUISE</u>				
25 MPH	--	7.9	--	1.5
35 MPH	--	12.0	--	0.8
48 MPH	--	16.6	--	0.5

DOWGIALLO

CONCLUSIONS

An Army assessment of electric vehicle applications should be conducted. Inclusion of hybrid fuel cell-battery power sources in the assessment is expected to provide the margin required for successful resolution of present lead acid battery deficiencies. This may accelerate the development of electric vehicles as a practical alternative to conventional gasoline vehicles.

REFERENCES

1. B.H. Rowlett and R. Murry, "Electric-Powered Passenger Vehicle Design Study Program," ERDA Contract No. E(04-3)-1213, Air Research Manufacturing Company, December, 1976.
2. J.B. O'Sullivan, E.J. Dowgiallo and I.R. Snellings, "Hybrid Power Source for Material Handling Equipment," 10th Intersociety Energy Conversion Engineering Conference Proceedings, Paper 759037, 1975.
3. Military Standard Test and Inspection of Trucks, Lift, Fork, Military Standard 268C, 17 July 1963.
4. E.J. Dowgiallo, Jr., J.B. O'Sullivan, I.R. Snellings, and R.B. Anderson, "High Power Facility for Testing Electrochemical Power Sources," Princeton, New Jersey: Journal of the Electrochemical Society, Vol. 121, No. 9, September 1974.
5. Comptroller General of the United States, "Potential for Using Electric Vehicles on Federal Installations," Washington, DC, Report LCD-76-206, March 1976.

SABOT DESIGN FOR A 105MM APFSDS KINETIC ENERGY PROJECTILE

*WILLIAM H. DRYSDALE, DR.
RICHARD D. KIRKENDALL, MR.
LOUISE D. KOKINAKIS, MS.
US ARMY BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MD 21005

Increased projectile muzzle velocity and extended range can be achieved with a given gun system through the use of a sub-caliber projectile with a discarding sabot. This improvement occurs because the bore area on which the gun pressure acts may be greatly increased with only a relatively modest increase in total projectile weight, and the small diameter flight body has less aerodynamic drag. Obviously, to obtain the optimum performance improvement, the sabot mass must be kept as low as possible within the constraints set by the structural requirements of the sabots.

Foremost of these requirements is that the sabot must ensure the in-bore operation of the projectile, i.e., it must be able to seal the gun tube against the hot propellant gas. Then, especially for long rod kinetic energy penetrators, the sabot must provide enough support to prevent the subprojectile from being permanently deformed in an undesirable manner during in-bore travel. Also, the sabot must constrain the subprojectile during in-bore travel and then discard in such a manner as to impart small initial yaw and yaw rate to the subprojectile during launch. These parameters are related to target dispersion and should be reduced as much as possible. Finally, the sabot will generally provide the means for assembly of the full round. Hence, such matters as fin encroachment into the cartridge case and location of crimping grooves so that the assembled round can withstand rough handling should be considered.

In the discussion to follow, we will assume that the penetrator diameter, length, and material, fins and nosecone (in

short, the inflight configuration) are given. In addition, the interior ballistic parameters of the gun system are known. In practice the sabot designer would have input to all these decisions. However, these interactions will not be treated so that some basic concepts of sabot design may be the primary focus. These ideas were applied during the recent Advanced Technology Tank Gun Initiative (ATTGI) Program which culminated in the December 1977 Trilateral Tank Gun Trials. At these trials the U. S. Armor Piercing Fin Stabilized Discarding Sabot (APFSDS) Projectile was very successful, so there has been experimental verification of the design features to be emphasized in the following.

The first concept is the use of a ramp-back sabot. This configuration has had an extensive development at the Ballistic Research Laboratories. Both the Silver Bullet and the 60mm Anti-Armor Automatic Cannon (AAAC) Technology Programs have a long conical taper as the aft configuration of the sabot. One reason for this feature is that a sabot is composed of a number of segments or petals around the circumference, rather than being a single piece. Thus, if the aft profile of the sabot is such that the base pressure acts in an outward direction over any portion of the sabot (as due to undercutting to remove material), the result is an opening of the sabot petals and/or separation from the subprojectile. Unless an additional structural seal is added to this conventional rear shape, the gas pressure seal can be lost and severe blowby occur. This is especially true when a condition of high secondary wear is present in the gun tube. The ramp-back causes gas pressure on its surface to be translated into compressive hoop and radial stresses in the sabot. Thus the interfaces between sabot petals and between the sabot and subprojectile are self-sealing.

Most of the force imparted to the sabot by the propelling gas must be transmitted to the subprojectile. The taper profile may be tailored so that the shear distribution between sabot and subprojectile is nearly uniform. This condition allows maximum load transfer over a given length of interface.

The second major concept is that of a centered rotating band. The conventional design methodology for discarding sabot projectiles locates the gas seal near the rearmost portion of the sabot. This has the effect of placing the rotating band well aft of the center of gravity of the in-bore projectile. As will be shown later, the result is an inherently unstable configuration. In order to avoid this difficulty, the rotating band must be located so that the center of gravity of the projectile is under it. The improved stability of this configuration substantially reduces the magnitude of the transverse moments applied to the projectile by its yawing motions. The

smaller restoring moments experienced during in-bore operation allow lighter weight sabots.

Locating the rotating band near the center of the projectile required a change in the long ramp-back sabot design of the 60mm AAAC projectile. The rear ramp was shortened by using a compound taper. To maintain the favorable uniform shear load transfer which is possible under a taper, a front ramp was added ahead of the centered rotating band. This configuration is the basis of the BRL Double-Ramp Sabot.

Calculation of Taper Profile. As mentioned previously, one of the primary reasons for adapting a ramp-back sabot is the self-sealing nature of this configuration. The propellant gas operating on the tapered surfaces creates very high compressive hoop and radial stresses. These stresses are higher than the gas pressure, so there is no tendency for pressure to leak into narrow splits between sabot petals, forcing them apart.

It has been realized for many years that a long ramp-back sabot configuration might enable subprojectiles to be launched without driving grooves. The high compressive radial stress at the sabot/subprojectile interface under a rear taper, in combination with a reasonable coefficient of friction, could allow the entire shear load to be transferred from the sabot by friction alone (1). A very recent study has attempted to determine the minimum weight taper profile which would transfer a given shear load from sabot to subprojectile without exceeding a specified friction coefficient (2). The minimum weight shape did not give a uniform shear transfer along the interface. However, the use of some very severe restrictions (e.g., a rigid subprojectile) make the practical application of this work doubtful.

The methodology of calculating the shape of a long rear taper for a cylindrical penetrator is based on elementary free-body stress analysis, combined with well known solutions of closely related problems. The basic procedure was developed by Dr. Bruce P. Burns of BRL during the 60mm AAAC Technology Program. The taper profiles are designed to provide a uniform shear transfer along the interface at the maximum chamber pressure condition.

The external base pressure and the projectile acceleration may be found from interior ballistic considerations based on an approximate projectile weight. The initial stress in the penetrator at the edge of the sabot is determined from the unsupported length of rod, weight of fins or nosecone, and the acceleration. In the derivations to follow, a subscript s will refer to the sabot and

subscript p will refer to the penetrator or subprojectile.

The free bodies shown in Figure 1 are formed from the projectile by planes perpendicular to the axis. The segment is thin enough that the outer contour may be replaced by a straight line, forming a truncated conical segment. Squares and higher powers of ΔZ will be ignored compared to one. In addition, the slope of the contour will be assumed to be small. It is further assumed that the axial stress in each component is constant over the cross section, or, alternately that only average axial stresses are considered. The shear τ is to be constant in the axial coordinate and may be assigned any desired value. The object is to determine $R(Z)$, the shape of the contour, which will accomplish this.

Summing the forces defined in Figure 1a acting on the sabot in the axial direction, with consideration of the mentioned assumptions, gives

$$(1) \quad (P_o + \sigma_{zs}) \frac{d\phi}{dZ} + \left(\frac{d\sigma_{zs}}{dZ} - \rho_s \ddot{Z} \right) \phi = 2\tau R_p$$

where the unknown is

$$\phi = R^2 - R_p^2 ; \phi = \phi_o \text{ at } Z = 0.$$

To be able to solve this relation for ϕ , the axial sabot stress must be known.

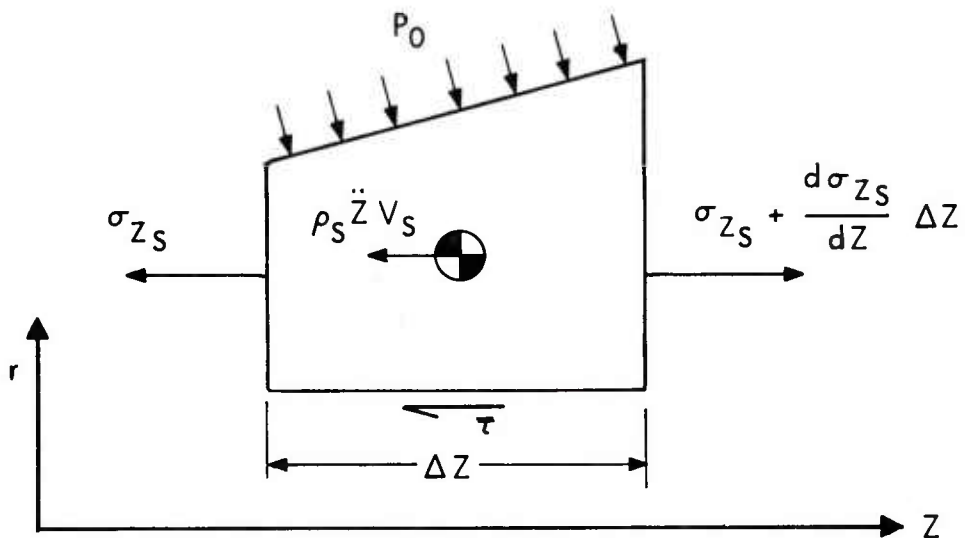
Along the interface between sabot and penetrator, the axial displacement is the same in both components (no slip condition). Hence the axial strain is equal across the interface. Utilizing Hooke's Law for elastic materials, the sabot axial stress may be found as

$$(2) \quad \sigma_{zs} = \frac{E_s}{E_p} [\sigma_{zp} - \nu_p (\sigma_r - \sigma_\theta)_p] + \nu_s (\sigma_r + \sigma_\theta)_s$$

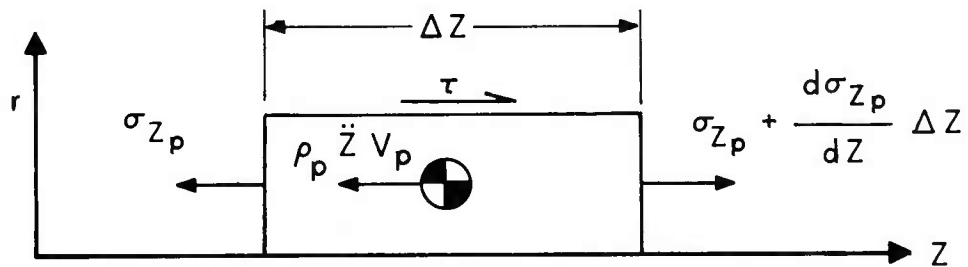
Summing forces acting on the penetrator, as shown in Figure 1b gives, after integrating

$$(3) \quad \sigma_{zp} = \left(\rho_p \ddot{Z} - \frac{2\tau}{R_p} \right) Z + \sigma_o$$

It remains to determine $(\sigma_r + \sigma_\theta)$ for the sabot and penetrator in terms of the known pressure and material properties. For the



a. Sabot Stress System



b. Penetrator Stress System

Figure 1. Free Body Diagram for Sabot and Penetrator Stresses

configuration under consideration, this would be a very difficult problem. The difficulty may be eliminated by appealing to the smallness of the taper ratio and the principle of superposition of stresses in order to consider only stresses in the $r - \theta$ plane. Then the stress state may be approximated by the Lamé solution for elastic circular cylindrical bodies of dissimilar materials acted on by external pressure. The pressure at the interface in the Lamé solution is determined by matching the radial displacements there as

$$P_i = 2R^2 P_o [(R^2 + R_p^2) + \nu_s (R^2 - R_p^2) + \frac{E_s}{E_p} (1 - \nu_p)(R^2 - R_p^2)]^{-1}$$

If this relation were used in the differential equation for R , the result would be a highly nonlinear expression due to the appearance of R^2 in the coefficients. To sidestep this difficulty, the limits of P_i as R approaches both R_p and large possible values are determined, and the numerical average used as the constant value of P_i for all values of R . This is

$$(4) \quad 2P_i = [3 + \nu_s + \frac{E_s}{E_p} (1 - \nu_p)] P_o [1 + \nu_s \frac{E_s}{E_p} (1 - \nu_p)]^{-1}$$

If equations (2), (3) and the Lamé solution for $(\sigma_r + \sigma_\theta)$, using equation (4) for interface pressure are substituted into the differential equation (1), the solution may be found by integration as

$$(5) \quad \phi = [\phi_o - \frac{2\tau}{\alpha} R_p] \left[\frac{\Sigma_o}{\Sigma_o + (\rho_p \ddot{Z} - \frac{2\tau}{R_p})Z} \right]^\beta + \frac{2\tau}{\alpha} R_p$$

$$\text{where } \Sigma_o = \sigma_o + 2\nu_p P_i + \frac{E_p}{E_s} (1 - 2\nu_s) P_o$$

$$\alpha = \frac{E_s}{E_p} (\rho_p \ddot{Z} - \frac{2\tau}{R_p}) - \rho_s \ddot{Z}$$

$$\beta = 1 - \frac{E_p}{E_s} \frac{\rho_s \ddot{Z}}{(\rho_p \ddot{Z} - \frac{2\tau}{R_p})}$$

$$\phi_o = R^2(0) - R_p^2$$

Up to the present the derivation has implicitly considered the rear taper. However, the same analysis applies to the forward taper. Since there is no propellant gas operating, the pressures are zero. In using the formula for ramp profile, the taper begins at the origin of the Z axis. Thus the slope of both front and rear tapers is positive. Care must be exercised with the sign of other quantities; however, especially the shear and acceleration, depending on the configuration under investigation.

The taper profile calculated from the formula is nonlinear. Within the accuracy of the analysis and when only the lower portion of the ramp is to be used in a compound taper, the actual R curve may often be replaced by a linear interpolation for ease of machining.

The formula as derived could certainly be improved. Ignoring the axial stresses while applying the Lamé solution is inconsistent at least. Axisymmetric stress solutions exist which improve the estimate of σ_o being constant over the penetrator cross-section. Use of axial stress values corrected for nonuniformity at the interface improve the taper predictions. However, it would be inappropriate to develop the approximate analysis too far. Its purpose is to give an initial estimate of a taper profile which permits uniform shear transfer at the sabot/penetrator interface. The predictions of the formula should be checked and improved if necessary by use of an axisymmetric, elastic, finite element code. Only by this procedure may the approximate analysis be justified.

The concept of the double-ramp as opposed to the more conventional sabot configuration may be illustrated by finite element analysis of two examples. In Figure 2 is seen a saddle-back sabot, with the pressure acting on the rear face of the sabot. To illustrate the shear load transfer qualities, no acceleration will be imposed on this example. The parameters used in the calculations are shown in Table I.

A dot is used in Figure 2 to locate each element of the sabot that has a tensile hoop stress. A segmented sabot cannot support hoop tension, so the seams must open in these regions. The areas of hoop tension form a path completely through the sabot, which would open to gas flow. Also shown is the shear stress variation at the interface. The result is very nonuniform, with a peak of nearly 344.75 MPa at the ends, falling to low values in the center. This type of variation is conducive to domino failures, with shear grooves failing sequentially along the interface.

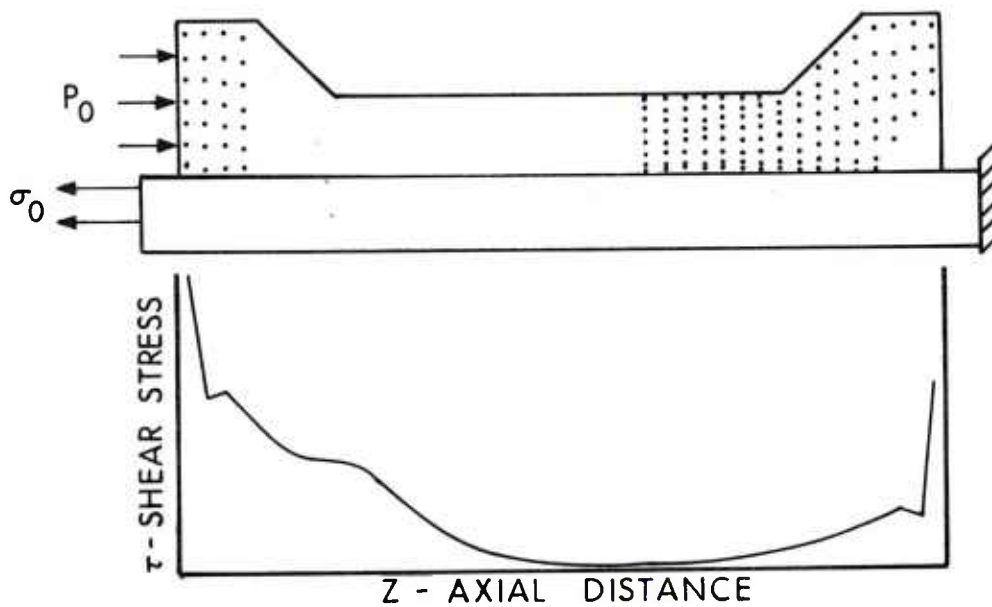


Figure 2. Conventional Saddle-Back Sabot Configuration

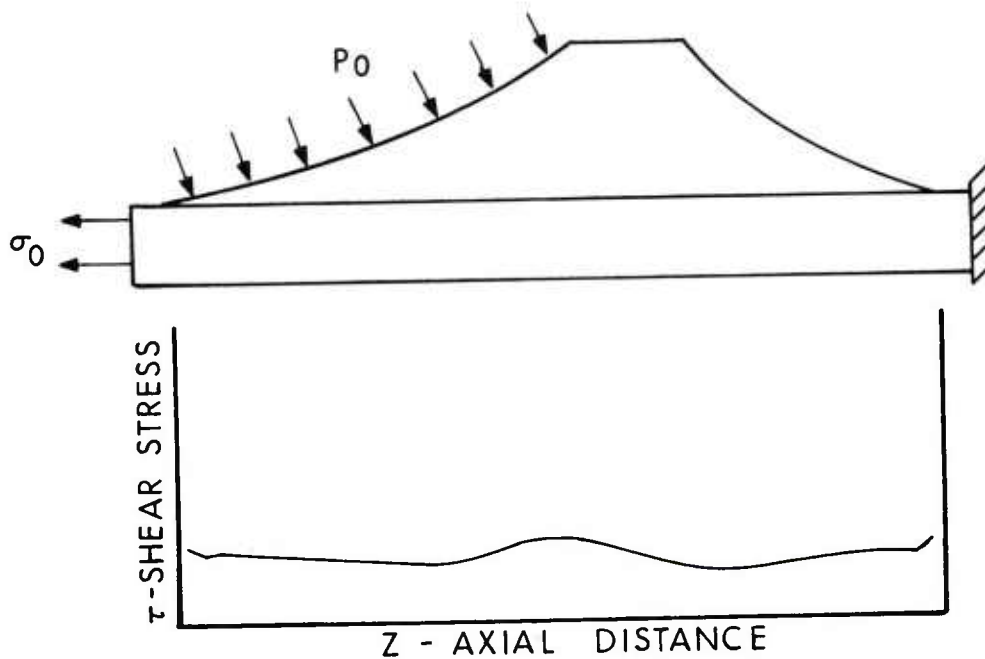


Figure 3. BRL Double-Ramp Sabot Configuration

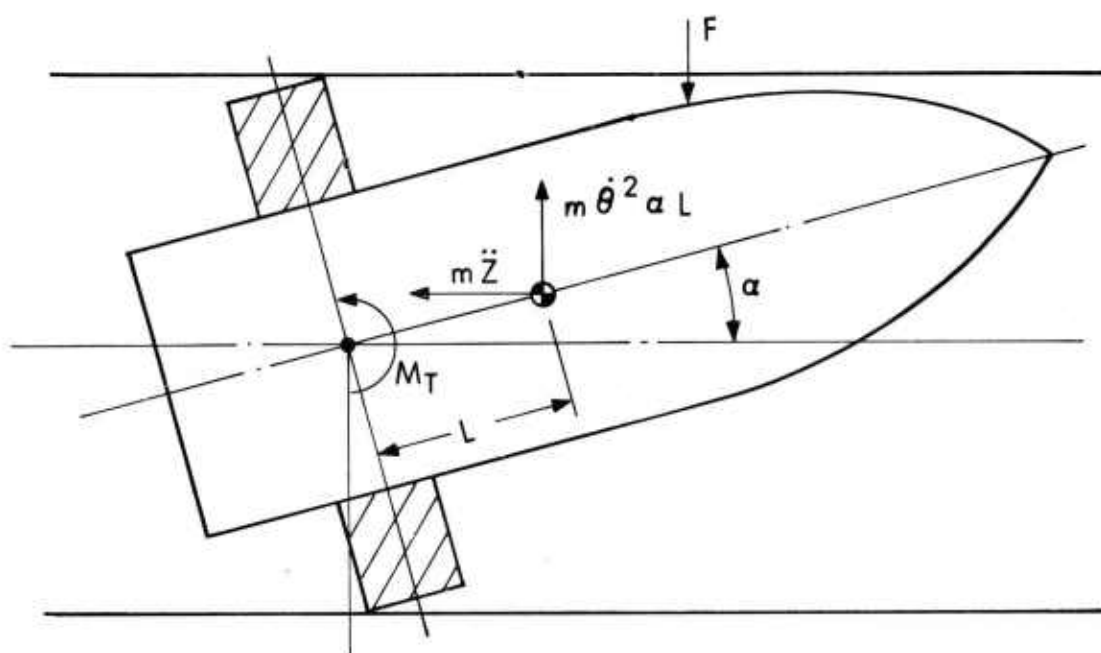
Table I. Properties of Sabot and Penetrator

$E_p = 275,800 \text{ MPa}$	$V_p = V_s = .3$
$E_s = 68,950 \text{ MPa}$	$R_p = .0254 \text{ m}$
$\ddot{Z} = 0$	$R_{\text{Bore}} = .0762 \text{ m}$
$P_o = 172.38 \text{ MPa}$	Length of Sabot = .254 m
$\sigma_o = 689.5 \text{ MPa}$	

The same material and loading parameters were used to calculate the stresses in a double-ramp sabot, shown in Figure 3. The desired shear stress at the interface was 68.95 MPa. Only two elements near the forward taper had hoop tension. The shear stress at the interface is relatively uniform, differing by less than 20.69 MPa from the projected value.

Both numerical examples were calculated with the SAAS II axisymmetric, elastic finite element code. They clearly present the advantages to be expected from a double-ramp sabot configuration with respect to self-sealing of gas pressure and uniform transfer of load from sabot to subprojectile.

In-Bore Rigid Body Dynamics. The lack of in-bore stability of a projectile has important consequences for its overall performance. It has been realized for at least 100 years that locating the rotating band at the projectile center of gravity improves in-bore stability (3). Reasons for this conclusion may be demonstrated using rigid body dynamics. Figure 4 is a conceptual diagram of a conventional projectile during its in-bore travel. It is assumed that the rotating band acts as a centering device and that, therefore, transverse yawing motion must be about an axis through the band. The center of gravity is located at a distance L ahead of the pivot axis. Clearances and eccentricities of parts cause the axis of the projectile to deviate from the centerline of the tube during in-bore travel and move the center of gravity away from the axis of the tube. The setback force, due to the acceleration of the projectile, will be a vector parallel to the gun axis, which will create an overturning moment M_t about the pivot. If the gun tube is rifled, the



$$M_T = m \ddot{Z} \alpha L + m \dot{\theta}^2 \alpha L^2$$

Figure 4. In-Bore Rigid Body Dynamic Configuration

centrifugal force due to the spin of the eccentric projectile will also produce a moment. The magnitude of the total moment is

$$M_T = MZ\ddot{\alpha}L + M\dot{\theta}^2\alpha L^2$$

and it is unstable in the sense that any yaw angle, no matter how small, generates a moment which operates to increase yaw. This unstable growth of yaw angle increases until the front bourrelet impacts the tube wall, after which the projectile will travel the remaining distance cocked at the maximum possible yaw attitude. The overturning moment is countered by the moment produced about the pivot by the force F applied to the forward bourrelet. This force is reduced by the adaptation of a longer wheelbase or distance to the front bore rider. Even so, this transverse force may be substantial. The structure supporting the front bore rider must suffer the weight penalty necessary to adequately resist the load.

As the length L approaches zero, the configuration of a centered rotating band appears. For this case, any small disturbance will not result in the unstable growth of yaw, because the moment arm is zero. This situation would normally be called neutrally stable, since as presented, there is no moment which would tend to eliminate an initial yaw. However, in a much more detailed dynamic analysis performed by L. H. Thomas (4), it was shown that, for the case $L = 0$, the gyroscopic moments due to spin would act to realign the projectile and tube axes.

To ensure that the rotating band was centered over the position of the c.g., use was made of two computer codes for calculating the rigid body properties of objects. The first is a locally derived code which finds the properties (weight, c.g., moments of inertia) of axisymmetric shapes or segments by numerical integration. The second code, known as MOMENTS II (5), can combine several basic shapes or defined shapes (such as are determined from the first code) into a total body. As any design changes affecting the taper angles, fin weight, etc. are made, the rotating band may be relocated over the center of gravity in an iterative manner. In this way the in-bore stability of the projectile is maintained.

Other Details. There is a multitude of other details that must be considered by the designer which will only be alluded to in this section. Since the double-ramp sabot is a centered configuration, the length of penetrator extending fore and aft of the sabot must be determined. When the penetrator length is known, this calculation also determines the minimum sabot length.

In order to determine the allowable stress level in a structural component in a triaxial state of stress, the effective or equivalent stress of plasticity theory is used. It is well known that this parameter correlates with plastic yield much better than maximum normal stress. The effective stress level in the free fore and aft length of the rod depends only on the acceleration forces acting on the subprojectile, since a uniform hydrostatic pressure has no effect on effective stress level. The unsupported lengths are chosen to give the desired stress immediately aft and forward of the sabot.

The above value of unsupported length must be checked for stability problems. One such check is for column buckling. Expressions for determining the critical length of a fixed-free column in a uniform acceleration field may be found in many handbooks. Likewise, the problem of the whirling of a shaft with an end mass (such as fins) will have some critical length which must not be exceeded for each value of projectile spin allowed.

Under the rear taper the ratio of interface radial stress to shear stress can be made nearly constant and less than some maximum allowable value of friction coefficient. In this area, no grooves or other load transfer devices are necessary (1). In the other portions of the sabot/penetrator interface, the resultant shear load must be carried by threads or grooves. The number of grooves per inch is determined by equating the product of allowable bearing stress and bearing area with the shear force per inch of interface to be transmitted. To obtain the highest possible groove strength in the dissimilar sabot and penetrator materials, equal shear area grooves should not be used. The areas of the root of a tooth in the sabot and rod is ratioed inversely to the ultimate strengths of the two materials so that both components may carry the same load.

After the initial design calculations have been performed, the candidate design and all further iterations should be thoroughly investigated with a reliable stress analysis code. Ideally a fully three-dimensional dynamic analysis simulating the interior ballistic process should be performed. We have settled for an axisymmetric, quasi-static analysis at peak pressure and acceleration.

One of the primary results of an analysis as described is a map of effective stress in the projectile. In the ATGTI program, this effective stress was required to be below the plastic yield stress of the respective material. It may be appreciated that this is an overly conservative requirement, since plastic flow does not of itself mean failure in a single use component. The adaption of a more sophisticated

failure criteria would be a very fruitful modification of the current design procedure.

Test Results and Discussion. Two separate models of the BRL double-ramp sabot were designed as part of the ATTGI program. A high-pressure version capable of operating at maximum chamber pressures of 730.87 MPa was initially tested, but not pursued. A low-pressure version operating at standard 105mm tank gun pressures was also designed and tested more extensively.

A sketch of the low-pressure BRL double-ramp projectile is shown in Figure 5. It has been successfully fired during high temperature tests to a peak chamber pressure of 572.29 MPa. Three of the early versions of this projectile suffered rotating band failures, which are believed to have been caused by excessive interference due to tolerance buildup. Even though the sabots suffered severe gas wash due to loss of the band, there was no gas leakage between sabot petals or at the sabot/penetrator interface. No rotating band failures have occurred since this band was redimensioned (28 rounds fired).

Included within the ATTGI program was the development of an advanced conventional saddle-back sabot. Effectively the same subprojectile was used in both this sabot and the BRL double-ramp, so comparisons of performance are easily made. For the full in-bore 105mm projectile, the BRL double-ramp weighed slightly over .4536 Kg less than the conventional sabot. This translates into a muzzle velocity increase of 57.3 m/s for the subprojectile. The initial yaw of the free flight subprojectile was measured from yaw cards approximately 40 m from the muzzle. These measurements showed nearly the same average value of yaw for the BRL double-ramp and the conventional saddle-back sabot. Likewise, the target dispersion at 1,000 meters, based on very small samples, was essentially equal.

It had been anticipated that the centered rotating band of the BRL double-ramp sabot would result in substantial reductions in initial yaw and, consequently, target dispersion, over the conventional configuration. Instead the values of these parameters are nearly equal for the two sabots. Development of the double-ramp, by the lengthening of the wheelbase and/or the use of full projectile spin, will continue in the future and should reveal the advantages in reduced yaw and target dispersion of which the stable in-bore configuration is capable.

The BRL double-ramp configuration has been demonstrated to be a very effective sabot for launching kinetic energy penetrators. However, the concepts involved are applicable whenever it is desired to obtain higher velocity or increased range by use of a sabot.

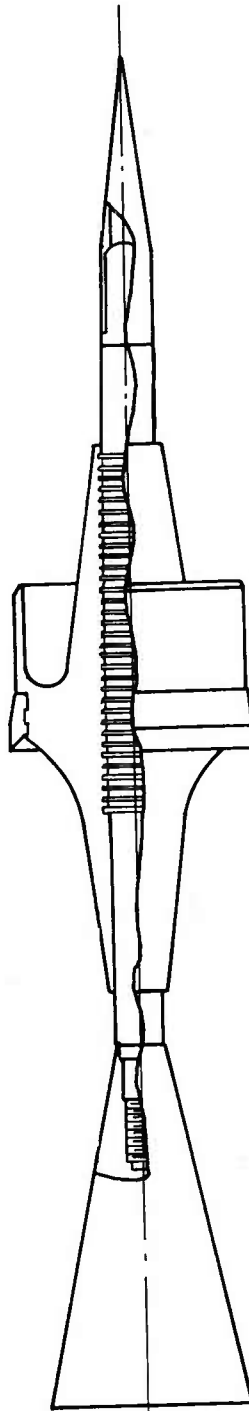


Figure 5. BRL Double-Ramp Sabot as Tested in 105mm ATTGI Program

References

- (1) "Development of a Special Type Small Arms Cartridge (Sabot Supported)" (U) (Confidential Report); Aircraft Armaments, Inc., Report No. ER-1414, July 1958, AD 300912.
- (2) Chun, Y. W. and Haug, E. J., "Two Dimensional Shape Optimal Design," Technical Report No. 38, Division of Materials Engineering, Univ. of Iowa, January 1978.
- (3) Breger, Captain M. P., "Position and Form of Bands for Projectiles," Memoires Militaires et Scientifiques, publies par le Department de la Marine, translated by Lt. C. C. Morrison, Ordnance Dept., U.S.A., in "Notes on the Construction of Ordnance, No. 27," Washington, D. C., June 1884.
- (4) Thomas, L. H., "The Motion of the Axis of a Spinning Shell Inside the Bore of a Gun," Report No. 544, USA Ballistic Research Laboratories, Aberdeen Proving Ground, MD, May 1945, AD PB22102.
- (5) Lacher, E. B., "Moments II - A Computer Program to Calculate Moments and Products of Inertia of Asymmetric Objects," Technical Report No. 4945, Picatinny Arsenal, NJ, May 1976.

ELECTRICAL CONDUCTIVITY OF SELECTED GRAPHITE INTERCALATION
COMPOUNDS IN THE RANGE $4K < T < 300K$

RUSSELL EATON, III, Ph.D.

*W. DAVID LEE, Ph.D.

U.S. ARMY MOBILITY EQUIPMENT RESEARCH AND DEVELOPMENT COMMAND
FORT BELVOIR, VIRGINIA 22060

INTRODUCTION

The Army's requirements for electric power in the field have increased dramatically in recent years as a result of increased use of special weapons, communications equipment, sensors, and other electric and electronic devices. The supply of the necessary power to the user is thus a matter of considerable importance to the Commander. At present the only systems available for transferring electrical power from source to user are wire and cable type systems. Rapid deployment and re-deployment capabilities, ease of handling and flexibility are essential characteristics of these systems for use in today's highly mobile tactical environment. The weight of presently available cable materials is a barrier to achieving these characteristics, especially for high power systems. Recent investigations have revealed a class of compounds, known as intercalated graphite, which appear promising as new conductor materials. Intercalation compounds are compounds of lamellar host materials which have atoms or molecules of a foreign chemical species inserted between layers. Graphite intercalation compounds are presently of interest in two major areas: in selective catalysis and as synthetic metals, especially for high infrared reflectance and high electrical conductivity. In particular, it has been shown that some of these materials exhibit higher electrical conductivity and have lower density than copper. These materials are of obvious interest to Army power programs. MERADCOM has initiated a cooperative effort with the University of Pennsylvania to investigate intercalated graphite compounds for possible application as an electrical conductor. MERADCOM's in-house program which supports this work is an investigation of the electrical conductivity of graphite and selected graphite compounds over the temperature range between

4 and 300 K. The subject of this paper is an initial investigation of graphite fibers.

BACKGROUND

A detailed knowledge of graphite itself is necessary for any investigation of graphite intercalation compounds. Consequently, some of the properties of graphite will be reviewed briefly along with a discussion of recent investigations of graphite intercalation compounds. Graphite crystallizes into a layered structure consisting of planes of carbon atoms arranged in a hexagonal pattern. A model of the ideal crystal structure of graphite is shown in Figure 1. At room temperature the separation distance between nearest neighbor atoms in the basal planes is 1.415Å, and successive basal planes are separated by 3.354Å(1). The difference in separation distances is due to the two different types of bonds found in the graphite structure. Carbon atoms in the basal planes are linked together by covalent bonds; adjacent layers are weakly bonded to each other by van der Waals forces. The approximate two orders of magnitude difference between the two different bonds is reflected in its anisotropic physical properties, e.g., the measured value of Young's modulus which is 10^{13} dyne/cm² along the a-axis is thirty times greater than the c-axis value. Most graphite crystals have a layer stacking sequence ABAB....

It has been established that graphite is a semi-metal in which the highest filled valence band overlaps the lowest empty conduction band by about 35meV. In contrast to a semiconductor, a semi-metal such as graphite will have empty valence states (holes) and filled conduction states (electrons) even at temperatures approaching absolute zero. Pure single crystal graphite is a moderately good electrical conductor with a resistivity parallel to the basal planes about twenty times that of pure copper at room temperature. (See Table I which is a chart showing comparisons between significant properties of graphite and a good metallic conductor, copper.). The relatively high conductivity of graphite is due to a mobility which is greater than any metal (10^5 cm²/V·sec), in spite of a relatively low carrier concentration ($n = p = 10^{19}$ cm⁻³). The high mobility is a consequence of the bond structure of the hexagonal graphite rings. Delocalized π electrons (molecular orbitals) are free to follow the carbon chains throughout an entire layer leading to enhanced electrical conductivity within the graphite layers. Measurements of c axis conductivity on natural crystals have resulted in a conductivity ratio σ_a/σ_c from 10^2 to 10^5 at ambient temperature; measurements on synthetic materials give values near 5×10^3 . The large spread in the c axis conductivity measurements is apparently due to the difficulty of preparing samples which are free from defects, especially fissure-like

defects parallel to the a axis. Many early uncertainties observed in the electrical properties arising from varied sources and types of graphite available, as well as sample history, have been resolved with the development of techniques for stress-annealing pyrolytic graphite. This process provides near-ideal material which is comparable to natural crystals(2)(3).

It has been known since the 1860s that certain elements and compounds cause graphite to swell and increase in weight. Not until the 1930s were systematic studies of this effect undertaken. It is now well established that this effect is due to the formation of graphite intercalation compounds. Studies(4)(5) have shown that while the distance between the carbon layers is increased due to the intercalated atoms or molecules and the stacking order of the carbon layers is changed, there is almost no change in the in-plane separation of carbon atoms. Graphite intercalation compounds exist in various "stages" of intercalation. "Stage" is defined as the number of carbon layers separating intercalant layers. In general, graphite intercalation compounds possess a high degree of order with definite crystal structures and chemical compositions.

At the present time, at least 60 atomic and molecular species have been intercalated into graphite. Two classes of intercalation compounds have been identified. Intercalation with Group I elements from the Periodic Table (and some other metals) produces donor compounds(6)(7) while acceptor compounds are formed by intercalation with bromine(8), bisulfate(9), and strong acids(10). Many of the resulting intercalation compounds are metallic, or semimetallic in their behavior as the intercalant modifies the semimetallic electronic properties of the graphite.

Many intercalation compounds show an enhanced electrical conductivity along the a -axis (σ_a). The magnitude of this effect is controlled by the intercalation species and the stage of the resulting compound. The mechanism for the increased conductivity appears to be due to the transfer of fractional negative charge from the intercalated species to the carbon layers in a donor compound or from the carbon layers to the intercalated species in an acceptor compound. At the same time, the mobility, compared to graphite, is reduced due to enhanced electron-phonon scattering associated with the enlarged Fermi surface or an increase in the effective mass associated with changes in band structure or both of the preceding effects. In any event, the decrease in mobility is more than offset by the increase in carrier concentration, and the net effect is an increase in σ_a due to intercalation.

The effect of intercalation upon the c-axis conductivity (σ_c) is more complicated and illustrates a fundamental difference between donor and acceptor compounds. For donors, σ_c increases, and in some cases σ_c increases more than σ_a , e.g., for the first stage potassium compound KC_8 , $\sigma_a/\sigma_c = 34$, two orders of magnitude smaller than pure graphite. The essentially three dimensional behavior of the electrical conductivity in donor compounds is due primarily to the strong interaction between the electron clouds of the graphite structure and the valence electrons of the intercalants. For acceptors, σ_c drops significantly, and σ_a/σ_c increases strongly, by as much as 300 for arsenic pentafluoride (AsF_5) relative to pure graphite. It is concluded that acceptor compounds exhibit a two dimensional character due to the decreased attraction between carbon layers and weak interaction between intercalant and the graphite structure. This conclusion is supported by observed tendency of acceptor compounds to exfoliate during intercalation.

The significant increases in σ_a of acceptor compounds make them promising candidates for lightweight electrical conductors. Recent investigations of graphite intercalated with strong acid fluorides, antimony pentafluoride (SbF_5) and arsenic pentafluoride (AsF_5), have produced samples with electrical conductivities exceeding pure copper at room temperature. The first experiment demonstrating the high conductivity of acid fluorides was performed by Vogel(11) et al. on composite wires consisting of a graphite powder core intercalated with SbF_5 in a copper sheath. Results were obtained from a standard four terminal dc resistance method on samples 0.1cm dia. by 10cm long. The resistivity of the core was found to be $1.0 \times 10^{-6} \Omega \cdot cm$ or approximately two-thirds that of pure copper. Subsequent investigations on highly oriented pyrolytic graphite (HOPG) crystals intercalated with SbF_5 by Thompson(12) et al. and Fuzellier(13) et al. yielded resistivities five times greater than the composite wire results. These results are difficult to reconcile because the preparation of the samples was significantly different. The composite wires were fabricated in open air by compacting graphite- SbF_5 mixtures in copper tubes, heating overnight at $105^\circ C$, swaging the tubes to a fraction of their diameters, and then annealing. The HOPG samples were made with distilled materials and under moisture- and impurity-free conditions. In both cases, the SbF_5 conductivity results were obtained from dc resistance measurements, and are suspect due to the non-uniform distribution of current in the samples caused by the extremely high anisotropy of the acid fluoride samples. Zeller(14) et al. have developed a contactless radio frequency (r.f.) technique for measuring electrical conductivity in samples with large anisotropy. A sample is inserted into an induction coil with the c-axis parallel to the magnetic field. Eddy currents induced in the sample

change the coil's inductance; this change causes a shift in the resonant frequency of an LC oscillator circuit. The conductivity of the sample is determined from a semi-empirical formula based on the resonant frequency shift. With the r.f. technique, Falardeau(15) et al. have measured conductivities (σ_a) exceeding silver at room temperature for HOPG crystals intercalated with AsF_5 . The contactless r.f. technique eliminates the problem of non-uniform current density which can occur in four terminal dc measurements and is therefore well suited for measuring conductivities of highly anisotropic graphite compounds.

The investigation of intercalated graphite compounds has recently been extended to graphite fibers. These fibers are of considerable technological interest, and highly conducting fibers have obvious electrical applications. Vogel and his associates have recently produced graphite fibers intercalated with AsF_5 with room temperature resistivities from $200\text{--}500\mu\Omega\cdot\text{cm}$. The low temperature behavior of the electrical conductivity of the fiber samples is being investigated at MERADCOM. The temperature dependence of the electrical conductivity provides fundamental information about the mechanisms contributing to the overall conductivity.

EXPERIMENTAL DETAILS

The experimental apparatus shown in Figure 2 consists of a liquid helium cryostat with a variable temperature insert. The system permits the sample temperature to be controlled over the range between liquid helium temperature and room temperature. The graphite samples are mounted (Figure 3) on a sample block of OFHC copper which aids in maintaining uniform sample temperature; this sample holder also provides a heat sink for instrumentation leads and a mounting point adjacent to the sample for the temperature sensor. The sample temperature is monitored using a diode with a digital readout. Electrical conductivity measurements are made using a four terminal dc technique. A constant current source supplies sample currents of 10 microamperes, and the voltage drop across the sample is measured with a nanovoltmeter. Power dissipation is held to less than a microwatt to minimize spurious effects caused by self-heating of the sample. In a typical experiment, the sample is cooled to the lowest temperature and allowed to warm quasi-statically to room temperature. For convenience in taking data, the analog outputs of the nanovoltmeter and temperature monitor are connected to an X-Y recorder to produce a continuous record of each run. The four terminal dc measurement technique requires some care in its application and interpretation since anisotropic characteristics of graphite can lead to non-uniform current density in the sample. Graphite fiber samples which have length to

thickness ratios greater than 1000 were used to assure uniform current distribution; individual filaments used in this experiment are nominally 8μ dia x 1 cm long. The work reported here was performed on Celanese GY-70 fibers with their high conductivity direction (a axis) oriented parallel to the fiber axis. Individual fibers are mounted on alundum wafers with platinum contacts baked on. Secure joints are made by soldering the fiber to these contacts with 80Au-20Sn solder paste and small pieces of platinum wire. The fibers were intercalated by immersing them in the intercalant, in the case of HNO_3 , or suspending them in AsF_5 vapor, under controlled temperature conditions. Observation of the fibers before and after intercalation with a microscope revealed that the fiber thicknesses were nominally unchanged by the intercalation process.

EXPERIMENTAL RESULTS

The reliability of the experimental apparatus was determined by measuring the temperature dependence of the resistivity of a piece of #36 AWG copper wire over the range 4.2-300K. The experimental data (open circles) are plotted in Figure 4 along with the Block-Grüneisen function (solid line) for simple metals. The semi-empirical Block-Grüneisen relation describes the ideal temperature dependence of the lattice or phonon contribution to the electrical resistivity at all temperatures. The relation gives the resistivity, ρ , $\propto T^5$ for $T < \Theta$ (Debye Temp.) and $\rho \propto T$ for $T > \Theta$, as required by theory. The agreement between the data and the theory is quite good indicating that the experimental apparatus provides reliable data. The systematic departure of the experimental data from the ideal theory is due to additional resistivity contributions, including faults due to annealing, impurities, and size effects especially at low temperatures, not accounted for in the theory.

The recent investigations of intercalated graphite fibers at MERADCOM have revealed two unexpected and not yet reported results in the temperature dependence of the electrical resistivity. First, for both pristine graphite fibers and the same fibers intercalated with either HNO_3 or AsF_5 , the resistivity increases monotonically with temperature from room temperature down to 4.2K, the boiling point of liquid helium at atmospheric pressure (low temperature limit of this investigation). In contrast, the basal plane resistivity(16) of natural and synthetic crystalline graphite decreases monotonically over the same temperature range. (This behavior of pristine HOPG graphite crystals has also been verified at MERADCOM as part of the experimental apparatus checkout.) Second, the dependence of fiber resistivity upon temperature is essentially linear, especially for the intercalated fiber samples.

The preceding results are clearly illustrated by plotting reduced resistivity ($\rho(T)/\rho(300K)$) data versus absolute temperature (K) for a pristine fiber and for the same fiber after intercalation. Figure 5 is such a plot for a fiber (F-1) intercalated with HNO_3 . Intercalation with HNO_3 was performed at MERADCOM with fibers supplied by the University of Pennsylvania. The resistivity of the fiber increased almost linearly to 140% of its room temperature value of $480\mu\Omega\cdot\text{cm}$ upon cooling to 4.2K. Intercalation with HNO_3 reduced the room temperature resistivity to $102\mu\Omega\cdot\text{cm}$ (almost a factor of 5 from the unintercalated value). The intercalated fiber's resistivity had a small, approximately linear temperature dependence (2% increase over the full temperature range). Several successive cyclings between 4.2K and room temperature reproduced the temperature dependence of the resistivity shown in Figure 5; changes in the absolute values of the resistivity were negligible indicating little de-intercalation of HNO_3 from the fiber.

Data obtained from another sample, Fig. 6, prepared from the same fiber and intercalated with HNO_3 confirmed the results obtained from sample F-1. Sample F-2 had nominally the same room temperature resistivity ($483\mu\Omega\cdot\text{cm}$) as F-1. The resistivity increased monotonically as the temperature was decreased to 4.2K; the resistivity at 4.2K was 136% larger than the room temperature value. Upon intercalation, the room temperature resistivity was reduced to $189\mu\Omega\cdot\text{cm}$ or approximately 40% of the unintercalated value. The resistivity of sample F-2 had a small, approximately linear temperature dependence (9% increase over the full temperature range), larger than the temperature dependence of sample F-1. Repeated temperature cyclings again verified the temperature dependence of the resistivity of the intercalated fiber with negligible effects due to de-intercalation.

Intercalation with AsF_5 produced the results shown in Figure 7. Samples intercalated with AsF_5 were supplied by the University of Pennsylvania, and consequently it was not convenient to measure resistivity before and after intercalating. The unintercalated curve is representative of data taken from other Celanese fiber samples at MERADCOM and is supplied only for qualitative comparison. The room temperature resistivity of the intercalated fiber was $241\mu\Omega\cdot\text{cm}$. The important result obtained by the AsF_5 intercalated fiber is that the temperature dependence of the resistivity was again linear over the range from 4K to room temperature. The magnitude of the change with resistivity of the AsF_5 intercalated fiber was larger (about 30%) than the magnitude of the change in the HNO_3 intercalated fibers (less than 10%). Again repeated cyclings between 4K and room temperature resulted in little change in the temperature dependence of the resistivity although de-intercalation caused measurable increases in the absolute resistivity values.

The preceeding results can be understood qualitatively in terms of a size dependence argument. When one or more dimensions of a sample becomes sufficiently small as in the case of a thin film or wire, collisions of the charge carriers with surfaces decrease the effective mean free path (mfp) of the sample. The effective mfp by Matthiessen's rule is the sum of independent components and is written as

$$\frac{1}{l_{\text{eff}}} = \frac{1}{l_{\text{ph}}(T)} + \frac{1}{l_{\text{def}}} \quad (1)$$

where $l_{\text{ph}}(T)$ is the temperature dependent m_{fp} due to scattering of electrons by thermal vibrations (phonons) of the lattice and l_{def} is the temperature independent mfp due to scattering from defects. The term due to defects may be written as a sum of terms due to scattering from surfaces l_{sur} and scattering from impurities l_{imp} . It will be assumed that contribution due to impurities is negligible compared to the surface contribution. For sufficiently thin samples or at low temperatures where l_{ph} is very large due to negligible scattering from phonons, l_{eff} is approximately equal to l_{sur} . Microscopic examination of individual graphite fibers reveals that they are composed of large numbers of tangled fibriles whose thicknesses are probably no more than 1μ or about one-tenth the fiber thickness. If the effective m_{fp} is limited by scattering at the fibrile surfaces or $l_{\text{ph}} \gg l_{\text{sur}}$, then the effective mean free path for a graphite fiber is equal to l_{sur} . The consequences of a size limited l_{eff} upon the fiber resistivity can be seen by writing the resistivity as:

$$\rho(T) = \frac{1}{n(T)e\bar{\mu}(T)} \quad (2)$$

where n is the temperature dependent carrier density, e the electronic charge and $\bar{\mu}(T)$ the temperature dependent mean mobility for both electrons and holes. The mean mobility can be represented approximately by a combination of relaxation processes as

$$\mu(T) = \frac{m^* v_F}{e l_{\text{eff}}(T)} \quad (3)$$

where m^* is the effective mass and v_F the Fermi velocity. Replacing l_{eff} by l_{sur} (for fibers) in Eq. (3), the resistivity can be written as

$$\rho(T) = \frac{m^* v_F}{n(T)e^2 l_{\text{sur}}} \quad (4)$$

In this simplified model, the temperature dependence of the resistivity is due only to the change in carrier density with temperature. For crystalline graphite, $n(T)$ is found to decrease with temperature. This fact implies that $\rho(T)$ will increase as the temperature decreases for this model. The resistivity data for the unintercalated fibers exhibits this behavior and consequently supports this model. The data for the intercalated fibers does not present a uniform trend. The resistivity of the fibers intercalated with AsF_5 exhibit considerably greater temperature dependence, than the fibers intercalated with HNO_3 . The lack of temperature dependence of the resistivity can be accounted for in the size dependence model with a temperature independent carrier density. This is consistent with results obtained by Zeller(17) which supports a temperature independent carrier density in graphite intercalated with HNO_3 . The experimental data for the fiber intercalated with AsF_5 shows that the resistivity does depend upon temperature implying a temperature dependent carrier density based upon the size dependent model.

CONCLUSION

The investigation of intercalated graphite fibers has produced two interesting results, viz., the resistivity is a monotonically and approximately linearly decreasing function of the temperature for both pristine and HNO_3 - and AsF_5 -intercalated graphite fibers. These results have been interpreted in terms of a size dependence model in which the temperature dependence of the resistivity is essentially due to the carrier concentration. Further work which will include such additional parameters as the magnetoresistance are being undertaken to define more completely the mechanisms involved in the electrical transport properties. The results of these investigations will be used to further clarify the role of size effects and carrier concentration on the electrical conductivity of graphite fibers. Finally, to illustrate the possible technological application of these highly conductive intercalated graphite materials to the Army requirement for lightweight electrical conductors, Figure 8 shows a section of wire prepared by the group at the University of Pennsylvania. This conductor consists of a copper jacket with a core of SbF_5 -intercalated graphite; it was prepared by forming the intercalation compound in a sealed copper cylinder and swaging the cylinder to the desired diameter. The resistivity of the core was found to be approximately $1.0 \mu\Omega\cdot\text{cm}$ which corresponds to a conductivity about 70% greater than pure copper.

REFERENCES

- (1) A. R. Ubbelohde and T. A. Lewis, Graphite and Its Crystal Compounds, (Oxford University Press, London, 1960).
- (2) A. W. Moore, A. R. Ubbelohde, and D. A. Young, Proc. Roy. Soc. A280, 153(1964).
- (3) I. L. Spain, A. R. Ubbelohde, and D. A. Young, Phil. Trans. Roy. Soc. (London) A262, 1128(1967).
- (4) E. D. Nixon, G. S. Parry. J. Phys. C2, 1732(1969).
- (5) W. Rüdorff and E. Schulze, Z. Anorg. Chem. 277, 156(1954).
- (6) A. R. Ubbelohde, Proc. Roy. Soc. (London) A309, 297(1969).
- (7) A. R. Ubbelohde, Proc. Roy. Soc. (London) A327, 289(1972).
- (8) F.C.F. Blackmond, J. F. Mathews, and A. R. Ubbelohde, Proc. Roy. Soc., A258(1960).
- (9) A. R. Ubbelohde, Proc. Roy. Soc. A321, 445(1971).
- (10) F. L. Vogel, Bull. Am. Phys. Soc. 21, 262(1976).
- (11) F. L. Vogel, J. Mat. Sci. 12, 982(1977).
- (12) T. E. Thompson, E. R. Falardeau, and L. R. Hanlon, Carbon 15, 39(1977).
- (13) H. Fuzellier, J. Melin, and A. Herold Carbon 15, 45(1977).
- (14) C. Zeller, A. Denenstein, and G.M.T. Foley, submitted to Rev. Sci. Inst.
- (15) E. R. Falardeau, G.M.T. Foley, C. Zeller, and F. L. Vogel, J.C.S. Chem. Comm., 1977, 389.
- (16) I. L. Spain, Chemistry and Physics of Carbon, edited by P. L. Walker and P. A. Thrower (M. Dekker, inc., New York, 1973) Vol. 8, p. 1.
- (17) G.M.T. Foley and C. Zeller, private communication (1978).

COMPARISON OF GRAPHITE PROPERTIES
(T = 300 K)

PROPERTY	GRAPHITE (BASAL PLANE)	COPPER
MASS DENSITY (g/cc)	2.25	8.90
RESISTIVITY ($\Omega \cdot \text{cm}$)	40.0×10^{-6}	1.67×10^{-6}
CARRIER DENSITY (#/cc)	10^{19}	10^{23}
MOBILITY ($\text{cm}^2/\text{V} \cdot \text{SEC}$)	7×10^5	35

TABLE 1

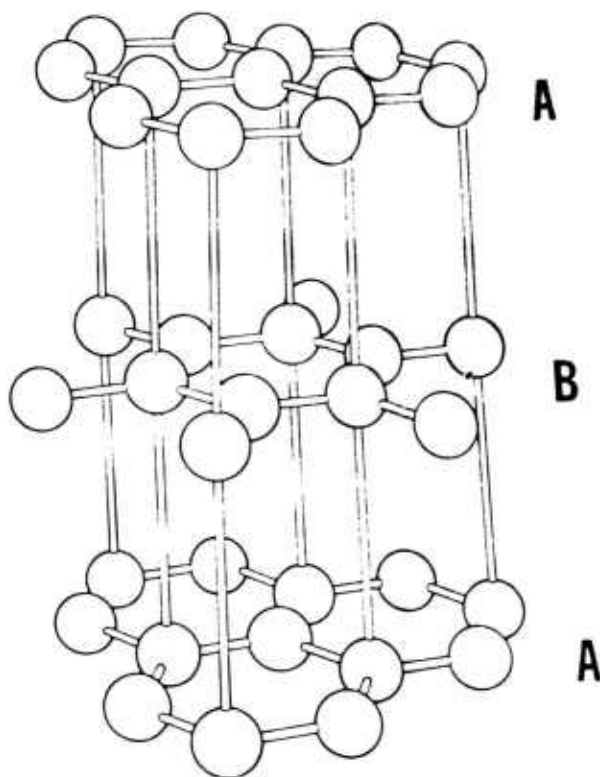


Figure 1. The crystal structure of graphite, showing the stacking sequence. ABAB....

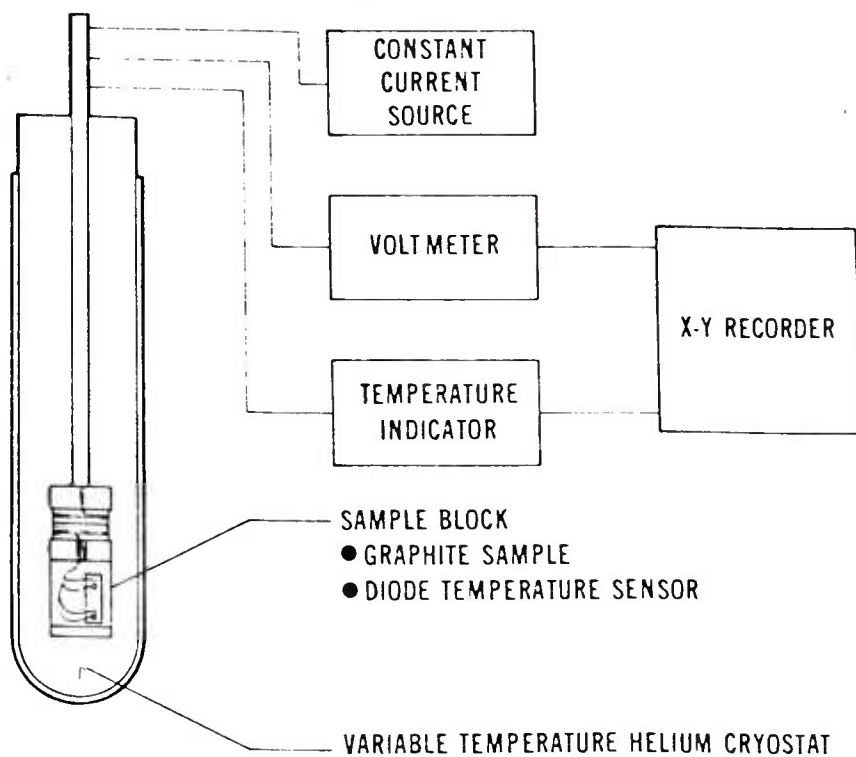


Figure 2. Schematic drawing of the experimental apparatus.

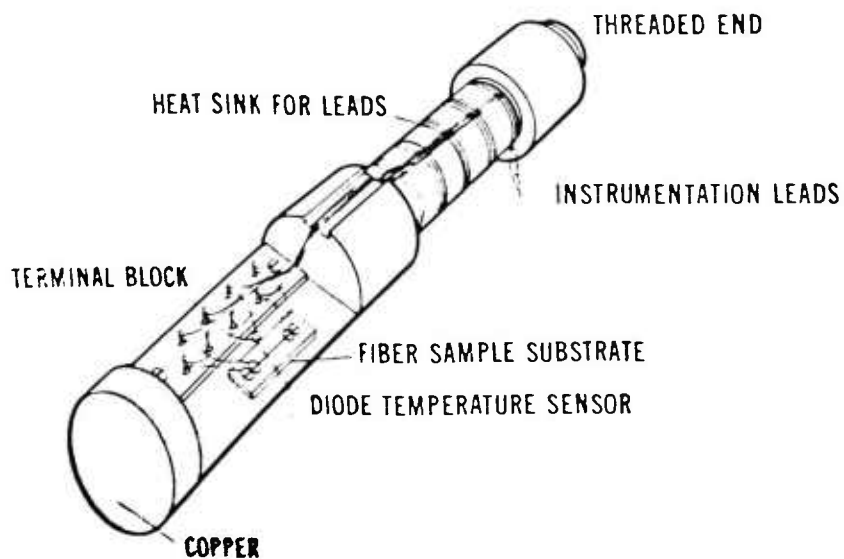


Figure 3. Artist's rendition of the sample block.

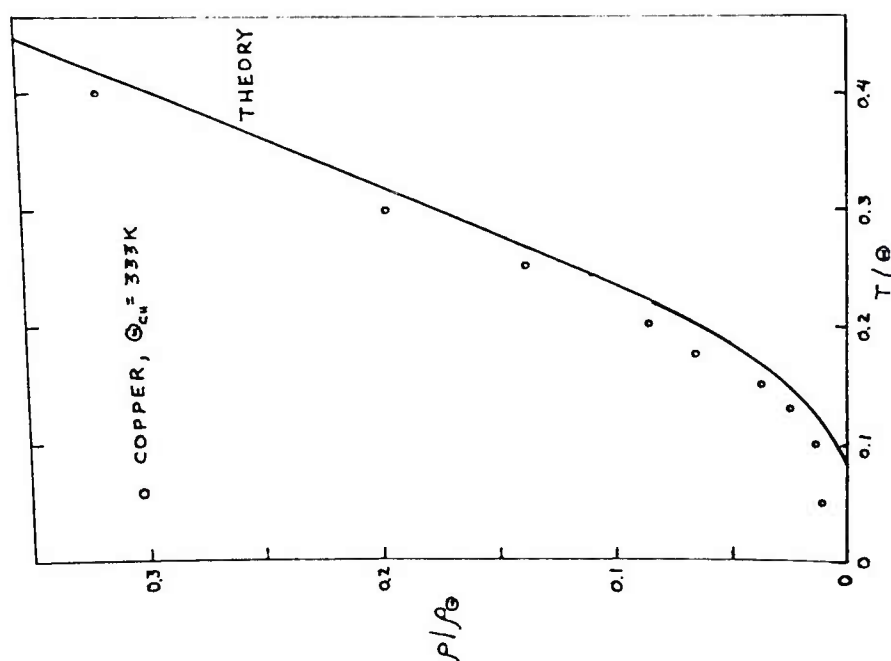


Figure 4. Resistivity versus temperature of a copper wire.

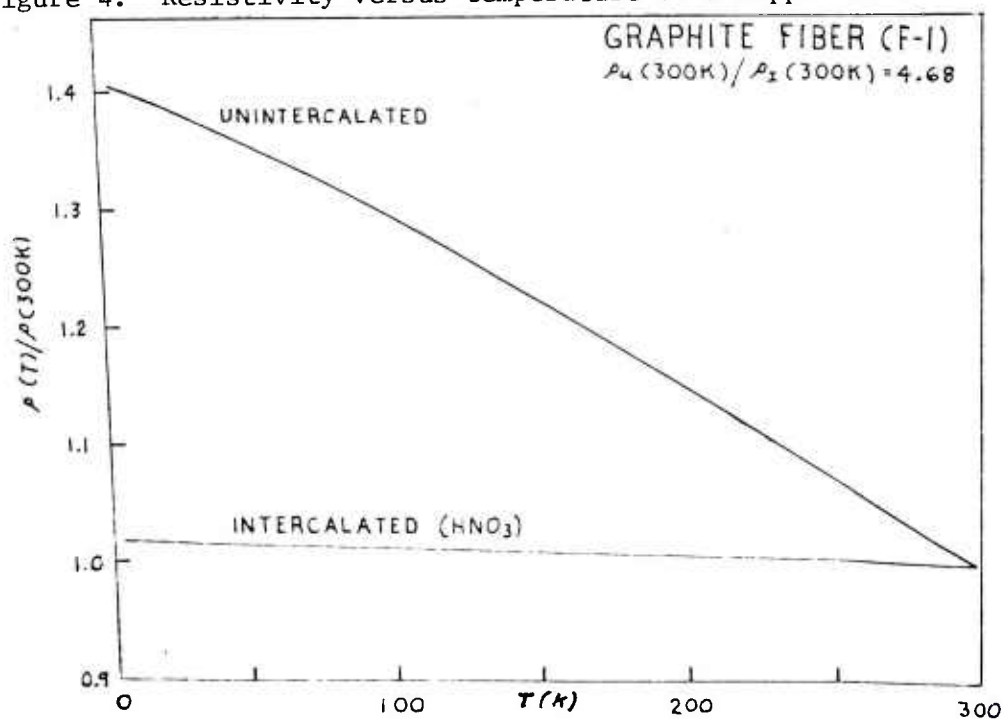


Figure 5. Resistivity versus temperature of graphite fiber (F-1).

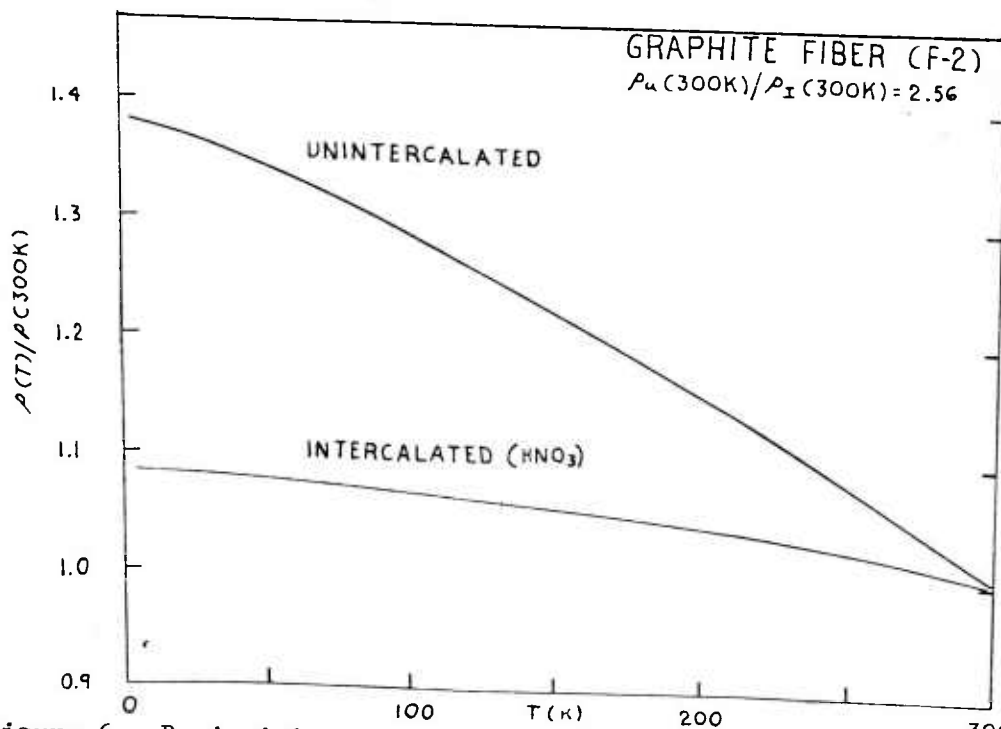


Figure 6. Resistivity versus temperature of graphite fiber (F-2).

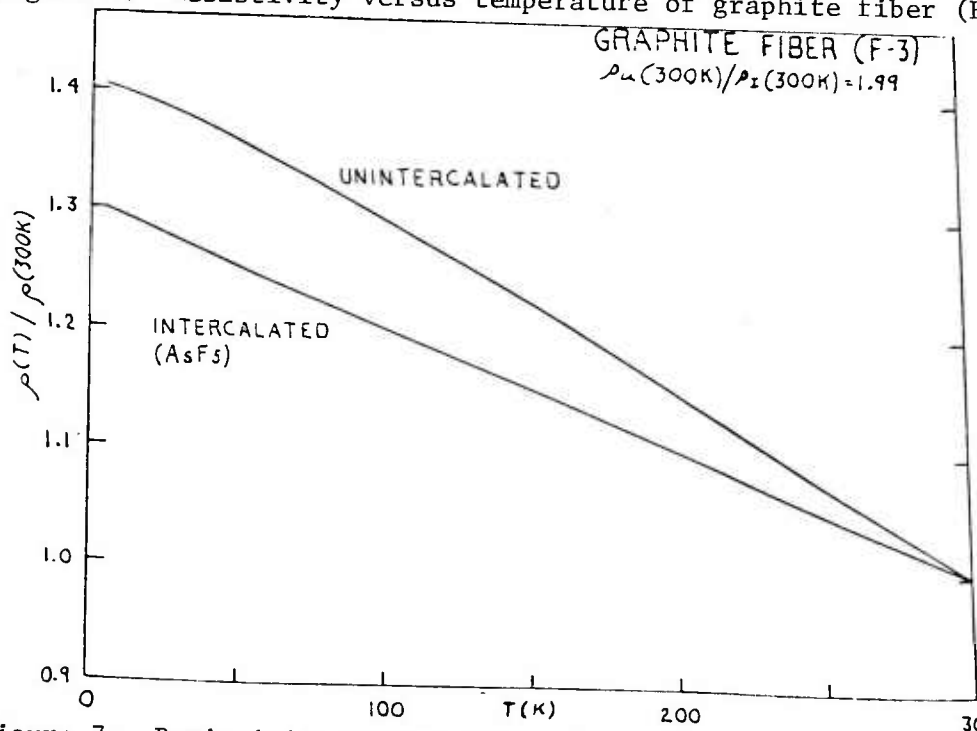


Figure 7. Resistivity versus temperature of graphite fiber (F-3).

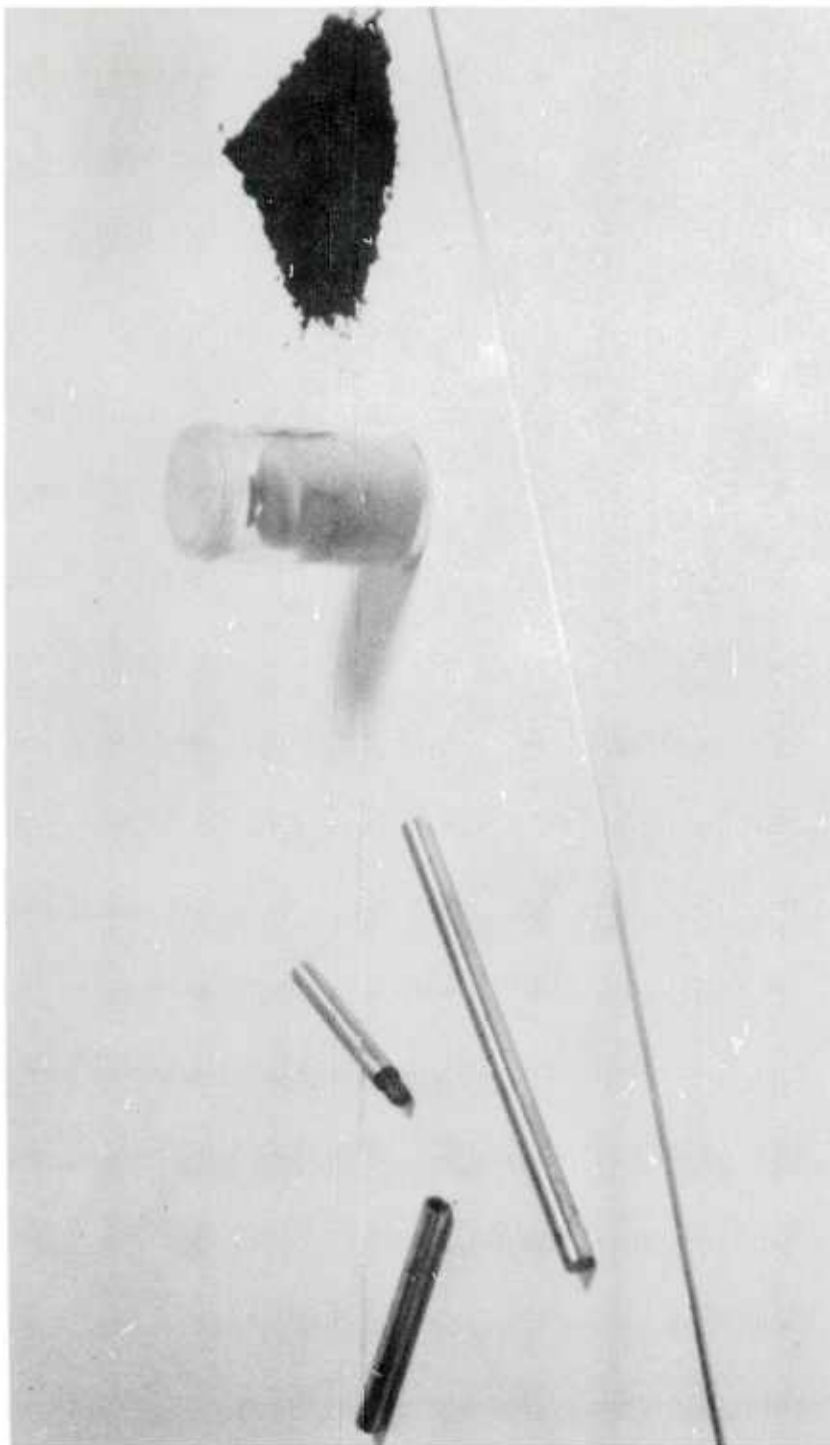


Figure 8. Photograph showing wire swaged from copper cylinder with an intercalated graphite core.

HIGHLY SURVIVABLE TRUSS TYPE TAIL BOOM

THOMAS F. ERLINE, MR.
BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MARYLAND 21005

INTRODUCTION. Successful completion of Army helicopter missions in future battle scenarios may well depend upon survival of the structure after battle damage. Survivability of a helicopter will depend significantly upon the structure's ability to retain structural integrity. The principle purpose of this study is to develop a structural concept which assures a high degree of confidence in the integrity of a structure that has received combat damage. This study has been pursued because the Army needs to meet and provide a solution to the ever escalating high explosive anti-aircraft threat to the helicopter tail boom.

The highly lethal 23mm high explosive projectile represents an existing widely deployed threat to Army helicopters. The more lethal 30mm high explosive projectile appears on the horizon as the potential future threat. Because of this potentially severe 30mm threat and the possible inadequacy of present semimonocoque designs to survive hits by the 30mm, the present study was initiated to develop a structural challenge to the 30mm and to improve upon the present designs.

The tail boom of a helicopter (for example, the present AH-1 and UH-1 models) presents a significant amount of vulnerable area, and due to the flight loads of the tail rotor and elevators, the tail boom is constantly in some stressed condition. The semimonocoque tail boom construction configuration consists of skins, longerons, stringers, and bulkheads. Four longerons provide the main bending support for the tail boom. Shear loads are carried by the skin structure which is locally supported against buckling by the stringers. Presently the semimonocoque structure is configured to the minimum weight design.

ERLINE

Ballistics tests using the 23mm high explosive against the minimum weight semimonocoque tail boom design have demonstrated its lack of damage tolerance.(1) Structural modifications have been shown to increase the damage tolerance of the structure but the amount of damage is predictably a function of confined volume and detonation distance to the surface.(1) Clearly the larger 30mm projectile reduces the survivability of the entire semimonocoque tail boom structure.

The solution for a structural challenge in this study took the form of a search for a highly redundant tail boom structure. A highly redundant structure is a structure that starts with a compact unit structure. The compact unit structure is interconnected within itself by comparatively small, relative to the entire structure, but stiff structural elements. The entire structure is then built up by a replication of the unit structure, scaling as desired or necessitated. The main reason for a high degree of redundancy is to build-in damage tolerance by attempting to keep damage strictly localized.

A possible engineering solution that can easily be made highly redundant is the truss type structure. Use of modern technology and standard elements can make a truss both practical and economical. Because of its potential to fulfill such characteristics, the truss type structure was selected for this study in place of the semimonocoque structure. This study not only utilizes the truss concept but also introduces the concept of complete imbedded substructures. Complete imbedded substructures are easily generated within a truss structure that has a base figure equivalent to a quadrilateral by including the interior diagonals in a simple open truss structure. The intent is to develop a truss type tail boom with complete substructures that is highly redundant so that it can absorb massive damage and yet still hold the aerodynamic loads of flight. The truss tail boom can reduce vulnerability while lowering the weight of the tail boom. The observable surface area drops significantly reducing visibility and radar echo. A bonus would be the possibility of mounting a recoilless rifle on the helicopter because the openness of the truss allows the passage of back blast.

The development was performed throughout by computer modeling. The aerodynamics loads can be simulated and a damage criterion established very easily by this technique. A damage criterion should reflect a maximum amount of damage that can be sustained by the structure. The design objective is to retain structural integrity after imposition of the damage criterion.

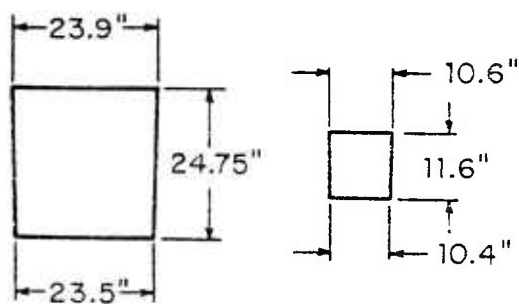
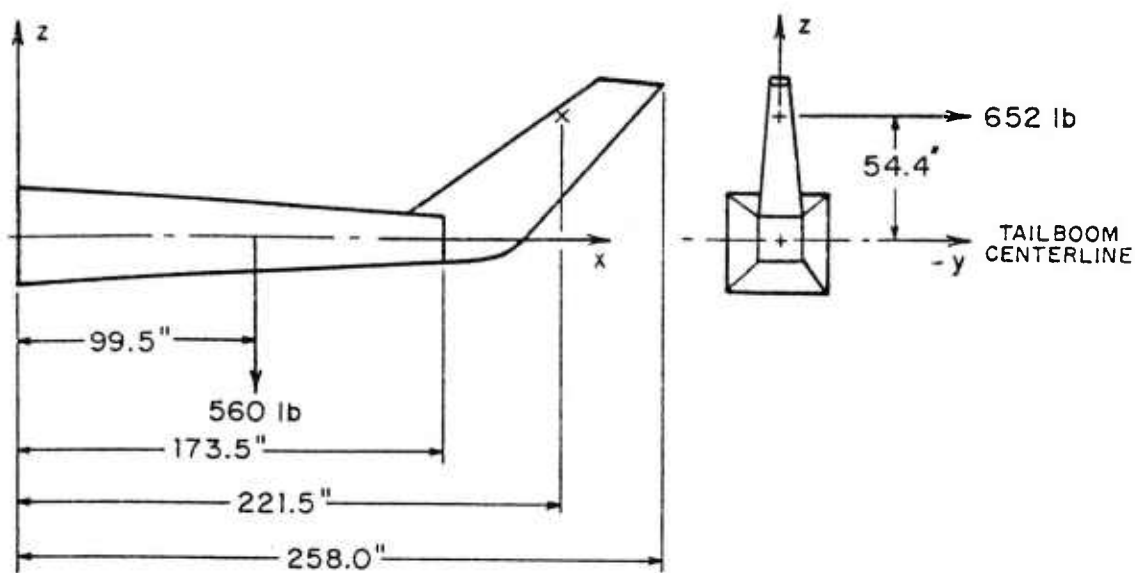
Damage to a truss structure would not be in terms of confined volume or surface distance as it is in the semimonocoque structure, but

damage would be in terms of loss of a member(s) or a loss of a joint. Not counting a completely destructive blast, the most catastrophic single event that could occur to a truss structure would be the destruction of a joint. The loss of a joint in a truss structure can be considered as massive damage because the loss of many members assembled at the joint is associated with loss of the joint. The demand that the truss sustain loss of a single joint and still retain structural integrity (not have other members buckle or fail) under flight loads is considered maximum survivability for the purpose of this study. Thus, loss of a joint is the damage criterion employed in this study.

Static and dynamic analysis of two truss design concepts were performed by the NASTRAN (NASA STRuctural ANalysis) program. One of these truss concepts is a simple open truss design. The other model incorporates the concept of complete substructures. The semimonocoque tail boom of AH-1 helicopter series is used as the basis for a replacement truss tail boom model. The AH-1 series helicopter presents a logical choice to develop a truss structure tail boom to replace a semimonocoque structure. This helicopter has been in the Army's arsenal for a while and will continue in service for a number of years.

PROCEDURES. The longerons are the longitudinal elements which constitute the basic configuration to be employed in the truss structure model development. The basic overall dimensions come from the AH-1G helicopter reported in reference (2) (see Figure 1). Longerons pattern at the larger (or base) end has been selected to conform to the bolt pattern of the AH-1G model at the tail boom - main fuselage manufacturing break line. The aerodynamic loading conditions were obtained from Bell Helicopter Company.(3) Common design parameters and formulas are listed in Table 1. Consideration of cost and logistics leads the ideal design concept to have all structural truss members with the same cross sectional dimensions. There is less cost involved buying large quantities of structural elements all the same size than small quantities of various sizes. The standard tube(4) structural element of 3.81cm (1 1/2 in) outside diameter and .159cm (1/16 in) thickness was chosen for this study because of its higher inertia over rods of the same cross sectional area.

Initial model goal was a high degree redundancy. The goal was achieved by orienting outside diagonals at 45° angles, which generated a large number of joints (44) within the dimensional bounds of Figure 1. Though truly highly redundant and damage tolerant, the geometric configuration is not a low weight design. To obtain a geometric configuration conducive to a low weight design without doing a geometry optimization analysis, this study utilizes only the joint locations described in reference (7). Thus, both models of this



BASE OF
TAILBOOM

END OF
TAILBOOM

Figure 1. Geometry of Helicopter Tail Boom.

ERLINE

TABLE 1.
Common Design Parameters and Formulas

Material:	Aluminum Alloy	
Modulus of Elasticity:	$E = 7. \times 10^{10} \text{ n/m}^2 (10.5 \times 10^6 \text{ psi})$	(4)
Density ρ :	$\rho = 2.7 \times 10^3 \text{ kg/m}^3 (0.1 \text{ lbs/in}^3)$	(4)
Margin of Safety: of Individual Member	$M.S. = \frac{\text{stress limit}}{\text{applied stress}} - 1.$	(5)

Stress Limits per member

Compressive	$CSL = .8 * \delta_{cr}$
Tensile	$TSL = \delta_{cr}$

Euler Column Buckling (6)

$$\delta_{cr} = \frac{P_{cr}}{CSA}$$

Cross Sectional Area of Tube: $CSA = \pi(r_o^2 - r_i^2)$

Compressive Critical Load with Hinged Ends (6)

$$P_{cr} = \frac{\pi^2 EI}{l^2}, \text{ where } l = \text{length of member}$$

study have 28 joints and 72 degrees of freedom. Model 1 is an open truss framework containing 114 members (see Figure 2). Model 2 is a substructured truss framework by the inclusion of the interior diagonals generating a total of 138 members (see Figure 3). The chosen standard structural tube element makes Model 1 weigh approximately 40.8 kg (90 lbs), and Model 2 weigh approximately 52.2 kg (115 lbs).

Static and dynamic analyses were performed on both models by NASTRAN.(8) NASTRAN is a large, comprehensive, general purpose, finite element, displacement method computer program. Design and analysis of all forms of airframes have been carried on by NASA, aerospace industries and other government agencies for a number of years by use of NASTRAN.

The loading that corresponds to the maximum, 130 kt. level flight condition(3) was applied to each complete model and its subsequent damage cases. The application of the 130 knot level flight load was

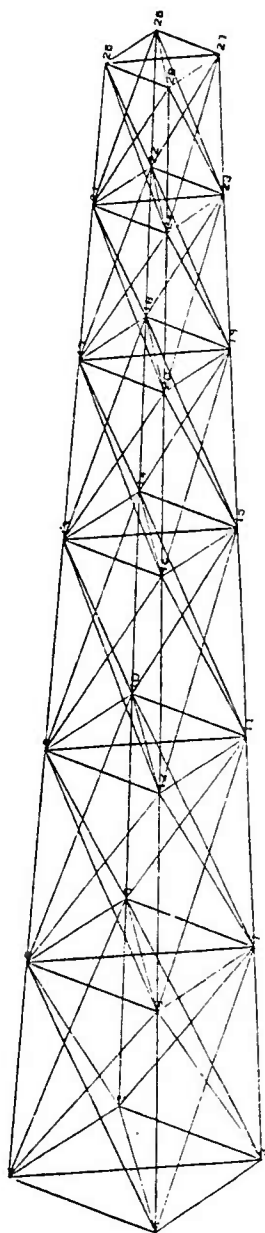


Figure 2. Computer Drawing
of Model 1.

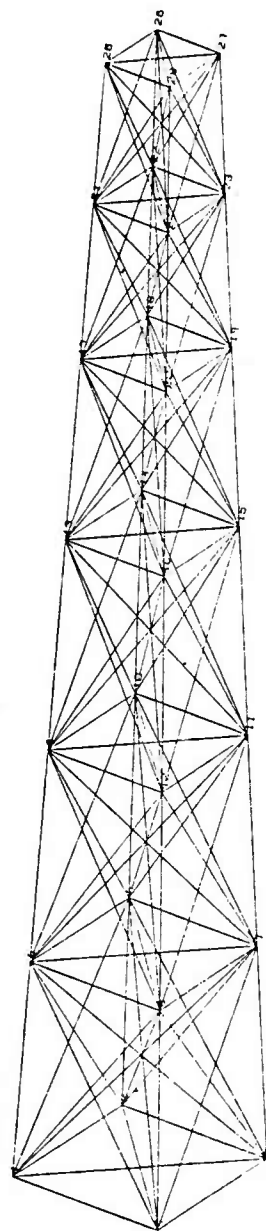


Figure 3. Computer Drawing
of Model 2.

ERLINE

simulated using NASTRAN rigid format 4, Static Analysis with Differential Stiffness outputting displacements, stresses and margins of safety. More accurate design operating stresses are calculated for analysis of buckling failure to individual members by this method(9) over NASTRAN rigid format 1, Static Analysis.

Buckling failure determined by margin of safety for each individual element is calculated in NASTRAN according to the relations given in Table 1. Buckling due to compressive stress occurs before failure of an element due to tension. An additional 20% safety factor is included for compressive stress limits. A margin of safety less than zero indicates failure. A margin of safety between zero and one indicates structural integrity and is acceptable. Margin of safety greater than one is preferred for the purposes of this study.

The complete models were analyzed first with the 130 knot flight loads applied. Then the imposition of the damage criterion was simulated by removing a joint and all elements connecting that joint. A vertical station in front, middle, and rear of the models were chosen to have their joints deleted one at a time and then the damaged structure was reanalyzed with the same applied flight loads. A total of twelve damaged cases were simulated.

Realistically, portions of members that enter a joint may remain affixed to the truss structure after loss of a joint. The effects of these members (or portions thereof) remaining are considered minor, and therefore are neglected in this study.

RESULTS. Maximum deflection constraint due to tail rotor driveshaft couplings (10) is 8.13cm (3.2 in). Table 2 shows maximum displacements of the truss models with and without damage imposed. These displacements lie comfortably within this maximum deflection constraint. The complete semimonocoque tail boom structure weighs approximately 90.72 kg (200 lbs)(3) and its deflections due to the 130 knot level flight load are 1.37cm (.539in) in y direction and .37cm (.146 in) in the z direction.(10) The complete truss models which weigh about half the semimonocoque tail boom weight, show as indicated by Table 2, a stiffness 12% greater in the y direction and 46% less in the z direction. Considering the weight difference the truss models stiffness is reasonably competitive. Table 2 shows a maximum increase in the y direction displacement of .82cm more than the undamaged case occurring in the case where joint 5 is deleted. It shows a maximum increase in the z direction displacement of .803cm more than the undamaged case occurring in the case where joint 7 is deleted.

TABLE 2.
Maximum Displacement at End of Boom Due to 100% Flight Load.

Damage	MODEL 1			MODEL 2		
	y	z		y	z	
none	1.09cm(.430 in)	.797cm(.314 in)		1.10cm(.433 in)	.784cm(.309 in)	
-5	1.91cm(.754)	1.55cm(.611)		1.89cm(.745)	1.53cm(.603)	
-6	1.10cm(.435)	.955cm(.376)		1.10cm(.435)	.935cm(.368)	
-7	1.83cm(.721)	1.60cm(.629)		1.84cm(.725)	1.53cm(.605)	
-8	1.17cm(.463)	1.01cm(.399)		1.13cm(.448)	.950cm(.374)	
-13	1.55cm(.610)	1.04cm(.411)		1.50cm(.591)	1.04cm(.411)	
-14	1.32cm(.522)	.630cm(.248)		1.33cm(.525)	.635cm(.250)	
-15	1.34cm(.529)	1.22cm(.482)		1.37cm(.541)	1.14cm(.449)	
-16	1.53cm(.603)	.81cm(.319)		1.46cm(.576)	.726cm(.286)	
-21	1.43cm(.565)	.889cm(.350)		1.36cm(.538)	.899cm(.354)	
-22	1.15cm(.454)	.726cm(.286)		1.19cm(.469)	.741cm(.292)	
-23	1.17cm(.462)	1.10cm(.436)		1.21cm(.479)	1.01cm(.399)	
-24	1.39cm(.550)	.924cm(.364)		1.33cm(.524)	.843cm(.332)	

ERLINE

Members in tension have been found to have margins of safety greater than 1.0. Compressive margins of safety under 1.0 do occur due to imposition of the damage criterion. (Table 3 is a key for Tables 4, 5 and 6). Since changes in margins of safety only occur locally, only the damage criterion in the neighborhood of the joint in question is presented. Tables 4, 5, and 6 show the compressive margins of safety and change in load path due to imposition of the damage criterion.

The compressive margins of safety (M.S.) shown in Tables 4, 5, and 6 indicate load path redistributions due to the damage cases. The comparison made between model 1 and model 2 shows the interior diagonals of model 2 to be working by taking on tensile and compressive loads. Though no members of model 1 or model 2 failed, Table 7 extracts four cases from Table 4 and 5 where M.S. is under 1.0 for model 1 and two cases for model 2. Table 7 also shows that substructured model 2 consistently has M.S. greater where needed than the open truss model 1. Also, to be noted in Table 7 is a 75% loading condition for two cases out of the four where the M.S. is less than 1.0 for model 1 and comparative M.S. over 1.0 for the substructured model 2.

Table 8 shows the maximum and minimum presented areas and their appropriate orientations. The semimonocoque structure presents a continuous surface whereas the truss models present non-continuous surface areas. For the maximum model 1 has a 47% reduction and model 2 has a 33% reduction in presented area compared to semimonocoque. Comparison of the minimum presented areas show: model 1 has a 80% reduction, and model 2 has a 75% reduction. These reductions show a clear reduction in visibility of the truss tail boom relative to the semimonocoque type tail boom.

TABLE 3 Key for Tables 4, 5, and 6

I.D. = interior diagonal; O.D. = outside diagonal
T.H. = transverse horizontal; T.V. = transverse vertical
T.D. = transverse diagonal; Long. = longeron
J-J = joint-to-joint connection
D = deleted member; T = member in tension
- = Nonexistent member; Blank space = compressive

Margin of Safety greater than 10.0

ERLINE

TABLE 4. Compressive Margins of Safety Under 10.0 Due to 130 knot Flight Load on Forward Vertical Sta with 100% Loading

DAMAGE MODEL MEMBER	J-J	No#	NONE		Joint 5		Joint 6		Joint 7		Joint 8	
			1	2	1	2	1	2	1	2	1	2
Long.	1-5	7	T	T	D	D	T	T	T	T	T	T
"	2-6	8			T	T	D	D	2.0	2.3		
"	3-7	9	3.6	3.6	1.4	1.3	4.1	4.1	.D	D	1.9	2.2
"	4-8	10	T	T	T	T	T	T			D	D
O.D.	1-6	11		8.0	8.4	10.	D	D	2.8	2.2		7.1
"	2-5	12	T	T	D	D	T	T	T	T	T	T
"	1-8	13	T	T	T	T	T	T	T	T	D	D
"	4-5	14			D	D	8.7	7.6	T	T	1.6	3.0
"	4-7	15	T	T	T	T	T	T	D	D	T	T
"	3-8	16	2.0	2.2	2.8	2.6	2.3	2.3	.36	.71	D	D
"	2-7	17	2.0	2.0	1.5	1.3	1.2	1.4	D	D	1.5	1.2
"	3-6	18	T	T	T	T	D	D	T	T	T	T
I.D.	1-7	19	-	T	-	T	-	-	-	D	-	T
"	2-8	20	-	-	-	T	-	-	-	3.2	-	D
"	3-5	21	-	-	-	D	-	T	-	-	-	3.7
"	4-6	22	-	T	-	-	-	D	-	T	-	T
T.V.	5-6	23			D	D	D	D				
T.H.	5-8	24			D	D					D	D
T.V.	7-8	25	T	T	T	T	T	T	D	D	D	D
T.H.	6-7	26	T	T	T	T	D	D	D	D	T	T
T.D.	5-7	27	T		D	D	T	T	D	D		
"	6-8	28			4.0	3.5	D	D	T	T	D	D
Long.	5-9	29	T	T	D	D	T	T	T	T	T	T
"	6-10	30	T	T	T	T	D	D				T
"	7-11	31	4.4	4.4	1.8	1.8	1.9	2.2	D	D		8.5
"	8-12	32			T	T	T		2.5	3.1	D	D
O.D.	5-10	33	8.6	7.0	D	D	1.6	2.4	T		6.9	4.9
"	6-9	34	T	T	T	T	D	D	T	T		T
"	5-12	35	T	T	D	D	T	T	T	T	T	T
"	8-9	36	6.0	5.7	3.7	5.6		5.7	2.2	1.9	D	D
"	8-11	37	T	T	T	T	T	T	T	T	D	D
"	7-12	38	1.7	1.9	1.3	1.1	1.2	1.1	D	D	.81	1.2
"	6-11	39	2.3	2.3	3.1	2.6	D	D	.44	.79	2.6	2.1
"	7-10	40	T	T	T	T	T	T	D	D	T	T
I.D.	5-11	41	-	T	-	D	-	5.1	-	T	-	T
"	6-12	42	-		-	T	-	D	-	3.0	-	
"	7-9	43	-		-		-	T	-	D	-	7.5
"	8-10	44	-	T	-		-	T	-	T	-	D

ERLINE

TABLE 5. Compressive Margins of Safety Under 10.0 Due to 130 knot Flight Load on Middle Vertical Sta with 100% Loading

DAMAGE MODEL MEMBER	J-J	No.#	None		Joint 13		Joint 14		Joint 15		Joint 16	
			1	2	1	2	1	2	1	2	1	2
Long.	9-13	51	T	T	D	D	T	T	T	T	T	T
"	10-14	52	T	T	T	T	D	D	5.1	7.5	T	T
"	11-15	53	6.2	6.2	3.0	3.0	T	T	D	D	1.5	1.9
"	12-16	54			7.6	8.5	5.6	6.1			D	D
O.D.	9-14	55	7.0	5.6	2.5	3.7	D	D	2.9	2.2	T	8.4
"	10-13	56	T	T	D	D	T	T	T	T	T	T
"	9-16	57	T	T	T	T	T	T	T	T	D	D
"	12-13	58	3.5	3.4	D	D	3.8	3.0	10.	4.7	.62	1.1
"	12-15	59	T	T	T	T	T	T	D	D	T	T
"	11-16	60	1.3	1.5	1.3	1.2	1.3	1.2	.24	.56	D	D
"	10-15	61	2.3	2.3	2.2	1.7	1.2	1.7	D	D	1.2	1.0
"	11-14	62	T	T	T	T	D	D	T	T	T	T
I.D.	9-15	63	-	T	-	T	-	-	-	D	-	T
"	10-16	64	-		-	T	-	-	-	3.2	-	D
"	11-13	65	-		-	D	-	T	-	-	-	2.9
"	12-14	66	-	T	-		-	D	-	T	-	T
T.V.	13-14	67			D	D	D	D				
T.H.	13-16	68			D	D	T	T			D	D
T.V.	15-16	69	T	T	T	T	T	T	D	D	D	D
T.H.	14-15	70	T	T	T	T	D	D	D	D		
T.D.	13-15	71			D	D			D	D	T	T
T.D.	14-16	72					D	D	T	T	D	D
Long.	13-17	73	T	T	D	D	T	T	T	T		
"	14-18	74	T	T	T	T	D	D	T	T	T	T
"	15-19	75	9.8	9.9	4.8	4.9	4.7	5.7	D	D		
"	16-20	76			T	T	5.9	5.8	2.6	3.1	D	D
O.D.	13-18	77	5.4	5.4	D	D	1.7	2.9		7.2	3.2	2.8
"	14-17	78	T	T	T	T	D	D	T	T	T	T
"	13-20	79	T	T	D	D	T	T	T	T	T	T
"	16-17	80	4.0	3.9	2.1	2.9	4.8	3.0	2.2	1.9	D	D
"	16-19	81	T	T	T	T	T	T	T	T	D	D
"	15-20	82	2.3	2.3	2.3	1.8	2.1	1.7	D	D	.87	1.2
"	14-19	83	3.3	3.4	3.3	2.8	D	D	.82	1.3	5.5	4.1
"	15-18	84	T	T	T	T	T	T	D	D	T	T
I.D.	13-19	85	-		-	D	-	5.9	-	-	-	T
"	14-20	86	-		-	T	-	D	-	4.2	-	
"	15-17	87	-	T	-	T	-	T	-	D	-	7.3
"	16-18	88	-	T	-		-	T	-	T	-	D

ERLINE

TABLE 6. Compressive Margins of Safety Under 10.0 Due to 130 knot Flight Load on End Vertical Sta with 100% Loading

DAMAGE MODEL			None		Joint 21		Joint 22		Joint 23		Joint 24	
			1	2	1	2	1	2	1	2	1	2
MEMBER	J-J	No#										
Long.	17-21	95	T	T	D	D	T	T	T	T	T	T
"	18-22	96	T	T	T	T	D	D			T	T
"	19-23	97			5.4	5.6	T	T	D	D	3.2	3.7
"	20-24	98			7.8	9.1	6.6	6.8			D	D
O.D.	17-22	99	6.5	6.5	2.6	4.1	D	D	3.6	3.2		9.1
"	18-21	100	T	T	D	D	T	T	T	T	T	T
"	17-24	101	T	T	T	T	T	T	T	T	D	D
"	20-21	102	5.1	4.9	D	D	5.0	4.1	6.5	6.5	1.4	2.0
"	20-23	103	T	T	T	T	T	T	D	D	T	T
"	19-24	104	2.9	2.9	2.9	2.4	3.0	2.4	1.1	1.5	D	D
"	18-23	105	4.2	4.3	4.1	3.1	2.3	3.3	D	D	2.5	2.2
"	19-22	106	T	T	T	T	D	D	T	T	T	T
I.D.	17-23	107	-		-		-	10.	-	D	-	T
"	18-24	108	-		-	T	-		-	6.7	-	D
"	19-21	109	-	T	-	D	-	T	-	T	-	6.1
"	20-22	110	-	T	-	9.6	-	D	-	T	-	T
T.V.	21-22	111			D	D	D	D				
T.H.	21-24	112			D	D	T	T			D	D
T.V.	23-24	113	T	T	T	T	T	T	D	D	D	D
T.H.	22-23	114			T	T	D	D	D	D		
T.D.	21-23	115		T	D	D			D	D	T	T
T.D.	22-24	116	T				D	D	T	T	D	D
Long.	21-25	117	T	T	D	D	T	T	T	T		
"	22-26	118	T	T	T	T	D	D	T	T	T	T
"	23-27	119			6.7	7.1	5.6	7.4	D	D		
"	24-28	120			T	T	7.8	7.3	3.4	3.9	D	D
O.D.	21-26	121	7.3	7.2	D	D	2.4	4.0			4.1	3.7
"	22-25	122	T	T	T	T	D	D	T	T	T	T
"	21-28	123	T	T	D	D	T	T	T	T	T	T
"	24-25	124	5.0	4.8	2.9	4.0	6.0	3.7	2.9	2.4	D	D
"	24-27	125	T	T	T	T	T	T	T	T	D	D
"	23-28	126	3.1	3.1	2.9	2.3	2.8	2.3	D	D	1.2	1.6
"	22-27	127	4.2	4.4	4.3	3.7	D	D	1.2	1.9	7.4	5.6
"	23-26	128	T	T	T	T	T	T	D	D	T	T
I.D.	21-27	129	-		-	D	-	6.2	-		-	T
"	22-28	130	-	T	-	T	-	D	-	6.3	-	
"	23-25	131	-		-	T	-	T	-	D	-	8.0
"	24-26	132	-	T	-		-	T	-	T	-	D

TABLE 7. Cases of Models 1 and 2 Where M.S., Less Than 1.0

Case Where Damage is - Joint 7				
Loading	Element #	J-J	Model 1 M.S.	Model 2 M.S.
100%	16	3-8	.36	.71
	39	6-11	.44	.79
75%	16	3-8	.82	1.3
	39	6-11	.94	1.4
Case Where Damage is - Joint 8				
100%	38	7-12	.81	1.2
Case Where Damage is - Joint 15				
100%	60	11-16	.24	.56
	83	14-19	.82	1.3
75%	60	11-16	.66	1.1
Case Where Damage is - Joint 16				
100%	58	12-13	.62	1.1
	82	15-20	.87	1.2

TABLE 8. Observable Presented Area

	Maximum Orientation	Minimum Orientation
AH-1G Semimonocoque	58,655cm ² (4546 in ²)Side	39,233cm ² (3506 in ²)Bottom
Truss Model 1	31,354cm ² (2430 in ²)Oblique	7,838cm ² (608 in ²)Side
Truss Model 2	39,019cm ² (3024 in ²)Oblique	9,754cm ² (756 in ²)Side

CONCLUSIONS. The truss type tail boom models of this study provide a reduction in weight of approximately 50% over the present semimonocoque tail boom structure of the AH-1G. The truss models are reasonably stiff structures compared to the semimonocoque structure.

Analyses of the highly redundant truss models under the aerodynamic loads of flight and with imposition of a massive damage criterion show:

- o Substantial retention of stiffness
- o Change of load path that is localized
- o No failure of elements due to tension or compression
- o Retention of structural integrity

The substructured truss model has more supporting structural elements to redistribute the load and consistently has margins of safety higher than the non-substructured truss model. The substructuring concept has vulnerability reduction built-in. The substructure concept assures a higher degree of confidence in the truss concept to retain structural integrity after imposition of the loss of a joint. The substructured truss model has at least a 33% reduction in presented area compared to the semimonocoque structure. The highly redundant substructured truss type tail boom is a highly survivable structure.

REFERENCES

- (1) D.F. Haskell, Damage Tolerance of Semimonocoque Aircraft, Paper at 41st Meeting of NATO-AGARD Impact Damage Tolerance of Structures Conference, Ankara, Turkey, October 1975.
- (2) J.D. Cronkite, V.L. Berry, J.E. Brunlsen, A. NASTRAN Vibration Model of AH-1G Helicopter Airframe, June 1974, AD A009482.
- (3) Private communication between D.F. Haskell of BRL and D. Reisdorfer, Bell Helicopter Company.
- (4) ALCOA STRUCTURAL HANDBOOK. A Design Manual for Aluminum Company of America, Pittsburg, PA 1958.
- (5) NASTRAN Programmer's Manual, NASA SP-221(03), July, 1976, Washington, D.C.
- (6) S. Timoshenko, Strength of Materials, Part I. D. VanNostrand Co. Inc., New York, N.Y. 1940, pp 244-254.
- (7) J.S. Arora, E.J. Haug, A.K. Govil, FAIL-SAFE, Design of An Open Truss Helicopter Tail Boom Technical Report #32, April 1977, Coll. of Engineering, University of Iowa.
- (8) Caleb, W. McCormick, Editor, The NASTRAN User's Manual, NASA SP 222(03), March 1976, Washington, D.C.
- (9) R.B. Bennett, NASTRAN Differential Stiffness, Analysis of an Aircraft Canopy, NASA TMX2378 pp 85-105, September 1971.
- (10) D.A. Reisdorfer, Tail Boom Vulnerability Reduction Test Program, Report No. 699-099-004, August 1975, Bell Helicopter Company, Fort Worth, Texas 76101.

FOG AND HAZE IN EUROPE AND THEIR EFFECTS ON
PERFORMANCE OF ELECTRO-OPTICAL SYSTEMS (U)

*OSKAR M. ESSENWANGER, DR., and DORATHY A. STEWART, DR.
US ARMY MISSILE RESEARCH AND DEVELOPMENT COMMAND
REDSTONE ARSENAL, ALABAMA 35809

INTRODUCTION

It is well known that aerosols which cause poor visibility affect infrared and submillimeter wave propagation, and water vapor causes significant additional attenuation of the near-millimeter wavelengths. Direct estimates of climatological probabilities for the degrading performance are difficult to obtain. These probabilities require the knowledge of the liquid water content and the drop-size distribution, and neither quantity is observed on a routine basis. In the past the liquid water content was often assumed to be a unique function of the visibility. This postulation is erroneous because the relationship between visibility and liquid water content depends upon the drop-size distribution.

The second quantity, drop-size distribution, is even more deficient because information is available only from sporadic measurements made during specific research projects of insufficient duration to develop a systematic climatology of drop-size distributions. Available information in the literature has recently been summarized by Stewart (12). There are large variations, but on the average fogs caused primarily by radiational cooling (radiation fogs) have smaller drops than other fogs.

We have attempted to develop a propagation climatology by inference from the frequency of occurrence of different types of fog. The methodology is presented in detail in the next 3 sections. The analysis resulted in the estimation of average attenuations of different wavelengths by fogs in Central Europe. Attenuation of 1250- μm energy is always less than attenuation of 870 μm in fog. Average

attenuation of $870\text{ }\mu\text{m}$ in dense fogs is much less than attenuation of $10.6\text{ }\mu\text{m}$ in all seasons. In summer $10.6\text{ }\mu\text{m}$ penetrates light fog better than $870\text{ }\mu\text{m}$, but $10.6\text{ }\mu\text{m}$ penetrates less well than $1250\text{ }\mu\text{m}$.

Before the details of our research effort are discussed, the occurrence of poor visibility in Central Europe will be reviewed. It is known that fog has a distinct diurnal variation with the highest chances of occurrence in the early morning hours. Ten stations were selected from Central Europe, and the frequency of fog (visibility $\leq 1\text{ km}$) was established. The result is depicted in Figures 1 and 2. We learn from Figure 1 that for most stations fall is the worst season having the highest amount of morning fogs up to 25% of the time, and winter is next. Had we included monthly results the frequency for the worst month at some stations would have been more than 30% of the early morning hours.

Fuerstenfeldbruck has been selected to illustrate the duration of fogs. As disclosed in Figure 2 one case with over 120 hours of consecutive fog existed, and fog was present in the morning hours on 13 consecutive days. Consequently, poor visibility occurs frequently enough to warrant the consideration of the degrading effect on electro-optical propagation. Further details are given by Essenwanger (2, 3, 4) and Essenwanger and Stewart (6).

THE FOG CLASSIFICATION MODEL

Because average drop sizes are usually smaller in radiation fog than in other fogs, electromagnetic energy with wavelengths in the middle infrared and longer should penetrate radiation fog better. Recent measurements published in the meteorological literature (e.g., Mészáros, 9, and Pilié et al., 11) contradict the widespread belief that fogs remote from coastal areas can be assumed to be pure radiation fogs. However, no probabilities of occurrence are available.

Unfortunately, fog is not classified by the observer at the time of measurement, and a detailed analysis of individual fog cases is tedious and time consuming. Consequently, the authors sought a computerized method which could be efficiently applied to large data collections. It is difficult to develop a computer model which separates all known types of fog because most fogs are mixtures of processes, and some radiative cooling occurs in nearly all of them. Therefore our goal was the separation of fogs into primary groups: those caused predominantly by radiational cooling and all other types. Although the individual fog may be a borderline case and be placed into the wrong group, these cases should balance out in a statistical sense.

FIG. 1. FREQUENCY OF VISIBILITY ≤ 1 KM (SEASONS)

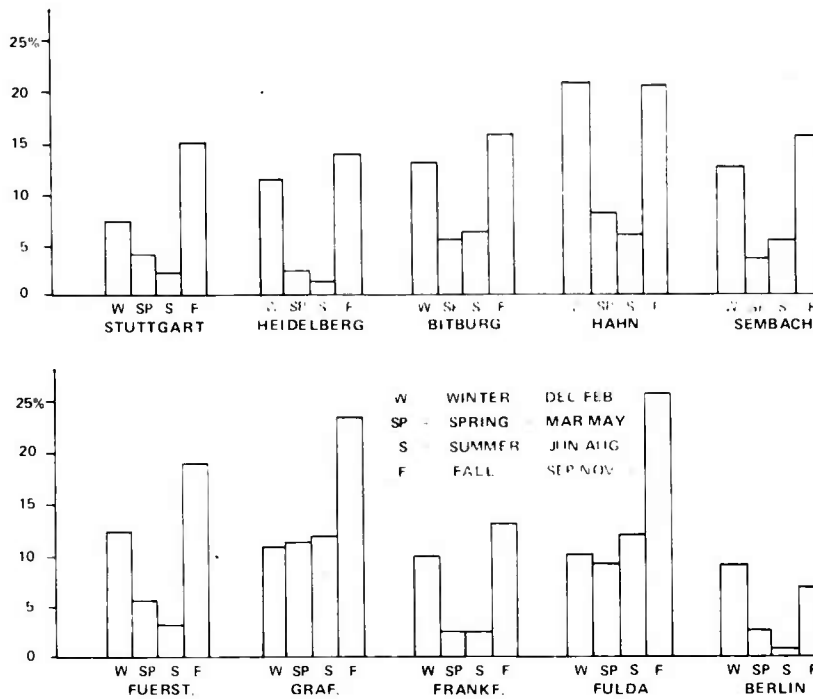


FIG. 2. MAXIMUM DURATION OF VISIBILITY ≤ 1 KM

FUERSTENFELDBRUCK

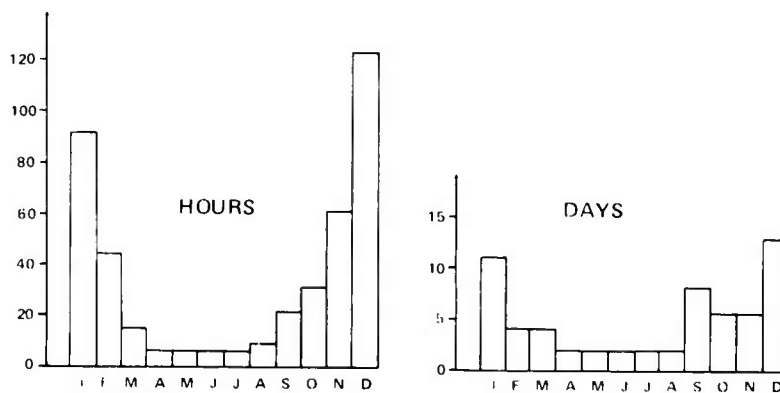


Figure 3 delineates the computer processing of the data at an individual station with observations in synoptic code form. The flow chart should be self-explanatory, and the grouping is described in detail in a forthcoming report by Essenwanger (5). The model is applicable to morning fogs, and is based on precipitation, cloud and temperature behavior in fog processes. A clear or partly cloudy night leads to radiation fog (class 11-13) while precipitation and an overcast sky imply the presence of predominantly non-radiative types (class 6-8). Class 9 remains a mixture of both but the data cannot be separated without a considerable increase in computer costs. A splitting of this class in half is very cost effective, and this simple rule has no significant impact on the final outcome. The model was robust to reasonable changes in cloud conditions. Table 1 provides an excerpt from the detailed tables in the report by Essenwanger (5), and illustrates the grouping of fogs for 10 stations in Central Europe.

The reasonable functioning of the fog classification model is supported by results depicted in Figures 4 and 5. These show the cumulative frequency of wind speeds and dew points by fog groups during fall and winter. As expected, radiation fogs display a lower wind speed and dew point than non-radiation fogs. More details can be found in Essenwanger (5).

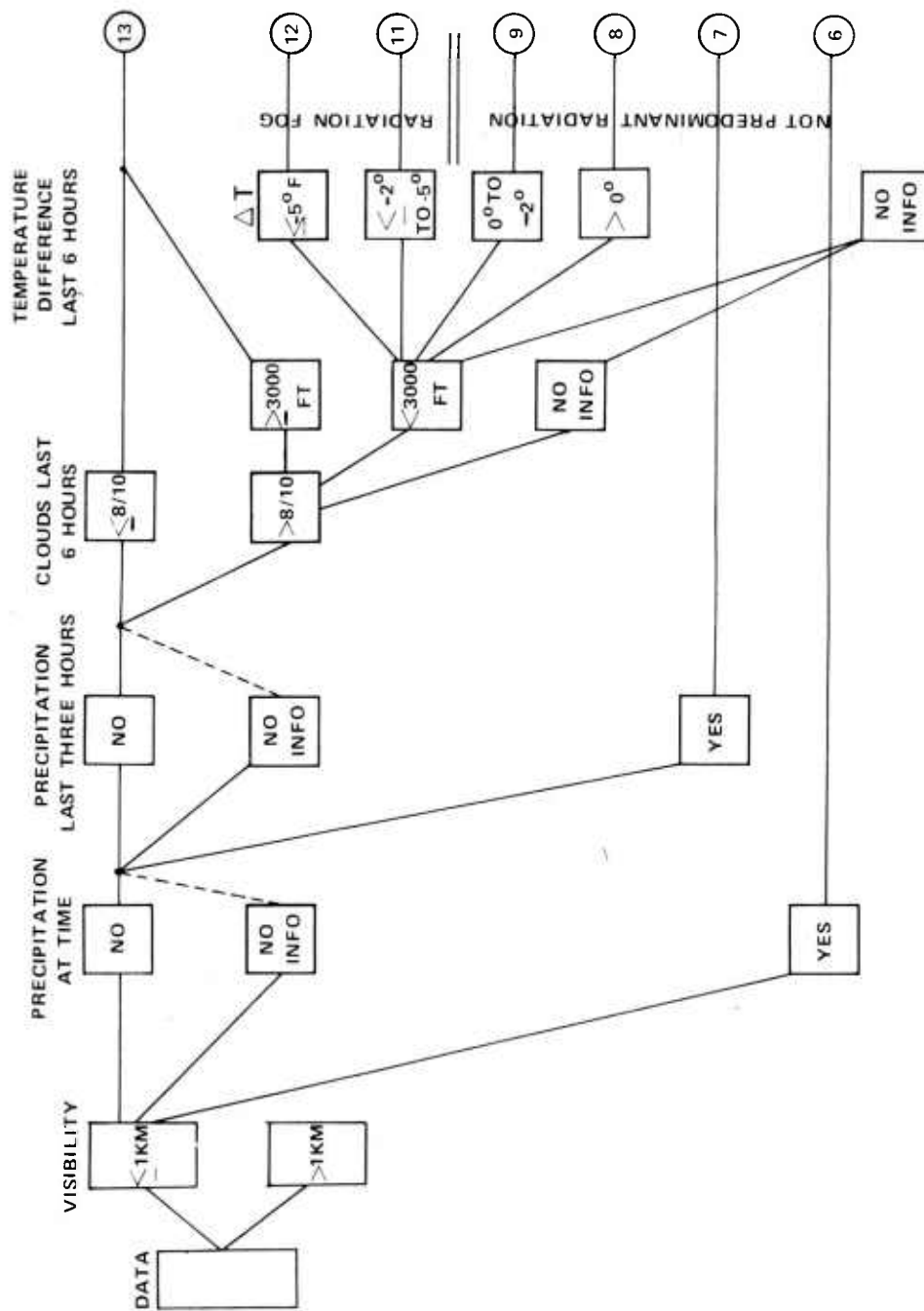
Table 1. Grouping of Fog From the Model in Figure 3 (in ‰ of Fog).

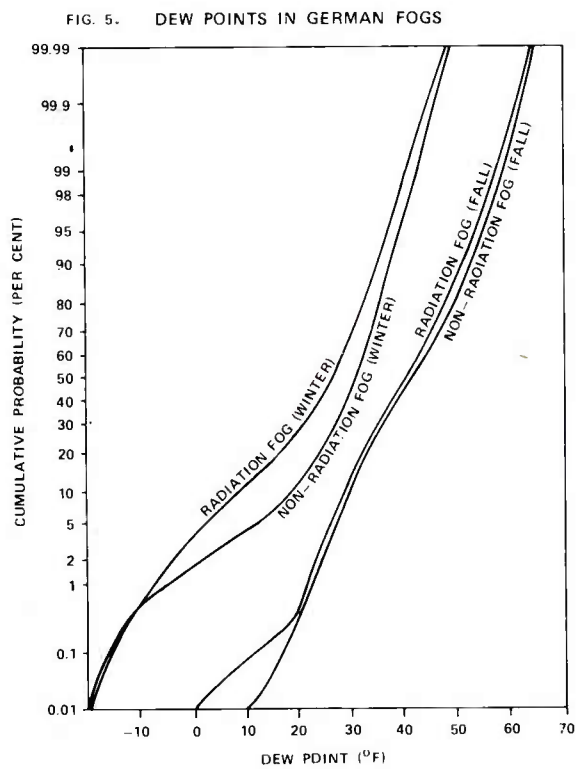
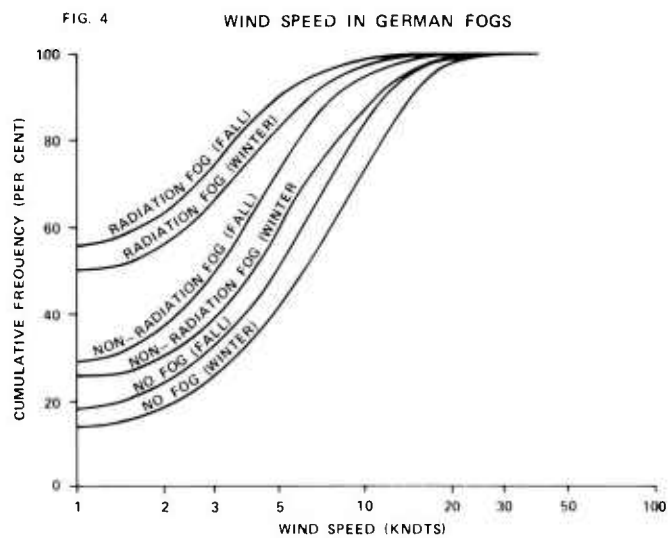
	Fall				Winter				Spring			
	NR	M	R	N	NR	M	R	N	NR	M	R	N
	‰	‰	‰		‰	‰	‰		‰	‰	‰	
Stuttgart	73	147	780	2234	154	256	590	2196	99	198	703	2208
Heidelberg	154	201	645	1546	246	347	407	1561	209	116	675	1623
Bitburg	206	206	588	1661	461	262	277	1654	180	157	663	1666
Hahn	245	315	440	1623	419	353	228	1292	403	258	339	1555
Sembach	114	127	759	1559	305	288	407	1499	64	149	787	1507
Fuerstenfeldbr.	218	243	539	1061	243	421	336	993	161	53	786	1081
Grafenwoehr	92	172	736	969	421	228	351	933	108	108	784	973
Frankfurt	268	271	461	2273	363	381	256	2192	241	241	518	2206
Fulda	246	0	754	799	-	-	-	-	125	83	792	840
Berlin	215	321	464	2275	346	318	336	2197	143	224	633	2269

NR = non-radiation fog (class 6-8). M = mixture (class 9).

R = radiation fog (class 11-13). N = total data base in days.

FIG. 3 . FOG CLASSIFICATION MODEL





THEORETICAL COMPUTATIONS

Extinction of electromagnetic energy by fog droplets depends upon the wavelength of the energy, the complex index of refraction of the drops for that wavelength, and the drop-size distribution. We chose to compute attenuations for the wavelengths of 0.55, 10.6, 870, and 1250 μm . The complex index of refraction of water for 0.55 μm was taken from Hale and Querry (7), and the value for 10.6 μm was interpolated from a table in the same article. Davies et al. (1) made measurements from which the index of refraction of water for 1250 μm could be computed, and the value for 870 μm was interpolated from their data. The computational procedure for obtaining theoretical attenuations based on Mie (10) theory is outlined by Stewart (12). She made an extensive literature survey, and her report lists the references for the drop-size distributions used in this study.

Figures 6-9 illustrate some of the results of our computations. Figure 6 is a plot of visibility as a function of attenuation at 10.6 μm , and one can readily see that there is a great deal of scatter in the data. An attenuation of 50 dB/km at 10.6 μm can be associated with visibilities over 300 m or considerably less than 100 m. Figure 7 depicts the relation of visible and 1250- μm attenuations, and these data contain even more scatter than the data in Figure 6. Figure 8 shows that 10.6- μm attenuation by fog droplets is more closely correlated with 1250- μm attenuation than it is with visible attenuation. Figure 9 contains a plot of 870 μm attenuation versus 1250- μm attenuation, and it is obvious that these attenuations are very closely correlated.

The data in Figures 6-9 were examined individually to determine which fogs could be classified as radiation or non-radiation fogs. Table 2 contains the expected mean attenuations of 10.6, 870, and 1250 μm by fog drops for three different visibilities.

Water vapor is another source of extinction of infrared and near-millimeter wavelengths. McCoy et al. (8) gave a formula for computing attenuation of the 10.6- μm wavelength by water vapor. Webster (13) developed a procedure for computing water vapor attenuation near 1 mm. The attenuation by water vapor depends primarily upon the amount of water vapor and the temperature, but there is also a slight pressure dependence. Figure 10 displays the cumulative frequency distribution of attenuation of 870 μm by water vapor for radiation and non-radiation fogs in fall and winter in Germany.

FIG. 6. PLOT OF VISIBILITY VERSUS ATTENUATION
AT $10.6\text{-}\mu\text{m}$

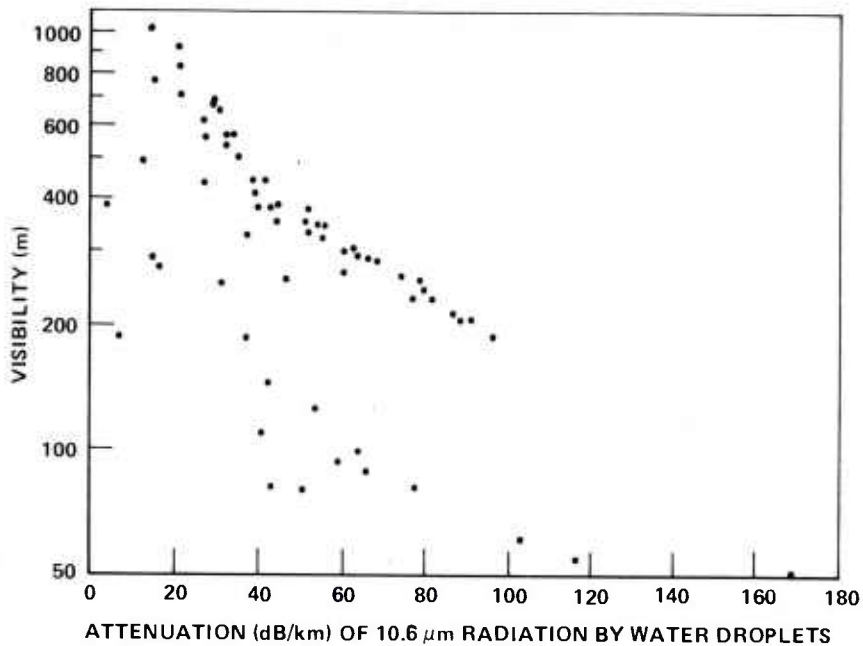


FIG. 7. COMPARISON OF ATTENUATIONS OF $1250\text{-}\mu\text{m}$
AND $0.55\text{-}\mu\text{m}$ WAVELENGTHS BY FOG DROPLETS

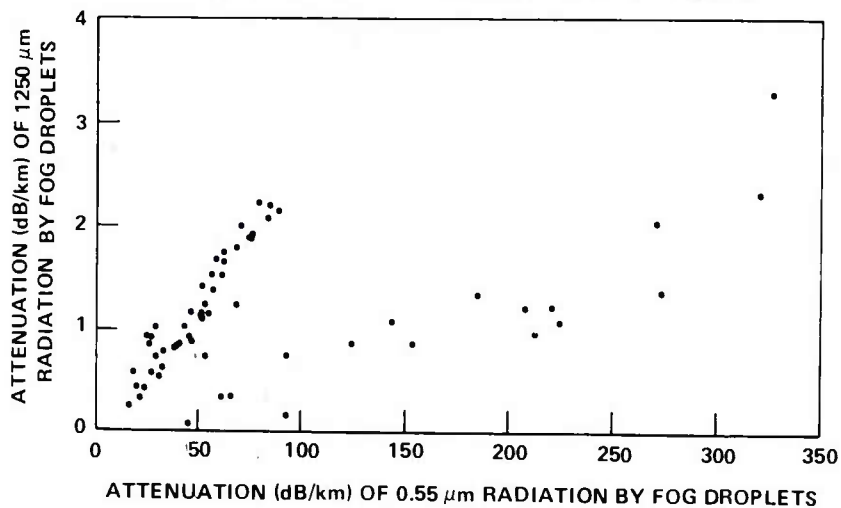


FIG. 8. COMPARISON OF ATTENUATIONS OF 1250- μm AND 10.6- μm WAVELENGTHS BY FOG DROPLETS

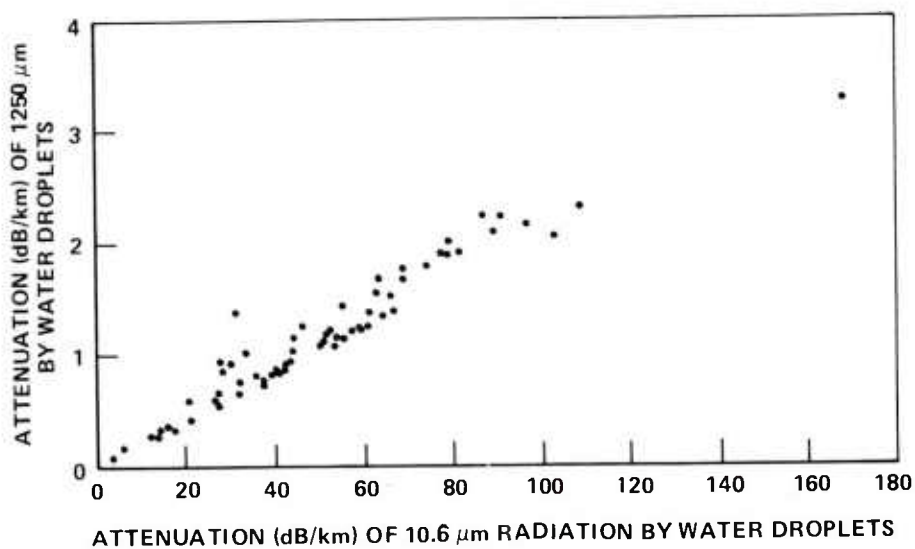
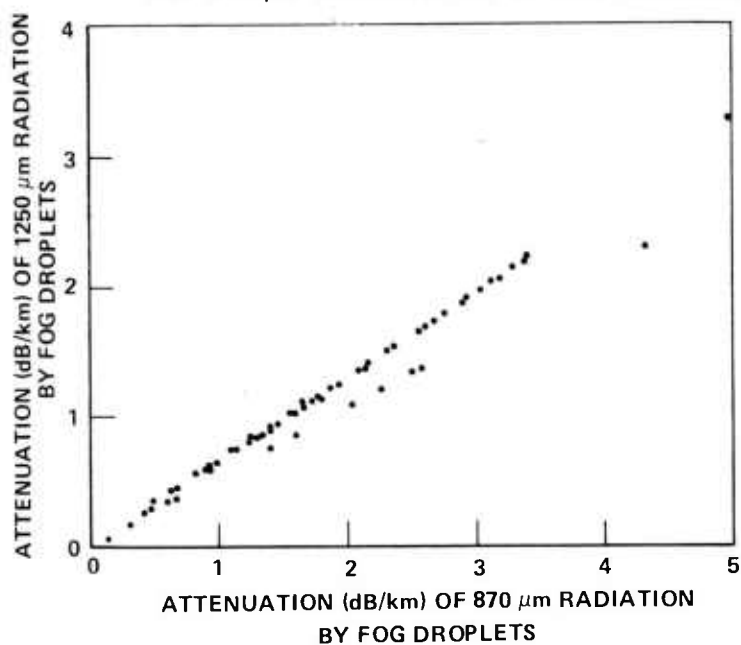


FIG. 9. COMPARISON OF ATTENUATIONS OF 1250- μm AND 870- μm WAVELENGTHS BY FOG DROPLETS



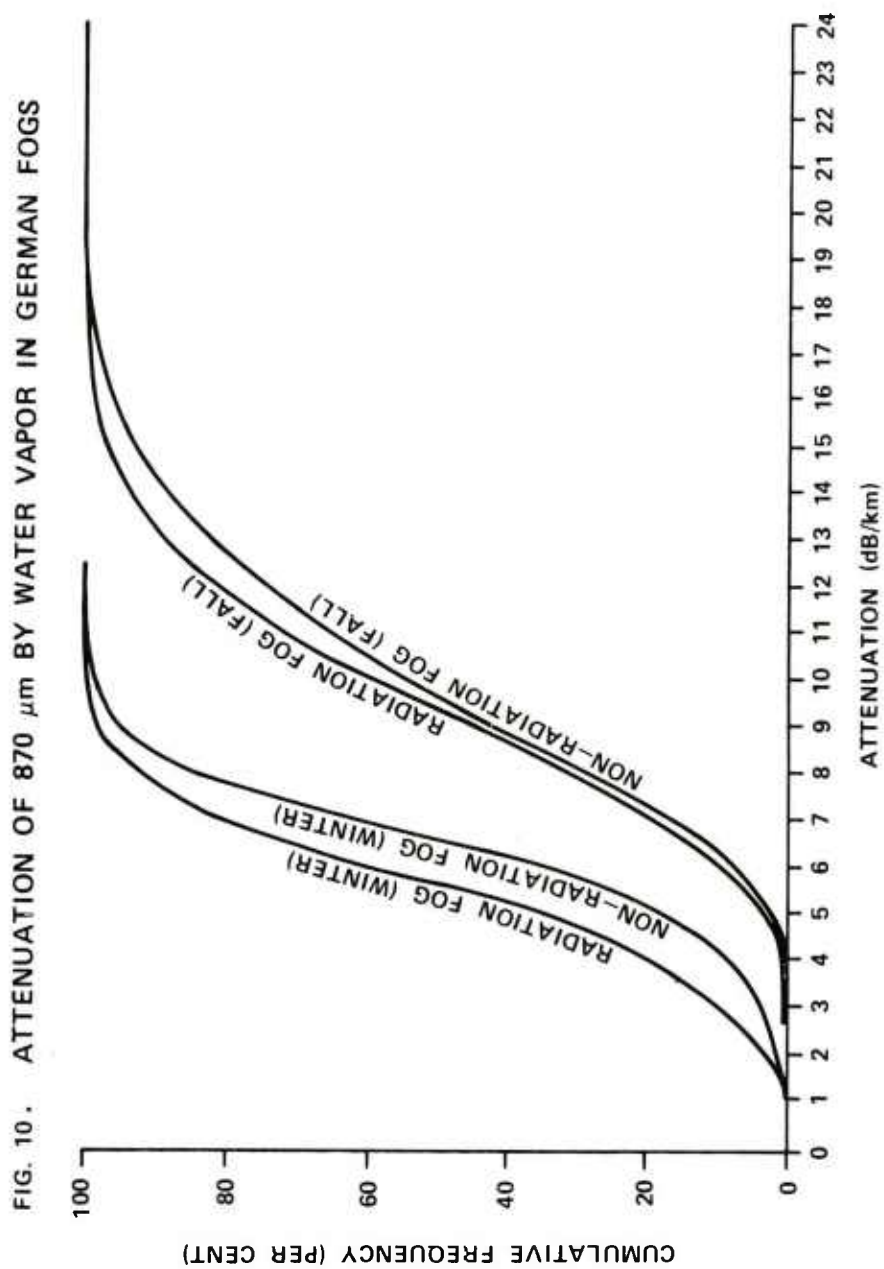


Table 2. Expected Mean Attenuation (dB/km)
by Fog Droplets.

	Visibility		
	<u>1000m</u>	<u>400m</u>	<u>200m</u>
0.55 μm	16.99	42.47	84.95
10.6 μm , radiation fog	5.95	14.87	29.73
10.6 μm , non-radiation fog	14.78	36.95	73.91
870 μm , radiation fog	0.21	0.52	1.04
870 μm , non-radiation fog	0.54	1.34	2.68
1250 μm , radiation fog	0.12	0.31	0.62
1250 μm , non-radiation fog	0.35	0.87	1.74

PROPAGATION CLIMATOLOGY

From the above theoretical framework and climatological information one can infer climatological attenuation characteristics for fogs in Europe. The mean attenuations by radiation and non-radiation fog droplets in Table 2 were weighted according to the number of fog cases in Table 3. Table 4 contains the mean attenuations by water vapor. These have been weighted according to the number of fog cases in Table 3 and added to the mean attenuations by fog droplets to obtain the mean total attenuations in Table 5. One sees that 1250 μm penetrates fog better than both 10.6 μm and 870 μm in all seasons and for all visibilities. This occurs because attenuations by both water vapor and fog drops are small at 1250 μm . At 10.6 μm attenuation by water vapor is always small but attenuation by fog drops may be quite large, especially in non-radiation fogs which typically have larger drops than radiation fogs. In winter when about half the fogs are non-radiation fogs, average attenuation of 10.6 μm is considerably larger than attenuation of 870 μm for all visibilities. In summer water vapor attenuation causes average 870- μm attenuation to be larger than 10.6- μm attenuation in light fogs, but 870- μm attenuation is less than 10.6- μm attenuation in dense fogs. In spring and fall 10.6 μm and 870 μm have comparable attenuations in light fogs, but 870 μm is attenuated less in heavy fogs. From 0600 to 0800 Greenwich Mean Time 64 per cent of fogs have visibilities less than 500 m in spring and 75 per cent in fall.

Table 3. Number of Cases of Fog From Ten German Stations at 0600 Greenwich Mean Time.

Type of Fog	Winter	Spring	Summer	Fall
Radiation	672 (49.8%)	477 (71.9%)	456 (77.8%)	1591 (71.3%)
Non-Radiation	677 (50.2%)	186 (28.1%)	130 (22.2%)	641 (28.7%)

Table 4. Mean Attenuation (dB/km) by Water Vapor in Germany at 0600 Greenwich Mean Time.

	Radiation Fog			Non-Radiation Fog		
	10.6 μm	870 μm	1250 μm	10.6 μm	870 μm	1250 μm
Winter	0.11	5.61	1.74	0.14	6.63	2.06
Spring	0.21	8.16	2.54	0.25	9.06	2.82
Summer	0.50	13.45	4.21	0.64	15.38	4.82
Fall	0.29	9.82	3.06	0.32	10.26	3.20

Table 5. Expected Mean Attenuation (dB/km) in German Fogs at 0600 Greenwich Mean Time.

Season	Wavelength (micrometers)	Visibility		
		1000m	400m	200m
Dec - Feb	10.6	10.5	26.1	52.0
	870	6.5	7.1	8.0
	1250	2.1	2.5	3.1
Mar - May	10.6	8.6	21.3	42.3
	870	8.7	9.2	9.9
	1250	2.8	3.1	3.6
Jun - Aug	10.6	8.4	20.3	40.1
	870	14.2	14.6	15.3
	1250	4.5	4.8	5.2
Sep - Nov	10.6	8.8	21.5	42.7
	870	10.3	10.7	11.5
	1250	3.3	3.6	4.0

SUMMARY AND CONCLUSIONS

It is generally agreed that propagation of electromagnetic energy is affected by fog drops and water vapor, but climatological probabilities of attenuation do not exist except for the visible range which has wavelengths from about 0.4 to 0.7 μm . Unfortunately, attenuation in the visible portion of the spectrum does not uniquely determine attenuation of other wavelengths. The ratio of middle infrared attenuation to visible attenuation is determined mainly by the sizes of the drops in a fog because water vapor attenuation is small in the middle infrared and negligible in the visible. Water vapor is an important factor to be considered near one millimeter (see Table 4).

Average drop sizes depend primarily upon the meteorological processes which cause condensation. Fogs caused primarily by radiational cooling (radiation fogs) typically have smaller drop sizes than other fogs. Droplet sizes for fogs of each type vary widely with a large dispersion around the average. The curve of droplet size distribution for non-radiation fogs is shifted toward larger droplet size relative to the curve for radiation fog. Some overlapping area exists, but in principle the two frequency curves are significantly different for large data collectives.

In order to handle large amounts of data for a statistical analysis, we developed a computerized model for morning fogs in which the classification depends upon precipitation, cloud cover and temperature changes during the previous six hours. Data for several years from ten stations in Central Europe were examined. Theoretical mean attenuations of 10.6, 870, and 1250 μm by water drops were calculated from a large sample of drop-size distributions in the literature. The theoretical attenuations by water vapor were calculated from the actual observed water vapor distributions (see Fig. 10). The two longer wavelengths are in window regions with respect to attenuation by water vapor. Our data indicated that average water vapor content in radiation fogs was less than in non-radiation fog (see Fig. 4, Table 4).

Attenuation of 1250 μm in fog is less than attenuation of 870 μm and 10.6 μm in both light, moderate, and heavy fog in all seasons. The relationship between 870 μm and 10.6 μm is more complex. In a light fog with a large water vapor content the attenuation of 870 μm by water vapor may be so large that the total attenuation of drops plus water vapor is larger at 870 μm than at 10.6 μm . In moderate and heavy fogs attenuation of 10.6 μm by water drops usually causes 10.6 μm to be attenuated more than 870 μm . With visibilities

of 200 m or less even radiation fogs attenuate $10.6\ \mu\text{m}$ by 30 dB/km or more, and non-radiation fogs have attenuations in excess of 70 dB/km. On the other hand the maximum attenuation of $870\ \mu\text{m}$ by water vapor in our large data sample was 24 dB/km (see Fig. 10). One must add the attenuation by fog drops to obtain the total attenuation. Fog drops in an average non-radiation fog of 50 m visibility would have an attenuation of 11 dB/km. Therefore, even an extreme attenuation of 35 dB/km for $870\ \mu\text{m}$ is less than the average attenuation of $10.6\ \mu\text{m}$ for visibilities of 200 m or less. According to Essenwanger (5) 49 per cent of German fogs in fall have visibilities less than 200 m. In winter, spring, and summer the corresponding percents are 36, 37, and 41, respectively. Therefore, one must conclude from our data that a system operating at $870\ \mu\text{m}$ penetrates most European fogs better than $10.6\ \mu\text{m}$.

ACKNOWLEDGMENTS

The authors wish to express their appreciation for the assistance by Mrs. Helen M. Boyd and Mr. Don McLain in establishing computer programs for the classification of fog. Mrs. Clara Brooks deserves our thanks for her careful typing of the manuscript.

REFERENCES

1. Davies, M., G. W. F. Pardoe, J. Chamberlain, and H. A. Gebbie, 1970: Submillimetre- and millimetre-wave absorptions of some polar and non-polar liquids measured by Fourier transform spectroscopy. Trans. Faraday Soc., 66, Part 2, 273-292.
2. Essenwanger, O. M., 1973: On the Duration of Widespread Fog and Low Ceiling in Central Europe and Some Aspects of Predictability. US Army Missile Command, Technical Report RR-73-9, 58 pp.
3. Essenwanger, O. M., 1973: On Spatial Distribution of Visibility and Clouds in Central Europe. US Army Missile Command, Technical Report RR-73-10, 63 pp.
4. Essenwanger, O. M., 1977: Estimation of Absolute Humidity During Morning Fog in Central Europe. US Army MIRADCOM, Technical Report TR-77-10, 22 pp.
5. Essenwanger, O. M., 1978: A Model of Fog Classification Applied to Central Europe. US Army MIRADCOM Technical Report (in preparation).

6. Essenwanger, O. M. and D. A. Stewart, 1976: Fog Climatology in Central Europe and Inferred Propagation Characteristics. Proceedings of the Optical-Submillimeter Atmospheric Propagation Conference sponsored by the Office of the Director, Defense Research and Engineering 6-9 December 1976, Vol. 1, Unclassified Papers, pp 165-179.
7. Hale, G. M. and M. R. Querry, 1973: Optical constants of water in the 200-nm to 200- μ m wavelength region. Appl. Optics, 12, 555-563.
8. McCoy, J. H., D. B. Rensch, and R. K. Long, 1969: Water vapor continuum absorption of carbon dioxide laser radiation near 10 μ . Appl. Optics, 8, 1471-1478.
9. Mészáros, A., 1965: Concentration et Distribution Dimensionnelle des Gouttelettes de Brouillards Atmosphérique. J. Rech. Atmos., 2, 53-64.
10. Mie, G., 1908: Beiträge zur Optik Trüber Medien, Speziell Kolloidaler Metallösungen. Annalen der Physik, 25, fourth series, 377-445.
11. Pilié, R. J., E. J. Mack, W. C. Kocmond, C. W. Rogers, and W. J. Eadie, 1975: The life cycle of valley fog, Part I: micro-meteorological characteristics. J. Appl. Meteor., 14, 347-363.
12. Stewart, D. A., 1977: Infrared and Submillimeter Extinction by Fog. US Army Missile R&D Command Technical Report TR-77-9, 51 pp.
13. Webster, D. W., 1973: Military Potential of the Extreme Infrared: Technology and Applications. Naval Weapons Center Report TP 5565. Catalogued under AD 915347.

UNSTABLE RESONATORS FOR ARMY LASER DESIGNATORS (U)

THEODORE F. EWANIZKY, Mr.
US ARMY ELECTRONICS RESEARCH AND DEVELOPMENT COMMAND
FORT MONMOUTH, NJ 07703

1. Introduction

The objective of this program was to investigate optical resonator properties relevant to the improvement of laser systems suitable for military designator/rangefinder applications. A study of the particular characteristics of a negative-branch, unstable resonator laser was performed. Laboratory experimentation revealed unique aspects which offer a new approach to solutions of existing technological problems of: beam quality, operational efficiency and reliability, and boresight stability.

In an initial effort to extend the results of the laboratory experimentation, an existing "Handheld Laser Designator" (HLD) was retrofitted with an unstable resonator laser module. A performance evaluation, and comparison with the standard unit using a conventional resonator system, revealed particular advantages in this preprototype design.

The conclusions drawn from this investigative effort highlight the unique characteristics of this new approach and offer insight to possible techniques for improving reliability of laser designator systems.

2. Technical Background

The primary function of a laser resonator is to maintain, through multiple reflection, the high intensity internal field necessary for the stimulated emission process. A fortunate

by-product of this process is the formation of a coherent, monochromatic, narrow beam of radiation which is the hallmark of a laser and distinguishes it from all other known sources of radiation.

However, this statement is only qualitative in nature, and it is not until the actual performance of real lasers is measured against specifications for real applications that the quantitative shortcomings of state-of-the-art systems can be appreciated. Military applications, in fact, probably generate the most stringent requirements and specifications.

Army or Tri-Service applications for lasers range from very small rangefinder units, lightweight handheld target designators to larger, sophisticated units which may be tripod or vehicle mounted and have multi-functional capabilities. Although these systems may be very different in operating characteristics, due to their different design requirements, they share common problem areas in which improvement is necessary. The areas to be covered in this work are those which specifically require advances or improvements in laser resonator technology.

Military designator/rangefinder systems which employ flashlamp excitation and crystalline laser media (such as neodymium-YAG), are typically Q-switched and generate pulses of radiation with megawatts of peak power in approximately 10-20 nanoseconds duration. They must be capable of reliable repetitive pulse operation, as well as single-shot, from a "cold" start through a continuous time period dictated by mission requirements. The problem areas generated by these operational characteristics, which still confront current laser systems, are identified in the following list and will be addressed in the text. (a) Overall operating efficiency: This factor is important for all systems but especially for lightweight units with limited power resources. (b) Beam Quality: This term encompasses not only far-field beam divergence, but must also be concerned with near-field spatial coherence and freedom from spurious radiation. This factor determines the facility by which the "raw" laser beam emitted from the oscillator can be subsequently optically processed and implies the degree to which the resonator discriminates against unwanted modes of oscillation. Resonator mode discrimination is important since spurious internal radiation can cause not only a loss in true operating efficiency, but also produce "hot spots" within the resonator which can exceed materiel damage thresholds and thus deteriorate optical components within a short operating time span. (c) Initial expense and long-term reliability: These factors are related not only to fabrication quality, but also to inherent

superiority and simplicity of design, since improvement may sometimes be accomplished by merely reducing the number of components, if a superior design eliminates unnecessary complication required by an inherently more sensitive optical configuration.

(d) Maintenance of accurate boresight: This factor is critical for military applications and represents a primary area of consideration for improvement by resonator technology. (e) Overall performance over temperature change: This factor encompasses the totality of performance of the individual operating characteristics of the laser system and represents a major challenge in meeting rigorous military specifications. Ideally, lasers designed for field application should be operable over a relatively large ambient temperature range and/or over an extended operating time with minimum deterioration of output energy, beam quality, or boresight accuracy. The degree to which a laser system has these qualities is related to the relative sensitivity of the resonator design to optical misalignment caused by thermal variation.

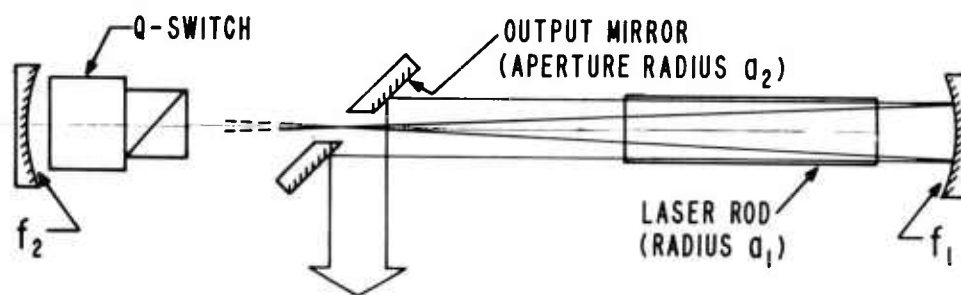
3. Technical Approach

This work was specifically concerned with improvement of the performance characteristics of Q-switch neodymium-YAG lasers operating in an energy category and repetition-rate suitable for designator/rangefinder applications.

Based on previously successful experience (1) with rhodamine 6G dye lasers, the negative-branch form of confocal, unstable resonator was investigated for use with the crystalline system. Although numerous publications describing unstable resonators have been presented in recent years, (2) their primary application has been virtually restricted to large-bore gas lasers - exploiting the inherent feature of arbitrarily large mode volume. However, more important to our interests, we found that the unstable resonator has the additional beneficial capacity of producing near diffraction-limited beam quality in short pulse-duration, high-gain lasers (3, 4, 5). An important consideration which particularly motivated the choice of the negative-branch configuration was provided by its unique feature of relatively large misalignment tolerance. Unstable resonator laser performance was not significantly degraded by mirror misalignment far beyond the point where laser action completely ceased with a conventional plane mirror-porro prism resonator (3).

4. Experimental Results

Figure 1 schematically illustrates the arrangement of optical components for the laboratory experimentation. Pertinent design parameters and formulas are defined (2,3). A $1/4 \times 2\ 1/2$ in. neodymium-YAG laser rod was pumped by a linear flashlamp in an elliptical pump cavity. Q-switching was accomplished with either a Pockels cell or a plastic sheet saturable absorber. Output coupling corresponding to an optical magnification of $m = 2$ was found to be optimum.



$$\text{OUTPUT COUPLING FRACTION } c = 1 - 1/m^2$$

$$\text{OPTICAL MAGNIFICATION } m = f_1 / f_2$$

$$\text{CONFOCAL RESONATOR LENGTH } L = f_1 + f_2$$

$$\text{OUTPUT MIRROR APERTURE RADIUS } a_2 = a_1 / m$$

Figure 1. Physical arrangement of the negative-branch, unstable resonator laser and definition of pertinent parameters.

4.1 Beam Quality Measurements

Figure 2 shows the quasi-far-field intensity profile of the laser beam as detected by a linear array of pyro-electric elements at a distance of 17 m. At this distance, the Fresnel Number is about 0.33, and the array pattern describes the far-field distribution with reasonable accuracy. A Fraunhofer pattern is

clearly evident, and the diffraction angles to the first and second minima were immeasurably close to the theoretical values of 0.174 and 0.405 mrad, respectively.

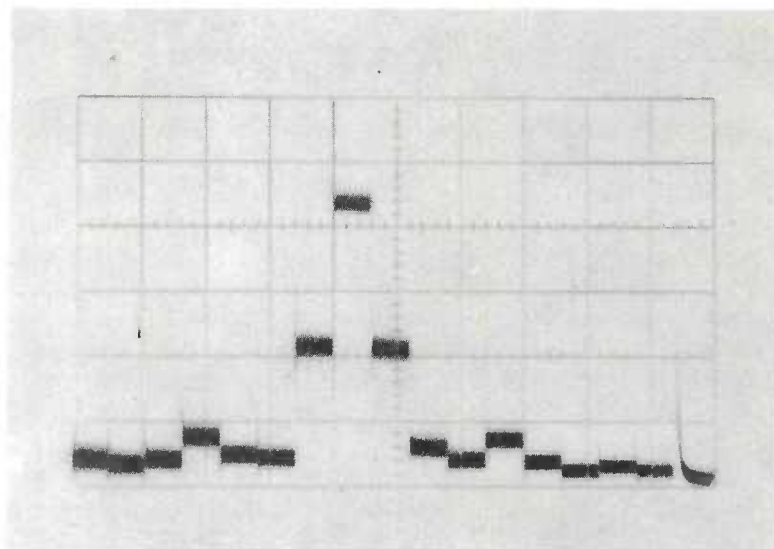


Figure 2. Far-field intensity profile of the unstable resonator laser beam as detected by a linear array of pyro-electric elements.

Figure 3 shows examples of laser burn patterns that were produced on Polaroid prints at various distances. In general, they were well resolved and clearly traced the evolution of the beam through the Fresnel diffraction regime. These observations implied that the output beam had excellent spatial-coherence characteristics and beam quality.

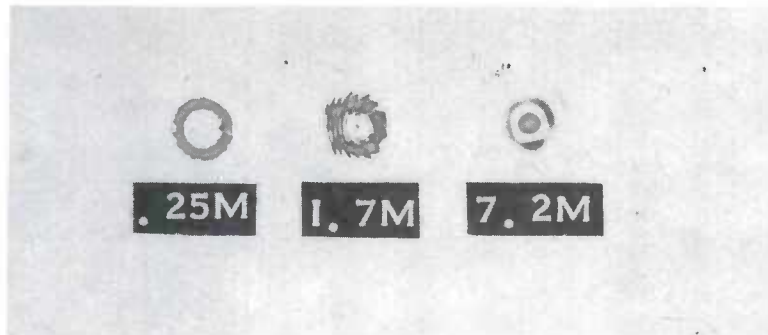


Figure 3. Laser burn patterns at 0.25 m, 1.7 m, and 7.2 m from the laser. A Fraunhofer pattern was observed to occur at a distance approximately 6 m, corresponding to diffraction theory.

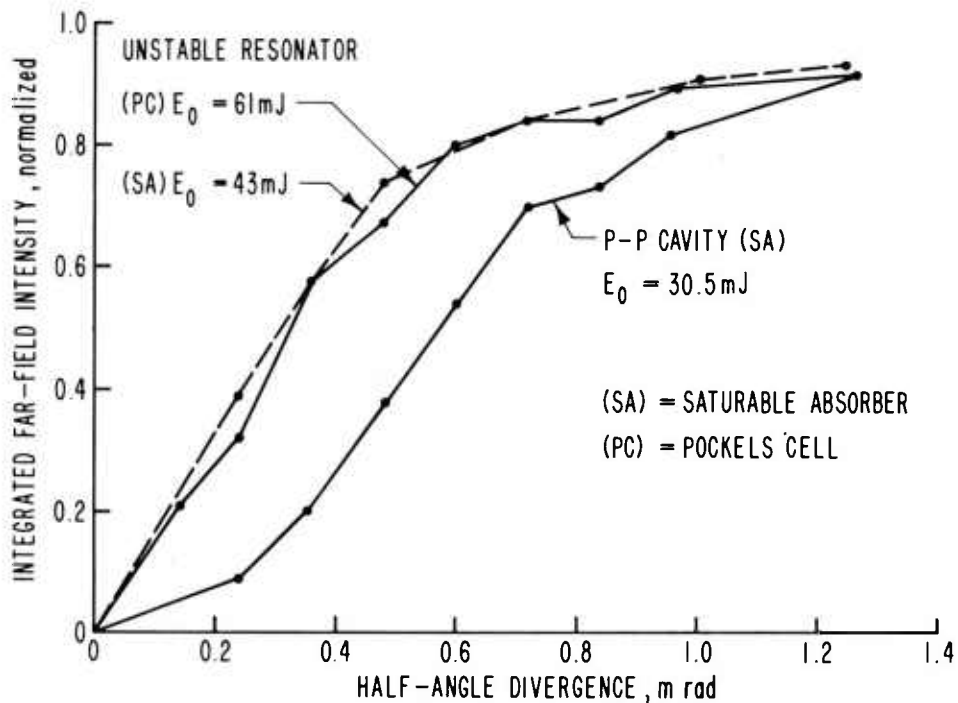


Figure 4. Comparison of energy distribution measurements for the laser with either an unstable or parallel-plane resonator.

Quantitative measurements of the emitted energy distribution were made by measuring the fractional transmission through apertures placed at the focal plane of a well-corrected plano-convex lens. Figure 4 shows the results of such measurements using either a Pockels cell or saturable absorber Q-switch. The results of corresponding measurements made with a parallel-plane mirror resonator are shown for comparison.

This data graphically illustrates the superior "brightness" achieved in the far-field beam by the unstable resonator. Note that although the divergence angle for the 90% energy distribution is only slightly less than that for the conventional resonator, the 20-70% energy points are achieved at significantly smaller angles.

4.2 Misalignment Tolerance

4.2.1 Since the optical axis defines the orientation of the resonator mode, it may be expected that angular mirror misalignment will result in beam steering. If the short focal length end mirror is misadjusted by an angle θ so that the optical axis makes some angle ϕ with respect to the geometrical axis of the resonator, the rate of beam steering is given by simple geometrical calculations to be $\phi/\theta = 2/(m + 1)$. It was found that actual laser performance closely concurred with the results predicted by these geometrical considerations. The beam steering rate was measured to be 0.61 mrad/mrad while the above equation gives the value 0.67. Over a measured range of θ as large as 4.5 mrad, output energy diminished by only about 1/3. Beam divergence characteristics changed significantly only when misadjustment was so severe that internal vignetting was eventually produced.

4.2.2 Actual resonator length was empirically determined by adjustment for optimum beam collimation. In general, length misadjustment was found to be not critical and a change up to about one-half percent produced no discernible difference in either output energy or beam divergence.

With respect to the radially expanding, outgoing wave, the resonator appears to be an expanding telescope. Changing its length merely focuses or defocuses the radiation. Consequently, this feature may be used as a simple form of adaptive optics that can compensate for thermal "lensing" of the laser rod or other such thermal effects in the resonator.

4.2.3 The major significance of these observations is that angular or length misalignment produces effects which can be explained by a simple geometric model and can be corrected by correspondingly simple means. For example, by readjusting only the short focal length mirror, both bore-sighting and collimation adjustments can be performed without necessity for additional adjustment in the other resonator optics and without loss in operating efficiency.

5. Packaged, Fieldable Prototype

The objective of this extension program was to investigate the characteristics of the unstable resonator design in an actual, packaged laser designator. An existing "Handheld Laser Designator" (HLD), shown in Figure 5, was retrofitted with the unstable resonator and a hybrid form of pump cavity (in which the flashlamp is conduction cooled, with the jacketed laser rod liquid cooled). The original HLD utilizes a conventional laser resonator and gas cooled pump cavity.

Figure 6 is a schematic illustration of the optical configuration actually packaged in the laser module. The raw beam from the laser module is directed through a pair of optical wedges, for bore-sight adjustment, to the standard HLD six-power telescope. Because the resonator module was retrofitted to an existing unit, physical constraints prevented realization of the new design concept for alignment optimization as described in the previous section. However, it was expected that this retrofitting would provide a fair comparison with conventional engineering approaches and thus strengthen the technological base for future development.

Table I summarizes the results of a performance comparison with a standard HLD unit. Both units were operated at 10 pulse/sec. In general, it is seen that the unstable resonator version produced superior beam radiance and equaled the standard version in operating efficiency.

Figure 7 shows a plot of beam divergence variation with pulse repetition frequency (PRF) for the raw beam output of the laser module. Comparison is made with a conventional resonator which employed an intracavity compensating lens necessary to reduce thermally induced deteriorating effects. Although the unstable resonator was fixed in adjustment and employed no explicit compensating elements, it gave a clearly superior performance and demonstrated its inherent misalignment insensitivity.

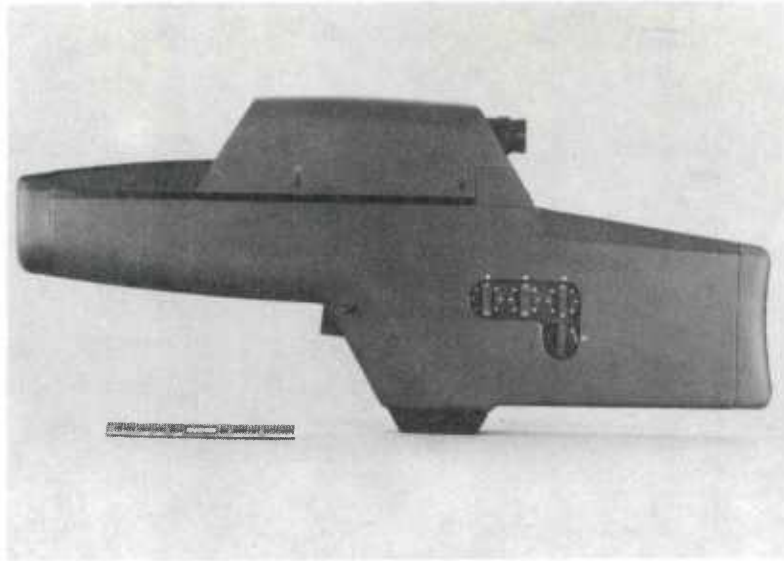
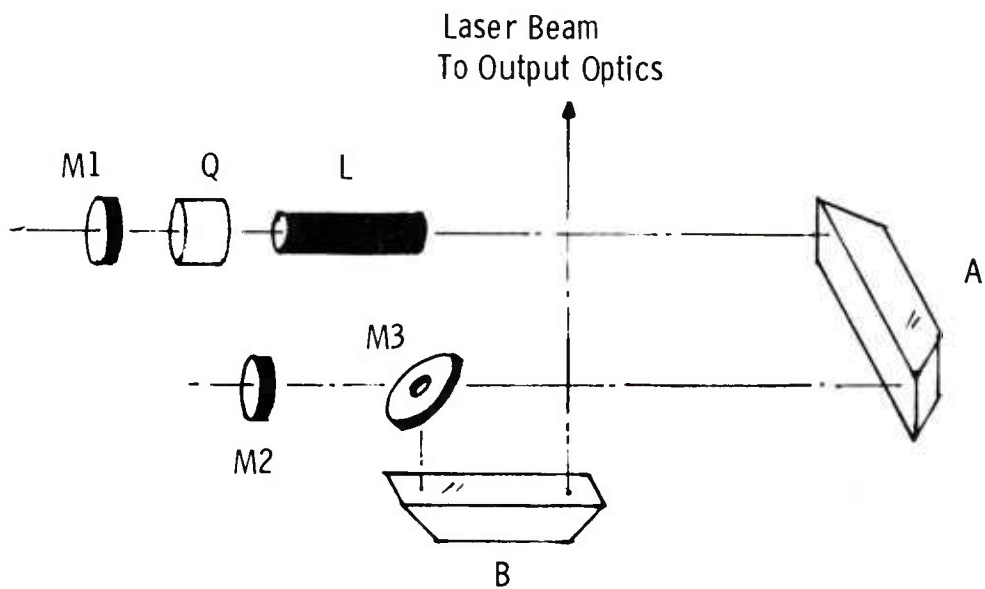


Figure 5. Handheld Laser Designator (HLD).



M1 - Long Focal Length Mirror
M2 - Short Focal Length Mirror
M3 - Output Coupling Mirror

Q - Q-Switch/Polarizer
A&B - Folding Prisms

Figure 6. Schematic illustration of the unstable resonator optical arrangement, as packaged in the retrofitted laser module.

Table 1. Performance comparison of standard HLD Designator with retrofitted Unstable Resonator/Hybrid Pump cavity version.

	HLD	Unstable Resonator
Energy/pulse	58 mJ	60 mJ
Pulse Duration (FWHM)	22 ns	14 ns
Peak Power	2.6 MW	4.3 MW
Full-Angle Beam Divergence		
50% Energy	0.39 mrad	0.37 mrad
75% Energy	0.58 mrad	0.48 mrad
90% Energy	0.82 mrad	0.65 mrad
Shot-to-Shot Amplitude Stability	$\pm 10\%$	$\pm 1\%$

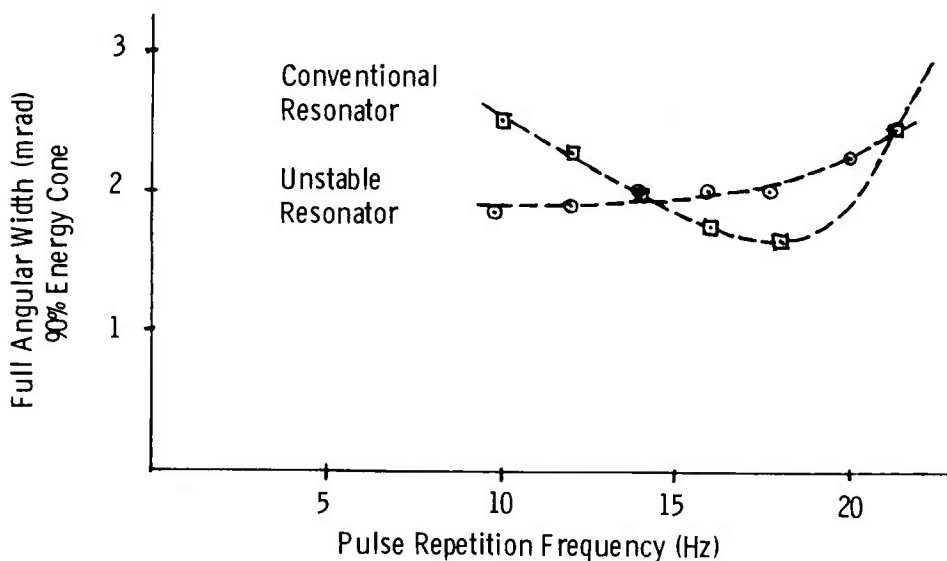


Figure 7. Comparison of "raw" beam divergence from the unstable resonator and the compensated conventional resonator as a function of pulse repetition frequency.

6. Conclusions

Experimental results of a Q-switched, neodymium-YAG laser in the designator energy category have been presented. A performance comparison between an unstable resonator, retrofitted designator and a conventional unit has been described. The experience and information gained have shown that the unstable resonator system can have a significant impact on laser technology in the following areas:

a. Beam quality: Improved beam divergence and high radiance can be achieved. Both of these parameters are important for improvement of designator target discrimination characteristics.

b. Resonator Mode Discrimination: This factor is important in producing good output beam quality. In addition, the dominant character of the internal flux reduces spurious oscillation which can produce "hot spots" responsible for both short and long term deterioration of optical components, with consequent loss of reliability.

c. Performance insensitivity to misalignment: It has been shown that the unstable resonator has inherent qualities which reduce sensitivity of output energy or beam divergence to optical misadjustment. Furthermore, because the confocal unstable resonator has focusing properties inherent to its configuration, it is feasible that this feature could be directly employed as an efficient mechanism for correcting beam divergence and output variation typically induced by either ambient or dynamic thermal variation. The approach presently employed to perform these tasks in conventional resonators is to introduce additional, corrective, intracavity optical elements. The problems associated with this approach might be obviated by the focusing characteristics of the unstable resonator.

d. The feasibility of extending the unstable resonator design concept to large bore, high Fresnel Number lasers operating to wavelengths as far as the submillimeter range would appear to have an established technological basis.

EWANIZKY

REFERENCES:

1. T. F. Ewanizky, "An Unstable Resonator Flashlamp-Pumped Dye Laser," Applied Physics Letters, Vol. 25, p. 295, Sept. 1974.
2. A. E. Siegman, "Unstable Optical Resonators," Applied Optics, Vol. 13, p. 353, Feb 1974, and many references contained therein.
3. T. F. Ewanizky and J. M. Craig, "Negative-Branch Unstable Resonator Nd:YAG Laser," Applied Optics, Vol. 15, p. 1465, June 1976.
4. T. F. Ewanizky, "A High Efficiency, Unstable Resonator, Dye Laser Module," R&D Technical Report ECOM-4388, Feb. 1976.
5. T. F. Ewanizky, "Polychromatic, Flashlamp-Pumped Dye Lasers," R&D Technical Report ECOM-4456, Jan 1977.

SCRUB TYPHUS: MECHANISM OF INFECTION
AT THE CELLULAR LEVEL

*EDWIN P. EWING, JR., M.D., AKIO TAKEUCHI, M.D.,
AKIRA SHIRAI, Ph.D., and JOSEPH V. OSTERMAN, Ph.D.
WALTER REED ARMY INSTITUTE OF RESEARCH
WASHINGTON, D. C. 20012

INTRODUCTION

Scrub typhus is an acute infectious disease of man that has caused significant morbidity in troops operating in Southeast Asia during World War II and the Vietnam conflict. Infection is caused by Rickettsia tsutsugamushi and is acquired through the bite of an infected larval mite. Rickettsiae, living organisms similar to but smaller than most bacteria, require an intracellular environment for multiplication. Rickettsiae must therefore leave infected cells and enter other cells efficiently for an infection to proceed and disease to be produced. A better understanding of the infection mechanism will hopefully lead to improvement of existing measures for prevention and treatment of scrub typhus.

Previous morphologic studies of scrub typhus rickettsiae in various types of cultured cells have revealed organisms protruding from the surfaces of undamaged cells (2,6), suggesting the possibility that rickettsiae can escape from cells by budding, leaving the host cells intact. It is known that susceptible mice as well as cultured cells can be experimentally infected (1,5), and that mesothelial cells lining the abdominal cavity of these animals will vigorously support rickettsial growth (4). In the present study, we focus on infected mouse mesothelial cells to reveal events occurring at the ultra-structural level as an infection progresses in the living animal.

MATERIALS AND METHODS

Adult female BALB/c mice (Flow Laboratories, Dublin, Va.) were divided into infected and control groups of four animals each. Infection was accomplished by injection into the abdominal cavity of 1000 50% mouse lethal doses of the Karp strain of Rickettsia tsutsugamushi in a volume of 0.2 ml. Control animals received the same volume and dilution of normal, uninfected yolk sac suspension. All animals were killed by cervical fracture on day 10 postinoculation. Portions of spleens bearing peritoneal mesothelium were collected, fixed in chilled half-strength Karnovsky's (3) glutaraldehyde, and processed for electron microscopy.

RESULTS

Rickettsiae were often numerous within the cytoplasm of mesothelial cells from infected mice. The organisms appeared as pleomorphic coccobacilli bounded by a double membrane made up of an outer cell wall and an inner cell membrane separated from each other by a narrow space. The interior had a mottled, granular structure with a loose network of fine fibrils. An occasional rickettsia was found bulging up beneath the host cell plasma membrane slightly (Fig. 1) or markedly (Fig. 2). Rickettsiae were also found free of any apparent cell connection, yet covered by an extra, third membrane indistinguishable from host cell plasma membrane (Fig. 3).

Rickettsiae bearing a third coat appeared within mesothelial cell surface invaginations (Fig. 4) and within membrane-lined vacuoles near the free cell surface (Fig. 5). On occasion, the extra membrane coat and the vacuole membrane appeared as discontinuous, electron-dense fragments encircling an intact rickettsia (Fig. 6). Rickettsiae showing central annular constriction characteristic of binary fission were always free within the cytoplasm (Fig. 7).

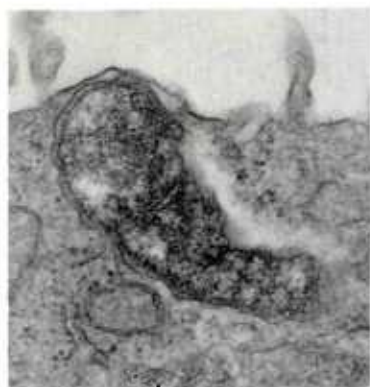


Fig. 1 A rickettsia has moved to a position just beneath the host cell plasma membrane. Double membrane of the organism is clearly visible. (x33,000)



Fig. 2 An organism protrudes from the free surface of a mesothelial cell, still covered by host cell plasma membrane. (x33,000)



Fig. 3 Rickettsia in extra-cellular environment, enveloped by third membrane derived from host cell. (x33,000)



Fig. 4 Phagocytosis of
enveloped rickettsia by
mesothelial cell.
(x33,000)

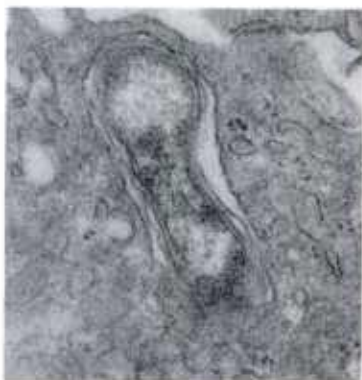


Fig. 5 Still enclosed by a
host membrane coat, an organism
appears within a phagocytic
vacuole near the cell surface.
(x33,000)

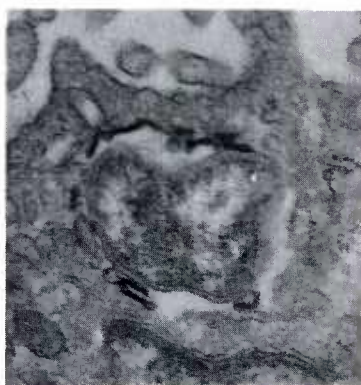


Fig. 6 The host membrane coat
and vacuole membrane are
disintegrating to liberate an
intact organism into the
cytoplasm of a host cell.
(x33,000)

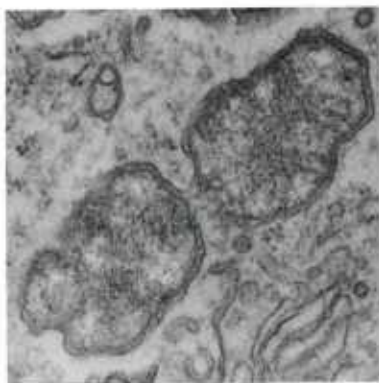


Fig. 7 *Rickettsia* free within host cell cytoplasm, undergoing binary fission.
(x33,000)

DISCUSSION

The plasma membrane of the host cells appeared to play an important role in the spread of rickettsiae from one host cell through the extracellular environment to another host cell. Organisms multiplied in the cytoplasm of mesothelial cells, moved to the cell periphery and acquired a host membrane coat as they budded from the cell surface. Free rickettsiae enveloped by this membrane entered other mesothelial cells, apparently by a phagocytic mechanism. Entry of rickettsiae lacking this coat was not observed. After internalization, organisms escaped from the phagocytic vacuole into the cytoplasm as the vacuole membrane and host membrane coat disintegrated. Rickettsiae devoid of extraneous membranes replicated by binary fission in the cell cytoplasm.

A host-derived membrane coat enveloping extracellular rickettsiae may stabilize the organisms, protect them from the immune system of the host and promote their entry into susceptible host cells by phagocytosis.

REFERENCES

1. Catanzaro, P. J., A. Shirai, P. K. Hildebrandt, and J. V. Osterman. Host defenses in experimental scrub typhus: Histopathological correlates. *Infect. Immun.* 13:861-875, 1976.
2. Higashi, N. 1962. Electron microscopic studies of Rickettsia orientalis, p. 145-156. In T. Tamiya (ed.), Recent advances in studies of tsutsugamushi disease in Japan. Medical Culture, Inc., Tokyo.
3. Karnovsky, M. J. A formaldehyde-glutaraldehyde fixative of high osmolality for use in electron microscopy. *J. Cell Biol.* 27:137A-138A, 1965.
4. Kokorin, I. N., C. Din Kyet, N. G. Kekcheeva, and E. D. Miskarova. Cytological investigation on Rickettsia tsutsugamushi infection of mice with different allotypic susceptibility to the agent. *Acta Virol.* 20:147-151, 1976.
5. Kundin, W. D., C. Liu, P. Harmon, and P. Rodina. Pathogenesis of scrub typhus infection (Rickettsia tsutsugamushi) as studied by immunofluorescence. *J. Immunol.* 93:772-781, 1964.
6. Schaechter, M., F. M. Bozeman, and J. E. Smadel. Study on the growth of rickettsiae. II. Morphologic observations of living rickettsiae in tissue culture cells. *Virology* 3:160-172, 1957.

HUMAN PERFORMANCE UNDER CLIMATIC STRESS AND THE FALLACY
OF THE "AVERAGE" SOLDIER: POTENTIALLY SERIOUS IMPLICATIONS
FOR MILITARY OPERATIONS IN EXTREME CLIMATES¹ (U)

*BERNARD J. FINE, PH.D., JOHN L. KOBRICK, PH.D.
U.S. ARMY RESEARCH INSTITUTE OF ENVIRONMENTAL MEDICINE
NATICK, MASSACHUSETTS 01760

Investigators in the life sciences in this country typically rely on normative concepts in evaluating the outcomes of experiments; that is, averages and/or variances are compared by means of statistical methods based on data for groups, and conclusions are drawn based on the probability of occurrence of the results obtained. The conclusions taken from research so produced usually are not tempered in any way as to their specific applicability to individuals within the group and, consequently, through a process of oversimplification, come to be interpreted as being applicable to people in general, or to the average person.

The concept of the average person is particularly embedded in the military milieu. It is implicit in the "can do" concept, based on the assumption that all soldiers can perform all tasks equally well under all conditions, and pervades most military actions, policies, and decisions. The military research establishment reinforces and perpetuates these policies by generating in-house or selecting from the scientific literature, research information which is a product of normative thinking. Almost all experimentally-derived human performance information included in military manuals and bulletins is based on normative concepts of average performance; although purportedly written for the individual soldier, they really refer to the average soldier. This is extremely important to recognize, since, otherwise,

¹Human research reported herein as accomplished at the U.S. Army Research Institute of Environmental Medicine was reviewed and approved, in protocol form, by the Office of The Surgeon General for The Department of The Army in accordance with Army Regulation 70-25.

the tendency is to form a false favorable impression of the extent to which the sciences can predict, understand, or explain behavioral phenomena of practical importance to the Army.

With reference to knowledge about the effects of climatic stress on human performance, the area of expertise within which this paper is written, thirty years of normative research have left even the simplest questions about individual human capabilities unanswered. At this time, one can do no better than to say that exposure to routinely occurring extremes of heat, cold, or altitude may adversely affect some people's performance of some tasks some of the time. One cannot specifically predict which people or what kinds of tasks will be affected, or when the effects, if any, will occur, let alone state the reasons for their occurrence.

The belief in and reliance on the concept of normative behavior by the military to account for troop performance and to anticipate future outcomes can be dangerously misleading. In fact, it can legitimately be termed the fallacy of the "average" soldier. This fallacy is founded on a gross misconception of the extent to which soldiers differ in all aspects of human functioning. The practical importance of these differences has been seriously underestimated by the military and this has led to untold numbers of military casualties, performance inefficiencies, accidents, and man-machine mismatches. The fallacy is even less appropriate, and much more serious, when applied to new concepts of a future army of individually trained specialists.

The normative approach to research and its opposite, the individual differences approach, do not differ substantially in basic experimental orientation. However, they do diverge significantly in research emphasis, as well as in assumptions made about human behavior.

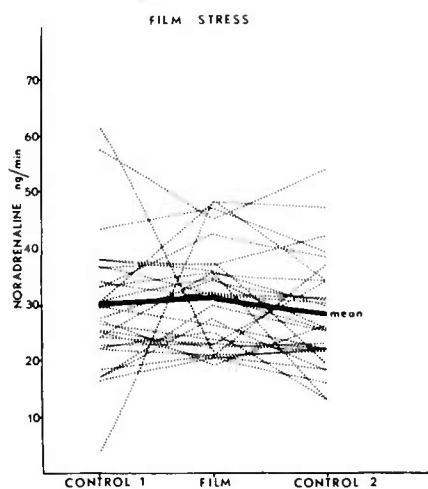
In the normative approach, similarities among people are assumed- in fact, emphasized - regarding the structure and function of mind and body. In substance, this approach assumes that "a body is a body is a body," and, therefore, that "on the average," men are interchangeable. Such ideas are probably generic to our society with its dependence on mass-production, making possible the interchangeability of parts. If machines, why not machine operators? While obvious interindividual differences such as age, sex, and weight are recognized and often taken into consideration in the design and analysis of normative experiments, the implicit assumption typically is made that the basic underlying processes which "govern" behavior, and, thus, the behavior itself, must be the same from person to person. The experiment now becomes a device to determine that performance which, once known, is assumed to be standard under the given circumstances for all individuals. Differences between individuals are assumed to average out, if one takes a random sample of subjects for study. In statistical analysis, the variance due to subjects is considered to be

unaccountable random error or "noise" in the system, which is basically irrelevant and antagonistic to the "science" of the research; i.e., it corrupts the picture of the "true" behavior. Again, random sampling is supposed to cancel the "noise."

On the other hand, the individual differences approach assumes that while people in general are grossly similar, they may differ quantitatively and qualitatively from one another in many ways. For example, some may be left-hemisphere dominant and some right-hemisphere dominant in brain activity. Some may respond to stress with increased and others with decreased excretion of the same hormone. Some may have more sensitive or stronger nervous systems than others. It is assumed that these kinds of differences can be categorized; in other words, that people can be classified into "types" on the basis of these and most other characteristics. Experiments are designed so as to maximize the possibilities of studying the differences between people and an awareness of the characteristics or "types" of test subjects is critical.

The issue is crystallized by questions such as: What kinds of information are obtained by the normative approach? Who is the "average" soldier? What is the meaning of an "average" response? The following examples bear on these questions. The first example, although taken from biochemistry, nevertheless relates to the behavioral area on which this paper is focused, and, thus, demonstrates that the principles being discussed here are general ones. Example 1 (1) shows the group mean (solid line) along with the individual excretion levels of the hormone noradrenaline (NA) of 25 soldiers measured before, immediately after, and several hours after viewing a two-hour film program depicting cruelty, violence, and torture.

Based on the group mean curve shown in Example 1, a reasonable normatively-oriented conclusion would be that the film had little effect on NA excretion. However, from the individual curves in Example 1, it is clear that: (1) approximately one-half of the subjects showed increases in NA excretion due to exposure to the film, while the other half showed decreases; (2) both initial- and post-control levels show large inter-individual differences under so-called "baseline" or non-stress conditions; and, (3) the initial levels show no relationship to,

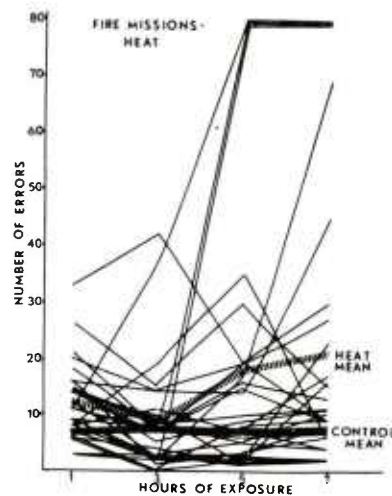


Example 1

the direction of influence of the film; that is, some subjects with initially high levels of NA nevertheless increased in NA excretion following exposure to the film, while some subjects with quite low initial levels decreased in NA excretion following film viewing.

An individual differences-oriented investigator would hesitate to conclude that there was no effect of the film, for only a few subjects actually showed no change. What, then, is the meaning of an "average" response in this situation? How many of the individual curves shown in Example 1 actually are of the same shape and magnitude as the derived "average" curve? Has the question of the effect of the film on NA excretion really been answered by the "average" curve presented in Example 1?

Turning to military performance, Example 2 shows the performance curves of 28 soldiers who participated in a heat stress study (2) at the U. S. Army Research Institute of Environmental Medicine (USARIEM). The task involved was analogous to one performed by artillery fire direction center personnel engaged in fire missions. The heavy solid line represents the average performance of the men over a period of seven hours under normal conditions (70°F, 50% RH). The heavy broken line represents the performance of the same men under severe heat stress (95°F, 88% RH). The difference between the two curves was shown to be highly significant on the basis of parametric statistics. Example 2 also depicts the individual performance curves (narrow lines) from which the averages were derived. Although only the individual curves for the heat stress are shown, it should be noted that considerable inter-individual variation in performance also occurred under normal conditions.



Example 2

Comparison of the group mean heat curve with the individual curves from which it was derived shows that the group mean curve fails to represent adequately the true behavior of the group. It is obvious that the very significant heat effect obtained was due to the reactions of only about 1/4 to 1/3 of the participants. The rest were either affected slightly or not at all.

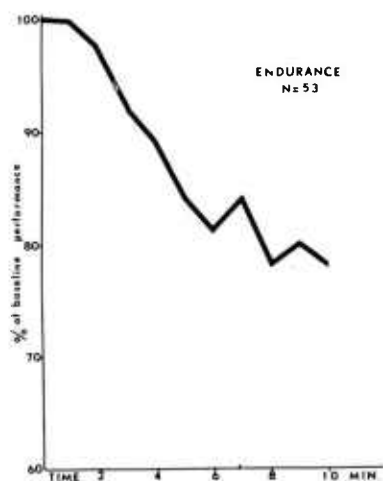
In the face of such discrepancies, one can legitimately

question the real value of average-curve performance data for serving the practical needs of the military field commander. Can he get from them a true estimate of what the effect of heat will be on the performance of his troops? When shown this kind of information, a commander typically will evince considerable interest, but then will ask: "Can you tell me ahead of time which men will be combat-effective, and which will fail?" This kind of question simply cannot be answered by use of a normative approach alone.

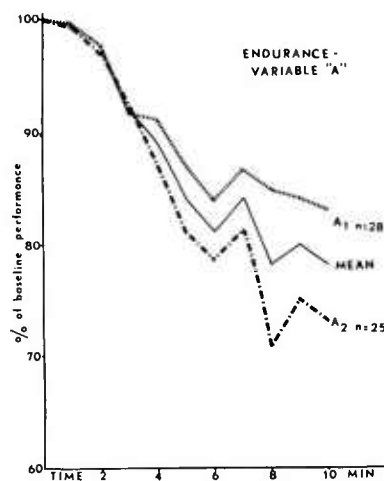
With questions like this, the commander is implicitly seeking information such as that shown in Examples 3, 4, 5, and 6. These examples are derived from actual (unpublished) data obtained in a research study at USARIEM.

Example 3 shows minute-by-minute group mean performance of 53 soldiers on an extremely trying test of physical endurance, that of attempting to squeeze a hand dynamometer for 10 minutes at a target level of tension previously determined to be virtually impossible. The data are expressed as percent of the target level achieved.

Example 4 illustrates the same performance, except that the data have been plotted for two sub-groups of the 53 subjects, separated on the basis of scores on a personality variable, referred to here as Variable A for illustrative purposes. Group A₁ are those subjects scoring above the median, and Group A₂ those scoring below the median of the "A" distribution. One can observe a large difference in performance between the curves of Groups A₁ and A₂.



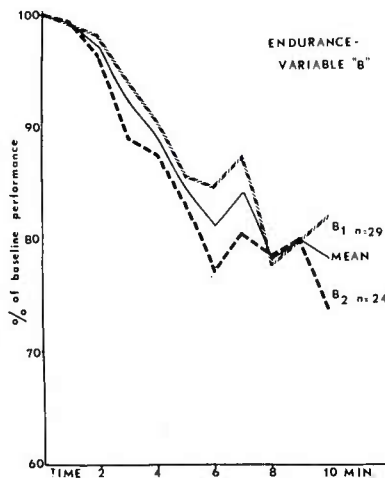
Example 3



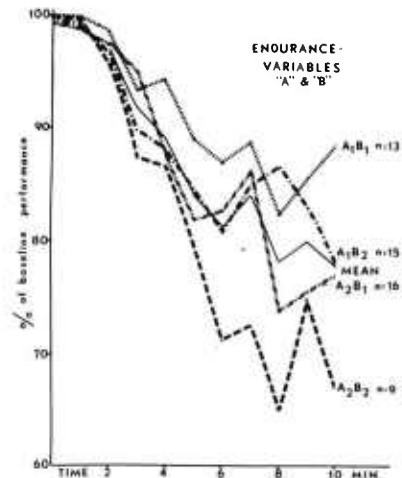
Example 4

In Example 5, the same index of performance is plotted for a second variable, "B". Sub-group B₁ involves those scoring above the median of the distribution of this personality variable, and sub-group B₂ includes those scoring below the median. Again, differences between the two groups can be seen, although not as large as in the case of Variable A.

Logically, it would be expected that subjects scoring above the median on both variable A and variable B would be the best performers, and those scoring below the median on both A and B would be the poorest performers. The data shown in Example 6 indicate that this actually was the case. The large differences in performance between groups A₁B₁ and A₂B₂ show that the original average-curve in Example 3 has little meaning.



Example 5



Example 6.

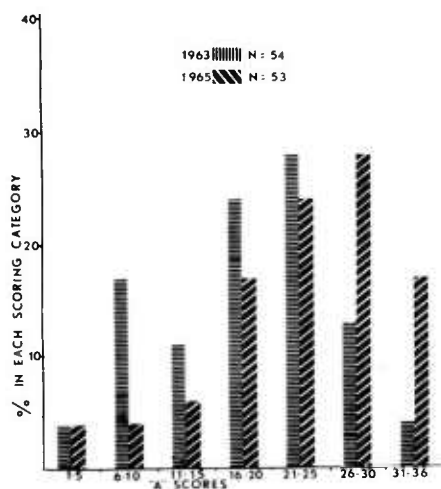
One should not conclude from the above discussion that all individuals perform consistently with a variety of tasks or stressors. The converse is more likely to be true. Individuals who excel on a task in the cold may be poor performers on the same task in the heat or at high altitude. Those who excel at high altitude may be poor performers in the heat, and so on. The problem of sorting out which kinds of individuals can best perform specific kinds of tasks in various climates will probably be best resolved by an application of the individual differences approach.

It has been noted previously that individual differences are assumed to average out in the normative approach, particularly when the group being studied is a random sample drawn from a large

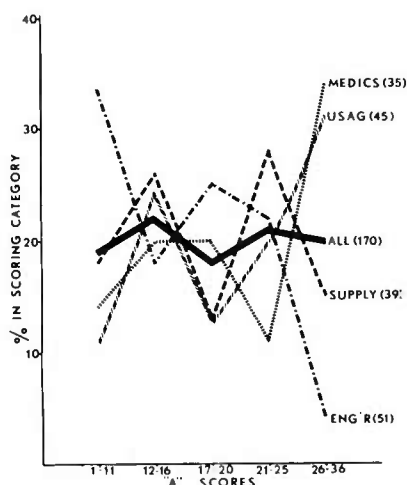
population. While this may be true in theory, such differences rarely average out in actual practice, particularly in the usual study which typically employs small numbers of subjects.

The risks inherent in this assumption are illustrated in Example 7 which shows the distribution of Variable A in two samples of nearly identical size, both drawn from Fort Devens, Massachusetts, one in 1963 and one in 1965. It is obvious that the two samples differ substantially. Given the relationship between Variable A and performance shown in Example 4, two quite different "average" curves would be obtained, depending upon which of the two samples, 1963, or 1965, was used in the study.

Example 8 shows yet another source of error which users of the normative approach are usually unaware of, or assume will average out. The average scores of 170 soldiers tested on Variable A at Fort Devens in 1972 are shown by the heavy solid line. It can be seen that the scores are evenly distributed throughout five scoring categories. However, when the subjects were divided according to the units from which they came, dramatic differences became apparent between units in the distribution of Variable A scores. Therefore, a sample of subjects selected from the Engineers unit will be very different from a sample selected from the Medical unit with respect to Variable A. If one is studying physical endurance, as depicted in Example 3, and the subjects are primarily from Engineer units, a very different average performance curve will be obtained than if the subjects were from Medical units.



Example 7



Example 8

All of these points would be academic except for one major consideration. Variables A and B above are actually psychological "individual difference" variables which have been shown to be related to many militarily relevant types of performance, such as; motor performance (3,4); vigilance (5,6); detecting booby traps in jungle terrain (7); identifying targets in aerial photographs (8); identifying targets in the field (9); motor vehicle accidents (10); monitoring visual information (11); and illness at altitude (12).

Failure to take into account these research findings has very likely resulted in many military personnel being exposed to avoidable hazards. For example, no military commander would send a soldier with very poor eyesight on a difficult reconnaissance mission. Yet some line troops with perfect visual acuity have been found to have considerable difficulty in detecting hidden or camouflaged targets (9). Undoubtedly such personnel are sent on patrol and other types of combat missions. Exposing those individuals to combat situations not only puts them at greater than usual risk, but also endangers other personnel who may be dependent upon them for critical perceptions.

While the studies cited above relate to military performance and are of the individual differences type, very few of them stem from programmatic research. Rather, they reflect the efforts of a relatively small number of investigators, the majority of whom are not allied with the military research establishment. In view of the issues raised above, what the Army appears to need is a strongly supported research effort incorporating the individual differences concept, and directed toward specific goals. Within the area of climatic-oriented research, the need for such an effort can be demonstrated particularly well by looking at cold weather military operations in Alaska, and by contrasting our situation and outlook in this regard with those of the Soviet Union (13).

Years ago, experience gained in previous military operations; e.g., World War II Europe, and Korea, was used to justify the need for more research on the effects of cold exposure on man. However, three significant aspects of current affairs now make the need for such research even more important:

1. The strategic value of the Alaska pipeline, in view of the world oil situation, and the need for its direct policing and security. The vulnerability of the pipeline was graphically demonstrated by the ease with which it was sabotaged on 15 February 1978.

2. Reports of Russian capability for military operations at Division strength in the Arctic for extended periods of time.

3. The unanimous opinions of Arctic military experts that the U.S. Army cannot presently conduct extended military operations in extreme cold.

A broader perspective on the comparative capabilities of the

U.S. and the Soviet Union for cold weather operations can be gained by considering the fact that the Soviet Union has had millions of people living under severe Arctic and sub-Arctic conditions for a substantial period of their history. For example, Murmansk, a city of over 500,000 population, is above the Arctic circle.

On the other hand, Alaska is only sparsely populated. Other areas of the United States which experience extremely cold weather also are not highly populated and, furthermore, lack the added stresses of the polar night and its attendant adjustment difficulties.

Consider not only the size differential between the cold-dwelling populations of the two countries, but also the length of time of residence of the inhabitants. Except for natives, trappers, and a few other hardy individuals, most cold-dwelling Americans are relative newcomers to such a climate, particularly when compared to the Russian people who have lived in the extreme cold for centuries. One might even expect the latter to have undergone some natural selection processes resulting in better cold survival. Reasoning from this, one must expect Russian cold-weather troops to be drawn from among these cold-dwelling peoples.

The comparative standards of living of the two countries should also be considered. The Russian standard is substantially lower than ours, and, as a result, the people are accustomed to expect extreme hardship in their daily lives. From this, they undoubtedly have acquired many simple and effective skills for coping with the environment, and for making the most out of their relationships with it. Coming from such a background, transition to the rigors of Arctic life, for the Russian, apart from family separation, would probably be relatively easy.

On the other hand, the American soldier assigned to the Arctic is transplanted abruptly from a temperate or even hot climate, leaving not only family but also comparative luxury and thermal comfort. There he is immediately subjected to many stressors not previously experienced. Despite all, he is expected to be highly motivated and productive.

These assignment practices stem directly from U. S. Army policy and doctrine which dictates that with leadership, training and support, the American soldier can function in any climate. But how effectively can he function? Can all men make this extreme transition?

The extremity of the transition is seriously underestimated. Individuals, most of whom dislike the extreme cold and some of whom already have psychological problems, are put into the field in small groups, usually isolated, under conditions of prolonged semi-darkness. They have little or no experience with the Arctic and only rudimentary survival training. They must contend with vehicles, weapons and communications systems which often become inoperable in Arctic cold. Furthermore, they feel no real purpose for being where they are.

Temperatures range from -30°F to -80°F combined with wind, a completely unforgiving environment where human flesh can freeze in less than one minute and where one mistake, theirs or someone else's, can mean loss of limbs or death. It should hardly be surprising to find that under these conditions some individuals literally cease to function, or begin to behave in bizarre ways.

To properly understand the problem, one must distinguish between the concrete and the abstract aspects of cold. Concrete aspects refer to effects of being cold on the individual's organic and behavioral functions, such as core and skin temperature, speed of chemical reactions and neural transmission, metabolism, psychomotor and mental performance. Most research, both laboratory and field, has focused on concrete aspects almost exclusively.

However, the abstract aspect of cold may be even more critical but has not been examined systematically. By abstract aspect of cold is meant that extreme cold, in the range of -40°F and below, represents a threat to a man's life which at the very least rivals the hazards of combat. The evidence, most of which is anecdotal rather than scientific, points to the fact that the effects of extreme cold resemble those of combat in that cognitive and motor aspects of the brain appear to become dissociated in some individuals.

Consider the following anecdote: A squad on an exercise is caught in a sudden temperature drop, down to -80°F. A man who thinks his feet are freezing tells the leader to stop. When the group halts to make camp, the leader orders the man with cold feet to gather wood and start a fire. The man simply stares back at him, apparently uncomprehending, hugging himself. After several further attempts to get the man moving, the leader gets the wood himself and starts a fire. When questioned later, the man indicated that he had heard the leader, understood everything that was said, realized that he would have been better off if he had followed orders and kept active, but was unable to move.

Other anecdotes differ in situation, but are similar in substance. Men have burned their survival gear to keep warm instead of leaving the fire to get wood which was readily available nearby; men, including commanders, have become virtually inert at temperatures of -80°F and have failed to perform chores necessary for their own survival, such as unloading nearby trucks laden with food and fuel supplies; and men who normally function well in the cold have started to have psychological problems when told to remove a mitten to work on a piece of equipment.

Rioch (14) has addressed this general problem, speaking of the "body failing to support the brain," and cites S.L.A. Marshall's Omaha Beach story: "...We had more beaten troops there than we had troops that were successful. It was only a small fraction of Americans that pulled us out of a great disaster. We had companies that

folded completely. Men who were obviously physically strong drowned because they did not have the strength to pull themselves out of the water which was only two or three feet deep. I'm convinced that we lost more men from drowning than we did from enemy fire." Rioch states: "Whether or not this was due to the sudden exposure to extreme danger following a long passage in a landing craft is a matter of speculation. It may be noted, however, that equivalent factors would be involved in flying paratroopers from distant fields into the Arctic for a dangerous jump."

Thus, danger seems to produce a separation of cognition from motor behavior in some men, while others can successfully withstand such stress. How do we identify these latter types of men? Could we use them as the nucleus of special stress-resistant troops for duty in the Arctic and other strategic areas of importance?

Must the disasters of Korea be repeated? Consider an incident in the Korean War, at a temperature of -18°F , as related by S.L.A. Marshall: "They decided to move on. Halfway up the next ridge the column stopped. Davis moved up front to see what had happened. Nothing had happened, except that they could not move anymore, at least they thought they couldn't. So Davis took the lead...and...they got to the top of this ridge and then, by the witness of Davis and his fellow officers, they saw happen what they never expected to see happen among Marines. As each company came over that ridge, the men fell flat on their faces in the snow and not a man would move...."

There is very little research relating directly to this kind of behavior under stress. The normative approach does not properly conceptualize this kind of problem, but, rather, seeks universal principles applicable to all men; unfortunately, there are very few such principles. A more individual-specific approach is needed, one which emphasizes individual differences.

In contrast, the Russians are strongly emphasizing the individual in their research. For example, a publication (15) about bioelectric activity of the brain and its relation to mental processes under extreme cold conditions concludes with the following (underlining is ours):

"This conclusion is confirmed by results of psychometric investigations if we take into consideration the link between the degree of introversion and the magnitude of nervous system strength....

All of the abovesaid confirms that theoretical studies of the dynamics of man's psychophysiological conditions should be determined taking into account the general and specific type of higher nervous activity of the examinee. An objective understanding

of changes in the functional properties of the CNS in the course of adaptation (to cold) may be achieved only on the basis of combined psycho- and neurophysiological investigations by quantitative methods; this will help solve the problem of psychological selection of personnel."

Notice that "introversion" referred to in the quotation is the same as Variable B in Example 5.

The Russians evidently consider selection of personnel for cold-weather duty to be a very serious issue (the article is entitled "Urgent Tasks of Psychophysiological Studies in the Antarctic"). The quotation concerns selection of scientific personnel for Antarctica. One might assume there must be even more sophisticated material pertaining to military operations in the cold.

An analysis of the differences between Soviet and American cultures presents an apparent paradox with respect to the behavioral science research approaches of the two countries. On the one hand, the Soviet Union, with its communist orientation, appears to western eyes to be a country in which the sameness of people is commonplace. Individual initiative is encouraged, but is directed collectively toward support of the state. Emphasis is placed on group identity and a classless society. People tend to dress, eat and live alike. Non-conformity in thought and expression is discouraged or punished.

On the other hand, American values seem to derive from differences between people, such that both self-expression and individual initiative are encouraged. Freedom of thought and choice are fundamental. Government exists to serve the people; individual rights are paramount and transcend all else. Dress, food preferences and life styles vary widely.

It appears paradoxical, then, that the Soviet Union, although oriented around the collective, supports an individual differences approach to research, while the United States, which encourages individuality, fosters predominately normative research.

This contrast appears paradoxical only when viewed through American eyes. If one looks analytically beneath the apparent uniformity of Russian life, a deep concern for individuality can be found, tempered by the political constraints of the totalitarian system. By similar analysis, beneath the apparent diversity and individuality of Americans, one can find considerable conformity in behavior, in the context of political freedom.

The basis for the above may lie in differences between the two countries in child-rearing practices. According to Bronfenbrenner, (16) Russia has become a matriarchal society since World War II, as a result of the decimation of its male population in that war. As one effect of this change, children now receive much affection, emotional

support, and guidance from both male and female adults. Even when children are taken from their homes at early ages and reared collectively (and politically indoctrinated at the same time), they, nevertheless, receive constant attention as individuals, and warm, loving, personal care by adults in the collectives. While in school, they compete for academic awards in teams, but heavy emphasis is placed on individuals as contributors to team success. The notion of the integrity and worth of children as individuals pervades the society; they are highly valued, and are dealt with as though small, but real, adults.

An opposite picture emerges with regard to American society. Here, while children may be thought of as individuals, they are not treated that way. In our patriarchal society, children implicitly are expected to fend for themselves, and to develop their own individualities. Open expressions of warmth, affection, and emotion are not sanctioned, particularly among males. The children receive open emotional support primarily from the mother. However, as Bronfenbrenner points out, in the typical middle-class household, the father's work schedule, the mother's social and/or work demands and the children's own activities combine to limit opportunities for maternal expression of affection.

As a result, children have developed a reliance on peer groups rather than on parents for finding strong emotional support and self-identity. Although peer group cultures superficially may give the appearance of children performing and behaving as individuals, in reality, they conceal a pervasive similarity among children. Thus, in the context of great political freedom, many of our youth tend to wander aimlessly, seeking themselves.

We are all products of our culture; it shapes our thoughts and actions, and determines how we think and approach our problems. It is not paradoxical, then, to see differences emerge between the two countries in their approaches to research; emphasis on individuality in Russia has resulted in a behavioral science with a distinctly individual flavor, while American science has evolved with a normative orientation, dealing mainly in generalities.

It is our belief that we are nearing the limits of what a normative approach to science can yield for the Army. A great deal of descriptive information about human behavior had been amassed, but it has little practical value for the selection of soldiers for duty in extreme climates. Such selection is essential if the Army is to remain competitive with other world powers. It would be sad, indeed, if we had to relive Omaha Beach, but now in deep snow at -40°F, because of an unnecessary lack of adequate information about human performance capabilities.

We have attempted to clarify the differences between the normative and individual differences approaches to research and to

indicate the direction in which military research should go. The implementation of such research by a few investigators is not sufficient. What is needed is a realization at the highest staff levels that serious problems exist with military operations in extreme climates, particularly in the cold, and that these problems can be solved only by a major change in research orientation.

References

1. Levi, L. Sympatho-adrenomedullary responses to emotional stimuli: methodologic, physiologic and pathologic considerations. In E. Bajusz (Ed.) An introduction to clinical neuroendocrinology. New York: S. Karger, 1967.
2. Fine, B. J. and Kobrick, J. L. Effects of high altitude and heat on simulated artillery fire direction center tasks. Proceedings of Army Science Conference, U. S. Military Academy, West Point, N. Y., June, 1976.
3. Venables, P. H. Change in motor response with increase and decrease in task difficulty in formal industrial and psychiatric patient subjects. British Journal of Psychology, 45, 101-110, 1955.
4. Discipio, W. J. Divergent thinking: a complex function of interacting dimensions of extraversion-introversion and neuroticism-stability. British Journal of Psychology, 62, 545-550, 1971.
5. Krupski, A., Raskin, D. C. and Bakan, P. Physiological and personality correlates of commission errors in an auditory vigilance task. Psychophysiology, 8, 304-311, 1971.
6. Moses, J. L. Field dependence and the prediction of vigilance performance. Dissertation Abstracts, 28, 1719, 1967.
7. Bucklin, B. L. Field dependence and visual detection ability. U. S. Army, Picatinny Arsenal, Engineering Sciences Laboratory, Dover, N. J., Technical Report 4137, 1971.
8. Thornton, C. L., Barrett, G. V. and Davis, J. A. Field dependence and target detection. Human Factors, 10, 493-496, 1968.
9. Fine, B. J. and Kobrick, J. L. Note on the relationship between introversion-extraversion, field-dependence-independence and accuracy of visual target detection. Perceptual and Motor Skills, 42, 763-766, 1976.

*FINE & KOBRICK

10. Fine, B. J. Introversion-extraversion and motor vehicle driver behavior. In H. J. Eysenck (Ed.) Readings in Extraversion-Introversion, Volume 2, New York: Halstead, 1971.
11. Conklin, R. C., Muir, W. and Boersma, F. J. Field dependency-independency and eye movement patterns. Perceptual and Motor Skills, 26, 59-65, 1968.
12. Fine, B. J. Personality traits as related to symptomatology and running performance at altitude under normal and drug (acetazolamide) conditions. Perceptual and Motor Skills, 27, 975-990, 1968.
13. Fine, B. J. Psychological Problems of Men in the Arctic. Psychological Program For AMRLA. Staff Report to Commander, USARIEM, Natick, Massachusetts, 5 December 1973.
14. Rioch, D. McK. Psychiatric problems of man in the Arctic. In F. R. Fisher (Ed.) Man living in the Arctic, Washington, D.C.: National Academy of Science-National Research Council, 1961.
15. Bundzen, P. V. Urgent tasks of psychophysiological studies in the Antarctic. In A. L. Matusov (Ed.) Medical research on Arctic and Antarctic expeditions. Russian translation published for National Science Foundation, Washington, D. C. by the Israel Program for Scientific Translation, Jerusalem, 1973.
16. Bronfenbrenner, U. Two worlds of childhood. New York: Russell Sage Foundation, 1970.

Mie Theory for Non Spherical Particles

Bruce W. Fowler, PhD.
USAMIRADCOM
Redstone Arsenal, AL 35809

I. INTRODUCTION

This work was motivated during a study of radiative transfer through aerosols¹ when it was realized that spherical aerosol cross sections and phase functions only could be provided by Mie theory. Because many aerosols are irregular in shape, another formalism is needed to provide the optical parameters needed for radiative transfer calculations. The literature of previous efforts on the scattering of electromagnetic waves from irregular particles is extensive², although many of these calculations were directed either to various approximate approaches or to the problem of conducting materials only. An example of both these limitations is the application of least-squares³ techniques to match electric fields at the surface of a conductor. How the least-squares solution is related to the boundary conditions is unclear.

Several recent efforts⁴⁻⁶ using integral equation formulations deserve special attention since not only are the boundary conditions properly accounted for, but numerical results are presented. References (4) and (5) are based on the scalar rather than the vector wave equation. Reference (6) is completely general, but was only applied to conducting particles. Another entirely different effort of note⁷ which will be addressed later was that of Chylek, Grams and Pinnick (hereafter Chylek et al.) Chylek et al. modify the Mie theory to account for particle irregularity on the empirical grounds that the phase functions of irregular particles do not exhibit glories. As a result, they modify the resonance behavior of the field expansion coefficients to model irregular particle behavior. Chylek et al. have presented calculations that enjoyed better agreement with experimental data than did Mie theory.

In this paper, the scattering of a plane electromagnetic wave by an irregular but cylindrically symmetric dielectric particle is solved by generalizing the original work of Mie.^{8,9} As in the original work, the expansion coefficients of the scattered and interior (to the dielectric particle) electric and magnetic fields are obtained by satisfying the boundary conditions. Because of the particle irregularity, the expansion coefficients are coupled and form a set of linear equations that must be solved numerically. Both the integral and differential equation approaches are equivalent in analogy to the calculation of the scattering amplitude in quantum mechanics by solving either the Schrodinger or the Lippman-Schwinger equation.¹⁰ Further, both approaches develop into solutions of linear algebraic equations.

It must be emphasized that both References (4)-(6) and this paper are limited in applicability to "slightly" irregular particles whose deviation from a smooth geometric shape is much less than the wavelength. This limitation on solving the scattering problem for "arbitrary" irregular particles is reviewed by Millar.² In essence the limitation arises from the fact that the scattered field solutions incorporate radial functions that become outgoing spherical wave representations in the far field, but not functions that represent scattering of light from one area on the surface of the particle to another area. Consequently, the problem of multiple scattering between areas on the particle have to be neglected. While the overall problem may be handled by dividing the particle into subparticles,¹¹ and satisfying the boundary conditions for and between all particles, this paper will be limited to only "slightly" irregular particles.

The differential equation approach is, of course, also valid for "slightly" irregular particles without cylindrical symmetry, but these calculations require more computer core space than is commonly available. In Section II of this paper, the Mie solution for a spherical particle is reviewed by developing the differential equation formalism. Section III is an extension of this formalism to cylindrically symmetric irregular particles by developing the four boundary condition equations into 4L equations that may be solved numerically for the expansion coefficients to arbitrary accuracy. In Section IV, comparisons with exact calculations for spheres and ellipsoids of revolution¹² are presented and the computation problems for more general particles are discussed. The effects of irregularity on the scattering phase functions for single particles are reviewed, and calculations performed for comparison with Reference (7) are presented and discussed.

II. PRELIMINARY

The Hertz vectors Π and Σ are related to the electric and

magnetic fields \vec{E} and \vec{H} by⁹

$$\vec{E} = i k \nabla \times \vec{\Pi} + k^2 m^2 \vec{\Sigma} + \nabla(\nabla \cdot \vec{\Sigma}) \quad (1)$$

and

$$\vec{H} = k^2 m^2 \vec{\Pi} + \nabla(\nabla \cdot \vec{\Pi}) - i k m^2 \nabla \times \vec{\Sigma} \quad (2)$$

where m^2 , the square of the complex refractive index is

$$m^2 = \epsilon + 4\pi i \sigma / \omega \quad (3)$$

ϵ is the dielectric constant, σ is the conductivity, ω is the radial frequency, k is the wave number of the monochromatic incident light, and $\omega = ck$ (c = speed of light.)

The requirement that \vec{E} and \vec{H} be transverse as $r \rightarrow \infty$ (the far field)², implies that only the radial components of $\vec{\Pi}$ and $\vec{\Sigma}$ need be considered,

$$\vec{\Pi} = \hat{r} \pi \quad (4)$$

$$\vec{\Sigma} = \hat{r} \sigma \quad (5)$$

where \hat{r} is the radial unit vector, and π and σ are known as Debye potentials.⁹ The vector wave equation satisfied by π and σ reduces to a scalar equation

$$(\nabla^2 + k^2 m^2) \begin{pmatrix} \pi \\ \sigma \end{pmatrix} = 0 \quad (6)$$

Equations (1), (2), (4), and (5) may be used to write the components of \vec{E} and \vec{H} in terms of the Debye potentials as

$$E_r = k^2 m^2 \sigma + \frac{\partial^2 \sigma}{\partial r^2} \quad (7)$$

$$E_\theta = \frac{ik}{r \sin(\theta)} \frac{\partial \pi}{\partial \phi} + \frac{1}{r} \frac{\partial^2 \sigma}{\partial \theta \partial r} \quad (8)$$

$$E_\phi = -\frac{ik}{r} \frac{\partial \pi}{\partial \theta} + \frac{1}{r \sin(\theta)} \frac{\partial^2 \sigma}{\partial \phi \partial r} \quad (9)$$

$$H_r = k^2 m^2 \pi + \frac{\partial^2 \pi}{\partial r^2} \quad (10)$$

$$H_\theta = \frac{1}{r} \frac{\partial^2 \pi}{\partial \theta \partial r} - \frac{ik}{r \sin(\theta)} \frac{\partial \sigma}{\partial \phi} \quad (11)$$

$$H_\phi = \frac{1}{r \sin(\theta)} \frac{\partial^2 \pi}{\partial \phi \partial r} + \frac{ik m^2}{r} \frac{\partial \sigma}{\partial \theta} \quad (12)$$

If the incident light is a plane wave propagating along the z axis, then

$$E_x = \exp(ikz) \quad (13)$$

$$H_y = i \exp(ikz) \quad (14)$$

which give rise to Debye potentials of the form

$$\frac{\sigma_i^1}{\pi^1} = k^{-2} \sum_{\ell} i^{\ell-1} \frac{2\ell+1}{\ell(\ell+1)} \psi_{\ell}(kr) P_{\ell}^1(\cos\theta) \begin{pmatrix} \cos(\phi) \\ \sin(\phi) \end{pmatrix} \quad (15)$$

where the superscript "i" indicates the incident wave. The general solution of Equation (6) has the form

$$S = \sum_{\ell, m} \{ \alpha_{\ell} \psi_{\ell}(kmr) + \beta_{\ell} \eta_{\ell}(kmr) \} P_{\ell}^m(\cos\theta) \{ \gamma_m \sin(m\phi) + \delta_m \cos(m\phi) \} \quad (16)$$

where $\psi_{\ell}(x) = x j_{\ell}(x)$, $\eta_{\ell}(x) = x y_{\ell}(x)$, $j_{\ell}(x)$ and $y_{\ell}(x)$ are spherical Bessel functions and P_{ℓ} is the m^{th} Legendre polynomial of order ℓ . The interior ("w") and scattered ("s") potentials may be written from Eq. (16) by noting from Eq. (15) that only $m=1$ terms will appear. Further, the interior potentials must converge to zero as $r \rightarrow 0$ so that $\beta_{\ell} = 0$, and the scattered potentials must converge to spherical waves as $r \rightarrow \infty$, so that $\beta_{\ell} = i\alpha_{\ell}$. This latter condition is equivalent to replacing ψ and η with $g_{\ell} = x h_{\ell}$, where h_{ℓ} is the spherical Hankel function of the first kind. These potentials may be written as

$$\frac{\sigma_w^1}{\pi^1} = k^{-2} \sum_{\ell} \left\{ \frac{a_{\ell} \cos(\phi)}{b_{\ell} \sin(\phi)} \right\} \psi_{\ell}(kmr) P_{\ell}^1(\cos\theta) \quad (17)$$

$$\frac{\sigma_s^1}{\pi^1} = k^{-2} \sum_{\ell} \left\{ \frac{c_{\ell} \cos(\phi)}{d_{\ell} \sin(\phi)} \right\} \rho_{\ell}(kr) P_{\ell}^1(\cos\theta) \quad (18)$$

For a sphere, the transverse continuity boundary condition⁴ requires that E_{θ} , E_{ϕ} , H_{θ} , and H_{ϕ} be continuous across the boundary of the sphere. Inspection of Eqs. (7)-(12) reveals that this boundary condition will be satisfied if

$$\frac{\partial}{\partial r} (\sigma^i + \sigma^s) \Big|_{r=a} = \frac{\partial}{\partial r} \sigma^w \Big|_{r=a} \quad (19)$$

$$\frac{\partial}{\partial r} (\pi^i + \pi^s) \Big|_{r=a} = \frac{\partial}{\partial r} \pi^w \Big|_{r=a} \quad (20)$$

$$(\sigma^i + \sigma^s)|_{r=a} = m^2 \sigma^w|_{r=a} \quad (21)$$

$$(\pi^i + \pi^s)|_{r=a} = \pi^w|_{r=a} \quad (22)$$

where a is the radius of the sphere. Equations (19)-(22) may be reduced by using the orthogonality of the P_ℓ^1 , $\cos(\phi)$, and $\sin(\phi)$ and the explicit forms of the Debye potentials to

$$i^{\ell-1} \frac{2\ell+1}{\ell(\ell+1)} \psi_\ell'(ka) + c_\ell \rho_\ell^1(ka) = \frac{1}{m} a_\ell \psi_\ell'(kma) \quad (23)$$

$$i^{\ell-1} \frac{2\ell+1}{\ell(\ell+1)} \psi_\ell'(ka) + d_\ell \rho_\ell^1(ka) = b_\ell \psi_\ell'(kma) \quad (24)$$

$$i^{\ell-1} \frac{2\ell+1}{\ell(\ell+1)} \psi_\ell(ka) + c_\ell \rho_\ell^1(ka) = a_\ell \psi_\ell(kma) \quad (25)$$

$$i^{\ell-1} \frac{2\ell+1}{\ell(\ell+1)} \psi_\ell(ka) + d_\ell \rho_\ell^1(ka) = \frac{1}{m} b_\ell \psi_\ell(kma) \quad (26)$$

where $\psi_\ell'(x) = \partial \psi_\ell(x) / \partial x$. Equations (23)-(26) are commonly solved for c_ℓ and d_ℓ and are called the Mie solution. These c_ℓ and d_ℓ are then used to calculate the cross sections and phase functions.

III. CALCULATIONS

The radius of an irregular, but cylindrically symmetric particle may be defined by

$$r = \sum_n s_n P_n(\cos\theta) \equiv a(\theta). \quad (27)$$

The normal vector to the surface is

$$\hat{n} = \frac{\hat{r} - \hat{\theta} \partial \ln(a(\theta)) / \partial \theta}{1 + \left(\frac{\partial \ln(a(\theta))}{\partial \theta} \right)^2}. \quad (28)$$

The boundary conditions are

$$\hat{n} \times \Delta \vec{E}|_{r=a(\theta)} = 0 \quad (29)$$

$$\hat{n} \times \Delta \vec{H} \Big|_{r=a(\theta)} = 0 \quad (30)$$

where

$$\Delta \vec{E} = \vec{E}^i + \vec{E}^S - \vec{E}^W \quad (31)$$

$$\Delta \vec{H} = \vec{H}^i + \vec{H}^S - \vec{H}^W. \quad (32)$$

These boundary conditions may be explicitly written as

$$\Delta E_\phi = 0 \quad (33)$$

$$\Delta H_\phi = 0 \quad (34)$$

$$\Delta E_\theta - \frac{\partial \ln a(\theta)}{\partial \theta} \Delta E_r = 0 \quad (35)$$

$$\Delta H_\theta - \frac{\partial \ln a(\theta)}{\partial \theta} \Delta H_r = 0. \quad (36)$$

Equations (33)-(36) differ from the Mie boundary conditions only in the ΔE_r and ΔH_r terms in Eqs. (35) and (36) that arise from the non spherical nature of the particle. These equations cannot be simplified in the same manner as the Mie solution by exploiting the orthogonality of the angular functions because the radius of the particle is itself a function of angle θ . Equations (31) and (32) may be used to rewrite Eqs. (33)-(36) as

$$E_\phi^S - E_\phi^W = -E_\phi^i \quad (37)$$

$$H_\phi^S - H_\phi^W = -H_\phi^i \quad (38)$$

$$E_\theta^S - E_\theta^W - \frac{\partial \ln a(\theta)}{\partial \theta} (E_r^S - E_r^W) = -E_\theta^i + \frac{\partial \ln a(\theta)}{\partial \theta} E_r^i \quad (39)$$

$$H_\theta^S - H_\theta^W - \frac{\partial \ln a(\theta)}{\partial \theta} (H_r^S - H_r^W) = -H_\theta^i + \frac{\partial \ln a(\theta)}{\partial \theta} H_r^i \quad (40)$$

The individual components of the electric and magnetic fields may be calculated in terms of the Debye potentials by use of Eqs. (7)-(12) and substituted into Eqs. (37)-(40). These resulting equations are functions only of the angles θ and ϕ , incorporating 4L unknowns; a_ℓ , b_ℓ , $c_{\ell,1}$ and d_ℓ . By calculating the overlap of Eqs. (37) and (39) with $P_\ell^1(\cos\theta) \cos(\phi)$ and Eqs. (38) and (40) with $P_\ell^1(\cos\theta) \sin(\phi)$, 4L equations in 4L unknowns may be formed. It may be noted that the ϕ integrations may be performed immediately. The resulting 4L

equations may be expressed in figurative form as

$$\sum_{\ell=1}^L B_{\ell \ell'} G_{\ell} = \sum_{\ell=1}^L F_{\ell' \ell} \quad (41)$$

where

$$B = \begin{pmatrix} A_{\ell\ell',10} & -iA_{\ell\ell',9} & -A_{\ell\ell',15} & -iA_{\ell\ell',14} \\ -imA_{\ell\ell',9} & -mA_{\ell\ell',10} & iA_{\ell\ell',14} & A_{\ell\ell',15} \\ -(A_{\ell\ell',6} + A_{\ell\ell',8}) & -iA_{\ell\ell',7} & (A_{\ell\ell',11} + A_{\ell\ell',13}) & iA_{\ell\ell',12} \\ -imA_{\ell\ell',7} & -m(A_{\ell\ell',6} + A_{\ell\ell',8}) & iA_{\ell\ell',12} & (A_{\ell\ell',11} + A_{\ell\ell',13}) \end{pmatrix} \quad (42)$$

$$G = \begin{pmatrix} a_{\ell} \\ b_{\ell} \\ c_{\ell} \\ d_{\ell} \end{pmatrix} \quad (43)$$

and

$$F = \begin{pmatrix} (iA_{\ell\ell',4} + A_{\ell\ell',5}) \\ -(iA_{\ell\ell',4} + A_{\ell\ell',5}) \\ -(A_{\ell\ell',1} + iA_{\ell\ell',2} + A_{\ell\ell',3}) \\ -(A_{\ell\ell',1} + iA_{\ell\ell',2} + A_{\ell\ell',3}) \end{pmatrix} \quad (44)$$

Each element $A_{\ell\ell',n}$ is an L by L array whose entries are the integrals given for $n = 1, 6, 11$ (by steps of five) by

$$A_{\ell\ell',n} = \int f_n(a(\theta)) \frac{\partial \ln a(\theta)}{\partial \theta} P_{\ell}^1(\cos \theta) P_{\ell'}^1(\cos \theta) d\cos \theta \quad (45)$$

where

$$f_1 = i^{\ell-1} \frac{2\ell+1}{\ell(\ell+1)} \{ \psi_{\ell}(ka) + \psi_{\ell}''(ka) \}$$

$$f_6 = \{\psi_\ell(kma) + \psi_\ell''(kma)\} \quad (46)$$

$$f_{11} = \{\rho_\ell^{-1}(ka) + \rho_\ell^{-1}''(ka)\}$$

and

$$A_{\ell\ell',n+1} = \int g_n P_\ell^1 P_{\ell'}^1 \frac{d\cos\theta}{\sin(\theta)} \quad (47)$$

$$A_{\ell\ell',n+2} = \int g_n \frac{\partial P_\ell^1}{\partial\theta} P_{\ell'}^1 \frac{d\cos\theta}{ka(\theta)} \quad (48)$$

$$A_{\ell\ell',n+3} = \int g_n \frac{\partial P_\ell^1}{\partial\theta} P_{\ell'}^1 \frac{d\cos\theta}{ka(\theta)} \quad (49)$$

$$A_{\ell\ell',n+4} = \int g_n P_\ell^1 P_{\ell'}^1 \frac{d\cos\theta}{\sin(\theta)} \quad (50)$$

where

$$\begin{aligned} g_1 &= i^{\ell-1} \frac{2\ell+1}{\ell(\ell+1)} \psi_\ell(ka) \\ g_6 &= \psi_\ell(kma)/m \\ g_{11} &= \rho_\ell^{-1}(ka). \end{aligned} \quad (51)$$

Equation (41) may be solved numerically for the a_ℓ , b_ℓ , c_ℓ , and d_ℓ , and the cross sections and phase functions may be calculated from the c_ℓ and d_ℓ using the same formulas as before.

IV. CALCULATIONS

A computer code was generated to perform the calculations described in the previous section. The integrals $A_{\ell\ell',n}$, Eqs. (45)-(51), were performed using a Gauss-Legendre Quadrature.¹³ Equation (41) was solved using a Gauss-Jordan elimination routine modified to treat complex entries.^{13,14} The cross sections and phase functions were calculated in the usual manner.^{15,16}

The theoretical development advanced in Section II is, of course, also valid for particles not cylindrically symmetric with respect to the incident plane wave. At this time, however, an extension of the theory and the computer code to consider these more general geometries is not feasible due to computer limitations. If a general particle or incident direction were to be considered, Eqs. (17) and (18) would include terms $\cos(m\phi)$ and $\sin(m\phi)$ $m=0$ to L .² The size of the array B would increase to approximately $4L^2$ by $4L^2$ since each current entry would increase by terms m and m' . Arrays G and F would increase in size from $4L$ to $4L^2$. Thus the total storage required would increase from $16L^2 + 8L$ to $16L^4 + 8L^2$. For $L^4 \gg L^2 \gg L$,

this is an approximate increase by a factor of L^2 . The increase in computer memory required for operational code (as compared to array storage) should be approximately proportional. The present code requires 2×8^5 words to initiate compilation, so that $L^2 \approx 8^4$ words and a general code will require approximately 2×8^9 words to initiate compilation. Computer memory of this extent is available only on a very few machines so that the implementation is not feasible at this time.

It should be noted that while this code requires a large amount of computer memory for initial compilation (2×8^5 words,) this does not restrict the utility of the code as much as might be expected. This code is normally exercised on the MICOM CDC 6600 computer using the SCOPE 3.4.2 compiler. This is a two pass compiler that performs some optimization of the compiled code. In this case, the optimization is significant as the computer memory required for the compiled code is only 6.5×8^4 words. Additionally, execution of the code is accomplished in about 15 seconds for a nominal particle of up to 19 terms in the radius (Eq.(17) and (18)) and Mie parameter of value nine, although compilation of the code requires about 60 seconds. Thus, the compiled code may be used to perform calculations with only moderate demands on computer memory and execution time in contrast to recompilation of the code for each particle to be considered.

The code was checked by exercising it for various sizes of spherical particles and various complex refractive indices. These results (c_p, d_p , cross sections and phase functions) were compared with the equivalent quantities calculated using a standard Mie code.¹⁷ The c_p and d_p for all calculations agreed to within the round-off error of the machine (fifteenth figure.)

An additional check was made by making calculations for prolate spheroids of semimajor to semiminor axis ratio of two. These calculations were compared with exact calculations.¹² There were minor differences between the approximate calculations of this paper and the exact calculations in terms of the fine structure of the phase functions, but the positions of relative maxima and minima and the relative magnitude of the curves agreed very well. These calculated phase functions are shown in Figure (1). While a precise estimate of accuracy is difficult, a comparison of calculated values with values from the exact curves at 10^0 increments gave an error bound of $\pm 2\%$ for value of $c = 2\pi a \sqrt{.75}/\lambda = 1, 3, 5, 7$ (a = semimajor axis) with an average error of approximately $8 \times 10^{-2} \%$. Because some of this error may be attributed to computer noise and the truncation error of the Gauss-Legendre integration, this error bound seems to be reasonable. The agreement shows that the code is useful despite the assumption of the validity of the Rayleigh Hypothesis.

The modification to Mie theory advanced by Chylek et al.

was examined.⁷ In this modification, based on the consideration that glories are not observed in experimental data, the particle irregularity is modeled by truncation of the resonances in the c_ℓ and d_ℓ in the region $2\pi r/\lambda \approx \ell$ when $\ell > 3$ in a standard Mie code. While this modification is attractive because it offers a simple, efficient computational treatment of non spherical particle scattering, the validity of the approximation has been questioned.¹⁸

As a starting point in considering the Chylek et al. modification, the cross sections and phase functions for several variations of particle size and shape were calculated. In terms of the Mie parameter,

$$x_0 = 2\pi r/\lambda, \quad (52)$$

particles of shape function

$$x(\theta) = x_0 \{1 \pm .1 P_n(\theta)\} \quad (53)$$

for $x_0 = 1, 3, 5, 7, 9$ and $n = 2, 3, 4$ were considered. A refractive index of $1.5 + 0i$ was used in this and all subsequent calculations. Cross sections both greater than and less than the equivalent Mie cross sections for the same value of x_0 were observed. The forward scatter and backscatter varied similarly, in some cases being greater and in other less. The structure of the phase functions were similar but not identical to the Mie phase functions. Although it is difficult to draw concise trends, it was noted that in general differences increased with increasing n and decreased with increasing x_0 .

Additionally, the effect of varying the strength of the deformation was examined. Shape functions of the form

$$x(\theta) = x_0 \{1 \pm a_n P_n(\theta)\} \quad (54)$$

for $n = 2$ and 3 , $a_n = .05, .1$, and $.15$ were considered. Two effects of this variation of a_n were noted; the values of the relative maxima and minima varied by as much as a factor of five although the positions of the maxima and minima varied only slightly, and the amount of near forward scattering (scattering angle less than 30°) varied by as much as a factor of five about the equivalent Mie curve. This latter effect would seem significant in terms of the experimental data reported by Chylek et al. The significance of the other effects is more difficult to assess due to the averaging process performed when the polydisperse phase functions are calculated.

The behavior of the c_ℓ and d_ℓ in the region $x_0 \approx \ell$ was studied to examine the validity of the modification of Chylek et al. Shape function of the form of Eq. (53) for $x_0 = 3$ to 4.5 by steps of $.1$, and $n = 2, 3, 4, 5$ were considered. The real and imaginary parts of c_3 and d_3 were compared with the c_3 and d_3 calculated using Mie theory. In seven of these eight calculations, the resonance curve shape is

preserved although it narrows by as much as a factor of two. Additionally, the center of the resonance curve shifts, apparently without trend, although the resonance curves (real parts) for the irregular particles always fall under the Mie resonance curves. In one case, that of $x(\theta) = x_0\{1 + .1P_3(\theta)\}$, the resonance curve is replaced by a slowly varying curve contained under the Mie curve (real part) whose real parts have a mean of approximately .3 and whose imaginary parts have a mean of approximately zero. In all other cases, the resonance curves narrow and fall under the Mie curves. The extent of narrowing appears to be approximately proportional to n . On the basis of these calculations, the modification of the c_ℓ and d_ℓ in the resonance region is not substantiated, but the numerical results of Reference (7) might be justified in view of the narrowing of the resonance curves.

Calculations were also performed for one of the polydisperse aerosol distributions reported by Chylek et al. (i.e. the first KCl.) Nine phase functions using

$$x(\theta) = x_0$$

$$x(\theta) = x_0\{1 \pm .05 P_2(\theta)\}$$

$$x(\theta) = x_0\{1 \pm .1 P_2(\theta)\}$$

$$x(\theta) = x_0\{1 \pm .15 P_2(\theta)\}$$

$$x(\theta) = x_0\{1 - .1 P_2(\theta) \pm .1 P_4(\theta)\}$$

were calculated and combined linearly with adjustable coefficients calculated using regression techniques to fit the data. Two different fitting techniques were used; one using all the data, and another using only those data for scattering angles greater than sixty degrees. A representative fit for each scheme is shown in Figure (2). These fits appear to be as good as that reported by Chylek et al. It is noted that the fit is not sensitive to the choice of combinations of phase functions used.

The insensitivity of the fits to variation in the phase functions used in the fitting calculations seems to indicate that a more elaborate calculation for more irregular shaped particles averaged over orientations would not significantly improve the results in terms of the cost of such a calculation. In the same vein, the Chylek et al. modification seems to be adequate for many calculations. It appears that more accurate calculations should be necessary for aerosols with much less random shape and orientation.

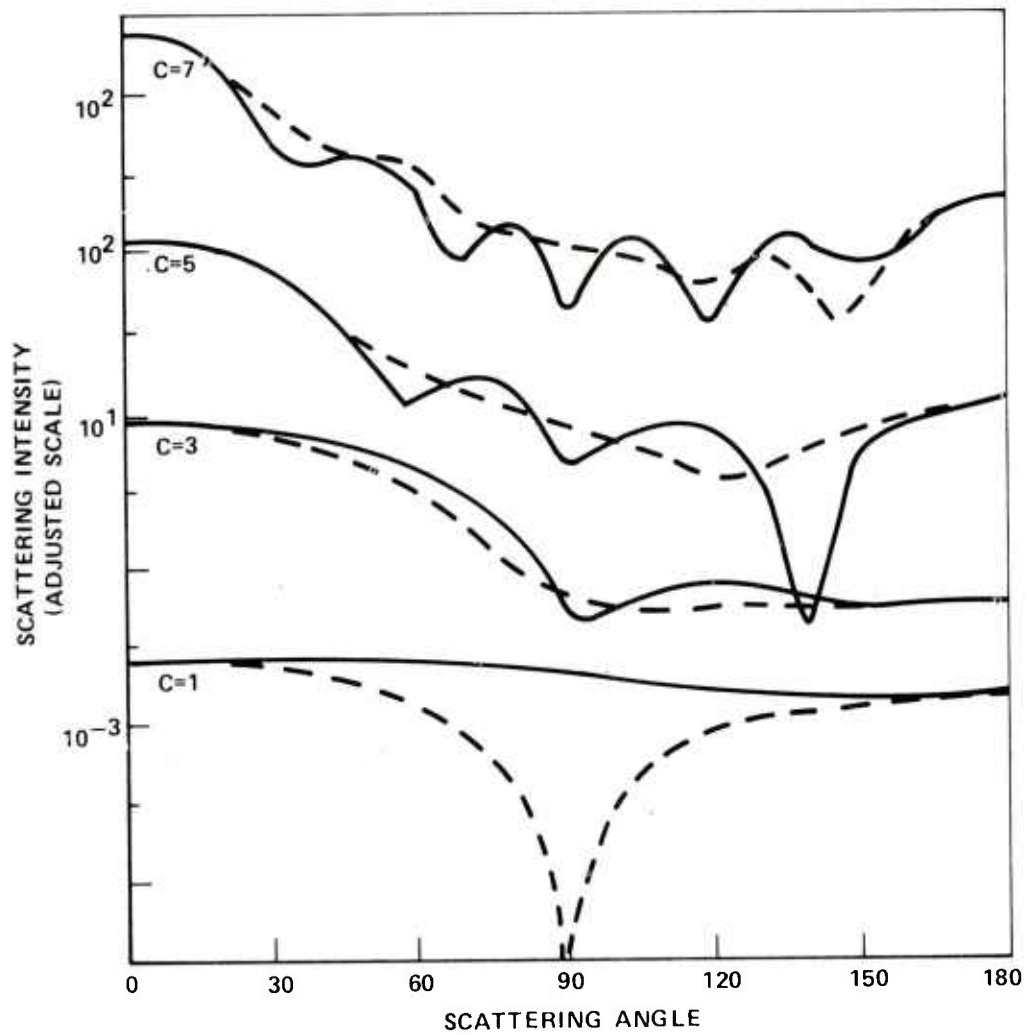


Figure (1). Calculated phase functions for prolate spheroids used for comparison with the exact phase function of Reference (12).

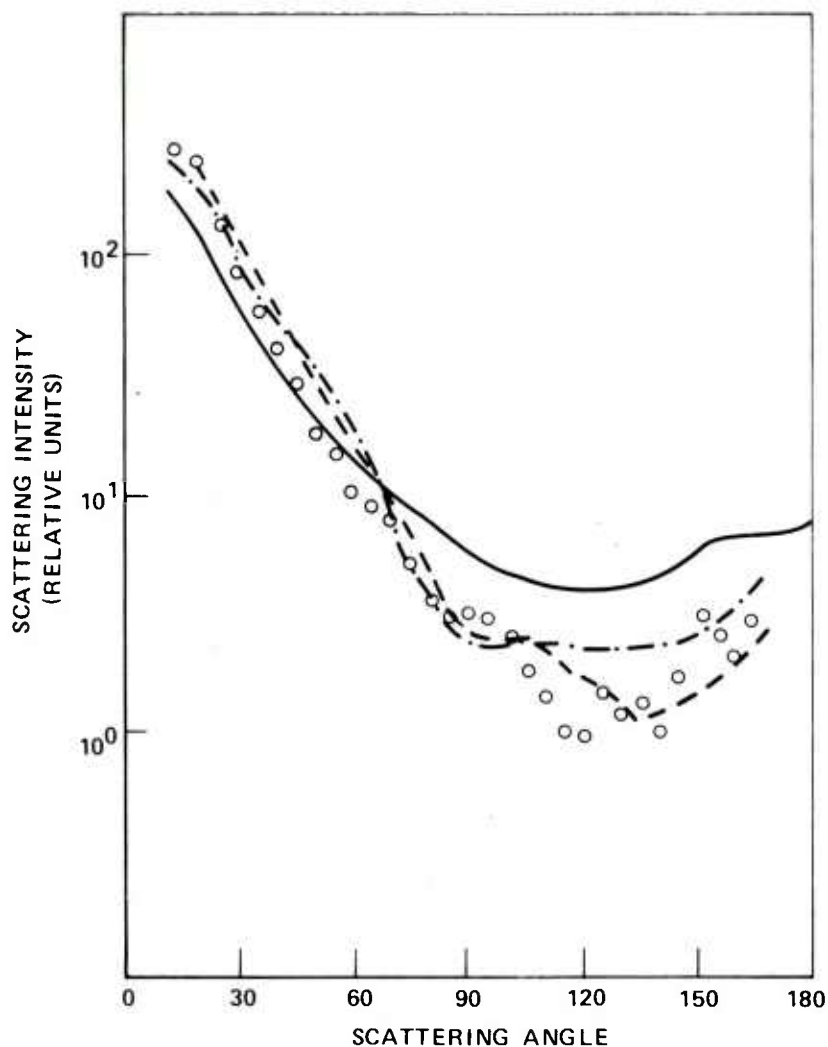


Figure (2). Comparison of non spherical experimental data (circles) with figurative Mie phase function (solid curve) and two composite non spherical phase functions (dashed curves.) The dashed curve is fitted over all data, and the dashed-dotted curve is fitted only over data for scattering angle greater than sixty degrees. Phase functions for shape functions incorporating $-.05$, $-.1$, and $-.15 P_2(\theta)$ were used in this fit.

REFERENCES

1. Fowler, B. W., and C. C. Sung, Appl. Opt., to be published circa July, 1978.
2. Millar, R. F., Radio Science, 8 785 (1971).
3. Davies, J. B., IEEE Trans. Microwave Theory, Tech., MTT-21(2) 99 (1973).
4. Eyres, L., and A. Nelson, Ann. Phys., 100 37 (1976).
5. Uzunoglu, U. K., and A. R. Holt, J. Phys. A., 10 413 (1973).
6. Waterman, U. K., Phys. Rev., D3 825 (1971).
7. Chylek, P., G. W. Grams, and R. G. Pinnick, Science, 193 480 (1976).
8. Mie, G., Ann. D. Physik. (4), 25 377 (1908).
9. Born, M. and E. Wolf, Principles of Optics, Pergamon Press, Oxford, 1975.
10. Schiff, L. I. Quantum Mechanics, McGraw Hill, New York, 1968.
11. Ward, G., "Scattering of Electromagnetic Radiation from Irregular dielectric Particles", unpublished.
12. Asano, S. and C. Yamamoto, Appl. Opt., 14 29 (1975).
13. Carnahan, B., H. A. Luther, and J. C. Wilkes, Applied Numerical methods, John Wiley and Sons, Inc., New York, 1969.
14. Sellers, W. R., and B. G. Gibbs, Descriptions-General Purpose Computer Subroutines, US Army Missile Command Report No. TR-WS-75-2, January 1977.
15. Van de Hulst, H. C., Light Scattering by Small Particles, John Wiley and Sons, Inc. New York, 1957.
16. Diermendjian, D., Electromagnetic Scattering on Spherical Polydispersions, Elsevier, New York, 1964.
17. Blattner, W., Utilization Instructions for Operation of the Mie Programs on the CDC-6600 at AFCRL, Research Note RRA-N7240, contract F19628-70-c-0156, Radiation Research Associates, Inc., 1972.

FOWLER

18. Kerker, M., private communication, 1977.

PROOF TEST PROCEDURES FOR CERAMIC MISSILE RADOMES

COLIN E. FREESE, MR., *DONALD M. NEAL, MR., and EDWARD M. LENOE, PhD
ARMY MATERIALS AND MECHANICS RESEARCH CENTER
WATERTOWN, MASSACHUSETTS 02172

INTRODUCTION

Currently ceramics are used for radome structures in several Army missiles because of their desirable dielectric properties, in addition to rain erosion resistance. The use of brittle ceramic materials as in radome structural components is relatively new. Since each radome developed requires a new design, it is important to thoroughly understand the materials capabilities. Ceramics rarely exhibit a single characteristic failure strength and therefore uncertainty exists regarding stress levels to which they can be safely subjected. The common concept of ceramic fracture is that of failure initiation due to tensile stresses acting at the size of a stress-intensifying flaw. Such flaws are either basic features of the microstructure, namely pores, inclusions, weak grain boundaries, or external scratches or cracks introduced during surface finishing or handling. In many structural ceramics variability of worst flaws leads to a size dependency of strength (1,2). Therefore recognition of variability in strength is important for proper structural design of these materials. This variability usually requires a probability-based failure criteria such as the Weibull representation. That is

$$P_f = 1 - \exp \left[-K \int_V (\sigma - \sigma_u / \sigma_0)^m dV \right] \quad (1)$$

where K is related to loading, V is the volume of material, σ is fracture stress, and σ_u is threshold stress; σ_u is defined as zero for the particular brittle materials considered in this paper; σ_0 and m are distribution constants determined from test data.

Consideration of time-dependent failure is also important. Fracture is delayed, occurring after stress has been sustained for some duration. Associated with the phenomena of static fatigue is an increase in strength with increase in stress or strain rates (3,4). For silicate glasses and most oxide ceramics it is known that the presence of moisture causes the environment to be reactive, even at ambient conditions (5). Numerous researchers have demonstrated the occurrence of subcritical crack growth in water and observed equivalent strength reductions in glass ceramics by a factor of one third and of alumina ceramic by one fourth. Thus the selection of a ceramic component should not be based on strength alone. The occurrence and rate of crack growth can be a decisive factor.

From the analytical viewpoint, theories are available and partially verified for treating the effects of time, size, and stress distribution on the likelihood of failure (6). From a pragmatic viewpoint, pre-service proof testing can also be applied. Effective proof testing must closely simulate not only the magnitude, but also the distribution of tensile stresses. In addition, proper account must be taken of any subcritical crack growth which might occur during attempted proof stressing.

DISCUSSION OF MATERIAL CHARACTERIZATION AND PROOF TEST PROCEDURES.

In component qualification, it is important to consider a typical ceramic radome product specification (7). Requirements for mechanical properties data include hoop tensile strength as determined via hydrostatic internal pressurization of a thin ring. The specification calls for two strength determinations on rings machined from the base of each radome. Density determinations are made on failed segments subsequent to burst testing. Figure 1 illustrates a typical ring specimen and includes a schematic of test apparatus. Hoop tensile strength is calculated from an elementary strength-of-materials formula. The modulus of rupture or flexure test is also a qualification requirement. Samples are fabricated in the same manner as the radome and from the same casting slip. Three specimens are required. It is of interest that both types of tests are performed at room temperature. Some mechanical properties data were available in the open literature for slip-cast fused silicate (SCFS). Tensile evaluations were performed in a gas bearing tension facility. Strength observations were completed over temperatures ranging from 70 F to 2200 F. Figure 2 summarizes tensile strength versus failure probability for the coupon specimens. Also shown in this illustration are results from hydroburst ring tests for one particular type of SCFS radome (Type I). Note the significant difference in tensile strength for the coupon versus ring tests, and the

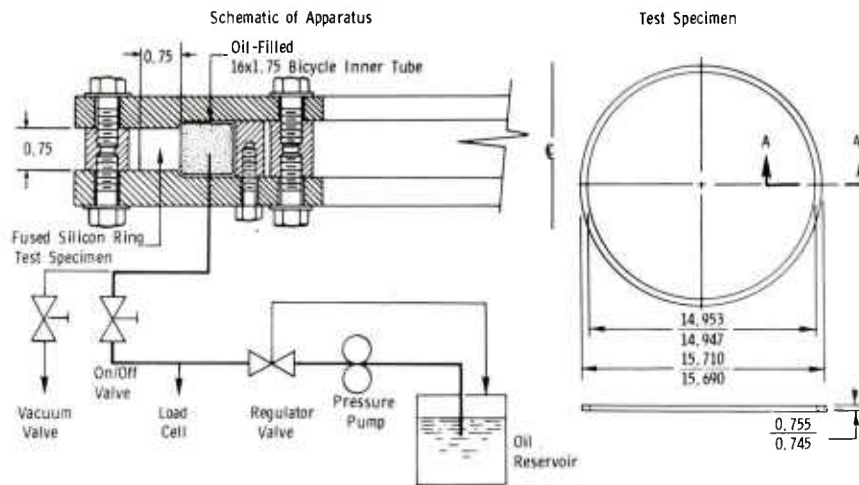


Figure 1. Hydroburst apparatus and test specimen

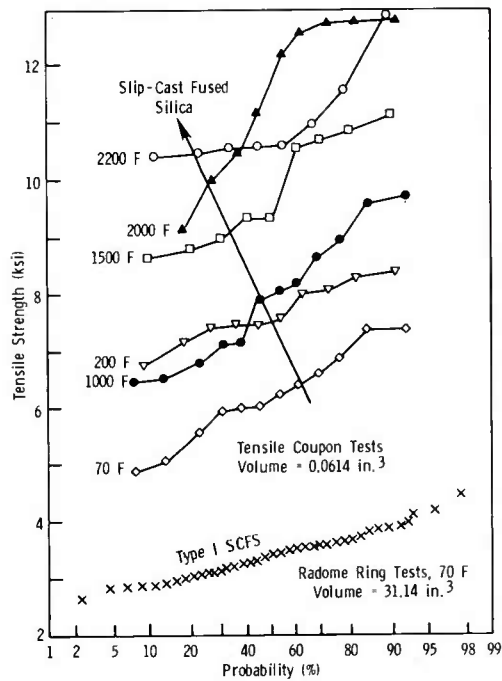


Figure 2. Temperature effects on tensile strength

difference in volume of the tension specimens. Interpretation of such phenomena has received a fair amount of attention (8,9). Based on the assumption that volume and/or surface area versus strength effects exist and furthermore can be well represented by Weibull statistics, numerous investigators have presented theoretical interrelationships to estimate the strength of a particular specimen geometry and load distribution from data obtained in different configurations. Suppose, for instance, we consider the case of uniform tension and assume strength is merely volume-dependent. Then at the same value of failure probability, for two distinct sets of statistical data,

$$\sigma_1/\sigma_2 = [(K_2V_2)/(K_1V_1)]^{1/m} \quad (2)$$

where σ_1, σ_2 are corresponding stresses, V_1, V_2 their associated volumes, and K_1, K_2 define the type of loading. Furthermore, a two-parameter Weibull representation has been assumed such that the probability of fracture at a given stress σ is taken as Eq. 1 with $\sigma_u = 0$ and the risk of rupture is defined as

$$R = \int_V (\sigma/\sigma_0)^m dV \quad (3)$$

Equation 3 can be evaluated for any given stress distribution appropriate to the experiment being conducted. Thus, simple interrelations can be developed for the variety of test methods applied to ceramics. Figure 3 summarizes data available for SCFS radomes. Data from each experimental apparatus is fit by the Weibull statistic and then the strength distribution of the other test method is estimated

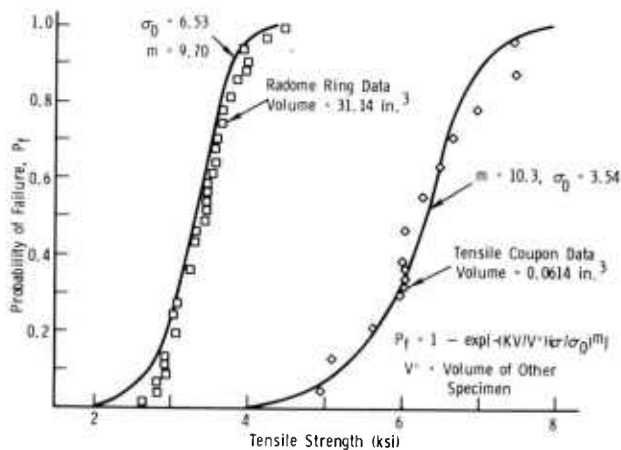


Figure 3. Size effects on tensile strength of slip-cast fused silica

and compared to observed response. Referring to Figure 3, it appears that the two-parameter Weibull model is a fairly good representation for the size/strength effect. As a further confirmation, experiments were conducted on Type I SCFS for three- and four-point flexure. Figure 4 summarizes these estimates and limited experimental data.

AEROTHERMAL TESTS

Typically aerothermal preflight certification tests are conducted on radome materials. Such tests are intended to simulate, to the extent possible, ascent and re-entry worst-case heating environments to screen candidate materials. Flight test risks on actual hardware rather than subscale models is used since numerous difficulties are inherent in aerothermal models. It is difficult, for instance, to accurately correlate subscale heat shield thicknesses, gaps, steps, protrusions, and boundary layer characteristics, with full-scale radome structures.

In operation, various gas combustors are used to impinge a high temperature exhaust with reasonable simulation of pressure and velocity characteristics. As an example, the ascent heating of the radome forebody is simulated by a relatively low preheat gas combustor and the radome removed and then allowed to equilibrate until a desired joint bond line temperature is achieved. Subsequently the radome is plunged into combustor exhaust. However, the gas is now at the appropriate temperatures to represent the re-entry phase of the simulation. Aerothermal tests are not considered in any detail in this report. Room temperature internal pressure tests such as illustrated in Figure 5 are considered in depth.

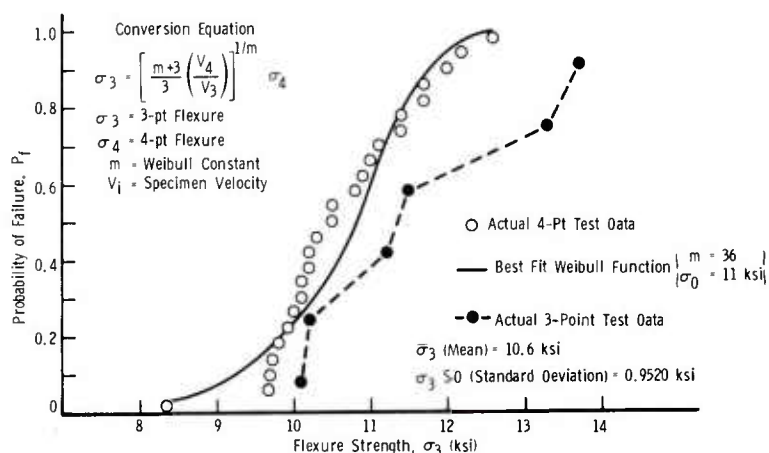


Figure 4. Estimates of three-point flexure strength from four-point flexure data

STATIC STRUCTURAL TESTS

Several types of static structural loadings are typically employed to verify load capacity, to assess structural integrity, or to proof test from the mechanical strength viewpoint. One experiment is depicted by Figure 5 wherein a radome is subjected to internal hydrostatic pressure. Figures 6 and 7 show experiments intended to verify load capacity of the radome and joint designs. In these instances the static tests employ whiffletree arrangements or simple off-angle loaded fabric straps. The load magnitudes are intended to simulate critical mission maneuver load magnitudes. The loads are applied quasi-statically and usually do not represent real mission times. The spatial distribution of loads is represented only in an approximate sense.

TIME DEPENDENCY

Time-dependent failure, or the possible existence of slow crack growth phenomena, had apparently not been investigated for SCFS typically used in missile radomes, therefore this aspect of material response was explored. Baseline data was obtained for Type I SCFS specimens prepared from radome fragments as well as specially prepared representative billets. Mechanical strength tests were performed at different loading rates. Differences in strength as a function of these stressing rates can be used to infer analytical models for slow crack growth. Based on such results, the so-called strength-probability-time (SPT) nomograph can be prepared (4). Such information is useful, for instance, in establishing proof test load limits and safety margins for long-term stress such as might be maintained by mismatch bonded or mechanical radome-to-substructure joints.

These beam specimens were subjected to ultrasonic and X-ray inspection and ultrasonic velocities were obtained as well as bulk densities. In this series of tests acoustic emission (AE) apparatus was additionally employed. The intention of the AE observations was to explore the possibility of detecting material damage prior to failure. These series of experiments did not yield any useful correlation between AE counts; neither total counts nor rate of emission gave a clear-cut trend with strength levels.

There were several objectives of the experiments on this SCFS. It was desired to thoroughly investigate slow crack growth at room temperature. Accordingly, flexure tests were conducted at each of two loading rates. This permitted construction of the SPT nomograph. In order to study the validity of this technique, a series of creep rupture tests were completed. These were compared to predicted failure times to substantiate the methodology, as discussed subsequently.

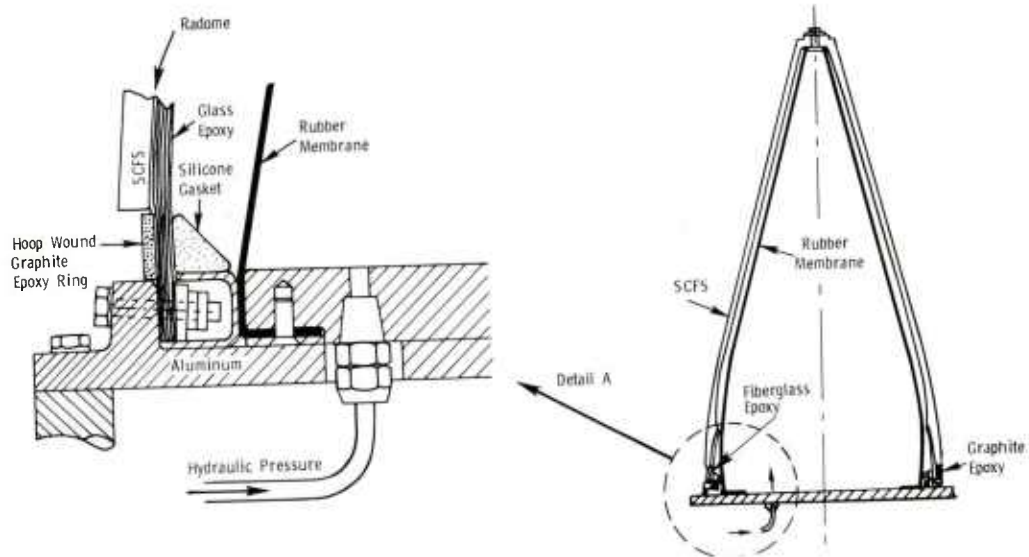


Figure 5. Internal pressurization radome proof test apparatus

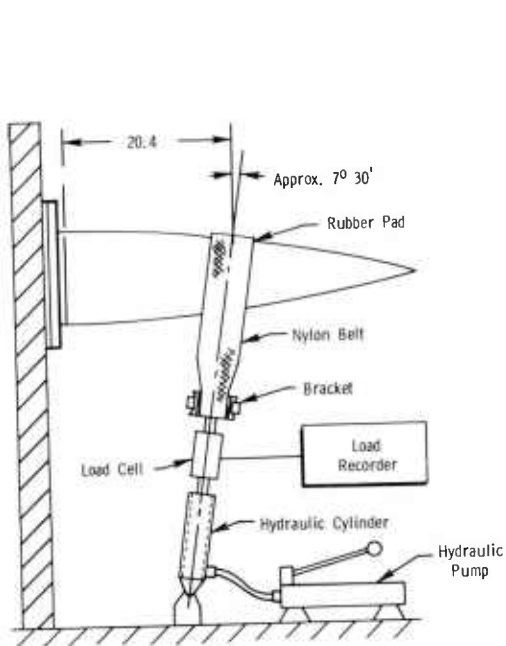


Figure 6. Schematic of off-axis joint proof test

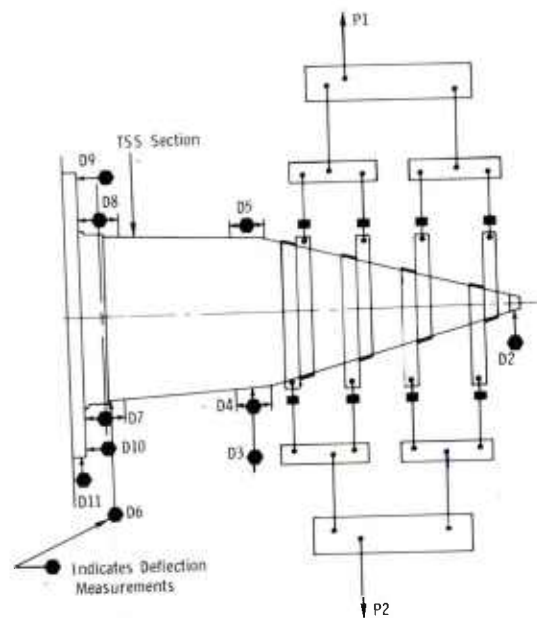


Figure 7. Schematic of whiffle-tree static test

Another objective for this series of experiments was to study possible proof stress damaging effects on strength and/or time to failure of SCFS at ambient conditions. Therefore a set of virgin, e.g., untested, beams was loaded to predetermined stress levels and tested to failure (see Figure 8).

CONSTRUCTION OF THE DESIGN STRENGTH-PROBABILITY-TIME NOMOGRAPH

Under constant stress σ in a delayed fracture test, the time to failure T is given, in an elementary fashion, by:

$$T = \int_{C_i}^{C_{Ic}} dc/V \quad (4)$$

where C_i is initial crack size, C_{Ic} critical crack size, and V crack velocity. Assuming that a simple power form of crack velocity versus stress intensity can be written $V = \alpha K_{Ic}^n$, then

$$T = 2K_{Ic}^{2-n}/[(\sigma y)^2 \alpha (n-2)] \quad (5)$$

and for typical large n values, K_{Ic} are negligible. For a given batch of N specimens with initial flaw size C_j

$$T = \frac{2y^{2-n} \sigma^{2-n} C_j^{(2-n)/2}}{(\sigma y)^2 \alpha (n-2)} = \frac{BC_j^{(2-n)/2}}{\sigma^n}$$

where y = geometric constant. Then, for specimens with the same initial flaw size and the same probability of failure,

$$T\sigma^n = \alpha_1. \quad (6)$$

Using Eq. 6, a family of lines can be constructed on the Weibull strength/probability of survival graph corresponding to increasing failure times. For an individual specimen stressed at σ_1 and failing

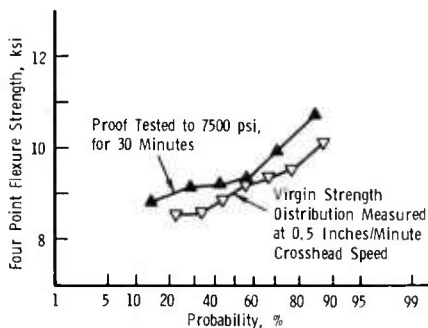


Figure 8. Effect of proof stress on flexure strength distribution

in T_1 , we now have a relationship for another specimen failing at stress σ_0 in a reference time T_0 . This permits construction of an SPT nomograph where a family of parallel lines are equispaced for equal logarithmic increases in failure time (4). According to this simple theory, creep rupture data can be related to instantaneous dynamic failure by use of the following relationship,

$$(\sigma_0/\sigma_j)^n = T_j/T_0 \quad (7)$$

DISCUSSION OF RATE EFFECTS

The SCFS specimens were machined to 0.11×0.15×2.0-inch sizes and tested in four-point bending (1.875-inch span) at room temperature, for a load rate of 0.5 in./min. Another group of 22 specimens was tested in the same manner except load rate was 0.01 in./min. The strength levels for each set of data were obtained. Results are plotted in Figure 9, indicating good fit to the Weibull distribution. It was determined from the test data at the 99% confidence level (10) that mean strength ($\sigma = 9.07$ ksi) for load rate of 0.5 in./min was significantly greater than $\bar{\sigma} = 8.14$ ksi for the 0.01 in./min rate. An SPT diagram, Figure 10, was constructed using the data described above in conjunction with the theory discussed next. Failure times of 1.27 sec and 57.0 sec were determined for strength values of 9.07 and 8.14, respectively.

The exponent $n = 35.12$ from Eq. 7 is obtained from the strength and time data, providing the necessary parameters for SPT diagram construction. The Weibull modulus $m = 15.06$, describing the slope of time-to-failure lines in the SPT diagram, was determined from averaging m values obtained from the two different load rates.

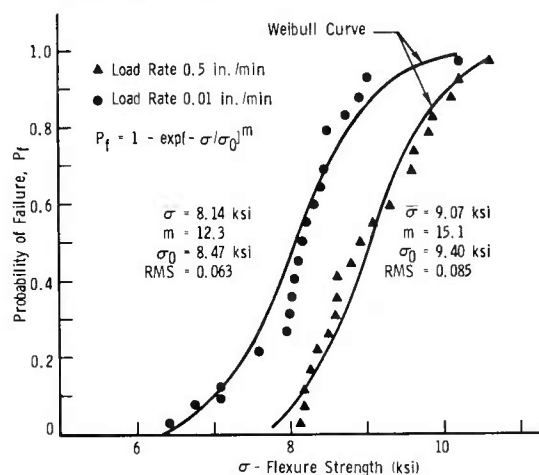


Figure 9. Load rate effects (SCFS)

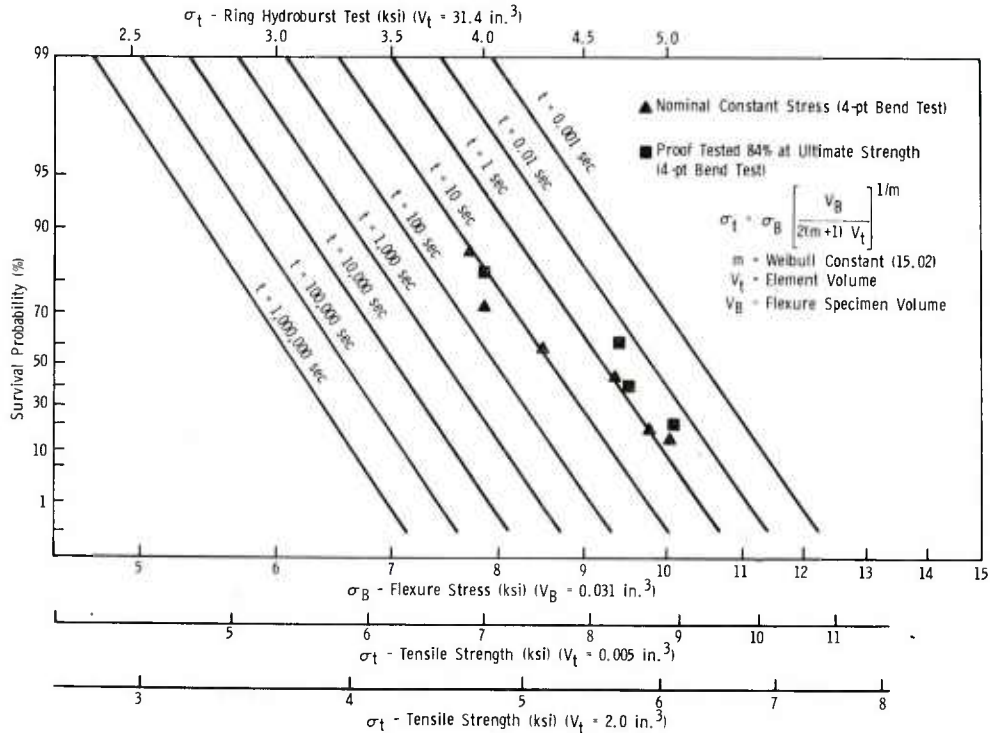


Figure 10. SPT diagram for SCFS superimposed delayed fracture data normalized to failure time of one second

An evaluation of the adequacy of the SPT diagrams for SCFS was made from delayed fracture data to estimate the equivalent failure stresses for a failure time of 1 second. Results in Figure 10 show excellent agreement between delayed fracture data and the predicted data, except for specimens failing within a short time frame ($T < 1$ minute). Figure 10 survival probabilities are represented as a function of both bend and tensile stress volumes (V_t). It should be noted that smaller V_t result in higher failure loads for equivalent P_s values. The flexure stress labelled with $V_B = 0.031 \text{ in.}^3$ represents the test results presented in this report. The equivalent data for a tension test were superimposed on the diagram to indicate the effects of volume changes. The σ_t values for $V_t = 0.005 \text{ in.}^3$ and 2.0 in.^3 are tabulated to provide ranges of probability versus stress for relatively small and large elements such as used in a finite element analysis. The label at the top of diagram where $V_t = 31.4 \text{ in.}^3$ represents the volume of the ring used in a hydroburst strength test to measure the quality of SCFS.

ANALYSIS OF THE INTERNAL PRESSURIZATION PROOF TEST

Thus far we have gained insight into the probability of failure and time dependency aspects of strength behavior. This information can now be used to assess the radome internal pressurization proof test depicted in Figure 5. For this purpose, an axisymmetric finite element analysis was performed using a unique set of finite element software modules. This approach has been shown to be both efficient and competitive (11) with other finite element procedures. The final result is a series of codes which address directly the particular structure in question and the types of information required. A primary concern with the effects of various end-restraint conditions isoparametric axisymmetric solid element was used to model the elastic behavior of the radome and a substructuring technique was implemented. The entire structure contains over 3000 degrees of freedom and is divided into four subregions, the details of which are shown in Figure 11.

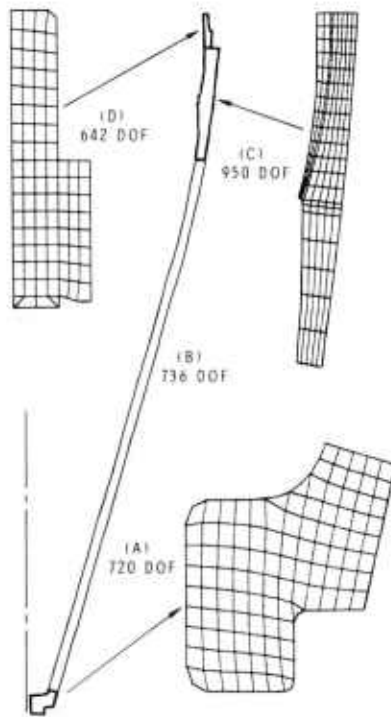


Figure 11. Finite element gridding

Several conditions were analyzed. First, the radome was assumed to be pressurized to 100 psi internal pressure (see Figure 5). This pressure was taken to act over the entire inner surface down to the aluminum restraining ring (see detail joint schematic, Figure 5). In the region of the retaining ring no radial movements were permitted on the outer surface of the radome. Two additional computations were completed wherein 7-mil and 15-mil radial displacements were imposed. These displacements are intended to simulate the effects of out-of-tolerance conditions which might occur and should be primarily applied to the inner surface of the restraint region due to the inability of the bolt-up configuration to transmit tensile loads. The entire structure can be treated for a unique set of boundary conditions in approximately six minutes (UNIVAC 1106) of computer time. A consideration of various restraint conditions applied to subregion (d) in Figure 11 can be achieved at a fraction of that cost.

Since axisymmetric conditions apply, the calculations of radial displacement effects are lower bound estimates. In reality, out-of-tolerance dimensions may be nonsymmetric.

Furthermore, in bolting the radome in place, the displacements would most likely be imposed sequentially on opposite ends of the diameter of the base. These facts, coupled with stress concentrations around bolt holes, suggest that somewhat larger stresses might exist due to mismatched tolerances.

ASSESSMENT OF PROBABILITY OF SURVIVAL

Results of the finite element stress analysis were used to estimate probability of survival under conditions of the proof test. Note that the radome joint consists of four types of material, namely SCFS, glass epoxy, graphite epoxy, and the joint adhesive. Accordingly, two analytical formulations were used. The finite element stress distributions were used in evaluating individual probability values of the finite volume elements of the SCFS structure and multiplying these probabilities to obtain P_s for the particular applied loads. The probability of survival P_s for individual stress components is written as

$$P_{s_i} = \exp [-KV_i/V^* (\sigma_{\max_i}/\sigma_0)^m] \quad (8)$$

where $K = 1$ for simple tensile stress, V_i = volume of elements, V^* = volume of test specimen, and σ_{\max_i} = maximum principal stress in element. The probability of survival P_s of the entire structure is

$$P_S = \prod_{i=1}^N P_{Si} \quad (9)$$

where N = number of elements in finite element solution.

Weibull strength parameters were $m = 10.30$ and $\sigma_0 = 3.55$ ksi, obtained from hydroburst tests on SCFS rings with 31.4 in.³ volume. Since fiberglass, graphite, and the adhesive materials are not assumed volume-dependent, reliability calculations for these materials were based on application of the Warner (stress-strength) diagram for each individual material element (12).

The reliability of these materials was determined from the basic concept that a no-failure probability exists when allowable stress S is not exceeded by applied stress s . That is,

$$R = \int_{-\infty}^{\infty} f(s) \left[\int_s^{\infty} f(S) dS \right] ds \quad (10)$$

where $f(s)$ is the probability density function obtained from knowledge of the design stress in the structural element and $f(S)$ represents the distribution of the material strength. The standard deviations for $f(s)$ and $f(S)$ are obtained from strength data and an assumed variability in design stress calculations. If density functions for strength and stress are assumed normal (12), then

$$R = [1/(\sigma_{\zeta}\sqrt{2\pi})] \int_0^{\infty} \exp \left[-1/2 \left((\zeta - \bar{\zeta})/\sigma_{\zeta} \right)^2 \right] d\zeta \quad (11)$$

where $\bar{\zeta} = \bar{S} - \bar{s}$, $\sigma_{\zeta} = \sqrt{\sigma_S^2 + \sigma_s^2}$, and $\zeta = S - s$. The σ_S and σ_s are standard deviations for strength and stress data.

P_S calculations were completed for each material component and used in Eq. 9 to determine reliability of the entire structure, which includes the four materials. Important results of the calculations are summarized in Table 1, where the location and material with corresponding maximum stress and associated stressed volume are indicated. Sensitivity of the P_S estimates was also evaluated by assuming coefficients of variation of 0%, 10%, and 15% in stresses.

Table 1. RELIABILITY CALCULATIONS FROM
PROOF TEST OF RA00ME

Displacements	CV (%) Design Stress	P_s Complete Radome	Material Peak P_f
0.000	0.000	0.99860	SCFS
.007	0.000	.99859	SCFS
.007	10.000	.99840	SCFS
.007	15.000	.99831	SCFS
.015	0.000	.96095	Graphite
.015	10.000	.89665	Graphite
.015	15.000	.79084	Graphite

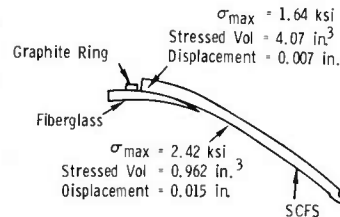
 P_s = Probability of Survival P_s = 0.99989

(MM Rohm & Haas)

 P_f = Probability of Failure

STRENGTH MATERIALS (%)

Materials	Mean Strength (ksi)	Standard Deviation (ksi)
SCFS	3.380	0.396
Adhesive	0.660	0.140
Graphite	80.000	8.000
Fiberglass	26.500	4.200



CONCLUSIONS

Table 1 indicates the influence of radial displacement on survival estimates. Referring to the case of 15-mil deflection, the maximum SCFS stress is 2.42 ksi and the associated stressed volume is 0.962 in.^3 . Eq. 7 and the SPT diagram (Figure 10) can be used to estimate, at 99% P_s , that such a stress could be sustained for more than 20 years.

It should be recalled, however, that the data used to construct the SPT diagram was obtained under ambient laboratory conditions and it is assumed these represent the radome environment. More severe environments would lead to shorter life estimates. So-called inert strength could be measured at higher strain rates in a benign atmosphere, whereas slow strain rate results could be obtained in aqueous or other detrimental environments. Such data would most likely lead to significantly shorter life estimates. Note that the SCFS material used to generate the SPT diagram was machined from flat billets, using similar slip and generally identical processing techniques as in radome manufacture. However, it is a moot point whether this SCFS is indeed representative of radome materials. Preliminary tests on Type I radome fragments conducted at loading rates varying from 0.01 to 1.0 in./min suggested a potential range of slow crack exponent of $14 \leq n \leq 35$. Using the Weibull exponent of 10.3, which resulted from a best fit of 44 hydro-burst tests on SCFS, and $n = 35.1$, the shortest life is estimated as 36 days for the 2.42 ksi stress maximum.

Referring to Table 1, use of the various coefficients of variation is intended to represent potential uncertainties in the analytical modeling and associated stress distributions. For the largest radial deflection, the P_S estimates are observably affected. This emphasizes the necessity to accurately describe stress distributions. It is apparent that other proof test methods should be analyzed in detail and compared to design condition stresses to assess the adequacy of these screening techniques.

In summary, techniques have been discussed which permit estimates of probability of survival under fast fracture as well as sustained stress conditions. These analytical computations are quite sensitive to values of m and n . Therefore, it is imperative that actual radome materials be thoroughly characterized.

ACKNOWLEDGMENTS

The authors are indebted to the following individuals: P. Ormsby and J. Pemberton of MIRADCOM for supplying the slip-cast fused silica; R. Peters and G. Driscoll of AMMRC for conducting the experiments; D. Mason of AMMRC for assisting in the calculations, and to the staff of the Technical Reports Office for excellent preparation of the manuscript.

REFERENCES

1. WEIBULL, W. Proc. Ing. Vetenskapsakad Akad., 1939, p. 151.
2. FREUDENTHAL, A. M., in *Fracture, An Advanced Treatise*, v. II, Academic Press, New York, 1968, p. 591-619.
3. CHARLES, R. J. J. Appl. Phys., v. 29, no. 11, 1958, p. 1549-1560; also v. 29, no. 12, 1958, p. 1657-1667.
4. DAVIDGE, R. W., McLAREN, J. R., and TAPPIN, G. J. Matl. Sci., v. 8, December 1973, p. 1699-1705.
5. WIEDERHORN, S. M., in *Fracture Mechanics of Ceramics*, Plenum Press, New York, 1973, p. 613-646.
6. *Ceramics for High Performance Applications*, Part II, 1977, in press.
7. Production Specification for the Impregnated or Non-Impregnated Ceramic Radome in a Semi-Finished Form, MIS-20074B Code Ident. 18876, 26 February 1974.
8. WEIL, N. A., and DANIEL, I. M. J. Am. Ceram. Soc., v. 47, June 1964, p. 268-274.
9. BATDORF, S. B., in *Proceedings of Intl. Symp. on Fracture Mechanics of Ceramics*, Pennsylvania State University, July 27-29, 1977.
10. HOEL, P. G. *Introduction to Mathematical Statistics*. 4th ed., John Wiley and Sons, 1971.
11. FREESE, C. E. *An Efficient Substructuring Procedure for Axisymmetric Finite Element Analysis*. In preparation.
12. HAUGEN, E. B. *Probabilistic Approaches to Design*. John Wiley and Sons, 1968.