LET
12
V1 - A090 350
V3 - A090 422




AD A090 337

# Army Science Conference Proceedings

17-20 June 1980

Volume II

Principal Authors E thru M

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

DDC FILE COPY

80 10 6 039

DEPARTMENT OF THE ARMY
OFFICE OF THE DEPUTY CHIEF OF STAFF
FOR RESEARCH, DEVELOPMENT, AND ACQUISITION
WASHINGTON, D.C. 20310

REPLY TO ATTENTION OF
DAMA-ARZ-D

21 July 1980

SUBJECT: Proceedings of the 1980 Army Science Conference

SEE DISTRIBUTION

1. The twelfth in a series of Army Science Conferences was held at the United States Military Academy, 17-20 June 1980. The conference presented a cross section of the many significant scientific and engineering programs carried out by the Department of the Army and provided an opportunity for Department of the Army civilian and military scientists and engineers to present the results of their research and development efforts before a distinguished and critical audience.

2. These Proceedings of the 1980 Army Science Conference are a compilation of all papers presented at the conference and the supplemental papers that were submitted. The Proceedings consist of four volumes, with Volumes I through III unclassified, and Volume IV classified.

3. Our purpose for soliciting these papers was:

a. To stimulate the involvement of scientific and engineering talent within the Department of the Army.

b. To demonstrate Army competence in research and development.

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

d. To promote the interchange of ideas among members of the Army scientific and engineering community.

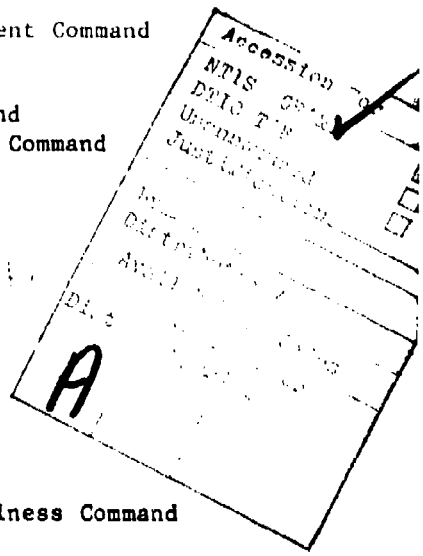
4. It is hoped that the information contained in these volumes will be of benefit to those who attended the conference and to others interested in Army research and development.

Donald R. Keith

DONALD R. KEITH
Lieutenant General, GS
Deputy Chief of Staff for Research,
Development, and Acquisition

DISTRIBUTION:
Director of Army Research
Office of the Surgeon General
Office, Chief of Engineers, ATTN: DAEN-ZA
Office, Chief of Engineers, ATTN: DAEN-ZC
Office, Chief of Engineers, ATTN: DAEN-RDM
Office, Chief of Engineers, ATTN: DAEN-RDZ-A
Office, Chief of Engineers, ATTN: DAEN-MPR-A
Office, Chief of Engineers, ATTN: DAEN-CWM-R
The Army Library, ATTN: ANRAL-RS
COMMANDERS/DIRECTORS
US Army Ballistic Missile Defense Systems Command
US Army Computer Systems Command
US Army Forces Command
US Army Intelligence and Security Command
US Army Materiel Development and Readiness Command, ATTN: DRCLDC
US Army Armament Materiel Readiness Command
Rock Island Arsenal
Watervliet Arsenal
US Army Armament Research and Development Command
US Army Ballistic Research Laboratory
US Army Chemical Systems Laboratory
Fire Control and Small Caliber Weapons Systems Laboratory
Large Caliber Weapons Systems Laboratory
Benet Weapons Laboratory
US Army Aviation Research and Development Command
US Army Aviation Center
US Army Avionics R&D Activity
US Army Research and Technology Laboratories (USARTL, AVRADCOM)
Aeromechanics Laboratory, NASA-Ames Research Center
Applied Technology Laboratory
Propulsion Laboratory, NASA-Lewis Research Center
Structures Laboratory, NASA-Langley Research Center
US Army Communications and Electronics Materiel Readiness Command
US Army Communications Research and Development Command
US Army Center for Communications
US Army Center for Technical Computer Sciences
US Army Electronics Research and Development Command
US Army Atmospheric Sciences Laboratory
US Army Combat Surveillance and Target Acquisition Laboratory
US Army Electronic Warfare Laboratory
US Army Electronics Technology and Devices Laboratory
US Army Harry Diamond Laboratories
US Army Night-Vision and Electro-Optics Laboratory
US Army Signals Warfare Laboratory
US Army Missile Command
Propulsion Directorate, US Army Missile Laboratory

US Army Mobility Equipment Research and Development Command
US Army Electrical Power Laboratory
US Army Natick Research and Development Command
US Army Tank-Automotive Materiel Readiness Command
US Army Tank-Automotive Research and Development Command
US Army Test and Evaluation Command
US Army Aircraft Development Test Activity
US Army Aberdeen Proving Ground
Materiel Testing Directorate
US Army Cold Regions Test Center
US Army Dugway Proving Ground
US Army Electronics Proving Ground
US Army Tropic Test Center
US Army White Sands Missile Range
Instrumentation Directorate
National Range Operations Directorate
US Army Yuma Proving Ground
US Army Troop Support and Aviation Materiel Readiness Command
US Army Electronics Materiel Readiness Activity
US Army Foreign Science and Technology Center
US Army HQ TRADOC Combined Arms Test Activity
US Army Human Engineering Laboratory
US Army Logistics Center
US Army Materials and Mechanics Research Center
US Army Materiel Systems Analysis Activity
US Army Research Office
US Army Research, Development and Standardization Group (Europe)
US Army Training and Doctrine Command
US Army Concepts Analysis Agency
US Army Operational Test and Evaluation Agency
US Army Japan, Camp Zama
US Army Inst for Research in Management Info and Computer Sciences
US Army Reserve Components Personnel and Administration Center
US Army Soldier Support Center
Lawrence Livermore National Laboratory, Univ of California, Livemore
Los Alamos Scientific Laboratory
US Army Health Services Command
US Army Institute of Surgical Research
US Army Medical Research and Development Command
US Army Medical Research Institute of Infectious Diseases
US Army Research Institute for the Behavioral and Social Sciences
US Army Research Institute of Environmental Medicine
Letterman Army Institute of Research
Walter Reed Army Institute of Research
US Army Aeromedical Research Laboratory
US Army Biomedical Laboratory
US Army Environmental Hygiene Agency
US Army Medical Bioengineering Research and Development Laboratory



Brooke Army Medical Center
Dwight D. Eisenhower Medical Center
Fitzsimons Army Medical Center
Madigan Army Medical Center
US Army Tripler Medical Center
Walter Reed Army Medical Center
Armed Forces Institute of Pathology
US Army Institute of Dental Research
US Army Coastal Engineering Research Center
US Army Cold Regions Research and Engineering Laboratory
US Army Construction Engineering Research Laboratory
US Army Engineer Topographic Laboratories
US Army Engineer Waterways Experiment Station
US Army Engineering Division, North Atlantic

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

COMMANDANTS
National Defense University
National War College
Industrial College of the Armed Forces
Academy of Health Sciences
US Army Field Artillery School
US Army Infantry School
US Army Intelligence Center and School
US Army Military Police Chemical School and Training Center
US Army Signal Center at Fort Gordon
US Army Transportation School

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

COPIES FURNISHED:
Office, Under Secretary of Defense for Research and Engineering
Office, Deputy Director for Science and Technology, Defense Nuclear Agency
Chairman, Defense Science Board
Defense Advanced Research Projects Agency
Defense Communication Agency
Defense Communication Engineering Center
Defense Documentation Center
Defense Science Board
Institute for Defense Analysis
Deputy Chief of Staff for Logistics, HQDA
Deputy Chief of Staff for Operations and Plans, HQDA
Deputy Chief of Staff for Personnel, HQDA
Assistant Chief of Staff for Automation and Communications, HQDA
Assistant Chief of Staff for Intelligence, HQDA
Office of the Chief of Public Affairs, HQDA

First United States Army
National Academy of Sciences
National Aeronautics and Space Administration (Code L-2)
National Bureau of Standards
National Institute for Occupational Safety and Health, HEW
National Institutes of Health
National Science Foundation
National Security Agency
Office of Personnel Management
US Army Acquisition Information Systems Agency
US Army Scientific Information Center
Air Force Institute of Technology
Air Force Office of Scientific Research
Air Force Systems Command
Naval Air Systems Command
Naval Materiel Command
Naval Research Laboratory
Naval Research Library (Code 2627)
Office of Naval Research (Code 102)
Office of Naval Research (Code 230)
HQ US Marine Corps (Code RD-1)
Science Advisor to the Commandant, US Coast Guard
Battelle Columbus Laboratories
Mathematics Research Center, Univ of Wisconsin, Madison
Southwest Research Institute

(6)

# PROCEEDINGS

OF THE

1980 ARMY SCIENCE CONFERENCE

UNITED STATES MILITARY ACADEMY

WEST POINT, NEW YORK

17-20 JUNE 1980

VOLUME II

Principal Authors E through M

(11) 21 Jul 80

(12) 17

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 1980 ARMY SCIENCE CONFERENCE


| Author | Title | Vol | Page |
|---|---|---|---|
| Abramovitz, I. J. | See Berg, N. J. | 4 | 1 |
| Ahn, B. H.<br>Clark, W. W., III | Second Harmonic Generation of Near Millimeter Wave Radiation by Non-Linear Bulk Material | 1 | 1 |
| Aiken, E. W. | The Effects of Control System and Display Variations for an Attack Helicopter Mission Through Piloted Simulation | 1 | 17 |
| Aitken, G. W.<br>Richmond, P. W., III<br>Albert, D. G. | Impact Fuze Performance in Snow | 1 | 31 |
| Albert, D. G. | See Aitken, G. W. | 1 | 31 |
| Albritton, G. E.<br>Balsara, J. P. | Response of Buried Vertically Oriented Cylinders to Dynamic Loading | 1 | 47 |
| Alexander, P. M. | Frequency Diverse Tracking/Guidance Millimeter Radar Adapted to Target Acquisition | 1 | 61 |
| Alley, B. J. | Low Temperature X-Ray Diffraction Investigation of Embrittled Minimum Signature Chaparral Propellant | 1 | 77 |
| Alster, J.<br>Slagg, N<br>Dewar, M. J. S.<br>Ritchie, J. P.<br>Wells, C. | A Reaction Mechanism in the Shock Initiation of Detonation. A Theoretical Study | 1 | 91 |
| Ashman, W. P.<br>Thornton, W.<br>Broome, P. H.<br>King, J. W.<br>Sacco, W. J. | Pattern Recognition Applications in Chemistry and Pharmacology: A 'Pharmacophore Acetylcholinoreceptor' Subunit Environment Model | 1 | 105 |
| Atkins, R. M. | See Heberlein, D. C. | 4 | 67 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Aucoin, T. R.<br>Savage, R. O.<br>Wade, M. J.<br>Gualtieri, J. G.<br>Schwartz, A. | Large High Quality Single Crystal Aluminum Phosphate for Acoustic Wave Devices | 1 | 121 |
| Aucoin, T. R. | See Ross, R. L. | 3 | 205 |
| Baladi, G. Y.<br>Rohani, B. | A Terrain-Vehicle Model for Analysis of Steerability of Tracked Vehicles | 1 | 135 |
| Ballato, A.<br>Lukaszek, T. | Combat Hardened Communications: High Precision Frequency Control Using Resonators Immune to Acceleration and Stress Fields | 1 | 151 |
| Ballato, A. T. | See Iafrate, G. J. | 2 | 295 |
| Balsara, J. P. | See Albritton, G. | 1 | 47 |
| Banderet, L. E.<br>Stokes, J. W.<br>Francesconi, R.<br>Kowal, D. | Simulated, Sustained-Combat Operations in the Field Artillery Fire Direction Center(FDC): A Model for Evaluating Biomedical Indices | 1 | 167 |
| Banister, G. H. | Testing of the Army's Interoperating Network of Tactical $C^3I$ Systems | 1 | 183 |
| Beatrice, E. S. | See O'Mara, P. A. | 3 | 31 |
| Beck, R. R. | See Hoogterp, F. B. | 2 | 263 |
| Beisel, W. R. | See Daniels, W. L. | 1 | 429 |
| Berg, N. J.<br>Abramovitz, I. J.<br>Lee, J. N.<br>Casseday, M. W. | Acousto-Optic Time Integrating Correlator for Detection and Characterization of Broad Band LPI Communications | 4 | 1 |
| Berry, P.<br>Keane, W.<br>Hatfield, G.<br>Filler, M. | The Biological Detector and Warning System, XM19/XM2 | 1 | 197 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Biggs, S. L. | See Zwick, H. | 3 | 461 |
| Blackman, G. R. | Geometric and Temporal Characterzation of Battlefield Smoke and Dust by Multispectral Digital Image Analysis | 1 | 211 |
| Bowden, C. M.<br>Howgate, D. W. | MICOM Program in Optical Bistability | 1 | 225 |
| Brandt, H. E. | See Kehs, R. A. | 2 | 339 |
| Britt, J. R. | Calculation of Ground Shock Motion Produced by Near Surface Airburst Explosions Using Cagniard Elastic Propagation Theory | 1 | 241 |
| Brodman, B. W. | See Devine, M. P. | 1 | 467 |
| Bromborsky, A. | See Kehs, R. A. | 2 | 339 |
| Broome, P. H. | See Ashman, W. P. | 1 | 105 |
| Brown, N. D. | See Doctor, B. P. | 1 | 487 |
| Bruce, C. W. | See Bruce, D. | 1 | 255 |
| Bruce, D.<br>Bruce, C. W.<br>Yee, Y. P.<br>Burket, H. | Experimentally Determined Relationship Between Extinction and Liquid Water Content | 1 | 255 |
| Brynjolfsson, A. | See Wang, C. P. | 3 | 357 |
| Burket, H. | See Bruce, D. | 1 | 255 |
| Buser, R. G. | See Rohde, R. S. | 3 | 195 |
| Bush, C. C. | See Sturek, W. B. | 3 | 281 |
| Bussell, N. E.<br>Miller, R. A.<br>Hawley, C. E. | Preliminary Studies on the Effects of Organophosphate Chemical Warfare Agents on Oral Cavity Physiology | 1 | 265 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Butler, D. K. | Microgravimetry and the Measurement and Application of Gravity Gradients | 1 | 279 |
| Cadwallender, W. K.<br>Kramer, K.<br>Jankowski, P. Z.<br>Kisatsky, P. J. | Full Field Interferometry Applications to Army Problems | 1 | 293 |
| Canonico, P. G.<br>Little, J. S.<br>Jahrling, P. B.<br>Stephen, E. L. | Mechanism of Action of Ribavirin: An Antiviral Drug of Military Importance | 1 | 309 |
| Casseday, M. W. | See Berg, N. J. | 4 | 1 |
| Cassidy, T. W. | See Moulton, J. R. | 4 | 111 |
| Charland, J. J. | Computer Graphics for the Army | 1 | 321 |
| Chen, F. | Materials Deterioration in Tropic Versus Conus Sites | 1 | 337 |
| Chen, P. F.<br>Seemuller, W. W. | Detection of Signal Signatures of Cartographic Features | 1 | 353 |
| Chernick, J. A. | See Scungio, R. C. | 3 | 219 |
| Choi, C. S. | See Prask, H. J. | 3 | 153 |
| Christensen, C. R.<br>Hartman, R. L. | Optical Correlation Seeker | 1 | 369 |
| Chubb, D. W. J.<br>Miller, J. A. | The Design and Implementation of Digital Demodulator for Use Against On-Off Keyed Signals | 4 | 17 |
| Chylek, P. | See Pinnick, R. G. | 3 | 101 |
| Clark, E. J. | See Neely, E. S. | 3 | 1 |
| Clark, W. W., III | See Ahn, B. H. | 1 | 1 |
| Clare, V. R. | See Posey, W. R. | 3 | 137 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Cline, J. H. | See Hammond, C. E. | 2 | 147 |
| Cohen, S. H. | See Walker, J. E., Jr. | 3 | 345 |
| Cole, J. E. | See Fifer, F. A. | 2 | 13 |
| Coleman, A. J. | The Adsorption and Electrooxidation of Simple Hydrocarbons for Direct Oxidation Hydrocarbon-Air Fuel Cells | 1 | 385 |
| Collett, E. | Spatial Coherence and Intensity Properties of Quasihomogeneous Optical Sources | 1 | 399 |
| Comas, J. | See Zavada, J. M. | 3 | 445 |
| Cook, C. F., Jr. | See Iafrate, G. J. | 2 | 295 |
| Corbin, N. D. | See McCauley, J. W. | 2 | 453 |
| Crowson, A. | Powder Metallurgy Steel Forgings for Small Arm Applications | 1 | 413 |
| Cuthbert, B. N. | See Graeber, R. C. | 2 | 73 |
| Cutright, D. E. | See Grower, M. F. | 2 | 105 |
| Cutright, D. E. | See Posey, W. R. | 3 | 137 |
| D'Agostino, J.<br>Lillie, T. | Overhead IR Target Signature Data Base | 4 | 43 |
| Daniels, W. L.<br>Wright, J. E.<br>Knapik, J. J.<br>Vogel, J. A.<br>Friman, G.<br>Beisel, W. R. | The Effect of Experimentally Induced Viral Infection on Physical Performance Capacity | 1 | 429 |
| Daniels, W. L. | See Wright, J. E. | 3 | 381 |
| Daum, G. R. | Broadband Absorption Studies | 4 | 55 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Dean, A. M., Jr. | Evaluation of Ice-Covered Water Crossings | 1 | 443 |
| Dehn, J. | The Particle Dynamics of Penetration | 1 | 455 |
| Devine, M. P.<br>Brodman, B. W. | The Effect of Chemical Interactions on the Small Arms Deterring Process | 1 | 467 |
| Dewar, M. J. S. | See Alster, J. | 1 | 91 |
| Dixon, S.<br>Jacobs, H. | Millimeter-Wave Self Mixing Oscillators | 1 | 473 |
| Doctor, B. P.<br>Brown, N. D.<br>Sleeman, H. K. | The Stability of Benactyzine.HCl Under Simulated Storage and Packing Conditions | 1 | 487 |
| Douglas, C. D.<br>Lewis, R. W. | Advanced Composite Applications to Large Caliber Weapons Systems | 1 | 503 |
| Drabo, M. J. | See Klarich, C. R. | 2 | 365 |
| Embury, J. F. | In Search of Strong Infrared Extinction in Aerosols | 2 | 1 |
| Faden, A. I. | See Holaday, J. W. | 2 | 233 |
| Fifer, R. A.<br>Cole, J. E. | Burning Rate Transitions for HMX Burned as a Binderless Propellant | 2 | 13 |
| Figucia, F. | Energy Absorption of Kelvar® Fabrics Under Ballistic Impact | 2 | 29 |
| Fileccia, R. J. | See Vogel, R. S. | 3 | 329 |
| Filler, M. | See Berry, P. | 1 | 197 |
| Francesconi, R. | See Banderet, L. E. | 1 | 167 |
| Frey, R. B. | The Initiation of Explosive Charges by Rapid Shear | 2 | 43 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Frickel, R. H. | See Stuebing, E. W. | 3 | 265 |
| Friman, G. | See Daniels, W. L. | 1 | 429 |
| Gavanis, T. J. | See Zavada, J. M. | 3 | 445 |
| Gilbert, A. L. | See Machuca, R. | 2 | 427 |
| Goldstein, S. | Interior Ballistic Modeling for Blank Ammunition | 2 | 59 |
| Graeber, R. C.<br>Cuthbert, B. N.<br>Sing, H. C.<br>Schneider, R. J.<br>Sessions, G. R. | Rapid Transmeridian Deployment: Cognitive Performance and Chronobiologic Prophylaxis for Circadian Dyschronism | 2 | 73 |
| Gray, J. | Army GPS-Doppler Hybrid Navigation System | 2 | 89 |
| Grower, M. F.<br>Cutright, D. E.<br>Russell, E. A., Jr. | Regeneration of Surgically Excised Segments of Dog Esophagus Using Biodegradable Pla Hollow Organ Grafts | 2 | 105 |
| Gualtieri, J. G. | See Aucoin, T. R. | 1 | 121 |
| Hackett, R. M. | Three-Dimensional Finite Element Combustion Instability Analysis | 2 | 119 |
| Hahn, F. E. | Experimental Chemotherapy: A Rapid and Simple Screening Method for Drug Binding to DNA | 2 | 133 |
| Hammond, C. E.<br>Cline, J. H. | On the Use of Active Higher Harmonic Blade Pitch Control for Helicopter Vibration Reduction | 2 | 147 |
| Hansen, F. V.<br>Pena, R.<br>Umstead, R. K. | Deliberate Air Pollution: The Art of Smoke Screening | 2 | 165 |
| Hartman, R. L. | See Christensen, C. R. | 1 | 369 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Harvey, J. F. | Analysis of Multipass Laser Amplifier Systems for Storage Laser Media | 2 | 181 |
| Hatfield, G. | See Berry, P. | 1 | 197 |
| Hawley, C. E. | See Bussell, N. E. | 1 | 265 |
| Heberlein, D. C.<br>Atkins, R. M. | Projection of Vehicle Magnetic Signatures for the Defeat of Magnetic Mine Influence Sensors | 4 | 67 |
| Heaston, R. J. | Redefinition of the Four Fundamental Forces | 2 | 203 |
| Helbert, J. N. | See Iafrate, G. J. | 2 | 295 |
| Henry, M. C.<br>Laible, R. C. | Improvements in CB Protective Clothing | 2 | 219 |
| Holaday, J. W.<br>Faden, A. I. | The Role of Endorphins in the Pathophysiology of Shock and the Therapeutic Benefit of Opiate Antagonists | 2 | 233 |
| Holst, G. C. | Infrared Transmission Measurements Through Screening Smokes: Experimental Considerations | 2 | 247 |
| Hoogterp, F. B.<br>Beck, R. R. | Vehicle Mobility or Firing Stability-a Delicate Balance | 2 | 263 |
| Houghton, W. W. | See Shuford, R. J. | 3 | 235 |
| Houle, M. J.<br>Janroga, S. | The Effect of EA 3834 on Plants | 4 | 81 |
| Howgate, D. W. | See Bowden, C. M. | 1 | 225 |
| Huber, W. A. | Coding and Processing for Reliable Data Transmission | 2 | 279 |
| Huddleston, R. L. | See Klarich, C. R. | 2 | 365 |
| Hunsperger, R. G. | See Zavada, J. M. | 3 | 445 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Iafrate, G. J.<br>Helbert, J. N.<br>Ballato, A. T.<br>Cook, C. F., Jr.<br>McAfee, W. S. | Fundamental Limitations of Electron Beam Lithography for Future Military IC Device Fabrication | 2 | 295 |
| Jacobs, H. | See Dixon, S. | 1 | 473 |
| Jahrling, P. B. | See Canonico, P. G. | 1 | 309 |
| Jankowski, P. Z. | See Cadwallender, W. K. | 1 | 293 |
| Janroga, S. | See Houle, M. J. | 4 | 81 |
| Jenkinson, H. A. | See Zavada, J. M. | 3 | 445 |
| Jennings, S. G. | See Pinnick, R. G. | 3 | 101 |
| Kascak, A. F. | The Response of Turbine Engine Rotors to Interference Rubs | 2 | 309 |
| Kayser, L. D. | See Nietubicz, C. J. | 3 | 17 |
| Keane, W. | See Berry, P. | 1 | 197 |
| Kearney, F. | Nondestructive Testing for Field Welds: Real Time Weld Quality Monitor | 2 | 325 |
| Kehs, R. A.<br>Brandt, H. E.<br>Bromborsky, A.<br>Lasche, G. | The Generation of Gigawatt Power Levels of Microwave Radiation | 2 | 339 |
| Kelso, D. H.<br>Marley, J. J. | Network Traffic Analysis Model | 2 | 351 |
| King, J. W. | See Ashman, W. P. | 1 | 105 |
| Kisatsky, P. J. | See Cadwallender, W. K. | 1 | 293 |
| Klarich, C. R.<br>Huddleston, R. L.<br>Drabo, M. J. | Electron Microscopy Used for Fracture Mode Identification | 2 | 365 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Klose, D. R.<br>Skudera, W. J. | Dual-Channel SAW Compressive Direction-Finding Techniques | 4 | 95 |
| Knapik, J. J. | See Daniels, W. L. | 1 | 429 |
| Knapik, J. J. | See Wright, J. E. | 3 | 381 |
| Kowal, D. | See Banderet, L. E. | 1 | 167 |
| Kramer, K. | See Cadwallender, W. K. | 1 | 293 |
| Kulpa, S. M. | See Weber, B. A. | 3 | 371 |
| Kuzanek, J. F. | Improved Methods for Computing Drag Corrected Missile Impact Predictions in Real Time | 2 | 381 |
| Lahoti, G.<br>Lee, F. M. | Application of Process Modeling to Shell Drawing Operations | 2 | 395 |
| Laible, R. C. | See Henry, M. C. | 2 | 219 |
| Lampo, R. G. | See Vogel, R. S. | 3 | 329 |
| Larson, D. C. | See Zavada, J. M. | 3 | 445 |
| Lasche, G. | See Kehs, R. A. | 2 | 339 |
| Lee, F. M. | See Lahoti, G. | 2 | 395 |
| Lee, J. N. | See Berg, N. J. | 4 | 1 |
| Lewis, R. W. | See Douglas, C. D. | 1 | 503 |
| Lillie, T. | See D'Agostino, J. | 4 | 43 |
| Link, L. E. | Terrain Thermal Modeling for Camouflage and Target Acquisition | 2 | 411 |
| Little, J. S. | See Canonico, P. G. | 1 | 309 |
| Lukaszek, T. | See Ballato, A. | 1 | 151 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Lund, D. J. | See O'Mara, P. A. | 3 | 31 |
| Machuca, R.<br>Gilbert, A. L. | Finding Edges in Noisy Scenes | 2 | 427 |
| Malik, R. J. | See Ross, R. L. | 3 | 205 |
| Marley, J. J. | See Kelso, D. H. | 2 | 351 |
| Martin, L. A. | High Performance Vehicles | 2 | 443 |
| McAfee, W. S. | See Iafrate, G. J. | 2 | 295 |
| McCauley, J. W.<br>Corbin, N. D. | Transparent, Polycrystalline Cubic Aluminum Oxide | 2 | 453 |
| McKnight, C. E. | Disposal of Soluble Salt Waste From Coal Gasification | 2 | 469 |
| Mikucki, W. J. | See Vogel, R. S. | 3 | 329 |
| Miller, J. A. | See Chubb, D. W. J. | 4 | 17 |
| Miller, R. A. | See Bussell, N. E. | 1 | 265 |
| Miller, R. J., Jr. | See Reichard, D. W. | 3 | 169 |
| Moore, R. L. | Non-Linear Least Chi-Square Algorithm-an Improvement on Non-Linear Least Squares | 2 | 483 |
| Moss, G. L. | Armor Design Based on Material Properties | 2 | 499 |
| Moulton, J. R.<br>Cassidy, T. W. | Grafenwoehr II Realistic Battlefield Sensors Trials | 4 | 111 |
| Mylin, D. C. | See Sturek, W. B. | 3 | 281 |
| Neely, E. S.<br>Clark, E. J. | Case Study Use of EDITSPEC, the Corps of Engineers Computer Aided Specification Preparation System | 3 | 1 |
| Nelson, J. F. | See Posey, W. R. | 3 | 137 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Nietubicz, C. J.<br>Kayser, L. D. | A Comparison of Navier-Stokes Computations With Experimental Data for a Projectile Shape at Transonic Velocities | 3 | 17 |
| Nomiyama, N. T. | See Rohde, R. S. | 3 | 195 |
| Obert, L. P.<br>Ratches, J. A. | Realistic European Battlefield Target Acquisition Model | 4 | 127 |
| O'Mara, P. A.<br>Stamper, D. A.<br>Beatrice, E. S.<br>Lund, D. J. | Laboratory Investigations of Biomedical Factors Influencing Laser Designator Operator Performance | 3 | 31 |
| O'Mara, P. A. | See Zwick, H. | 3 | 461 |
| Oscar, K. J. | Effects of Low Power Microwaves on the Local Cerebral Blood Flow of Conscious Rats | 3 | 45 |
| Owens, F. J. | Molecular Level Modeling of the Mechanism of Shock Initiation of Solid Explosives | 3 | 55 |
| Patton, J. F. | See Vogel, J. A. | 3 | 313 |
| Patton, J. F. | See Wright, J. E. | 3 | 381 |
| Pena, R. | See Hansen, F. V. | 2 | 165 |
| Peterson, E. G.<br>Salomon, L. L. | Methodology for Evaluation of Obscuration | 3 | 71 |
| Perkins, J. S. | Laser Interaction with TBR Materials | 3 | 85 |
| Pinnick, R. G.<br>Jennings, S. G.<br>Chylek, P. | Relationships Between Extinction and Mass Content of Atmospheric Fog and Military Smokes | 3 | 101 |
| Pleasants, W. A., III<br>White, G. T., III | Status of Improved Autorotative Landing Capability Research | 3 | 117 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Posey, W. R.<br>Cutright, D. E.<br>Russell, E. A., Jr.<br>Nelson, J. F.<br>Clare, V. R. | The Effects of High Velocity Variable Mass Projectiles on the Maxillofacial Complex | 3 | 137 |
| Prask, H. J.<br>Choi, C. S.<br>Trevino, S. F. | Nondestructive Testing of Armament-System Components by Means of Neutron Diffraction | 3 | 153 |
| Ratches, J. A. | See Obert, L. P. | 4 | 127 |
| Redwinski, R. J. | See Smith, R. C. | 4 | 171 |
| Reichard, D. W.<br>Miller, R. J., Jr. | Chemiluminescence Immunoreactive Assay (CLIA): A Rapid Method for the Detection of Bacterial and Viral Agents - Francisella Tularensis, Live Vaccine Strain (LVS) and Venezuelan Equine Encephalomyelitis Vaccine Strain (VEE TC-83) | 3 | 169 |
| Richmond, P. W., III | See Aitken, G. W. | 1 | 31 |
| Ritchie, J. P. | See Alster, J. | 1 | 91 |
| Rittenbach, O. E. | A New Technique for Doppler Frequency Analysis of Radar Signals | 3 | 181 |
| Robbins, F. M. | See Walker, J. E., Jr. | 3 | 345 |
| Rohani, B. | See Baladi, G. Y. | 1 | 135 |
| Rohde, R. S.<br>Buser, R. G.<br>Nomiyama, N. T. | Pulse Code Modulation of $CO_2$TEA Laser Pulse | 3 | 195 |
| Ross, R. L.<br>Aucoin, T. R.<br>Savage, R. O.<br>Winter, J. J.<br>Malik, R. J. | Semi-Insulating Gallium Arsenide for Millimeter Wave and High Speed IC Device Applications | 3 | 205 |
| Rubel, G. O. | See Stuebing, E. W. | 3 | 265 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Russell, E. A., Jr. | See Grower, M. F. | 2 | 105 |
| Russell, E. A., Jr. | See Posey, W. R. | 3 | 137 |
| Sacco, W. J. | See Ashman, W. P. | 1 | 105 |
| Salomon, L. L. | See Peterson, E. G. | 3 | 71 |
| Sampson, J. B. | See Wright, J. E. | 3 | 381 |
| Savage, J. J.<br>Shaffer, R. E. | Multispectral Screening Agent Studies | 4 | 143 |
| Savage, R. O. | See Aucoin, T. R. | 1 | 121 |
| Savage, R. O. | See Ross, R. L. | 3 | 205 |
| Schneider, R. J. | See Graeber, R. C. | 2 | 73 |
| Schwartz, A. | See Aucoin, T. R. | 1 | 121 |
| Scungio, R. C.<br>Chernick, J. A. | Performance Evaluation for Direct Fire Systems | 3 | 219 |
| Seemuller, W. W. | See Chen, P. F. | 1 | 353 |
| Sessions, G. R. | See Graeber, R. C. | 2 | 73 |
| Shaffer, R. E. | See Savage, J. J. | 4 | 143 |
| Shank, E. B.<br>Thein, B. K. | MOUT/Assault Weapon Status | 4 | 159 |
| Share, S.<br>Wasilik, J. | Nuclear Radiation Effects in Fiber-Optic Waveguides | 3 | 223 |
| Shuford, R. J.<br>Houghton, W. W. | Acoustic Emission as a NDE Technique for Determining Composite Rotor Blade Reliability | 3 | 235 |
| Sing, H. C. | See Graeber, R. C. | 2 | 353 |
| Skudera, W. J. | See Klose, D. R. | 4 | 95 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Slagg, N. | See Alster, J. | 1 | 91 |
| Sleeman, H. K. | See Doctor, B. P. | 1 | 487 |
| Smith, R. C.<br>Redwinski, R. J. | PATRIOT Survivability Study | 4 | 171 |
| Soicher, H. | Correlation and Prediction of Propagation Time-Delays Along Earth-Space Links | 3 | 251 |
| Stamper, D. A. | See O'Mara, P. A. | 3 | 31 |
| Stephen, E. L. | See Canonico, P. G. | 1 | 309 |
| Stokes, J. W. | See Banderet, L. E. | 1 | 167 |
| Stuebing, E. W.<br>Frickel, R. H.<br>Rubel, G. O. | Recent Research on Phosphorus Smoke | 3 | 265 |
| Sturek, W. B.<br>Mylin, D. C.<br>Bush, C. C. | Computational Parametric Study of the Aerodynamics of Spinning Slender Bodies at Supersonic Speeds | 3 | 281 |
| Tarbell, A. B. | Noise Performance of a New Type of Low Noise FM Detector | 3 | 297 |
| Thein, B. K. | See Shank, E. B. | 4 | 159 |
| Thornton, W. | See Ashman, W. P. | 1 | 105 |
| Trevino, S. F. | See Prask, H. J. | 3 | 153 |
| Umstead, R. K. | See Hansen, F. V. | 2 | 165 |
| Van Sice, C. W. | See Zwick, H. | 3 | 461 |
| Vogel, J. A.<br>Wright, J. E.<br>Patton, J. F. | Development of New Gender-Free Physical Fitness Standards for the Army | 3 | 313 |
| Vogel, J. A. | See Daniels, W. L. | 1 | 429 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Vogel, J. A. | See Wright, J. E. | 3 | 381 |
| Vogel, R. S.<br>Fileccia, R. J.<br>Mikucki, W. J.<br>Lampo, R. G. | Studies in the Identification of Hydrocarbon Products in Wastewater | 3 | 329 |
| Wade, M. J. | See Aucoin, T. R. | 1 | 121 |
| Walker, J. E., Jr.<br>Robbins, F. M.<br>Cohen, S. H. | The Effects of Catheptic Enzymes on Muscle Proteins | 3 | 345 |
| Wang, C. P.<br>Brynjolfsson, A. | Heat Conduction in Finite Cylinders and the Computer-Aided Calculation of Bacteria Survival in Heat Sterilization | 3 | 357 |
| Wasilik, J. | See Share, S. | 3 | 223 |
| Weber, B. A.<br>Kulpa, S. M. | The Extension of [Hg,Cd] Te Detector Technology to the Near-Milimeter Spectral Region | 3 | 371 |
| Wells, C. | See Alster, J. | 1 | 91 |
| White, G. T., III | See Pleasants, W. A., III | 3 | 117 |
| Winter, J. J. | See Ross, R. L. | 3 | 205 |
| Wright, J. E.<br>Vogel, J. A.<br>Sampson, J. B.<br>Patton, J. F.<br>Daniels, W. L.<br>Knapik, J. J. | Physiological Work Capacity and Performance of Soldiers Following Transatlantic Deployment | 3 | 381 |
| Wright, J. E. | See Daniels, W. L. | 1 | 429 |
| Wright, J. E. | See Vogel, J. A. | 3 | 313 |

| Author | Title | Vol | Page |
|---|---|---|---|
| Wright, T. W. | Penetration With Long Rods: A Theoretical Framework and Comparison With Instrumented Impacts | 3 | 397 |
| Wu, J. J. | The Initial Boundary Value Problem of Gun Dynamics Solved by Finite Element-Unconstrained Variational Formulations | 3 | 413 |
| Yalamanchili, R. | New Concepts in Recoil Mechanisms | 3 | 425 |
| Yee, Y. P. | See Bruce, D. | 1 | 255 |
| Zavada, J. M.<br>Jenkinson, H. A.<br>Gavanis, T. J.<br>Hunsperger, R. G.<br>Larson, D. C.<br>Comas, J. | Ion Implanted Guided Wave Devices for Army Fire Control | 3 | 445 |
| Zwick, H.<br>Bigge, S. L.<br>O'Mara, P. A.<br>Van Sice, C. W. | A Solid State Dark Adaptometer | 3 | 461 |

# IN SEARCH OF STRONG INFRARED EXTINCTION IN AEROSOLS

JANON F. EMBURY, Ph.D.
Chemical Systems Laboratory, USAARRADCOM
Aberdeen Proving Ground, MD 21010

## INTRODUCTION

There is no optical region more amenable to theoretical treatment of nonspherical and compound particles than the Rayleigh region. Here it is possible to calculate with relative ease, the extinction due to any particle which can be represented by an ellipsoid[1] or a confocally coated ellipsoid[2]. This is not very restrictive because nearly all convex particles can be accurately described by an ellipsoid envelope, with the exception of polyhedra. Coated spheres, needles and discs are accurately represented. Recently lattice dynamical and dielectric continuum calculations were made predicting extinction by Rayleigh cubes and polyhedra[3-5]. It is even possible to model accurately the extinction caused by a collection or irregular shapes, the most common form for solid aerosols. From the standpoint of producing strong extinction per unit mass, the Rayleigh region is the most fruitful for even a moderately absorbing particle. Here extinction as a function of size attains a high plateau which remains independent of size so long as we stay within the region. Here we do not have to depend, as is often the case elsewhere, on narrow extinction resonances which all but disappear when particle size, shape, composition or orientation change only slightly. The Rayleigh region holds the most promise for yielding particles engineered to produce strong broad electromagnetic extinction, and therefore it should be explored as extensively as possible.

## DISCUSSION

The Rayleigh region is constrained by definition to be located

where particle size is small with respect to wavelength both inside and outside the particle. Here there is a close connection linking the static polarization of a particle and its extinction per unit mass. Before describing this connection first we will take a closer look at the extinction per unit mass which will be referred to simply as the extinction. Extinction, $\gamma$, determines the fraction of incident radiation, T, which passes a distance, L, through an aerosol cloud having concentration, C, in accordance with Beers Law.

$$T = e^{-\gamma CL}$$

A convenient self-consistent set of units puts extinction in square meters per gram of aerosol, concentration in grams of aerosol per cubic meter of air, and pathlength in meters. The extinction coefficient depends on the geometric cross section, G, optical extinction efficiency factor, Q, and particle weight, W, in the following way.

$$\gamma = \left\langle \left\langle \frac{GQ}{W} \right\rangle_\theta \right\rangle_m$$

The inner brackets represent an average over solid angle to take into account random orientations experienced by particles in the cloud while the outer brackets represent an average over particle mass (size) distribution in the cloud. This double integral simplifies to the following single integral expression because $\langle GQ/W \rangle_\theta$ is independent of particle size for an absorbing particle in the Rayleigh region and of course particle weight is independent of orientation.

$$\left\langle \left\langle \frac{GQ}{W} \right\rangle_\theta \right\rangle_m = \frac{\langle GQ \rangle_\theta}{W}$$

The fundamental extinction theorem relates the extinction cross section, GQ, to the real part of the scatter amplitude in the forward direction, $\mathrm{Re}\{S(0)\}$, for radiation of wavelength, $\lambda$.

$$GQ = \frac{\lambda}{\pi} \mathrm{Re}\{S(0)\}$$

The Rayleigh theory relates the static complex polarizability, $\alpha$, of an absorbing particle to its forward scatter amplitude.

$$S(0) = i \frac{2\pi}{\lambda^3} \alpha$$

Substituting this value into the fundamental extinction theorem we find for a single particle at one orientation,

$$GQ = \frac{8\pi}{\lambda^2} \mathrm{Re}\{(i\alpha)\}$$

The extinction cross section resulting from a collection of randomly

oriented ellipsoidal particles can be expressed as a function of a complex polarizability which is just one third the sum of the polarizabilities, $\alpha_j$, along the major ellipsoidal axes.

$$\alpha = (\alpha_1 + \alpha_2 + \alpha_3)/3$$

The major axis polarizabilities of an ellipsoidal particle are directly proportional to particle volume, V, and depend on complex refractive index, n, and depolarization factors, $L_j$, along the major ellipsoidal axes a, b, and c.

$$\alpha_j = \frac{V(n^2-1)}{L_j(n^2 - 1) + 1}$$

where $n = N + iK$

$$L_1 = \int_0^\infty \frac{abc\ ds}{2\ (S+a^2)^{3/2}\ (s+b^2)^{1/2}\ (s+c^2)^{1/2}}$$

with appropriate cyclical changes for $L_2$ and $L_3$. The three depolarization factors have the properties that each is nonnegative and their sum is one.

The depolarization factors of spheroids depend only on the aspect or axial ratio of the axis of symmetry with respect to either one of the other two equal major axes. For a prolate spheroid with axial ratio, A, defined always to be greater than one;

$$L_1 = \frac{1-e^2}{e^2}\left(-1 + \frac{1}{2e}\ln\frac{1+e}{1-2}\right)$$

$$L_2 = \frac{1 - L_1}{2} \qquad L_3 = L_2 \qquad e^2 = 1-1/A^2$$

For an oblate spheroid with aspect ratio, A, also defined to be greater than one;

$$L_1 = \frac{1 + f^2}{f^2}\left(1 - \frac{1}{f}\ \text{arc tan}\ f\right)$$

$$L_2 = \frac{1 - L_1}{2} \qquad L_3 = L_2 \qquad f^2 = A^2 - 1$$

We may now write the extinction for a cloud of randomly oriented Rayleigh ellipsoidal or spheroidal particles.

$$\gamma = \frac{8\pi^2}{3\lambda W}\sum_{j=1}^{3} \frac{V\ 2NK}{\left\{L_j\ (N^2-K^2) + (1-L_j)\right\}^2 + (2NKL_j)^2}$$

When the wavelength and density, $\rho$, are moved over to the left hand side of the equation, shape dependence is more clearly evident on the right hand side. In this way, wavelength and density, which are altogether unrelated to shape, no longer appear as independent variables and extinction values will not be tied to a single wavelength and density.

$$\gamma_{\rho\lambda} = \frac{16\pi^2}{3} \sum_{j=1}^{3} \frac{NK}{\{L_j(N^2-K^2) + (1-L_j)\}^2 + (2NKL_j)^2}$$

It is this quantity $\gamma_{\rho\lambda}$ that we choose to plot on the contour maps. This frees us to select any wavelength, density combination and interpret extinction isopleths accordingly. One self-consistent set of units convenient to use in the infrared holds the extinction in square meters per gram, puts density in grams per cubic centimeter, and wavelength in micrometers. In this system of units, extinction · density wavelength contour maps may be said to represent, for example, extinction at one micrometer and unit density or at half a micrometer and a density of two.

When the shape and dielectric properties of the particle combine to reduce the value of the denominator in the previous equation, there is an increase in extinction or equivalently a resonance[6]. The shape dependence of this resonance enters through the depolarization factor. A sphere has depolarization factors equal to 1/3 for fields applied along any orthogonal set of three radial axes. A prolate spheroid approximating a thin needle has $L_j \to 0$ for fields applied parallel to its length and $L_j \to 1/2$ for fields in the plane of symmetry. An oblate spheroid approximating a thin disk has $L_j \to 1$ for fields applied perpendicular to the plane of symmetry and $L_j \to 0$ for fields in the symmetry plane.

A resonance occurs when the following conditions are satisfied:

$$L_j (N^2 - K^2) + (1-L_j) = 0 \qquad 2NKL_j = 0$$

Remembering that all depolarization factors are nonnegative, we find that for a given shape and depolarization factor the resonant values for the optical constants which satisfy both conditions are

$$N = 0 \qquad K = (1/L_j - 1)^{1/2}$$

A close look at extinction in the limit $n \to 0$ and K equal to the above value reveals that this in indeed an extinction pole. All three depolarization factors for a sphere are equal to 1/3 and its resonance therefore occurs at N=0 and $K = \sqrt{2}$.

The special case where $L_j \to 0$ warrents some discussion. This can occur, as mentioned before, for selected orientations of needles and disks with respect to the electric field vector. Taking this limit in the previous equation relating extinction to complex refractive index and depolarization factor, the terms in the summation corresponding to depolarization approaching zero become simply

$$\gamma_{\rho\lambda} = \frac{16\pi^2}{3} \sum_{L_j=0} NK + \text{terms with } L_j \neq 0$$

These terms tend to dominate over terms which correspond to depolarization which is nonzero over most of the complex refractive index plane.

Figure 1 shows extinction density wavelength contours for a sphere over a complex refractive index range typical of metals at visible wavelengths and semiconductors at visible and infrared wavelengths. Future references to extinction density wavelength contours will be made simply to extinction for brevity. There is one resonance or region of strong extinction centered at N=0 and $K=\sqrt{2}$. This has been identified as the first electrostatic surface polariton mode. The region where the imaginary component of the refractive index is greater than the real component is the restrahl region. There must always be absorption here in order to satisfy the Kramers-Kronig relationships. Any point lying on the K axis violates this requirement. Therefore the K axis may only be approached, and extinction will always remain bounded. The closest approach to this resonance will be made by a material with maximal oscillator strength in a single oscillator within sum rule limits and with minimal oscillator damping[7].

In figure 2 the single resonance of a sphere has split into two resonances for an oblate spheroid at an aspect ratio of ten. One pole moves down the K axis toward the origin without noticeable changes in the extent of its contour values while the second resonance climbs steadily up the K axis, spreading its influence in the form of high extinction over more and more of the complex index plane while aspect ratio grows. A similar situation applies to prolate spheroids except that the resonance moving down the K axis converges on the value one.

A metal is the aerosol material best suited to take advantage of high extinction produced by the resonance moving up the K axis. A simple model[7] for the optical constants of a metal, the Drude theory model, puts the real parts of the refractive index equal to the imaginary index at wavelengths greater than ten microns. Extinction produced by metal prolate spheroids obeying the Drude theory appears in

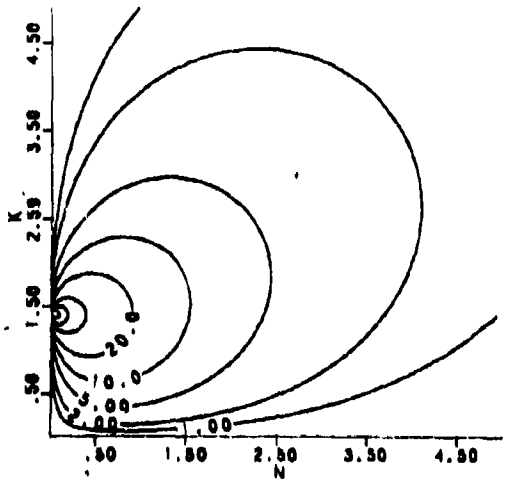



Figure 1 Extinction isopleths for a Rayleigh sphere as a function of the complex refractive index.

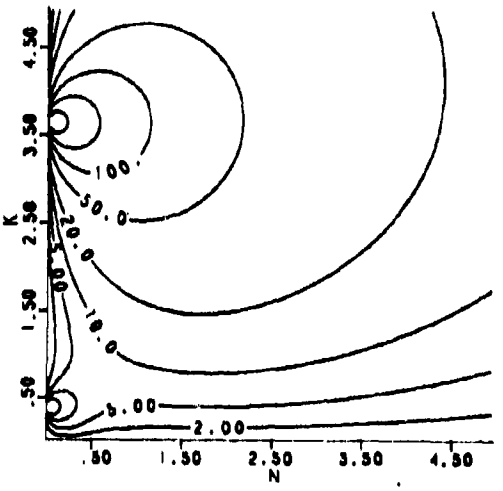



Figure 2 Extinction isopleths for a Rayleigh oblate spheroid with an aspect ratio of ten as a function of complex refractive index.

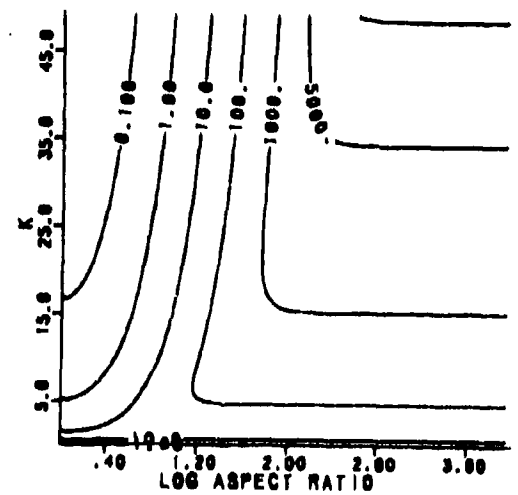



Figure 3 Extinction isopleths for a metal (N=K) Rayleigh prolate spheroid as a function of the base ten log of the aspect ratio and the imaginary part of the complex refractive index.

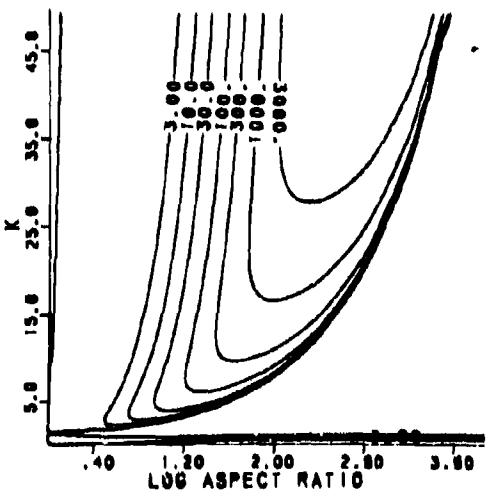



Figure 4 Extinction difference isopleths between metal (N=K) spheroids, prolate minus oblate compared at equal aspect ratio and complex refractive index. Negative isopleths are deleted.

figure 3 as a function of the base ten logarithm of the aspect ratio and optical constants. This range of optical constants is typical of metals in the infrared. Metal oblate spheroids generate a similar appearing set of contours. In figure 4 extinction produced by a metal oblate spheroid is subtracted from that produced by a prolate spheroid having the same aspect ratio and optical constants. Negative contour values are not plotted. In figure 5 prolate spheroid extinction is subtracted from oblate, so the region where an oblate spheroid is superior to a prolate spheroid appears in the area which is contoured out to an aspect ratio of ten thousand. Increasing metallic refractive index and increasing aspect ratio work to produce high extinction as can be seen in figure 3. When comparing which shape causes higher extinction at a given complex index we see from figure 4 and 5 that the prolate shape is superior for aspect ratios less than some value while the oblate is superior for aspect ratios greater then that value. In comparisons at a fixed aspect ratio, we see that the oblate spheroid is superior below certain complex indices and for higher complex indices a prolate spheroid is better.

The original ellipsoidal theory of Rayleigh has been extended in recent years to include confocally coated ellipsoidal particles. The major axis polarizabilities, $\alpha$ , of a confocally coated ellipsoid are directly proportional to overall ellipsoid volume, V, and depend on inner (core) complex refractive index, $n_i$, outer (shell) complex refractive index, $n_o$, core volume fraction, $V_i/V$, inner depolarization factor, $L_i$, and outer depolarization factor, $L_o$.

$$\alpha_j = \frac{V}{4\pi} \frac{(n_o^2-1)\left\{L_i(n_i^2-n_o^2) + n_o^2\right\} + \frac{V_i}{V}(n_i^2-n_o^2)\left\{L_j(1-n_o^2)+ n_o^2\right\}}{\left\{(n_i^2-n_o^2)L_i+ n_o^2\right\}\left\{1+L_o(n_o^2-1)\right\} + \frac{V_i}{V} L_o(1-L_o)(n_i^2-n_o^2)(n_o^2-1)}$$

Where the depolarization factors and complex indices are defined as before for solid ellipsoids and spheroids. The confocal constraint requires that the difference between the square of the outer axes and the square of the inner colinear axes remain constant, resulting in a thinner coating on larger axes. The value for this constant is determined by the core ellipsoid volume fraction. It is important to recognize that a confocal constraint removes a degree of freedom from simultaneous specification of inner depolarization factor, outer depolarization factor and core volume fraction. Once any two of these variables are specified, the third is automatically determined by the constraint.

If the coated ellipsoid is a coated prolate spheroid, both the inner and outer spheroids will be prolate with two foci points in

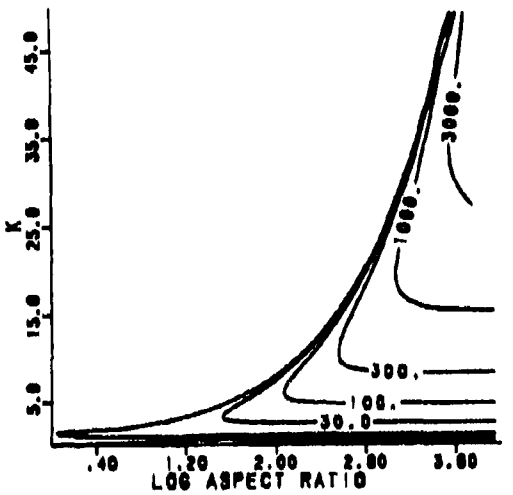



Figure 5. Extinction difference isopleths between metal (N=K) spheroids, oblate minus prolate, compared at equal aspect ratio and equal complex refractive index. Negative isopleths are deleted.

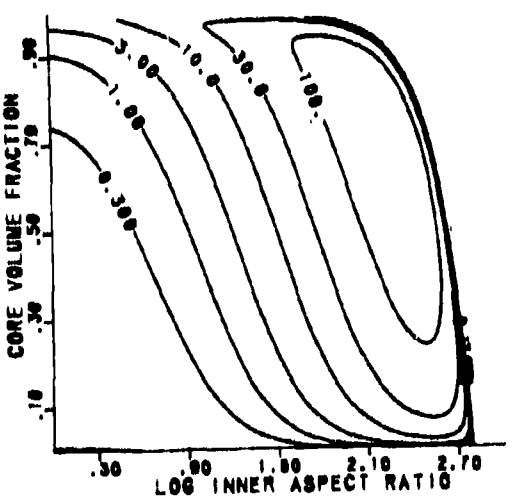



Figure 6. Extinction difference isopleths coated minus homogeneous, between a metal mantle (N=20, K=20) dielectric core (N=1.5, K=.01) oblate spheroid and a homogeneous oblate spheroid of the same metal. Aspect ratios of the core and homogeneous spheroid are equal. Negative contours are deleted.

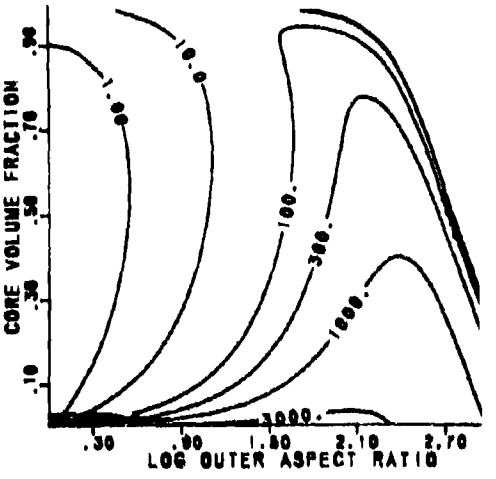



Figure 7. Extinction difference isopleths, coated minus homogeneous, between a metal mantle (N=20, K=20), dielectric core (N=1.5, K=.01) oblate spheroid and a homogeneous oblate spheroid of the same metal. Aspect ratios of the mantle outer surface and homogeneous spheroid are equal. Negative contours are deleted.

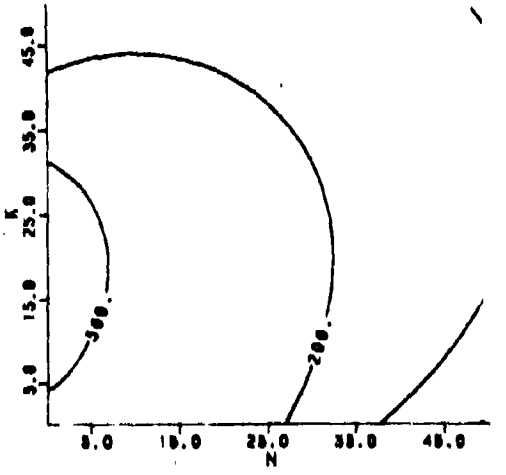



Figure 8. Extinction isopleths for a metal mantle (N=20, K=20) oblate spheroid at an outer aspect ratio of 100 and a core volume fraction of 0.5 as a function of core complex refractive index.

common, and if it is a coated oblate spheroid then both will be oblate with a foci ring of points in common. Both inner and outer depolarization factors depend upon aspect ratios as they did for solid spheroids. However, due to a constraint a choice must be made as to which two of the variables, inner aspect ratio, outer aspect ratio or core volume fraction will be specified. Inner aspect ratio is always greater than or equal to outer aspect ratio. The ratio of inner aspect ratio over outer aspect ratio becomes larger as the core volume fraction becomes smaller.

Altogether seven variables are used to characterize a confocally coated spheroid. These variables are inner and outer complex refractive indices, core volume fraction, inner aspect ratio and outer aspect ratio. As a result of the confocal constraint only two of the last three variables are independent leaving six independent variables governing extinction. Projecting contours onto a two dimensional plane was an effective technique to locate strong extinction produced by a solid spheroid where there were only three governing independent variables but not when there are six variables. The appropriate procedure for seeking out extinction maxima in six dimensions is to employ a function extrema search algorithm. Such a search is being undertaken and it is too early to describe the results, however it is valuable to portray in two dimensions what happens to the strong extinction observed earlier due to high aspect ratio solid metal spheroids when a dielectric confocal coating is applied over the outside.

Calculations have been made for metal core dielectric coated spheroids and it was found that dielectric coatings reduce extinction at all aspect ratios not close to one for both oblate and prolate spheroids. On the other hand it has been discovered that applying a metal coating over a dielectric core can significantly increase extinction for both prolate and oblate spheroids. Figures 6 and 7 map the extinction difference between coated and uncoated oblate spheroids. Extinction due to solid oblate spheroids with real index 20 and imaginary index 20 is subtracted from the extinction due to coated oblate spheroids with dielectric core real index 1.5 imaginary index 0.01 and metal coating complex refractive index equal to that of the solid spheroid. In figure 6, the solid spheroid aspect ratio is set equal to the inner aspect ratio while in figure 7 it is equal to outer aspect ratio. These contour maps are similar to those of a prolate spheroid except that the contour marking the boundary where extinction of both coated and uncoated become equal occurs at an aspect ratio nearly an order of magnitude smaller. A substantial increase in extinction is evident in figures 6 and 7 as a result of the dielectric core up to aspect ratios of nearly one thousand for core volume fractions between 0.1 and 0.9. At larger aspect ratios the solid spheroid

is superior, however this cutoff depends on the optical constants of the metal.

In figure 8, we can see what happens to this coated spheroid at an outer aspect ratio of one hundred and core volume fraction of 0.5 when we adjust the core refractive index to determined what improvements are possible over the dielectric core just discussed. The original dielectric core with real index 1.5 and imaginary index 0.01 is in an area of relatively strong extinction but figure 8 shows that by increasing the core imaginary index above 0.01 extinction will get larger. The optimal core would have a small real index and an imaginary index between 5 and 30; values to be expected in a metal a frequencies just below the plasma frequency at visible wavelengths.

As mentioned earlier the Rayleigh ellipsoidal theory is not a good approximation for particles that have edges and vertices such as polyhedra. Fortunately dielectric continuum and lattice dynamical calculations predict the absorption properties of cubes and other polyhedra small compared to wavelength. The results of such calculations for a cube appear in figure 9. Once again resonances emerge along the K axis but now there are six resonances located between imaginary indices of one half and two with the strongest resonance appearing furthest up the axis. Extinction far removed from these resonances or outside the restrahl region is almost indistinguishable from that of a small sphere.

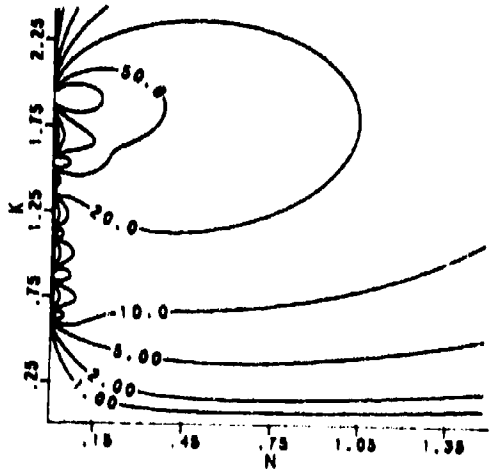



Figure 9. Extinction isopleths for a cube as a function of complex refractive index.

## CONCLUSIONS

When particle size becomes small compared to wavelength, extinction not only depends on geometric cross sections but also becomes a very strong function of complex refractive index. Regions of resonant extinction were located in the complex refractive index plane for a variety of spheroidal and confocally coated spheroidal particles using the Rayleigh ellipsoidal approximation. The strongest most extensive is found for high aspect ratio metal prolate and oblate spheroids approximating thin needles and disks. Even stronger extinction was found to occur when metal needles and disks have a dielectric core. On the other hand a dielectric coating reduces extinction. The unbounded peaks of extinction resonances, located where the real part of the complex refractive index becomes zero, were proven to be inaccessible to any material obeying the Kramers-Kroenig relation which requires absorption at restrahl. Finally the six resonances of a cubic particle were explored in the complex refractive index plane. Because a cube is an example of where the Rayleigh ellipsoidal approximation fails most severely, the extinction was predicted using a result taken from the dielectric continuum and the lattice dynamical theories.

## REFERENCES

1. Van de Hulst, H. C. "Light Scattering by Small Particles". Wiley, N.Y. 1957.

2. Gilra, D. P., Ph.D. Thesis. University of Wisconsin, Madison, Wisconsin. 1972

3. Chen, T. S., et. al. Phys. Rev. B, 18 958 (1978).

4. Fuchs, R. Phys. Rev. B, 11 1732 (1975).

5. Langbein, J. Phys. A. 9 627 (1976).

6. Huffman, D. R., 1977, Advances in Physics, 26, 129.

7. Ziman, J. M. "Principles of the Theory of Solids", Cambridge University Press, London. 1972.

# BURNING RATE TRANSITIONS FOR HMX BURNED AS A BINDERLESS PROPELLANT

*ROBERT A. FIFER, DR.
JAMES E. COLE
US ARMY BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MD 21005

## INTRODUCTION

The burn rates of propellants used in gun and rocket applications can usually be fit to an empirical equation of the form

$$R = AP^n \qquad (1)$$

where R is the burn rate, P is the pressure and A and n are constants. For a typical solid propellant, the pressure exponent n is about 0.8, and the burn rate increases from about 0.5 cm/s at a pressure of 1 MPa (150 psi) to roughly 70 cm/s at 500 MPa (72,500 psi). Rocket motors usually operate at a constant pressure of 69 MPa (10,000 psi) or less. Guns typically operate at peak pressures of 345 MPa (50,000 psi) or higher, although the pressure is much lower than the peak pressure for most of the time the projectile travels down the gun barrel.

Composite (non-homogeneous) propellants containing crystalline HMX (cyclotetramethylene-tetranitramine) or the similar compound RDX (cyclotrimethylene-trinitramine) imbedded in a polymeric binder are gradually replacing homogeneous nitrocellulose-based propellants in many gun and rocket applications. Propellants containing HMX or RDX are commonly called nitramine propellants.

A number of studies have been carried-out to determine the burn rate of pure HMX (1-4), using either large single crystals or pressed powder samples. Single crystal studies have only been successful to about 10 MPa, due to the tendency of the crystals to crack from the thermal stresses associated with burning at pressures higher than this. With pressed powder samples burn rates at pressures up to about

35 MPa (5000 psi) have been successfully measured, and then only by using extremely fine HMX particles (< 5 μm) pressed to very high densities. In most cases it was reported that sample "break up" occurred at a threshold pressure, preventing the burn rate from being measured at higher pressures.

We also have investigated the burn rate behavior of pressed binderless HMX. We have found that the "break up" reported by previous investigators is not random crumbling of the sample but rather an orderly transition from "normal" slow burning to "super-fast" burning. The regression rate increases by roughly three orders of magnitude at the transition pressure, causing the sample to appear to suddenly disappear in a cloud of smoke when observed by conventional photographic techniques. We have developed techniques for measuring very fast regression rates with sufficient accuracy to characterize the super-fast regression beyond the transition point. Using these techniques, we have shown that the regression attains velocities up to 6000 cm/s at high pressure, with pressure exponents as low as 0.3, depending on sample preparation.

These results are significant for certain applications that require extremely rapid gas generation rates. One such application being investigated at this laboratory is the "traveling charge gun" concept. In a traveling charge gun, a single solid cylinder of propellant (or a stack of different propellants) would be attached to the base of the projectile and burn rocket-style as the projectile moved down barrel, burning-out just before muzzle exit. If a near constant pressure were maintained at the propellant (i.e., projectile) base, extremely high ballistic efficiency would result leading to muzzle velocities considerably higher than for a conventional gun cycle for the same maximum gun pressure. Since burning takes place only at the base of the charge, extremely high burn rate propellant (several thousand cm/s or more depending on gun pressure and muzzle velocity desired) is required in order for the gas generation rate to be high enough to maintain constant pressure.

Prior to this study, the only propellant materials known to have burn rates high enough for traveling charge gun operation were certain carborane-based compositions. These materials are very expensive. HMX, on the other hand, is already being produced at low cost for propellant and explosive applications, and its use should make the traveling charge concept economically feasible.

## EXPERIMENTAL

### Preparation of HMX Samples

The pressed samples were prepared from three different particle sizes of military grade powdered HMX: as received class E (< 44 μm) and class C HMX screened into two size ranges, 149-297 μm and 105-149 μm. The samples were pressed at 110 MPa (16,000 psi) into parallelepiped strands having the dimensions 3.2 x 0.6 x 0.16-0.19 cm. The density of each parallelepiped was determined from its weight and measured dimensions. The percent theoretical maximum density (TMD), based on a density of 1.90 g/cc for crystalline HMX was 82% for class E, 92% for 105-149 μm class C and 95% for 149-297 μm class C.

One end of each sample was cemented to a holder which mounted securely inside the strand burner (for burn rate measurements) or window chamber (for high-speed photography). To assure one-dimensional (end-to-end) regression during burning, the four large sides of each sample were "inhibited" with a thin coating of epoxy to prevent flame spreading down the sides ahead of the regressing surface.

For the strand burner experiments, the regression rate was determined using two fuse wires -- one passing through the sample near the top, and the other near the base. To accommodate these wires, small (0.35 mm) holes were drilled through the 0.6 x 3.2 cm faces of the sample. The finest size fuse wire that could be handled without breaking (one quarter amp, 0.114 mm diameter) was used in order to achieve as rapid melting as possible. The fuse wire spacing (typically about 2.0 cm) was measured to $\pm$ 0.25 mm (1.5% or better). During an experiment a small current is passed through these wires. Burning of the propellant melts the fuse wires sequentially, generating electrical signals. The regression rate is then determined from the measured time interval and inter-wire distance. The HMX strands were ignited by a small cylindrical piece of JPN propellant that was mechanically attached to the top of the HMX sample and ignited by a heated wire through its center. This technique presumably produced more uniform ignition of the surface.

For the window chamber tests, the samples were not fitted with fuse wires. The samples were directly ignited by a heated wire on the end of the strand to prevent optical interference (flame, smoke, etc.) from burning pieces of JPN propellant.

Strand Burner, Window Chamber and Recording Equipment

In the propellant community, the term "strand burner" has come to mean any reactor in which a single "strand" of propellant can be burned one-dimensionally at constant, or near constant, pressure. Generally, constant pressure is achieved by simply keeping the volume of the system large compared to the mass of sample being burned. (In a "closed bomb", by way of contrast, a number of propellant grains are burned simultaneously on all exposed surfaces, and one measures the self-generated pressure as a function of time).

The strand burner used in these experiments is of conventional design. The main chamber is cylindrical, with a 4.5 cm internal diameter, a 36 cm internal length, and a free volume of 400 cc when assembled with the sample holder. Electrical feed-throughs are provided for the ignition current and time of event (fuse) wires. Compressors are used to prepressurize the burner with nitrogen a few minutes before igniting the sample.

The electrical signals resulting from melting of the fuse wires were recorded on separate channels of a magnetic tape, together with 0.1, 1.0 or 10 ms calibrated time marks. The signals were also used to start and stop a digital counter which measured the corresponding time interval to the nearest microsecond.

The window chamber used for the photographic measurements has a volume of 2700 cc and is equipped with windows sufficiently large to photograph the entire 3.2 cm length of the pressed strands. Movies have been obtained of samples burning at three different pressures: 3.45 MPa (500 psi), 13.8 MPa (2000 psi) and 34.5 MPa (5000 psi). Kodak type 7242 film was used, with a framing rate up to 11,000 frames per second for the fast-burning samples. All movies were obtained with no external illumination of the sample. The camera was started about one second before igniting the sample in order to enable it to reach full speed before sample burning. A blinking light inside the camera was used to put framing rate calibration marks along the edge of the film.

## RESULTS

Strand Burner Experiments

The results of the strand burner regression rate measurements for the three particle size pressed HMX samples are shown plotted in Fig. 1.

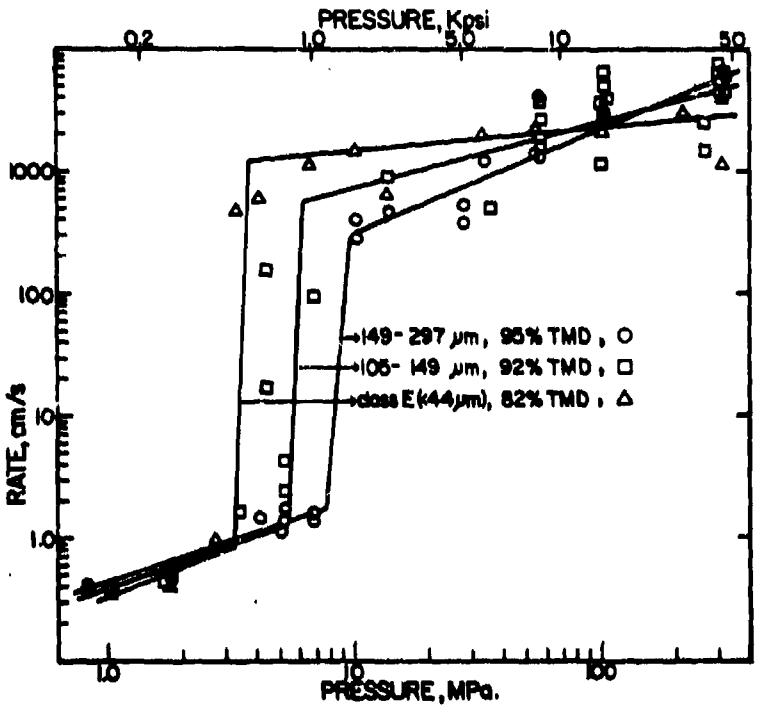



Fig. 1. Regression rate vs pressure for pressed binderless HMX

These samples can be seen to exhibit the following behavior: Below about 3.4 MPa (500 psi), all of the samples exhibit the "normal", slow HMX burn rate measured by other investigators. At higher pressures, each type sample undergoes a transition to super-fast regression. The transition pressure can be seen to increase with increasing particle size and sample density. Over a very small pressure range, the regression rates increase from about 1 cm/s to about 400 to 1000 cm/s. A second break then occurs and the regression rate increases slowly with increasing pressure. At pressures just above the transition pressure, the regression rate is greater, the smaller the HMX particle size. The three curves appear to converge to a rate of about 2500 cm/s at about 100 MPa (14,500 psi), and fall in the approximate range 3000-6000 cm/s at 345 MPa (50,000 psi). In the high pressure region, the apparent burn rate pressure exponent (see Eq. (1)) is less than 0.8 for the coarsest HMX particle size, and less than 0.3 for the finest HMX particle size material.

The solid lines in Fig. 1 were drawn in by simple visual inspection of the plotted data points. More sophisticated data analysis does not appear to be justified in view of the scatter in the data. As might be expected, reproducibility is worst in the transition and high pressure regions. The most likely explanation for the scatter is the finite melt time characteristics of the fuse wires. For a regression rate of 3000 cm/s and a typical inter-wire distance of 2.0 cm, the expected time interval is only 0.7 ms. For the fuse wires not to contribute to data scatter, they would have to melt on a time scale short compared to this (e.g., 0.1 ms), which is very unlikely. There appears to be much less scatter at high pressure for the class E HMX samples. This will be discussed below in connection with the proposed mechanism.

Window Chamber Experiments

The initial reason for carrying-out the window chamber experiments was to confirm the regression rates measured using the fuse wire technique. For the three pressures investigated (3.45, 13.8 and 34.5 MPa), the regression rates measured photographically agreed with those shown in Fig. 1. Note that at 11,000 frames per second, only 35 frames would record the regression of a 3.2 cm long sample burning down at 1000 cm/s. Much higher framing rate cameras would clearly be required to measure burn rates of 5,000 cm/s or higher.

In addition to confirming the measured strand burner regression rates, the high-speed movies provided valuable clues regarding the mechanism of the super-fast regression above the transition point. At 3.45 MPa (below the transition) the samples were observed to burn slowly with an orange flame, the height of which ($\sim$ 5 mm) is small compared to the length of the sample. At 13.8 or 34.5 MPa (above the transition), the samples burn very rapidly with a white to bluish flame that is very high -- extending beyond the top of the field of view even when the strand has burned down most of the way. (For "normal" burning, flame height would be expected to decrease, not increase, with increasing pressure). Moreover, luminosity fills the volume being observed for some time after the sample has burned down to its base. Finally, it was observed that the regressing surface above the transition pressure is very diffuse and poorly defined. These observations, together with a theoretical analysis, provide the basis for the mechanism proposed in the next section.

## DISCUSSION

### Mechanism of the High-Speed Regression

Above a certain transition pressure, the pressed HMX samples studied in this investigation exhibited very high regression rates, and therefore very high mass burning rates, M (g $cm^{-2}s^{-1}$). Mass burning rate can be related to the density of the propellant $\rho_s$ (g/cc), burn rate R (cm/s), burning surface area A ($cm^2$ per square cm of surface) and regression rate S (cm/s) by

$$M = \rho_s RA = \rho_s S \qquad (2)$$

Consequently, the primary question is whether the transition to high regression rates indicates a dramatic increase in burn rate, or in surface area. That HMX undergoes a transition at low pressures to super-high burn rates is not consistent with the following observations:

1. Other investigators have reported "normal" (slow) burning for their (finer particle size and/or higher density) pressed samples to pressures higher than the transition pressures of our samples.

2. We have burned unconsolidated (loose powder) HMX of different particle sizes in a closed bomb ($P_{max} \sim 190$ MPa). The measured pressurization rates were approximately what would be expected for the estimated surface area and a "normal" burn rate law.

3. The high-speed movies obtained in this study suggest a surface-area-increase mechanism, which is discussed in the following paragraphs.

It follows that the transition to very high mass burn rates is a reflection of an increase in burning surface area. There is more than one way by which this could occur. A number of reports in the literature discuss "convective burning", defined as in-depth burning due to diffusion of hot product gases into the porous surface of the propellant to create a greatly increased burning surface area.

Taylor (5), for example, has studied the burn rate behavior of HMX of different particle sizes, loosely packed to low densities ($\sim$ 1.05 g/cc or 55% TMD) in paper tubes and burned in a strand burner. He showed that these samples underwent a transition toward high apparent burning rate, the transition pressure being greater, the smaller the particle size of the HMX. (This trend is the opposite of that shown in Fig. 1). He interpreted his results in terms of

convective burning, and concluded that the transition occurs when a molten layer becomes too thin in relation to the pore diameter to provide a barrier to hot gas penetration.

Bobolev, et al. (6), have performed similar studies on low density (1.00-1.04 g/cc) RDX, as well as a number of other propellant materials having widely different melting temperatures. Since all behaved similarly, they concluded that disappearance of a melt layer was not the primary cause of the transition for their low density samples. They proposed instead that convective burning commenced when the pressure reached a critical value relative to the effective diameter of the pores at the surface.

Neither Taylor nor Bobolev attempted to measure regression rates above the transition pressures for their low density samples -- the highest regression rates reported were about 60 cm/s. We have shown that high density pressed HMX attains regression rates up to 3000-6000 cm/s at high pressure. Assuming a normal burning rate law for HMX, this corresponds to an increase in burning surface area of about three orders of magnitude, and is clearly inconsistent with an in-depth convective mechanism -- there is not enough internal surface area near the surface of the sample to explain such an increase.

We therefore propose that the increased surface area does not exist in the solid sample, but is generated by a progressive release of the particles which then burn in a jet flow moving away from the regressing surface, i.e., a "progressive deconsolidation" mechanism operates. In support of this mechanism, we found in our photographic observations a diffuse regression front, a dramatically increased flame height above the transition point, and luminosity persisting for some length of time after the sample has burned down. All these are consistent with such a mechanism for surface area generation. Figure 2 shows an idealized representation of our "progressive deconsolidation" mechanism.

The diffuse regression front results from the fact that there is no sharp boundary between solid and gas at the regressing surface -- particles are continuously breaking loose and accelerating away from the sample. The flame height corresponds to the distance of the particles from the sample when they burn out. The persisting luminosity is simply a reflection of the fact that some of the suspended particles are still burning even though the original sample has long since "disappeared".

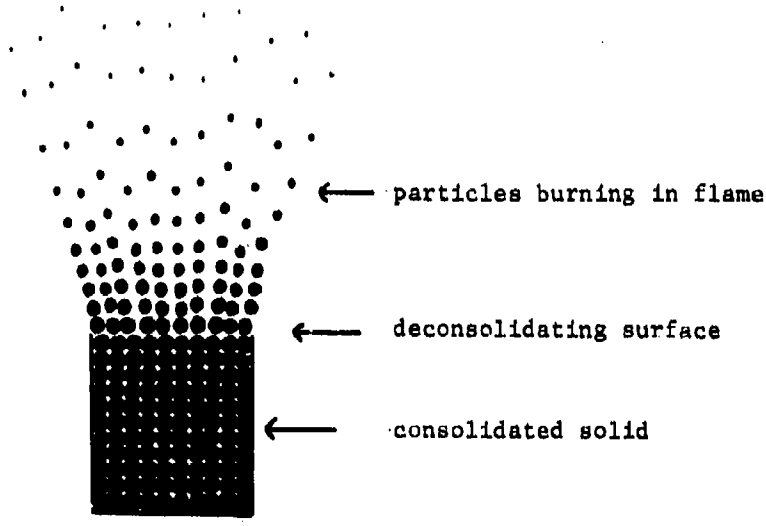



Fig. 2. Idealized representation of progressive deconsolidation mechanism

Theoretical Analysis

We have carried-out modeling calculations to predict the expected flame height (and particle burn-out time) for pressed HMX samples burning at constant pressure by a progressive deconsolidation mechanism. The following simplifying assumptions have been made: a) constant pressure through the flame, b) spherical particles of a single size, c) instantaneous ignition of all particles upon separation from the propellant, d) HMX flame dimensions small compared to the dimensions of the propellant flame, e) no velocity difference between particles and gas, and f) one-dimensional flow, i.e., sample burned in a tube to prevent radial expansion such as shown in Fig. 2.

The equations are as follows. The unit area mass flow rate M ($g\ cm^{-2}s^{-1}$) is given by Eq. (3) as

$$M = S\rho_s \tag{3}$$

where S is the regression rate and $\rho_s$ is the propellant density. By the law of mass conservation, M is constant at all distances from the

regressing surface (the velocity merely increases when the solid particles convert to gas as they move away from the propellant). For the burn rate R (cm/s) of HMX we take

$$R = 0.264\ P^{0.9} \tag{4}$$

(P in MPa) based on low pressure results in the literature (3). The density of the burned gas $\rho_g$ (g/cc) is determined from

$$1/\rho_g = \frac{I}{P} + b \tag{5}$$

where I is the mass impetus for HMX (1360 Joules/g) and b is the covolume (1.084 cc/g). This equation is based on the Abel-Noble equation-of-state: P(V-b)=RT. The instantaneous density $\rho$ in any region of the two-phase flow is given by

$$\rho = \frac{\rho_c \rho_g}{X\rho_g + (1-x)\rho_c} \tag{6}$$

where $\rho_c$ is the crystalline density of HMX (1.9 g/cc) and X is the mass fraction converted from solid to gas (calculated from the instantaneous sphere volume). From the law of mass conservation, the flow velocity U (cm/s) relative to the regressing surface at any point in the flame is given by

$$U = \frac{M}{\rho} \tag{7}$$

The principal differential equations to be solved involve the rate of change of particle diameter D with respect to time, and with respect to distance from the regressing propellant surface:

$$\frac{dD}{dt} = 2R \tag{8}$$

$$\frac{dD}{dx} = \frac{dD}{dt}\frac{dt}{dx} = \frac{2R}{U} \tag{9}$$

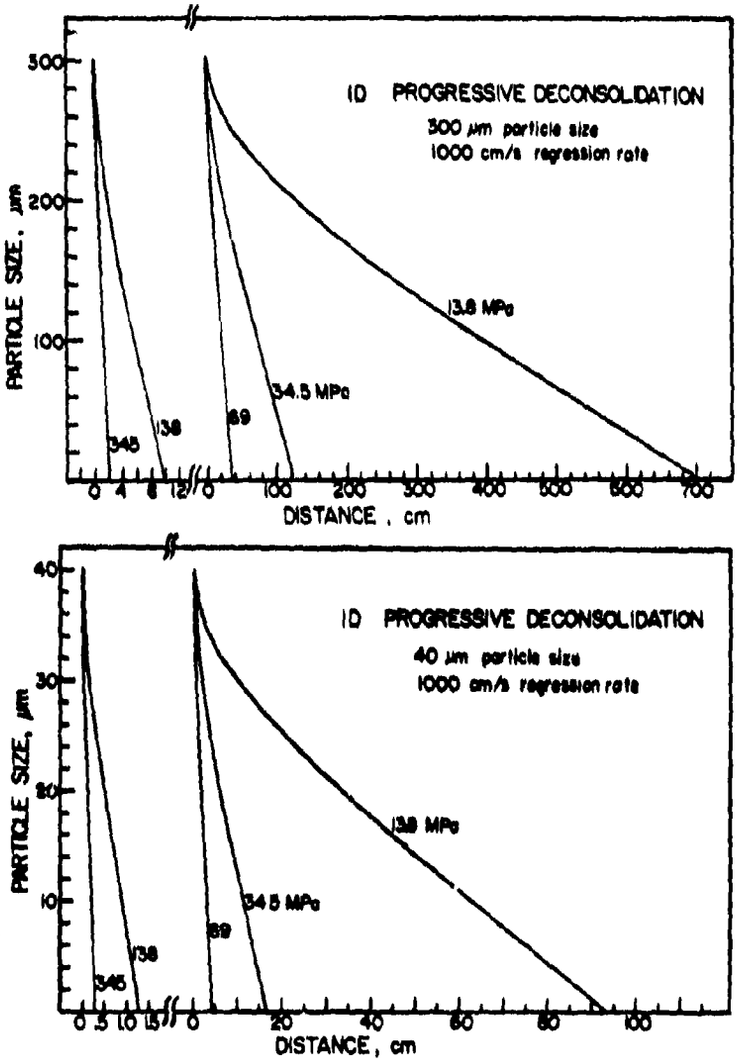
Equation (8) alone is sufficient to calculate the particle burn-out time, and hence the duration of luminescence after the propellant has regressed to its base. For constant pressure, Eq. (8) is a constant, and the calculation is trivial. For example, for 300 micron particles the calculated burn-out time is 5.4 ms at 13.8 MPa, and 2.3 ms at 34.5 MPa. These times are almost as large as the total regression time for the 3.2 cm propellant strands. Actually, luminescence was observed to last for longer than these calculated times, possibly due to agglomeration of some of the particles.

The solution of Eq. (9) in order to determine the burn-out distance (i.e., flame height) is non-trivial, and requires that the instantaneous two-phase density and flow velocity be recalculated using Eqs. (6) and (7), respectively, at each point in the integration. The results of computations at several pressures are shown in Figs. 3 and 4, for 300 μm and 40 μm particles, respectively. A 1000 cm/s regression rate was assumed in these computations. (Computed flame height scales directly with regression rate, as seen in Table I below).

It can be seen that the calculated flame heights are quite large -- a couple of orders of magnitude larger than those expected for "normal" burning where the solid-to-gas conversion occurs at the propellant surface. At the lower pressures (13.8 and 34.5 MPa) where the window chamber photographic experiments were carried out, the calculated flame heights range from 17 to 700 cm, depending on particle size and pressure. Thus, it is not surprising that the top of the flame could not be seen in the photographs. Both the theoretical burn-out times and flame heights for the progressive deconsolidation mechanism are therefore qualitatively consistent with experimental observations.

Note that the calculated flame heights for 300 μm particles are about 7.5 times larger than for 40 μm particles. The greater energy release close to the propellant surface for the finer particles should have led to quicker fuse wire melting in our strand burner experiments. This is the most likely explanation for the greater reproducibility for the class E samples shown in Fig. 1.

The results of this theoretical analysis are shown in Table 1, which gives the calculated flame heights, as well as the (gas) density and flow velocity (relative to the regressing surface) for each calculation.




Figs. 3 and 4. Calculated particle size vs distance from propellant for 1D progressive deconsolidation model. Top: 300 μm particle size; Bottom: 40 μm particle size

Table I. Theoretical Results for One-Dimensional Progressive Deconsolidation

| P(MPa) | S(cm/s) | d(cm) | ρ(g/cc)* | U(cm/s)* |
|---|---|---|---|---|
| | | 40 μm particles | | |
| 344.8 | 1000 | .275 | 0.198 | 8,821 |
| 137.9 | 1000 | 1.32 | 0.091 | 19,210 |
| 68.95 | 1000 | 4.64 | 0.048 | 36,520 |
| 34.50 | 1000 | 16.76 | 0.025 | 71,090 |
| 13.79 | 1000 | 93.91 | 0.010 | 175,000 |
| | | 300 μm particles | | |
| 344.8 | 1000 | 2.02 | 0.198 | 8,821 |
| 137.9 | 1000 | 9.87 | 0.091 | 19,210 |
| 68.95 | 1000 | 34.76 | 0.048 | 36,520 |
| 34.50 | 1000 | 125.7 | 0.025 | 71,090 |
| 13.79 | 1000 | 704.2 | 0.010 | 175,000 |
| 344.8 | 2000 | 4.05 | 0.198 | 17,640 |
| 137.9 | 2000 | 19.75 | 0.091 | 38,410 |

**at burn-out; d is the flame height.*

Since the sound speed in gaseous HMX combustion products is about 125,000 cm/s, it can be seen that the flow is supersonic at the lowest pressures. This result is independent of the assumed mechanism since $U = S\ (\rho_s/\rho_g)$ at flame burn-out.

Mechanism of the Transition to High-Speed Regression

Although there is considerable evidence for the progressive deconsolidation mechanism, the reason for the sudden transition to this mechanism at a certain pressure is not obvious. The process responsible for the transition from normal burning to progressive deconsolidation obviously depends on the driving force for the high pressure deconsolidation and this would have to be known before the process responsible for the transition could be understood.

Two possibilities exist to explain the driving force for the deconsolidation process. The first is that one of the normal subsurface processes associated with a burning propellant deconsolidates the charge ahead of the flame front. These include the "thermal wave" (probably too thin compared to particle dimensions to be effective), in-depth absorption of radiation, or a "precursor pressure wave". If one of these subsurface pressure or thermal effects ahead of the

flame front deconsolidate the sample one would expect the mechanical properties of the pressed charge to be important. Since the class C HMX samples are structurally stronger than the class E samples, and since strength increases with density as well, the relative transition pressures for the three samples would not be unexpected.

The second possibility is that burning in the pores very close to the propellant surface breaks the particles loose. This is quantitatively different from in-depth convective burning, and is not inconsistent with the high density of our samples. There is some evidence in the literature for such a mechanism for particle deconsolidation. For example, Andreev, et al. (7), and Belyaev, et al. (8), have both reported transitions from "normal" slow burning toward faster burning for a large variety of pressed propellant and explosive materials. These investigators showed that transition pressure correlates roughly with the permeability of the samples. (Permeability is a measure of the rate of gas flow through a porous sample for a certain pressure difference across the sample). These results suggest that the onset of hot gas penetration into the surface may be responsible for the process of deconsolidation in our samples.

The fact that a wide range of materials exhibit accelerated burning above a critical pressure, together with our results showing that for HMX the resulting super-fast regression is consistent only with a "progressive deconsolidation" mechanism, suggests that progressive deconsolidation may be a general phenomenon. We propose that most and perhaps all materials that have been described as burning by a "convective" mechanism actually burn by progressive deconsolidation, with convection playing a role at most only in the region just below the surface as the possible driving force for the deconsolidation.

## CONCLUSIONS, FUTURE WORK

We have shown that pressed high density binderless HMX undergoes a transition to extremely fast regression, which at high pressure attains values of 1000-6000 cm/s with a low pressure exponent. Such regression rates require a greatly increased burning surface area. The mechanism for this appears to be "progressive deconsolidation", rather than "convective" or in-depth burning. This mechanism may be characteristic of all porous propellants which exhibit unusually high apparent burning rates.

A number of additional experimental tests for this mechanism are possible. For example, in a "hybrid" strand burner-closed bomb, pressure would be expected to continue to rise for some time (depending

on particle size) after the sample has burned down, and holography should permit the observation of the particles in the flame.

Although pressed HMX appears to be an attractive, low cost propellant for certain gun applications requiring very high gas generation rates, further characterization is required before this material could be exploited. The mechanical strength of samples prepared in different ways will have to be determined. Since HMX is a "secondary" explosive, the effects of confinement will also have to be investigated. Finally, we will have to determine if samples burn the same way under closed bomb conditions as they do in a constant pressure strand burner. Based on our results to date, pressed HMX looks like a very promising material, and these additional studies would appear to be warranted.

## REFERENCES


1. Boggs, T. L., Price, C. F., Zurn, D. E., Derr, R. L., and Dibble, E. J., "Temperature Sensitivity of Deflagration Rates of Cyclotetramethylenetetranitramine (HMX)", 13th JANNAF Combustion Meeting, CPIA Publication 281, Dec 76, Vol I, pp. 45-56.

2. Isom, K. B., "A Window Bomb Study of HMX Combustion", 11th JANNAF Combustion Meeting, CPIA Publication 261, Dec 74, Vol I, pp. 243-246.

3. Boggs, T. L., Eisel, J. L., Price, C. F., and Zurn, D. E., "High Pressure Burning Rates of Cyclotetramethylenetetranitramine (HMX)", 15th JANNAF Combustion Meeting, CPIA Publication 297, Feb 79, Vol I, pp. 241-251.

4. Taylor, J. W., "A Melting Stage in the Burning of Solid Secondary Explosives", Combustion and Flame 6(2), 93-102 (1962).

5. Taylor, J. W., "The Burning of Secondary Explosive Powders by a Convective Mechanism", Trans. Faraday Soc., 58, 561 (1962).

6. Bobolev, V. K., Margolin, A. D., and Chuiko, S. V., "Stability of Normal Burning of Porous Systems at Constant Pressure", Combustion, Explosion and Shock Waves, 2(4), 15-20 (1966).

7. Andreev, K. K. and Chuiko, S. V., "Transition of the Burning of Explosives into an Explosion. 1. Burning of Powdered Explosives at Constant High Pressures", Russian J. Phy. Chem. 37(6), pp. 695-699 (1963).

8. Belyaev, A. F., Korotkov, A. I., and Sulimov, A. A., "Breakdown of Surface Burning of Gas-Permeable Porous Systems", Combustion, Explosion and Shock Waves, 2(3), 28-34 (1966).

# ENERGY ABSORPTION OF KEVLAR® FABRICS UNDER BALLISTIC IMPACT (U)

FRANK FIGUCIA
U. S. ARMY NATICK RESEARCH & DEVELOPMENT COMMAND
NATICK, MASSACHUSETTS 01760

Protection of the soldier from enemy threats is a multi-faceted research problem which requires study in many specialized areas of life support necessary for survival under combat conditions.

This paper considers one of these specialized areas, the prevention of debilitating wounds from fragmenting munitions. Specifically, it deals with protection by the use of flexible personnel armor made from fibrous materials.

Traditionally, the ballistic resistance of textile materials has been defined by laboratory measurement of ballistic limit ($V_{50}$), the velocity at which a material stops a simulated threat. This method has been widely accepted and it continues to serve the needs of the ballistic community; it is, however, very expensive and time consuming.

A new methodology has been developed at the U.S. Army Natick Research & Development Command (NARADCOM) which greatly reduces the cost and time necessary to develop equally reliable data. It generates a Ballistic Performance Indicator (B.P.I.) which can be used to predict the $V_{50}$ ballistic limit, or to measure the relative ballistic usefulness of candidate materials.

This paper describes the new test methodology, compares experimental B.P.I. with actual $V_{50}$ for five Kevlar materials, and suggests, through the use of B.P.I., fabric constructions for improved protection against fragmentation threats.

---

Kevlar® is the commercial designation for a polyaramid fiber manufactured by E. I. duPont deNemours & Co., Inc.



PRECEDING PAGE BLANK-NOT FILMED

FIGUCIA

## DEVELOPMENT OF TEST METHODOLOGY

The B.P.I. developed herein is based on an analysis of previously compiled data [1] for Kevlar materials subjected to ballistic impact. This data summarized $V_{50}$ ballistic limit velocities over a wide range of areal densities, as shown in Figure 1.

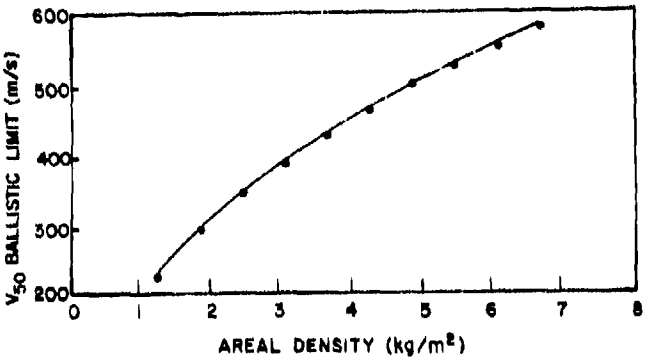



Figure 1. $V_{50}$ Ballistic Limit vs. Areal Density For Kevlar Fabrics.

Conversion of the ordinate values of velocity to kinetic energy, by K.E. = $mV^2/2$, indicated that the energy absorbed at ballistic limit velocity is linear over the range of target densities examined, as seen in Figure 2.

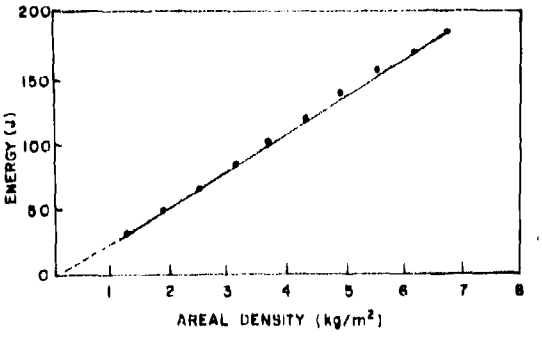



Figure 2. Energy Absorption vs. Areal Density For Kevlar Fabrics.

FIGUCIA

It was recognized that this trend could be effectively utilized, provided that this relationship was valid in the low energy, low areal density region, as extrapolated in Figure 2. Research scale equipment could then be used to generate energy absorption data with which to characterize behavior at higher levels.

An existing test facility, used for ballistic testing of yarns, was adapted for this purpose. The facility uses compressed helium gas to propel the standard 1.1 gram (17 grain) fragment simulating projectile. Electronic lumiline screens are placed before and after the target to provide time flight data for missile velocity calculations. These velocity data are then used, with appropriate corrections for aerodynamic drag between the screens and the target to calculate the energy absorbed by the target.

The target specimens are held between heavy aluminum plates in a specially designed fabric clamping device [2], shown in Figure 3. The device may be moved vertically and rotated, so that the five circular target openings cut from the plates are sequentially introduced into the path of the missile. Boundary conditions for all five impact points are equalized by this design. Therefore, anomolous variations in energy absorption, noted for other clamp designs investigated, are eliminated.

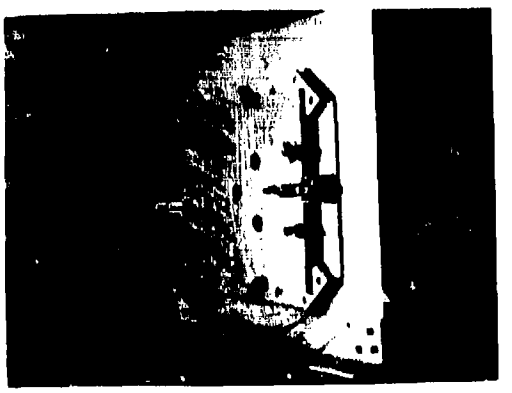


Figure 3. Rotatable Clamp For Testing Ballistic Fabric.

The following criteria were established to standardize test procedures.

(1) A complete screening consists of test firings at three striking velocities: 213, 274, and 366 m/s.

(2) At each velocity, tests are conducted starting with one layer of target material. Areal density is then varied by increasing the number of layers. The test sequence is continued until the target resistance approaches 50-60% of the available missile energy. Above this level, variability of individual readings increases significantly.

(3) Five replicate firings are used to generate one data point for a given test condition, i.e., number of layers/striking velocity.

## RESULTS AND DISCUSSION

The new test procedure was first applied to a 170 $g/m^2$ Kevlar fabric made in a satin weave construction from 44 tex yarn. Tests were performed at the three velocities prescribed. The energy absorption for the various numbers of layers examined at each striking velocity are shown in Table 1.

Table 1. Laboratory Data For 170 $g/m^2$ Satin Weave Kevlar Fabric.

| Striking Velocity (m/s) | Number of Layers | Areal Density ($kg/m^2$) | Energy Absorbed (J) |
|---|---|---|---|
| 213 | 1 | 0.17 | 5.56 |
| 213 | 2 | 0.34 | 12.57 |
| 274 | 1 | 0.17 | 4.87 |
| 274 | 2 | 0.34 | 11.00 |
| 274 | 3 | 0.51 | 18.67 |
| 274 | 4 | 0.68 | 27.04 |
| 366 | 1 | 0.17 | 6.90 |
| 366 | 2 | 0.34 | 12.14 |
| 366 | 3 | 0.51 | 18.54 |
| 366 | 4 | 0.68 | 24.12 |
| 366 | 5 | 0.85 | 28.42 |
| 366 | 6 | 1.02 | 35.20 |

Reproducibility of individual values was excellent at low and medium energy absorption levels (coefficient of variation approximately equal to 3%). However, at target energy absorptions of 50% or more of total available missile energy, variability was observed to increase. The test sequence was therefore terminated when this level was approached.

The data from Table 1 were analyzed graphically and statistically to test for linearity. It is seen in Figure 4 that energy absorption and areal density are directly proportional over this low areal density range. It appeared, therefore, that the assumption of linearity in the extrapolated portion of Figure 2 was valid, and further investigation was warranted.

FIGUCIA

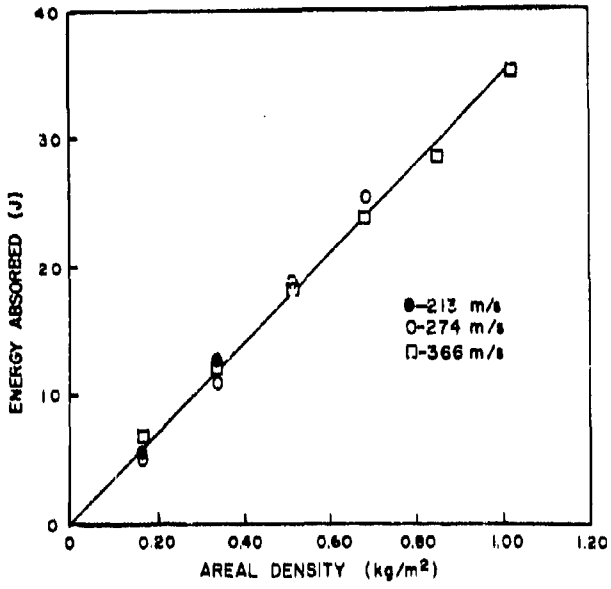



Figure 4. Energy Absorption of 170 $g/m^2$ Satin Weave Kevlar Fabric by Laboratory Screening Method.

A least squares fit regression of the data yielded a slope of 35.1, with a correlation coefficient of 0.99. It is this numerical value of the slope which is defined as the Ballistic Performance Indicator. It represents the energy absorbed per unit increase in areal density.

Similar data were generated for four additional Kevlar materials. Linearity was obtained in all cases, with a high degree of statistical confidence. The resultant B.P.I.'s obtained are recorded in Table 2.

FIGUCIA

Table 2. Ballistic Performance Indicators Obtained by Laboratory Analysis for Five Kevlar Fabrics.

| Identification Number | Fabric | | | B.P.I. |
|---|---|---|---|---|
| | Weight($g/m^2$) | Weave | Yarn Tex | ($J/kg/m^2$) |
| 1 | 170 | Satin | 44 | 35.1 |
| 2 | 294 | Plain | 111 | 29.7 |
| 3 | 278 | Satin | 111 | 32.9 |
| 4 | 464 | Basket | 111 | 24.9 |
| 5 | 464 | Basket | 167 | 22.6 |

Use of the B.P.I. to predict performance at higher levels is illustrated in Figure 5 for fabrics 1 and 4. The laboratory data are extrapolated to areal density levels at which conventional $V_{50}$ tests were performed on the same materials. The projected energy absorptions compare closely to those calculated from actual $V_{50}$ values.

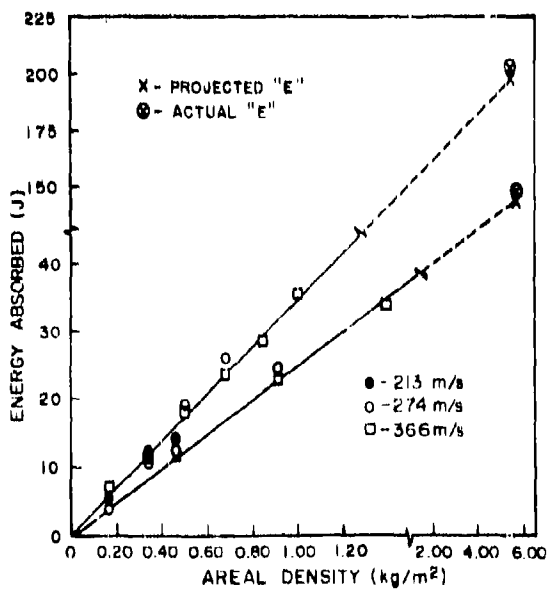



Figure 5. Projected and Actual Energy Absorption at Full Areal Density.

Table 3 summarizes projected and actual results for all five materials in terms of $V_{50}$ ballistic limit velocities. A high degree of accuracy is indicated, with no actual value varying from the predicted by more than 2.1%.

Table 3. $V_{50}$ Ballistic Limits of Five Kevlar Fabrics Obtained by Laboratory Prediction and Actual Test.

| Identification Number | B.P.I. ($J/kg/m^2$) | A.D. ($kg/m^2$) | $V_{50}$ Ballistic Limit (m/s) Predicted | Actual | % Difference |
|---|---|---|---|---|---|
| 1 | 35.1 | 5.68 | 602 | 610 | 1.4 |
| 2 | 29.7 | 5.83 | 561 | 551 | -1.8 |
| 3 | 32.9 | 5.72 | 585 | 573 | -2.1 |
| 4 | 24.9 | 5.75 | 510 | 518 | 1.7 |
| 5 | 22.6 | 6.97 | 535 | 543 | 1.5 |

Use of the B.P.I. methodology to predict $V_{50}$ provides significant advantages in time, material usage, and cost, over conventional methods. Table 4 compares expenditures for an in-house B.P.I. and a single-panel $V_{50}$ test performed in the customary manner, by outside contract.

Table 4. Comparison of In-House and Customary Methods of Obtaining $V_{50}$ Data.

| | Outside Contract (Single-Panel $V_{50}$) | In-House (Complete B.P.I.) |
|---|---|---|
| Elapsed Time (Days) | 14 | 1 |
| Material Required ($m^2$) | 2 | 1 |
| Test Cost ($) | 250 | 150 |

Estimates for performance by outside contract are very conservative in both time and money. The time is often increased due to higher priorities of the contracted agency. Also, it is not uncommon for material usage to be increased by submission of more than one test panel to validate results. This adds not only to material costs, which are substantial for Kevlar, but also to testing cost as well. Finally, charges for performance of a single $V_{50}$ test vary considerably, depending upon the particular agency doing the work. The test cost estimate in Table 4 is the lowest currently charged.

FIGUCIA

The B.P.I. is also useful for comparing ballistic performance to fabric constructional parameters. An example of this is demonstrated in the relationship between B.P.I. and fabric weight for the five Kevlar materials.

B.P.I. vs. Fabric Weight

It is shown in Figure 6, that the ballistic resistance, as measured by B.P.I., falls off as the nominal fabric weight increases, showing that lightweight fabrics are the most efficient on an energy absorption to weight basis. This information provides a practical guideline for use in armor design.

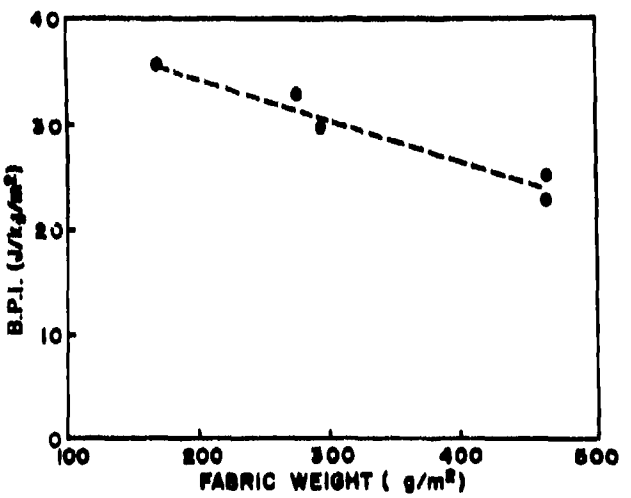



Figure 6. B.P.I. vs. Nominal Fabric Weight For Five Kevlar Materials.

Since layered armor systems are ordinarily restricted by weight limitations, this data would assist in the selection of the most efficient materials with which to achieve the design weight of an item; namely, use of more layers of light material as opposed to fewer layers of heavier material. Naturally, other considerations such as cost and ease of fabrication also influence the selection.

The information in Figure 6 has practical value due to its expression in terms of fabric weight, the most commonly used and easily understood constructional parameter. However, the observed trend is reflective of a more purely derived relationship between B.P.I. and the less popular parameter, fabric cover.

B.P.I. vs. Fabric Cover

Equation (1) is a convenient expression for cloth cover factor, when identical yarns are used in both warp and filling directions.

$$K_c = [(n_w + n_f)\, d - n_w n_f d^2]100 \qquad (1)$$

Where: $K_c$ = Cloth cover factor (%)

$n_w$ = Number of warp yarns per unit length

$n_f$ = Number of filling yarns per unit length

$d$ = Common yarn diameter

It gives the percentage of surface covered if viewing from a point normal to the fabric. The areas of double coverage which occur at each yarn crossover are eliminated by subtraction of the second term.

Proper analysis of fabric penetration by a ballistic missile should consider the resistance offered not only by the surface yarns, but also by these backup yarn areas at the crossovers. Therefore, a bulk cover factor, $K_B$, which includes the cover at the crossovers, will be used to represent the actual cover effective against missile penetration. It is defined by:

$$K_B = (n_w + n_f)\, d \qquad (2)$$

A $K_B$ of 1.0 represents a fabric made up of sufficient yarn to cover the entire surface, if placed side by side with no interlacings. It can be shown that a $K_B$ = 1.0, or 100% $K_B$ cover, is equivalent to $K_c$ = 75%.

Additional cover above this level might be expected to contribute more weight than ballistic resistance and reduce B.P.I. This is examined in Figure 7 for the five materials having $K_B$'s approximately between 1.0 and 2.0. A sharp decline in B.P.I. is observed with increasing $K_B$ over this range.

FIGUCIA

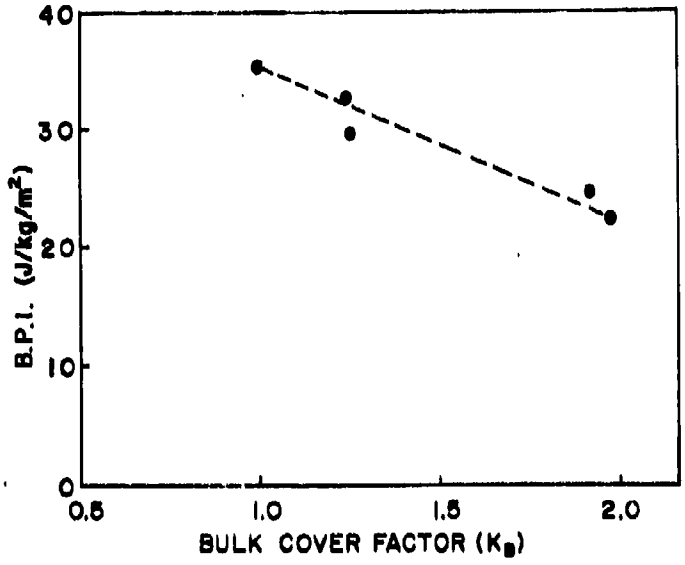



Figure 7. Effect of Bulk Cover Factor on Ballistic Performance For Five Kevlar Fabrics.

As was previously stated, the reduction in B.P.I. with increasing weight (Figure 6) reflects the influence of fabric cover illustrated in Figure 7. The similarities are due to a direct relationship between fabric weight and $K_B$.

It is suggested that, within practical limitations of weaveability and use, fabrics designed with a $K_B$ approximately equal to 1.0 would provide ballistically effective alternatives to those currently in use. Fabrics made from the four commercially available Kevlar yarns would have the following weights when constructed to a $K_B = 1.0$.

| Kevlar Yarn Tex | Fabric Weight ($g/m^2$) $K_B = 1.0$ | Current Use |
|---|---|---|
| 22 | 133 | - |
| 44 | 173 | - |
| 111 | 219 | 271 |
| 167 | 227 | 475 |

A more comprehensive analysis of fabric performance is obtained by the inclusion of data from eight supplementary fabrics which were not tested for B.P.I., but for which $V_{50}$/A.D. information was available. A B.P.I. was estimated for each supplementary material - a reversal of the application for predicting $V_{50}$. The B.P.I.'s for all materials are shown in Figure 8 as a function of $K_B$. In this plot, each material is identified by weave form.

Not only is the expected downward trend in performance again observed, but a clear indication of the effect of fabric weave unfolds, with the satin weave form showing superiority at all $K_B$ levels examined. This relationship is described by:

$$B.P.I. = 41.9 - 6.9\ K_B \qquad (3)$$

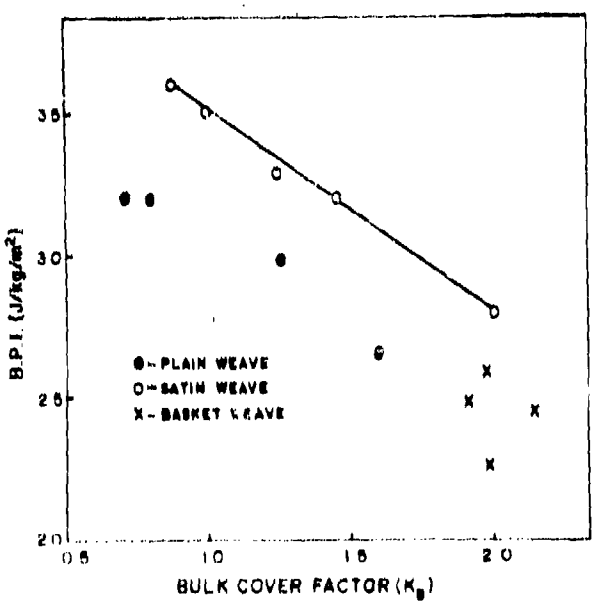



Figure 8. Effect of Weave Form and Bulk Cover Factor on Ballistic Performance Indicator.

Separate trend lines for the basket and plain weave forms are not distinguishable, but it appears that a relationship of the form given in equation (3) would apply as an estimate of the combined behavior of the two weave types.

It is noted that the material showing the greatest B.P.I. has a $K_B < 1.0$. The possibility may exist of advantages in cumulative cover for test panels made up of light weight materials (3). Most fabrics having $K_B$ less than one are impractical for use in body armor because of fabrication and other problems. Consequently, the value of further investigations is questionable. Based on knowledge obtained to date, the application of equation (3) should be limited to $K_B \geq 1.0$.

Weave Effect

The superior performance of the satin weave fabrics is attributed to the lateral mobility potential inherent in the satin construction. Observation of the representative weave cross-sections in Figure 9 shows long lengths of yarn which "float" across the fabric between interlacings for the 8-harness satin form. It is speculated that these provide greater yarn mobility and transverse deformation than the more tightly constructed plain and basket weaves, which results in higher energy absorption.

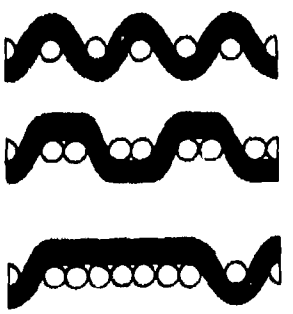



Figure 9. Generalized Cross-Sections of Three Weave Forms.

The suggestions made earlier for improved ballistic protection were based on idealized cover factor levels without regard for weave. Based on the higher B.P.I. values of the satin weaves, it is believed that the proposed fabric weights, woven into an 8-harness satin form, would offer additional advantages leading toward optimized ballistic protection.

## CONCLUSIONS

The kinetic energy absorbed at $V_{50}$ velocity for a large number of Kevlar test panels increases in direct proportion to the areal density of the panels. Based on this observed linearity, a test methodology was developed which characterizes the energy absorption at very low areal densities, through the use of research scale laboratory equipment. This relationship, which corresponds to energy absorption per unit areal density, is defined as the Ballistic Performance Indicator (B.P.I.). It can be extrapolated to predict $V_{50}$ at practical areal densities, with high accuracy.

Use of the B.P.I. methodology to predict $V_{50}$ provides distinct savings in time, materials, and money over conventional methods.

The use of B.P.I. to assess the effect of major fabric variables on their performance has been demonstrated. Based on relationships with parameters such as cover factor and weave, fabric forms offering improved ballistic performance have been projected. Some compromises may be necessary when practical factors such as cost, weaveability and structural integrity are considered.

## REFERENCES

(1) Mascianica, F. S., "Ballistic Technology of Lightweight Armor", AMMRC-TR 79-10, U.S. Army Materials and Mechanics Research Center (February 1979).

(2) Figucia, F., "Rotatable Clamp for Ballistic Testing of Fabric", TR-80-008, U.S. Army Natick Research and Development Command (September 1979).

(3) Figucia, F., "The Effect of Kevlar Fabric Construction on Ballistic Resistance", TR 75-103CE, U.S. Army Natick Research and Development Command (April 1975).

# THE INITIATION OF EXPLOSIVE CHARGES BY RAPID SHEAR (U)

ROBERT B. FREY, DR.
BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MD 21005

## I. Introduction.

Friction and/or shear have been suggested by many people as possible sources of ignition in solid explosives. Although these two mechanisms are usually considered to be distinct, on a microscopic scale they are essentially the same. On a macroscopic scale, friction involves sliding interfaces and is described in terms of the coefficient of friction and the stress normal to the interface. The rate of heat generation at the surface is the product of the sliding velocity times the coefficient of friction times the normal stress. On a microscopic level, friction is caused by surface irregularities (asperities) which deform as the surfaces slide. The frictional heating is caused by viscoplastic work on the asperities. In this paper, I will consider the conditions required for ignition as the result of shear deformation and viscoplastic heating.

A cursory examination of this question leads one to the conclusion that the deformation must be localized in some way if ignition is to occur. The plastic work per unit mass resulting from the deformation of a material is given by the following relation [1]:

$$W_p = \frac{1}{\rho} \sum_i \int_o^{e_i^f} S_i \, de_i \quad , \tag{1}$$

where $S_i$ is the principal deviatoric stress, $e_i$ is the principal



PRECEDING PAGE BLANK-NOT FILMED

deviatoric strain, $e_i^f$ is the final value of the deviatoric strain and $\rho$ is the density. Consider the homogeneous deformation of a cylinder of explosive so that the length decreases and the radius increases. Assume that the Tresca yield condition holds. Then $e_r = e_\theta = \frac{-1}{2} e_z$, $S_r = S_\theta = \frac{-1}{2} S_z$, and $S_r - S_\theta = Y$, where Y is the yield strength of the explosive, and the usual notation of cylindrical coordinates has been used. The change in temperature of the explosive is given by

$$\Delta T = -\int \frac{Y}{\rho C} de_z = -\int \frac{Y}{\rho C} \frac{dz}{z} = \frac{Y}{\rho C} \ln \frac{z_o}{z_f}$$

where C is the heat capacity, $Z_o$ is the initial length, and $z_f$ is the final length. For composition B, Y is about 0.07 GPa and C is about 1.25 joules/$^oC$. To obtain a 100$^oC$ change in temperature requires $z_o/z$ obtain a value of approximately 28. Clearly homogeneous deformation can result in initiation only in extreme circumstances.

There are several circumstances which may lead to very high shears in localized regions. Some of them are the following:

1. Sliding friction. As explained above, on a microscopic scale sliding friction involves shear deformation, and very high strains may occur in narrow layers. In particular, friction between a rotating shell and the explosive fill has been suggested as a cause of premature ignition in artillery shells [2]. It has also been suggested as a mechanism for the propagation of reaction during nondetonative explosions in confined explosive charges [3].

2. The collapse of spherical cavities. During ductile cavity collapse, a thin layer of material at the boundary of the cavity sees very high strains. This situation has been considered by Carroll and Holt [4]. Their analysis indicates that, in the absence of melting, extremely high temperatures can be obtained in small volumes around a collapsed cavity. This has been suggested as an ignition mechanism during shock wave initiation of explosives.

3. Adiabatic shear bands. Adiabatic shear bands in metals have been discussed by Recht [5]. A shear band is likely to form in a deforming material when thermal softening exceeds work hardening. In this case, deformation in any plane causes the material in that plane to weaken and tends to concentrate further deformation in the same plane. In metals, thin bands of material can be heated to the

melting point. Winter and Field [6] have applied Recht's analysis to explosives and concluded that explosives should be highly susceptible to shear banding. Afanasev [7] and co-workers have observed them in nonexplosive organics.

4. Extrusion of explosive into cracks. If a crack develops in the metal confinement around an explosive while the explosive is under pressure, ignition could result when the explosive extrudes into the crack. This has been suggested as a possible ignition mechanism when confined explosives are impacted by projectiles at velocities too low to cause shock wave initiation [3].

In all of these circumstances, melting tends to limit the temperature which can be achieved. Once the melting point has been reached, any further heating must be strictly by viscous processes. The melting point increases with pressure, and melting is a more important consideration at low pressure than at high pressure. From data in the International Critical Tables, the rate of increase of melting point with pressure for many organic materials is 20 to 30° per kilobar. For instance, for TNT the melting point increases about 25°/kilobar at atmospheric pressure. The rate of increase decreases as pressure increases. To obtain a crude estimate of the effect of melting on hot spot ignition, I assumed that the melting point of several explosives increased at about 20°/kilobar and computed the time to explosion at the melting point as a function of pressure. Kinetic constants were taken from Rogers [8]. The results are shown in Figure 1. For reactions which occur in 1 microsecond or less, which is typical of the shock initiation regime, melting is important in preventing initiation at pressures below 1.0 GPa in PETN, RDX, or HMX and for pressures below 3.0 GPa in TNT. For reactions which occur in a millisecond time frame, which might be typical of artillery setback conditions, melting is significant for pressures below 0.4 GPa in RDX and 1.5 GPa in TNT. In this paper I will investigate the conditions under which viscoplastic heating can lead to ignition when melting is considered.

## II. Ductile Cavity Collapse.

Carroll and Holt treated spherical cavity collapse in an elastic-plastic fashion [4] and in an elastic-viscoplastic fashion [9]. Application of their equations shows that, in the absence of melting, extremely high temperatures are produced in a thin layer around the collapsing cavity. Figure 2 shows the temperature as a function of applied pressure for a Comp-B charge where the initial porosity (actual volume/fully compacted volume) is 1.03. Extremely high temperatures are predicted in a very thin layer around the collapsing cavity. At a

pressure of 2.0 GPa, the surface of the cavity reaches a temperature of about 600°C. However, only a very thin layer is significantly heated. Figure 2 also shows the temperature of the isotherms enclosing the hottest 0.01% and the hottest 1% of the material. These isotherms are much cooler.

The analysis of Carroll and Holt does not include thermal softening, melting, or heat conduction. Since the heated region is close to the surface of the cavity, where the pressure is low, one can suspect that the inclusion of these factors might alter the results significantly. We have modified Carroll and Holt's formulation slightly to include these effects and also added a viscosity which varies with temperature and pressure. For simplicity, we use a rigid-plastic model (i.e., we assume that the elastic strains are negligible). The principal assumptions of the Carroll and Holt model are the following:

1. The flow in the vicinity of the cavity is spherically symmetric.

2. The material is incompressible.

3. The following constitutive relation applies:

$$\dot{e}_{ij} = \frac{1}{2G} S_{ij} + \frac{1}{2\mu} (1 - Y/\sqrt{3I}) S_{ij} ,$$

where $\dot{e}_{ij}$ is a component of the deviator strain rate tensor, $S_{ij}$ is a component of the deviator stress tensor, G is the shear modulus, $\mu$ is the viscosity, Y is the yield strength, and I is the second invariant of the deviator stress. For the spherically symmetric, rigid-plastic case considered here, this reduces to

$$S_r = 2\mu\dot{e}_r + \frac{2}{3} Y$$

$$S_\theta = 2\mu\dot{e}_\theta - \frac{1}{3} Y ,$$

where $S_r$, $S_\theta$, $\dot{e}_r$, $\dot{e}_\theta$ refer to the principle deviator stresses and principle deviator strain rates, and the usual notation for spherical coordinates has been used.

Carroll and Holt considered a hollow sphere with outer radius b, inner radius a, and porosity $\alpha$, defined as the ratio of the

actual volume to the fully compacted volume, where $\alpha = b^3/(b^3-a^3)$. The equation of motion for this system is

$$\frac{\partial \sigma_r}{\partial r} + \frac{2}{r}(\sigma_r - \sigma_\theta) = \rho \ddot{r},$$

where $\sigma_r$ and $\sigma_\theta$ are the principle stresses, r is the radius, and dots imply differentiation with respect to time. Following Carroll and Holt, we note that

$$\sigma_r - \sigma_\theta = S_r - S_\theta = \mu \dot{e}_r + Y,$$

and apply the boundary conditions,

$$\sigma_r = 0 \text{ at } r = a$$

$$\sigma_r = -P \text{ at } r = b,$$

where P is the applied pressure.

We also note that

$$\dot{e}_r = + \frac{2}{3} \frac{\dot{B}}{r^3},$$

where $B(t) = a_o^3 - a^3 = b_o^3 - b^3 = a_o^3 (\alpha_o - \alpha)/(\alpha_o - 1)$

and $a_o$, $b_o$, and $\alpha_o$ are the initial values of a, b, and $\alpha$. The equation of motion can then be integrated with respect to the radius and the result expressed as the following differential equation for $\alpha$:

$$P - \int_a^b \frac{2Y}{r}\,dr - \int_a^b \frac{4\mu}{3} \frac{\dot{B}}{r^3} dr = F(\ddot{\alpha},\dot{\alpha},\alpha)$$

$$F(\ddot{\alpha},\dot{\alpha},\alpha) = \frac{\rho a_o^2}{3(\alpha_o - 1)^{\frac{2}{3}}} \left[ \frac{\dot{\alpha}^2}{6} \left[ (\alpha-1)^{\frac{-4}{3}} - \alpha^{\frac{-4}{3}} \right] - \ddot{\alpha} \left[ (\alpha-1)^{\frac{-1}{3}} - \alpha^{\frac{-1}{3}} \right] \right]. \quad (2)$$

The rate of plastic work at any point in the flow may be computed from equation (1), giving the following result:

$$\dot{W}_p(r) = \frac{2}{3} Y \frac{\dot{B}}{r^3}, \quad (3)$$

where $\dot{W}_p$ is the rate of plastic work per unit volume. We have integrated equation (2) numerically. Simultaneously, we integrate the heat equation for the material around the cavity,

$$\frac{\partial T}{\partial t} = \frac{K}{\rho C}\left[\frac{\partial^2 T}{\partial r^2} + \frac{2}{r}\frac{\partial T}{\partial r}\right] + \frac{\dot{W}_p}{\rho C} \qquad (4)$$

where T is temperature, t is time, K is heat conductivity, $\rho$ is density, and C is heat capacity. This equation was also integrated using finite differences and a computational mesh which concentrated most of the points in the region close to the surface of the cavity. At each time step, equation (2) was integrated to obtain $\alpha$, $\dot{\alpha}$, $\ddot{\alpha}$, a, and b. Then $\dot{W}_p$ could be determined and equation (4) integrated to determine T(r) at the next time step.

Good data on the temperature dependence of the yield strength and on the temperature and pressure dependence of the viscosity of common explosives are not available. We assumed that the melting point increased with pressure at the rate of 200°/GPa and that the yield strength was zero above the melting point and decreased to zero in a linear fashion over a range of 30°C below the melting point. We assumed that the viscosity varied in the following way with temperature and pressure:

$$\mu = \mu_o \exp\left(\frac{P}{P_o}\right) \exp\left(\frac{E}{T} - \frac{E}{T_o}\right), \qquad (5)$$

where $\mu_o$ is the viscosity at temperature $T_o$, E is an experimentally determined activation energy for viscous flow, and $P_o$ is an experimentally determined constant. Pressure and temperature relations of this type are discussed by Frenkel [10] and have some theoretical support. The pressure dependence shown in equation (5) agrees well with data from the International Critical Tables. Some data is available for TNT, and we will use this material in all of the calculations reported here. We determined an approximate value for E, 3,880° K, on the basis of two data points [11]. We estimated $P_o$ to be 0.165 GPa, a value which is typical of organic liquids and fits the experimental data reported in the International Critical Tables for nitrobenzene. Other required parameters are shown in Table I.

TABLE I - MATERIAL PROPERTIES USED IN CALCULATIONS

| Parameter | Value |
|---|---|
| Y | 0.07 GPa |
| $\nu_o$ | $1.39 \times 10^{-2}$ Kg/(msec) at $85^oC$[35] |
| C | $1.29 \times 10^5$ joule/(Kg$^oC$)[35] |
| K | 0.262 joule/(m$^oC$ sec)[35] |
| $\rho$ | $1.64 \times 10^3$ Kg/m$^3$ [35] |
| E | 3,880$^oK$ |
| $P_o$ | 0.165 GPa |

Figure 3 shows the results of the calculation for a case where the pressure increased from zero to 2.0 GPa over a period of 20 nanoseconds (this simulates a shock wave in a porous medium). The peak temperature never exceeds the normal (one atmosphere) melting point. In this situation, the deformation is concentrated in the region close to cavity where the pressure and melting point are low. Viscosity is not sufficient to elevate the temperature significantly above the melting point. Therefore, ignition by this mechanism is unlikely in TNT. Although we have not performed calculations on other explosives, we can infer that they would not be heated much above the normal melting point either. Since RDX has a rather long thermal explosion time at the melting point, it is probably not susceptible to ignition by this mechanism in a millisecond time scale.

III. Friction and/or adiabatic shear bands.

We have already commented on the similarity of friction and adiabatic shear bands at the microscopic level. In each case, large deformations are concentrated in a narrow region. In a shear band, the concentration occurs because thermal softening exceeds work hardening, as explained above. At a sliding interace, the concentration occurs because the interface region is weaker than the rest of the material. We have calculated the time evolution of temperature in a shear band or at a sliding interface when melting occurs. Our intention was to perform the simplest possible calculation while retaining the essential features of the real world.

In the model, we impose a velocity gradient across an arbitrary thickness of explosive. The magnitude of this thickness does not influence the results of the calculation. We arbitrarily induce a shear band by specifying the initial temperature in a thin layer to be high enough so that thermal softening occurs. The time evolution of the system is followed by making use of the following assumptions:

1. The shear stress is constant in all planes parallel to the shear band at any instant; i.e.,

$$\mu(x)\frac{dv}{dx} + \sigma(x) = f(t) \tag{6}$$

where $\mu$ is viscosity, x is distance perpendicular to the shear band (perpendicular to the direction of flow), v is velocity, $\sigma$ is shear strength, and f is the shear stress applied at the shear band. f is independent of x but may vary with time. This assumption is a quasi-steady state assumption. It requires that the flow at any instant must have the form it would have at steady state given the temperature distribution which exists at that time, but it permits the flow to change as the temperature distribution changes.

2. The total shear velocity (the integral of dv/dx with respect to x) is independent of time. This condition and the previous one determine f via the following equation:

$$\int \frac{dv}{dx}\,dx = \int \frac{f-\sigma(x)}{\mu(x)}\,dx = v_t\ , \tag{7}$$

where $v_t$ is the total shear velocity. At any time if $\sigma(x)$ and $\mu(x)$ are known, this equation can be used to determine f, and equation (6) can be used to compute dv/dx as a function of x. As the shear band develops, the velocity gradient will become concentrated in a narrow region and dv/dx will be zero over most of the space. In this situation, the integrals in equation (7) must be evaluated only over the region where dv/dx is positive. This is accomplished by choosing the limits on the second integral, which is evaluated numerically, so that the shear stress f is minimized.

3. The solid is rigid-perfectly plastic with a shear strength which decreases to zero at the melting point. As in the previous discussion, the shear strength is assumed to decrease linearly to zero in a 30°C range of temperature below the melting point.

4. The viscosity varies with temperature and pressure in the manner described by equation (5).

5. The rate of viscoplastic work per unit volume is given by the following expression:

$$\dot{W} = \frac{1}{2}\mu\left(\frac{dv}{dx}\right)^2 + \sigma\frac{dv}{dx} \quad .$$

6. The heat generated per unit mass by chemical decomposition can be computed on the basis of simple Arrhenius kinetics, as follows:

$$\dot{Q} = QA \exp\left(-E_a/T\right)$$

where Q is the heat of reaction per unit mass, A is the frequency factor, and $E_a$ is the activation energy.

7. One dimensional heat equation applies; i.e.,

$$\frac{\partial T}{\partial t} = \frac{K}{\rho C}\frac{\partial^2 T}{\partial x^2} + \frac{\dot{Q}}{C} + \frac{\dot{W}_p}{\rho C} \quad . \tag{8}$$

Equation (7) was integrated numerically using finite differences. At each time-step, $\sigma(x)$ and $\mu(x)$ were determined from the known temperature distribution and dv/dx was determined from equations (6) and (7). Melting is handled in one of two ways. In one procedure, all of the material is considered to be a single phase, but the shear strength goes to zero when the temperature exceeds the melting point. This procedure neglects the heat of fusion. In the other procedure, two phases and a phase boundary are included in the calculation. The velocity of the phase boundary (the rate of growth of the melt layer) is determined from an energy balance at the boundary:

$$V_b \rho Q_f = K_e \frac{\partial T_e}{\partial x} - K_s \frac{\partial T_s}{\partial x} \quad ,$$

where $V_b$ is the velocity of the boundary, $Q_f$ is the heat of fusion, and subscripts e and s refer to the liquid and solid phases. When this procedure is used, a dynamic mesh is used in both phases to avoid the necessity of rezoning as the boundary moves. This is accomplished by introducing a variable, y, defined as follows:

$$y = \frac{x_{bo}}{x_b} x \quad ,$$

where $x_{bo}$ is the initial position of the boundary and $x_b$ is the position of the boundary at any time. x and its derivatives are expressed in terms of Y and substituted into equation (8). The resulting equation

is solved by finite differences. When expressed in terms of y, the computational grid automatically expands or contracts to adjust to the moving boundary. No significant differences were noted in the results of the two types of calculation. Surprisingly, the second procedure gave slightly higher temperatures, but this may be due to numerical error. All of the results reported here were accomplished with the second procedure.

Calculations were performed using material properties for TNT as listed in Table I. In addition, the following values for Q, A, and E were taken from Rogers [8]:

$$Q = 300 \text{ Cal}/(^{\circ}\text{C g})$$

$$A = 2.51 \times 10^{11} \text{ sec}^{-1}$$

$$E_a = 34.4 \text{ KCal/mole}$$

Since the model is only an approximation of reality and since many parameters are not known precisely, the results should be considered in a qualitative rather than a quantitative sense. Figure 4 shows how the computed temperature at the center of the shear band varies with time for a case where the shear velocity was 0.2 km/sec and the pressure was 1.03 GPa. The temperature asymptotically approaches a maximum value. The time required for significant heating to occur is very short. The initial width of the heated region, which triggers shear band formation, affects the time history at the shear band but does not affect the maximum temperature. Figure 5 shows the time history for several calculations with the same pressure and shear velocity but different initial widths. There is an optimum initial width which gives the most rapid heating rate. With smaller initial widths, the temperature in the shear band may decrease for a while before it begins to increase. For larger initial widths, the temperature increases more slowly. In the absence of reaction, the maximum attainable temperature is a strong function of pressure. Figure 6 shows how the maximum temperature varies with shear velocity and pressure. At low pressures, melting limits the attainable temperature even for very high shear velocities. High pressure increases both the melting point and the viscosity and permits much higher temperatures to be attained at lower shear velocities.

On the basis of these calculations, it appears very unlikely that friction could be responsible for a premature ignition in a TNT shell during gun launch. The maximum pressure during gun launch is less that 0.2 GPa, and the sliding velocity between the rotating shell and the explosive is less than 30 m/sec. Under these conditions, the

temperature would not rise significantly above the melting point (of course, the presence of high melting point grit would change the picture drastically). We have not done any calculations on RDX because we do not have data for its viscosity. However, we can make a rough estimate of what would happen in RDX by using the TNT parameters but adjusting the melting point and $T_o$ (from equation (5) for the viscosity) upwards to account for the difference in melting point between TNT and RDX. If we do this, we obtain the same curves which we obtained for TNT but with all of the temperatures higher by 124°C (the difference in melting point between RDX and TNT). On this basis, it appears that RDX should also be insensitive to friction under setback conditions. These conclusions should be checked by experiment, and a program to do so is currently in progress at BRL. At present, the experiments support the theoretical conclusions.

Under shock initiation conditions, shear bands would be expected to form in the vicinity of collapsing cavities. Figure 7 shows in a schematic fashion how this might happen. In this situation, both the pressure and the shear velocity could be much higher than in the setback situation. An estimate of the maximum shear velocity can be obtained by taking the difference between the free surface velocity and the particle velocity behind the shock. This difference is approximately equal to the particle velocity. A 1.03 GPa shock in TNT has a particle velocity of about 0.2 km/sec. Figure 8 shows how the computed shear band temperature varies with time in this case. Thermal explosion occurs after about 50 nanoseconds. However, if the shear stops before thermal explosion occurs, the temperature in the shear band decreases rapidly due to the steep thermal gradients, and reaction does not occur. This is also shown in Figure 8. An estimate of the maximum slip distance at a shear band of this type is the dimension of the cavity in the direction of shock motion. The duration of the shear is this distance divided by the shear velocity. Consequently, we can estimate, very approximately, the critical cavity size required for initiation by this mechanism at any shock pressure. Table II shows these estimates for cavities in TNT.

TABLE II - CRITICAL CAVITY SIZES FOR SHOCK IGNITION IN TNT, COMPUTED ON THE BASIS OF SHEAR BAND IGNITION

| Shock Pressure (K bar) | Particle Velocity (Km/sec) | Time to Explosion (Microsecond) | Critical Cavity Size |
|---|---|---|---|
| 4.9 | 0.1 | 4700 | 4.7cm |
| 7.5 | 0.15 | 1.75 | 0.26mm |
| 10.3 | 0.20 | 0.05 | 0.1mm |

Liddiard's data [12] showed a critical pressure of about 0.7 GPa for the ignition of TNT by long duration shocks. Since the maximum cavity size in a TNT charge of reasonable quality would be about 0.1 to 0.3 mm, the computed numbers are in reasonable agreement with experiment.

## IV. Summary and Conclusions.

The role of shear in the initiation of reaction in high explosive has been examined theoretically. The calculations show that, for initiation by shear to occur under realistic conditions, some localization of the deformation must occur. When localized shear occurs, pressure is a critical parameter which controls the occurrence of ignition. High pressure during shear deformation enhances the chances for ignition by raising both the melting point (thereby increasing the heating due to viscoplastic work) and the viscosity of the melt (hence, increasing viscous heating). This pressure effect is extremely important in causing an explosive to be sensitized to shear deformation, and should be kept in mind when one assesses explosive hazards. For a low melting explosive, such as TNT, pressures in excess of 0.5 GPa are required for ignition at any shear velocity likely to occur. For higher melting explosives, lower pressures can lead to initiation.

The calculations show that typical pressures encountered during gun launch are unlikely to cause ignition of either TNT or RDX by this mechanism. However, when ignitions occur as a result of other causes, the resulting higher pressures make this mechanism a serious potential source of secondary ignitions which lead to violent explosions. This mechanism is also considered to be an important pathway by which interround propagation occurs, when munitions within a storage array are subjected to multiple fragment impacts and severe crushing. (See reference 3.)

Ductile cavity collapse does not lead to conditions which cause ignition. During ductile cavity collapse, the strain is concentrated in regions where the pressure and viscosity are low, and the temperature which may be obtained is limited by the melting point. Brittle cavity collapse, such as might be expected under shock loading conditions, can lead to the formation of shear bands and high temperatures can be attained. This is also expected to occur when the explosive is extruded into cracks in the confining steel case while the munition is under pressure. This latter sequence of events is viewed as a very probable source of ignition when the casing fails under launch or when subjected to crushing.

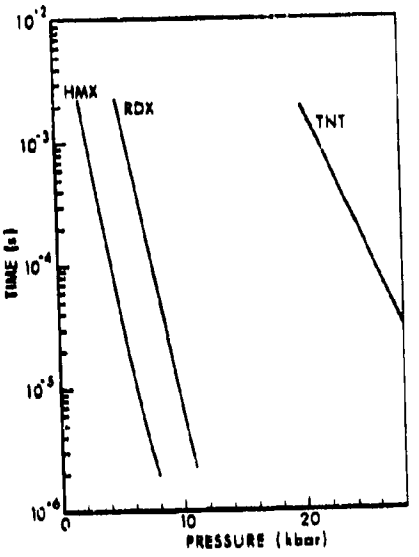



Figure 1 - Thermal Explosion Time at the Melting Point vs Pressure.

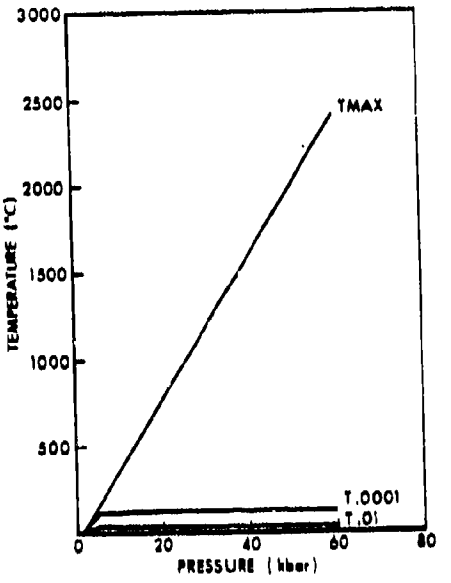



Figure 2 - Temperature vs Pressure After Cavity Collapse (Carroll & Holt's Model).

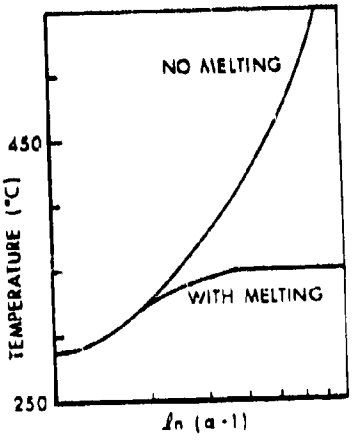



Figure 3 - Temperature vs Porosity During Cavity Collapse When Melting And Heat Conduction Occur.

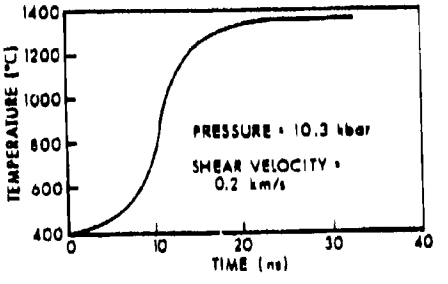



Figure 4 - Temperature vs Time in a Developing Shear Band.

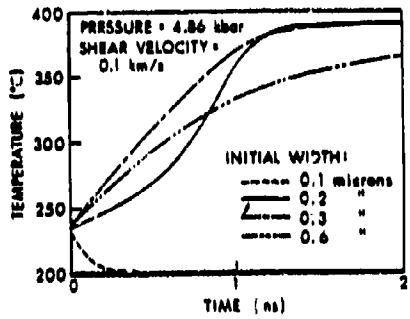



Figure 5 - The Time History of a Shear Band as a Function of the Width of the Initial Heated Region.

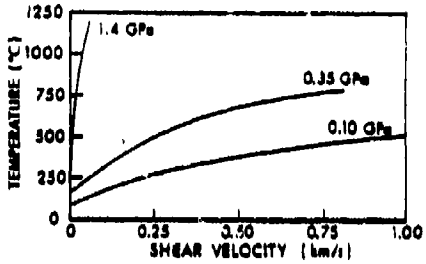



Figure 6 - The Effect of Shear Velocity and Pressure on Peak Temperatures in a Shear Band.

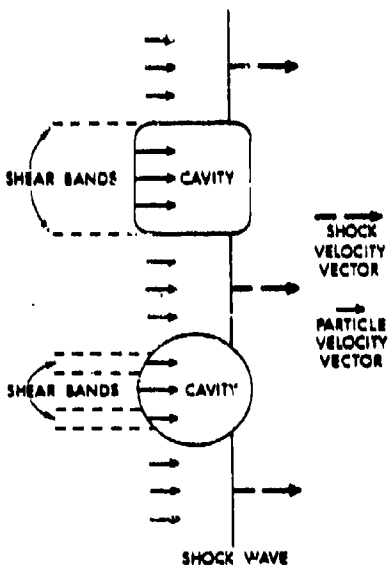



Figure 7 - Schematic Drawing Indicating How Shear Bands Could Be Formed in a Shocked Explosive.

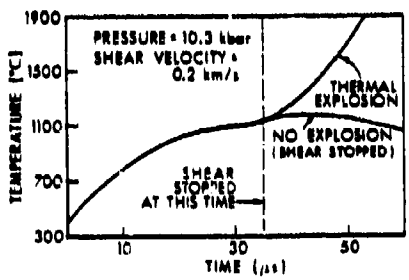



Figure 8 - Temperature vs Time in a Shear Band.

## REFERENCES


1. L. M. Kachanov, Foundations of the Theory of Plasticity, North Holland Publishing Company, Amsterdam (1971).

2. Malfunction Investigation of 8-Inch HE, Rocket Assisted Projectile, XM650E4, TECOM Project No. 2-MU-003-650-020 (1976).

3. R. Frey, P. Howe, G. Melani, J. Trimble, "Initiation of Explosive Charges by Projectile Impact," ARBRL-TR-02176 (1979).

4. M. M. Carroll and A. C. Holt, "Static and Dynamic Pore Collapse Relations for Ductile Porous Materials," Journal of Applied Physics 43, 4 (1972).

5. R. Recht, "Catastrophic Thermoplastic Shear," Journal of Applied Mechanics, Vol 31, Series E, No. 2 (1964).

6. R. E. Winter and J. E. Field, "The Role of Localized Plastic Flow in the Impact Initiation of Explosives," Proceedings of the Royal Society of London A 343 (1975).

7. C. T. Afanasev, U. K. Bololev, Yu. A. Kazarova, and Yu. F. Karabanov, "Formation of Local Hot Spots During the Fracture of Thin Layers Under Shock," Fiziha Goreniza i Vzryva, Vol 8, No. 2 (1972).

8. R. N. Rogers, "Thermochemistry of Explosives," Thermochimica Acta, 11 (1975).

9. M. M. Carroll, A. C. Holt, and B. M. Butcher, Application of a New Theory for the Pressure Induced Collapse of Pores in Ductile Materials," Proceedings of the International Symposium, Pore Structure and Properties of Materials, Praque (1973).

10. J. Frenkel, Kinetic Theory of Liquids, Dover Publications, New York (1955).

11. Engineering Design Handbook - Properties of Explosives of Military Interest, AMCP 706-177 (1967).

12. T. Liddiard, "The Initiation of Burning in High Explosives in Shock Waves," Fourth Symposium on Detonation.

# INTERIOR BALLISTIC MODELING FOR BLANK AMMUNITION

SIDNEY GOLDSTEIN, MR.
U. S. ARMY ARMAMENT RESEARCH AND DEVELOPMENT COMMAND
DOVER, NEW JERSEY 07801

The 1973 Yom Kipper War emphasized that adequate field training was essential for an effective modern mechanized army. There evolved, therefore, the requirement [1] to fire the .50 caliber M2HB and M85 machine guns during training exercises. As a result, blank ammunition for .50 cal weapons is currently being developed as part of the Multiple Integrated Laser Engagement System (MILES) program. This paper describes a computer model which would aid in the design of the ammunition and Blank Firing Attachment (BFA) for these systems.

## EXPERIMENTAL PROCEDURE

### Weapon Cycling

The .50 cal M85 machine gun was chosen as an example. Figure 1 depicts the functioning of blank ammunition in this weapon, which is recoil operated.

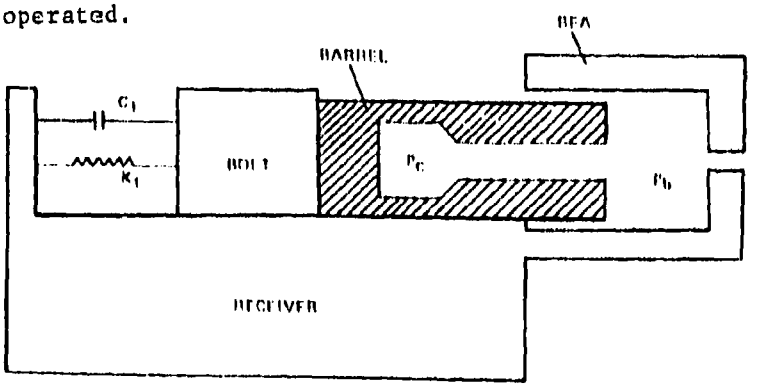



Figure 1. Bolt-Barrel Recoil.



PRECEDING PAGE BLANK-NOT FILMED

Burning propellant in the chamber produces a pressure $P_c$. At some point in time, the mouth of the cartridge case opens and gas flows through the barrel producing pressure $P_b$ in the BFA. Chamber pressure in the breech and BFA pressure on the muzzle of the barrel accelerate the bolt and barrel.

The bolt and barrel initially act as a single body during cycling of the weapon. But within a few milliseconds an accelerator separates the bolt from the barrel, and the weapon cycles independent of the gas pressure in the chamber and BFA. This model is concerned only with interior ballistics during the time the bolt and barrel remain in contact, where $K_1$ is the effective spring constant and $C_1$ is the spring damping constant.

Computer time-displacement for the bolt and barrel of the M85 machine gun when ball ammunition is fired [2] has been obtained. The bolt-barrel contact time lasts only a few milliseconds.

## Ball Ammunition Performance

Pressure-time traces .50 cal M33 ball cartridge were taken at the chamber, case mouth and middle of the barrel locations. Peak force due to chamber pressure was about 10,000 lbs; ballistic cycle lasted about 2 to 3 milliseconds; and impulse was 10-11 lb-sec. The blank ammunition and BFA should be capable of duplicating this performance.

## Test Set-Up

Because it was too difficult to modify the receiver in the M85 machine gun, a heavy-walled .50 cal test barrel was modified instead, and BFA simulator constructed to obtain pressure traces of .50 cal blank ammunition. Pressure stations were drilled at mid-chamber ($P_c$), case mouth ($P_{cm}$), midbarrel ($P_{mb}$), and the BFA ($P_b$) (Figure 2).

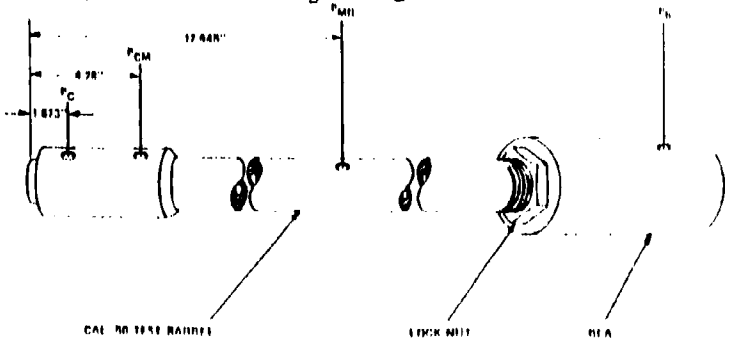


Figure 2. Cal .50 Test Barrel and BFA.

## Shock Waves

Examination of the pressure-time traces for the chamber, midbarrel, and BFA (Figures 3, 4, and 5) showed shock waves occurring between the chamber and BFA.

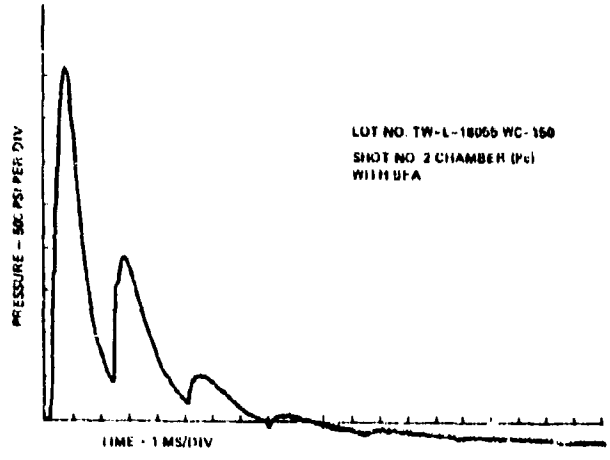



Figure 3. Chamber Pressure Versus Time.

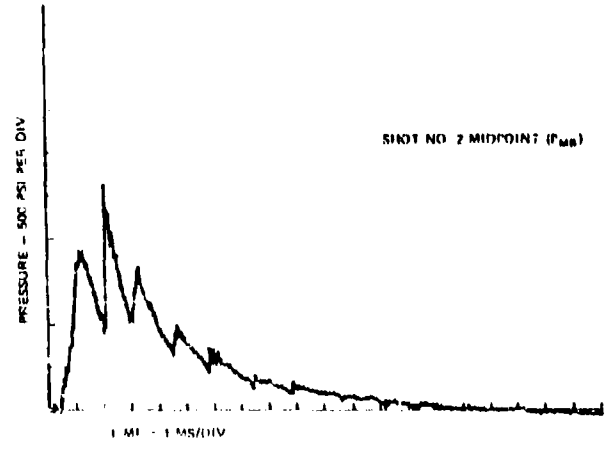



Figure 4. Midbarrel Pressure Versus Time.

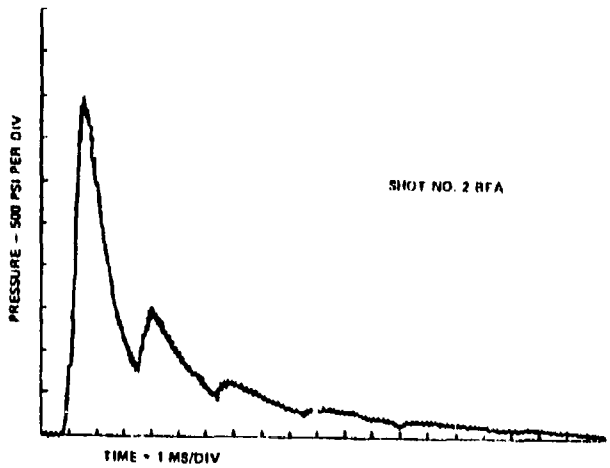



Figure 5. BFA Pressure Versus Time.

Further examination of the first return shock wave, however, indicates it does not occur immediately upon arrival of the leading shock at the BFA, but about 1 millisecond afterwards (Figure 6). Meanwhile, shock wave theory [3] states that a shock wave forms at the muzzle as soon as

$$\frac{P_b}{P_e} > \frac{2\gamma M_e^2 - \gamma + 1}{\gamma + 1} = r_{crit.} \qquad (1)$$

where $M_e$ is the Mach number of the gas at the muzzle and $\gamma$ is its ratio of specific heat.

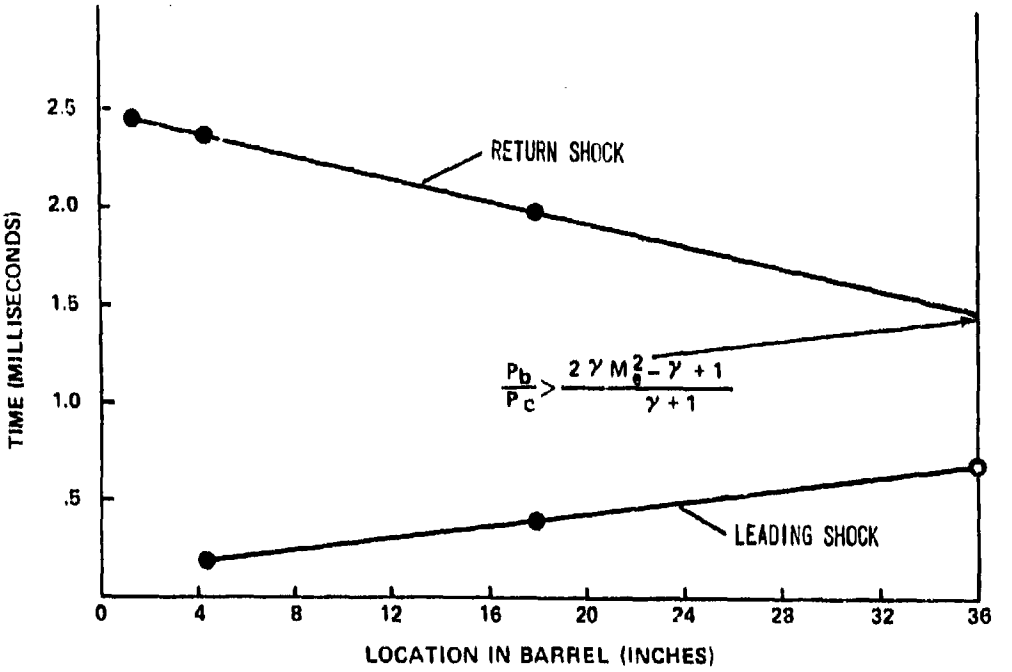



Figure 6. Shockwave Location Versus Time.

This computer program monitors the ratio $P_b/P_e$, and when it becomes larger than $r_{crit}$ the gas flow to the BFA is cut off. The pressure ratio across the leading shock [4, 5] wave depends on the ratio of chamber pressure to ambient pressure at the case-mouth opening, on the ratio of the speed of sound in the gas to the speed of sound in the air, and on the values of $\gamma$ in both gases (propellant gas and air).

## THEORETICAL MODEL

### Three Phases

Propellant combustion and gas flow occur in three separate phases (Figure 7). Phase I occurs in the chamber and includes:

a. Isochoric or constant volume combustion until the mouth of the cartridge case opens.
b. Quasisteady combustion and two-phase flow.
c. Flame quenching.
d. Quasisteady isentropic flow through a converging nozzle.

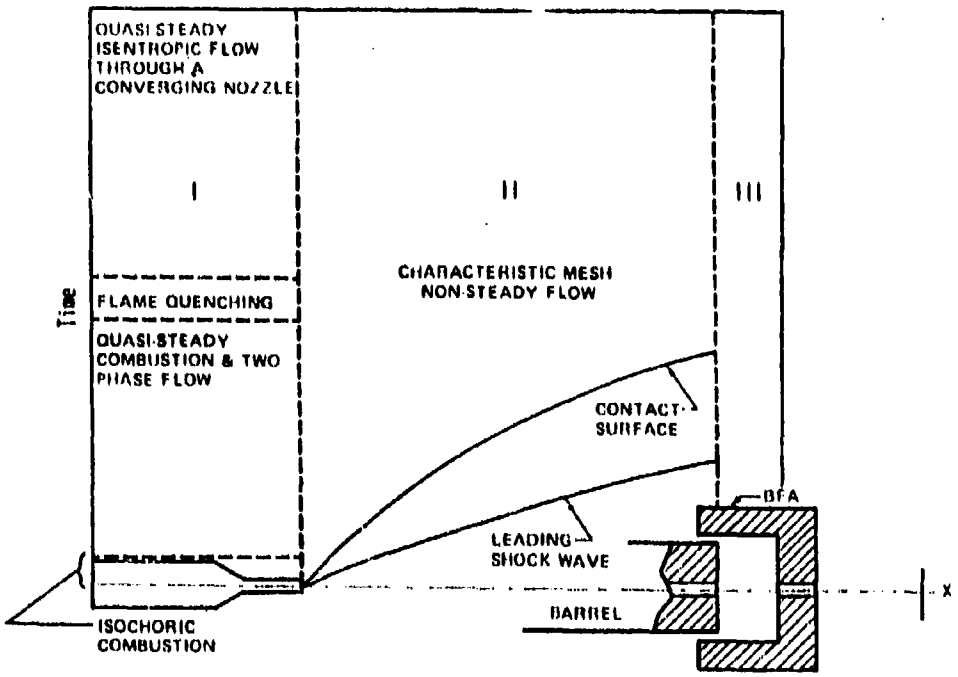



Figure 7. Three Phases of Blank Operation.

For Phase I a lumped ballistic model was used wherein all gas properties within the cartridge are assumed to be uniform. This phase also supplies the initial and boundary conditions for Phase II. Phase II describes the nonsteady flow in the barrel and includes the leading shock wave, contact surface and any compression or rarefaction waves. The method of characteristics [5] is used to solve the fluid flow equations for Phase II. Finally, Phase III defines the nonsteady flow entering the BFA [6], the pressure buildup in the BFA, and the motion of the bolt-barrel assembly. Interpolating along the characteristics at the muzzle, it is possible to determine the state of the gas entering the BFA. This flow is assumed to cease when the return shock forms.

## Development of Equations

Phase I. An important parameter in the Phase I equations is the solid mass fraction $\varepsilon$ [7]. For low values of loading density (characteristic of blank ammunition) the solid mass fraction $\varepsilon$ may be approximated by

$$\varepsilon = \frac{\rho_p}{\rho_p + \rho_g} \qquad (2)$$

Accordingly, it is then possible to define a discharge coefficient $C_d$ for the mixture.

$$C_d = \frac{\dot{m}}{APc} \qquad (3)$$

For a diabatic nozzel (constant temperature)

$$C_d = \frac{1}{\sqrt{F(T_c/T_o)\ (1-\varepsilon)\ e}} \qquad (4)$$

If one assumes an initial gas velocity ($v_o$) due to gas flow from the primer

$$C_d = \frac{\sqrt{v_o^2 + \frac{2FT_c}{T_o}\ (1-\varepsilon)\ \left(1/2 - \frac{v_o^2}{2F(T_c/T_o)\ (1-\varepsilon)}\right)}}{F(T_c/T_o)\ (1-\varepsilon)\ e\ [1/2 - v_o^2/(2F(T_c/T_o)(1-\varepsilon)]} \qquad (5)$$

For perfect heat transfer between the phases

$$C_d = \left\{\frac{2\ C_T}{(1-\varepsilon)\ \frac{F\ T_c}{T_o}\left[2\ C_T - (1-\varepsilon)\ F/T_o\right]}\right\}^{1/2}\left[\frac{2\ C_T - 2\ (1-\varepsilon)\ F/T_o}{c\ C_T - (1-\varepsilon)\ F/T_o}\right]^{\frac{C_T\ T_o}{(1-\varepsilon)F}-1} \qquad (6)$$

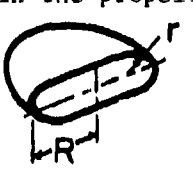
For no heat transfer between the phases

$$C_d = \left[\frac{2\ \gamma\ T_o}{(\gamma-1)\ (1-\varepsilon)\ F\ T_c}\right]^{1/2}\left[\left(\frac{2}{\gamma+1}\right)^{\frac{2}{\gamma-1}} - \left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{\gamma-1}}\right]^{1/2} \qquad (7)$$

With these two parameters defined, it is now possible to formulate a set of 13 differential equations describing the interior ballistics (7, 8, 9) for Phase I during the quasisteady combustion and two-phase flow. A nondeterred, rolled-ball propellant is used in the propellant charge.

1. Burning rate equation is

$$\frac{d\delta}{dt} = BP_c^n \qquad (8)$$




2. Rate of change of surface area for a propellant grain is derived by

$$\frac{d\ Si}{dt} = -2\pi\left[\pi(R-r) + 4\ (r-\delta)\right] BP_c^n \qquad (9)$$

3. Rate of change of grain volume equals

$$\frac{dVi}{dt} = -2\pi(R - r^2)^2 - 2\pi^2 (r - \delta)(R - r) - 4\pi (r - \delta)^2 BP_c^n \quad (10)$$

4. Rate of change of number of propellant grains in cartridge becomes

$$\frac{d N_h}{dt} = \frac{-\varepsilon \; C_d \; P_c \; A_c}{\rho s \; Vi}$$

5. Rate of discharge of propellant gas from cartridge case is

$$\frac{d C_o}{dt} = (1 - \varepsilon) P_c A_c C_d \quad (11)$$

6. Rate of increase of propellant gas in cartridge is solved by

$$\frac{d C_i}{dt} = \rho_s S_i N_B BP^n \quad (12)$$

7. Rate of change in amount of gas in cartridge is

$$\frac{d C_c}{dt} = \frac{d C_i}{dt} - \frac{d C_o}{dt} \quad (13)$$

8. Rate of change of gas temperature in cartridge is derived by

$$\frac{d T_c}{dt} = \frac{1}{C_c} \left[ - \gamma T_c \frac{d C_o}{dt} + T_o \frac{d C_i}{dt} - T_c \frac{d C_c}{dt} \right] \quad (14)$$

9. Rate of change of gas pressure in the cartridge is

$$\frac{d P_c}{dt} = \frac{F}{T_o V_c^2} \left[ V_c T_c \frac{d C_c}{dt} + V_c C_c \frac{d T_c}{dt} - C_c T_c \frac{d V_c}{dt} \right] +$$

$$\frac{Fig \; Cig}{Tig} \left[ \frac{1}{V_c} \frac{d T_c}{dt} - \frac{T_c}{V_c^2} \frac{d V_c}{dt} \right] \quad (15)$$

10. Rate of change of gas density in cartridge case is found by

$$\frac{d \rho g}{dt} = \frac{T_o}{F} \frac{1}{T_c} \frac{d P_c}{dt} - \frac{P_c}{T_c^2} \frac{d T_c}{dt} \quad (16)$$

11. Rate of change of solid mass density is

$$\frac{d\rho p}{dt} = \frac{- C_c P_c A_c - C_c}{U_o} \quad (17)$$

12. Rate of change of solid mass fraction is determined by

$$\frac{d\varepsilon}{dt} = \frac{\rho_g}{(\rho_p + \rho_g)} \frac{d\rho_p}{dt} - \frac{\rho_p}{(\rho_g + \rho_p)^2} \frac{d\rho_g}{dt} \qquad (18)$$

13. Rate change of free volume is

$$\frac{d\ V_c}{dt} = - \frac{u_o}{\rho_s} \frac{d\rho_p}{dt} - \eta \frac{d\ C_c}{dt} \qquad (19)$$

These 13 simultaneous differential equations were solved using a fourth order Runge-Kutta integration.

Flame Quenching: Flame quenching [10, 11] occurs when the pressure starts to drop rapidly. Subsequently, the flow is assumed to be quasisteady-insentropic as through a converging nozzle. Solution of these equations is shown in Figure 8 for the case where $\varepsilon$ = variable ($v_o \neq 0$) and $\varepsilon = 0.5$ ($v_o \neq 0$). Figure 8 shows the experimental curve. Since a BFA was not used, no shock waves occurred.

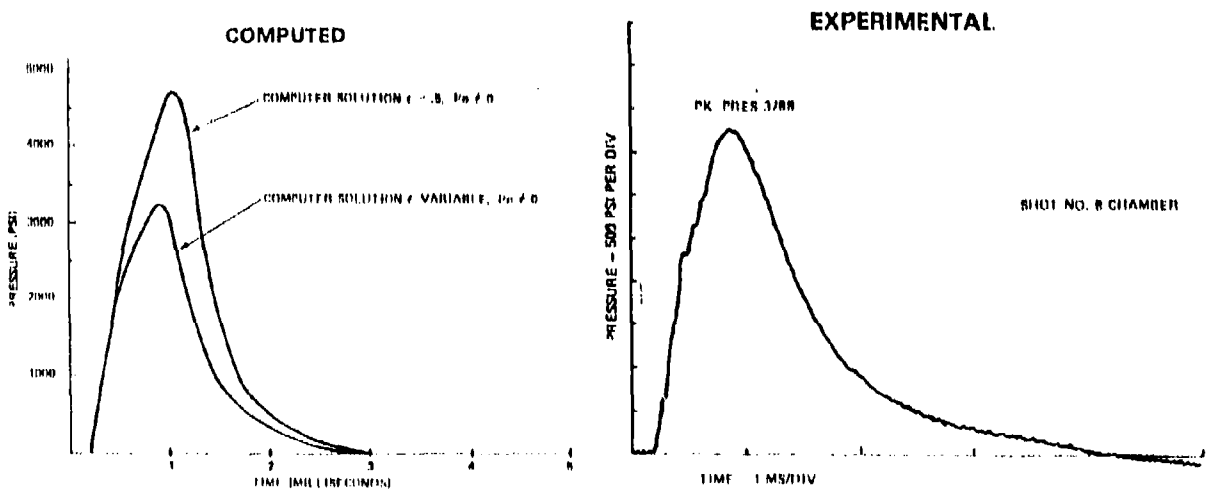



Figure 8. Chamber Pressure Versus Time.

Gas Leakage. Gas leakage in the BFA has a significant effect on the pressure This leakage (Figure 9) occurs through the forward end of the BFA (ABFA), through the clearance between the BFA and the barrel ($A_{BLK}$) and after the barrel has recoiled sufficiently through the vent port $A_{PRT}$.

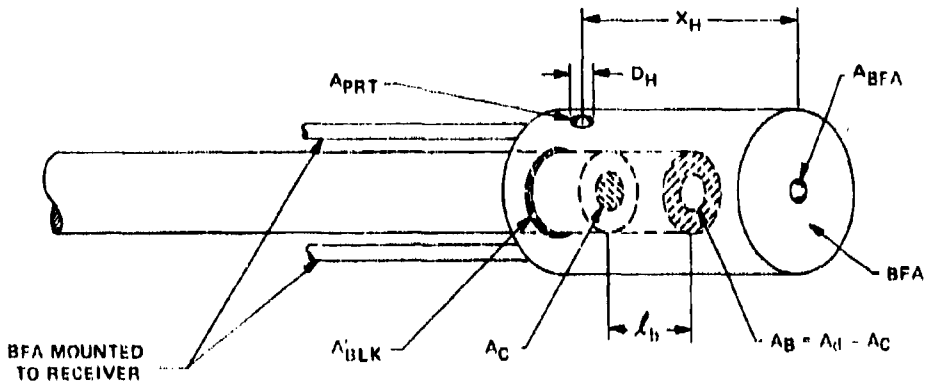



Figure 9. Cal .50 Barrel and BFA.

Equation of Motion. The equation of motion for the bolt-barrel assembly is determined from applied forces. Pressure in the BFA and in the cartridge case supply the accelerating forces. Five simultaneous differential equations must, however, be solved to determine this motion.

1. Equation of state for the BFA is solved by

$$\frac{d\,P_b}{dt} = \frac{m_b R}{U_b + A_d \ell_b}\,\frac{d\,T_b}{dt} + \frac{R\,T_b}{U_b + A_d \ell_b}\,\frac{dm_b}{dt} - \frac{R T_b\, A_d\, m_b}{(U_b + A_d \ell_b)^2}\,\frac{d\ell}{dt} \qquad (20)$$

2. Energy equation for BFA becomes

$$\dot{m}e\,(1-\beta)\left[C_p\,T_e + \frac{ve^2}{2gJ}\right]\left(1 - \frac{A_{BFA}}{A_c}\right) = \frac{P_b\,A_B}{J}\,\frac{d\ell_b}{dt} + \frac{d}{dt}\,(m_b g C_v T_b) +$$

$$C_p\,T_b\,\dot{m}ob \qquad (21)$$

or solving for gas temperature

$$\frac{d\,T_b}{dt} = \frac{\dot{m}e\,(1-\beta)}{C_v\,mb}\left[C_p\,T_e + \frac{ve^2}{2gJ}\right]\left(1 - \frac{A_{BFA}}{A_c}\right) - \frac{P_b\,A_B}{gJC_v\,mb}\,\frac{d\ell_b}{dt} -$$

$$\frac{C_p\,T_b\,\dot{m}ob}{mb\,C_v} - \frac{T_b\,\dot{m}b}{b} \qquad (22)$$

where $\dot{m}_e$ is the rate at which gas enters the BFA and

$$\dot{m}_{ob} = P_b \, (A_{BLK} + N_H \, A_{PRT}) \sqrt{\frac{\gamma}{R \, T_b} \left(\frac{2}{\gamma + 1}\right)^{\frac{\gamma + 1}{\gamma - 1}}} \qquad (23)$$

3. Equation of motion is found by

$$P_c \, A_c + P_b \, A_B - K_1 \, (\ell_b + X_o) - C_1 \, V - F_1 - F_2 = M_t \frac{dv}{dt} + 16.56 \, K_n \frac{dv}{dt} \qquad (24)$$

4. Conservation of propellant gas mass is computed by

$$\dot{m}_b = \dot{m}_e - \dot{m}_{ob} \qquad (25)$$

5. Equation of barrel-bolt displacement is

$$\frac{d\ell_b}{dt} = V \qquad (26)$$

BFA Pressure-Time Curves. The computed and experimental pressure-time curve for the BFA and the experimental pressure-time curves are shown in Figure 10. The effect of the reverse shock wave is evidenced by the decrease in pressure following the peak. Meanwhile, the rise in pressure for the experimental curve following the peak is due to the arrival of the return shock from the breech.

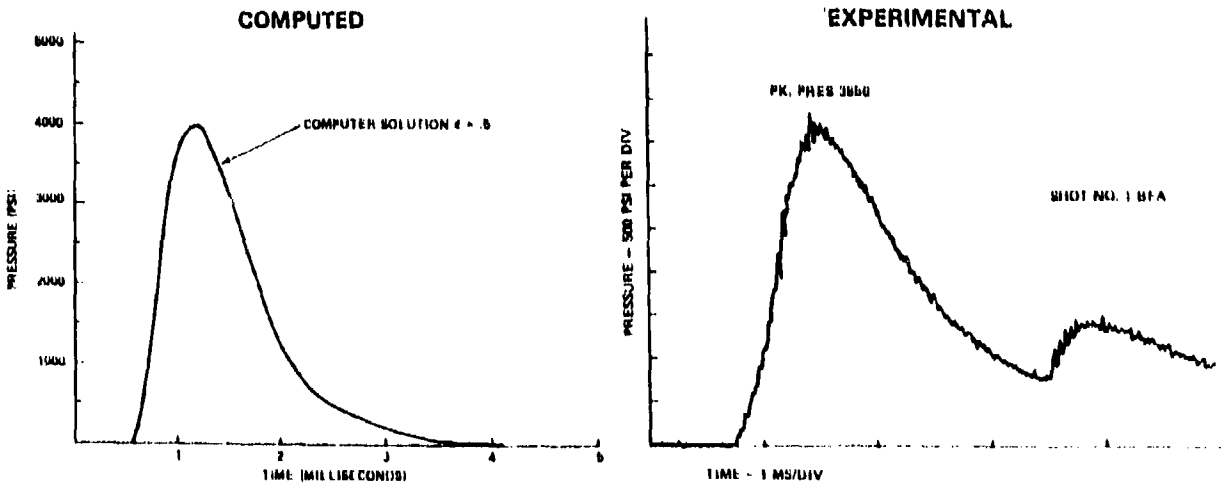



Figure 10. BFA Pressure Versus Time.

## CONCLUSIONS

1. This model appears adequate for describing the quasisteady combustion and two phase flow from the .50 cal blank cartridge. It can also simulate the leading shock wave, compression wave and pressure buildup in the BFA.

2. The effective time-averaged solid mass fraction ε of about 0.5 seems reasonable for simulating the peak pressure in the cartridge case.

3. Quenching of the propellant flame immediately following the start of depressurization of the cartridge accounts for the shape of the pressure-time curve in the cartridge following peak pressure.

4. The pressure drop in the BFA chamber occurs when the first return shock wave forms and travels back toward the breech.

5. The method of characteristics can be used to solve equations describing nonsteady isentropic gas flow from cartridge case to BFA.

## REFERENCES

1. R. F. Schwegler, "Evaluation of Two Blank Firing Attachments for the M2HB Caliber .50 Machine Gun," R-TR-76-022, General Thomas J. Rodman Laboratory (SARRI-LS-P), Rock Island Arsenal, Rock Island, IL 61201, June 1976.

2. H. E. Weidner , "M85 Math Model, Kinematic Model of the M85 Cal .50 Machine Gun," RETR-71-83, Small Arms Laboratory, US Army Weapons Command Research and Engineering Directorate, Rock Island, IL, December 1971.

3. G. Rudinger, Nonsteady Duct Flow: Wave Diagram Analysis, Dover Publications, Inc., New York, NY 1969.

4. S. Goldstein, "Qualitative Investigation of the Functioning of 7.62MM Blank Ammunition," Frankford Arsenal Technical Note TN-1141, June 1969.

5. J. A. Owczarek, Fundamentals of Gas Dynamics, International Publishing Co., 1964.

6. S. Goldstein, "Study of the Flow Through the Pressure Port in a Gas Operated Small Arms Automatic Weapon," Frankford Arsenal Report FA-TR-76059, November 1976.

7. S. Goldstein, "Analytical Design Study of Caliber .50 Blank Ammunition and Weapon Interface," US Army Armament Research and Development Command, Technical Report (to be published).

8. S. Goldstein, "A Simplified Model For Predicting The Burning Rate and Thermochemical Properties of Deterred Rolled Ball Propellant," Frankford Arsenal Technical Note TN-1184, December 1973.

9. AMCP 706-150, Engineering Design Handbook, Ballistic Series, Interior Ballistics of Guns, US Army Materiel Command, February 1965.

10. C. L. Merkle, S. L. Turk, and M. Summerfield, "Exinguishment of Solid Propellants by Depressurization Effects of Propellant Parameters," AIAA Paper No. 69-176 (1969). Also AMS Report No. 880, Princeton University, Princeton, NJ 1969.

11. K. K. Kuo, "Review of Dynamic Buring of Solid Propellants In Gun and Rocket Propulsion Systems," 16th International Symposium On Combustion, pages 1177-1192, August 15-20, 1976.

## LIST OF SYMBOLS (7)

$A_d$ = cross-sectional area of BFA
$C_p$ = specific heat for gas
$C_S$ = specific heat for particles
$C_f$ = $(1 - \varepsilon) C_p - \varepsilon C_S$
$C_1$ = spring damping constant
$e$ = 2.718 -- base of natural logs
$F$ = propellant impetus
$F_1$ = bolt receiver friction force
$F_2$ = barrel receiver friction-force
16.50 $K_n$ = effect of number of rounds in the ammunition belt
$K_1$ = effective spring constant for bolt-barrel assembly
$\ell_b$ = travel of bolt-barrel assembly
$m_b$ = mass of the gas in BFA
$M_t$ = total moving mass
$P_b$ = pressure in BFA
$R$ = gas constant
$T_c$ = gas temperature in the cartridge
$T_0$ = isochoric adiabatic flame temperature
$U_b$ = initial volume of BFA
$X_0$ = initial compression of spring
$\eta$ = gas covolume
$\rho_g$ = density of gas in the mixture
$\rho_p$ = density of solid particles in the mixture

# RAPID TRANSMERIDIAN DEPLOYMENT: COGNITIVE PERFORMANCE AND CHRONOBIOLOGIC PROPHYLAXIS FOR CIRCADIAN DYSCHRONISM (U)

*R. CURTIS GRAEBER, CPT, MSC
BRUCE N. CUTHBERT, CPT, MSC
HELEN C. SING, MRS.
ROBERT J. SCHNEIDER, MAJ, MSC
G. RUFUS SESSIONS, MAJ, MSC
WALTER REED ARMY INSTITUTE OF RESEARCH
WASHINGTON, D. C. 20012

Rapid deployment of combat units to overseas locations is a primary concern of today's strategic planners. Such movements require the airlifting of units across multiple time zones. Numerous studies have documented the adverse physiological and behavioral consequences accompanying the rapid crossing of three or more time zones (1). These effects result from the requirement that the body must adjust its circadian rhythms to the new local time.

Under normal conditions, these daily cycles are synchronized by the external Zeitgebers (i.e., time-givers) of the local environment. The sudden shifting of these Zeitgebers causes the shifting at different rates of the body's physiological, biochemical, and behavioral rhythms. While some circadian rhythms adjust quite rapidly, others adjust very slowly. Consequently, the passenger's circadian system is not only out of synchrony with the environment but is also internally desynchronized. It is the latter condition, circadian dyschronism, which is particularly responsible for the fatigue and malaise typically reported as "jet lag" during the first several days following rapid transmeridian flight.

While the physiological and behavioral consequences of such flights are a common experience for airline travellers, their impact poses a potentially serious problem for troops required to display maximal combat effectiveness upon arrival at a hostile destination. The high-level cognitive performance required by the modern sophisticated battlefield may only serve to exacerbate the problem. Any



PRECEDING PAGE BLANK-NOT FILMED

reduction of the adverse effects of rapid deployment would enhance combat readiness. Consequently, we designed a set of experiments to test a series of chronobiologic countermeasures (CM) which may hasten physiological and behavioral adaptation to new time zones.

Previous attempts to develop chronobiotics have not been successful. These efforts were limited to the use of a pharmacologic agent, i.e., a corticosteroid or a combined tranquilizer and central neurotransmitter depletor (2,3). In contrast, we chose to manipulate a number of different Zeitgebers simultaneously. Selection of specific CMs was based upon their suitability to the operational requirements of emergency military airlifts and their potential chronobiological effectiveness. Hence, control of the following variables was established: mealtiming, dietary constituents, caffeine and theophylline consumption, light-dark (LD) cycle, rest-activity pattern, and social-psychological time cues.

Support for the potential effectiveness of these interventions are found in the current chronobiologic literature. Ehret and his colleagues (4) have demonstrated that injections of methylated xanthines i.e., theophylline or caffeine, in rats can advance or delay the daily maximum for body temperature. If they are administered just before or during the early active phase of the circadian cycle (i.e., rising body temperature), a phase delay results, whereas if they are administered during the late active, early inactive phase (i.e., just before or after the thermal peak), a phase advance results. These investigators have also induced more rapid phase adjustment of the temperature rhythm to a shift in LD cycle by (a) fasting a rat on the day prior to the shift and (b) restoring food coincidental with the first active phase of the new LD cycle. Presumably, this chronobiotic effect is mediated by the depletion of liver glycogen during the fast followed by the reinitiation of feeding at the chronotypically appropriate time in the revised LD cycle. Others have demonstrated the importance of mealtiming as a synchronizer of circadian rhythms in humans (5).

Related work by Wurtman and Fernstrom (6,7) forms the basis of the dietary manipulations. They have shown that fasted rats exhibit a significant increase in brain tryptophan and serotonin within 1 hr. after a high carbohydrate, low protein meal. The effect is mediated by an increase in serum tryptophan elicited by insulin secretion. Fasted rats also manifest a rapid increase in brain catecholamine levels, particularly norepinephrine, following a meal rich in protein. This enchanced catecholamine synthesis can be traced directly to increases in brain tyrosine levels, the amino acid precursor of these neurotransmitters. Since a dramatic rise in brain

catecholamine levels occurs at the onset of the active phase of the circadian rest-activity cycle and is associated with increased alertness (8), consuming high-protein meals in the morning and at lunchtime on the day of arrival should facilitate the rise in brain catecholamines appropriate to the active phase of the shifted circadian cycle. Conversely, a large, high-carbohydrate dinner eaten at a time in synchrony with the destination populace should facilitate the increase in brain serotonin which typically precedes sleep and therefore should hasten sleep adjustment.

Aschoff's group (1) has used an underground bunker to demonstrate the expected importance of the LD and rest-activity cycles as synchronizers of human circadian rhythms; however, their studies have also revealed the special role of social interaction in determining the speed of phase adjustment to shifted LD schedules. Similarly, post-flight participation in outdoor group activities can hasten the adaptation process following transmeridian flight (9).

In designing "jet lag" CMs for eastward deployment, we combined some of Ehret's (4) suggestions with manipulation of the factors just described. The operational requirements of a potential large-scale emergency military airlift limited the extent and duration of experimental interventions to those which could be instituted on the day of departure and carried out with minimal disruption to mission accomplishment. Likewise, in both studies described herein, operational considerations required that data collection be restricted to relatively few days before and after the flight with minimal interference in the subjects' ability to carry out their military duties. The first study evaluated the effectiveness of the CMs on troops being permanently transferred via chartered commercial airliners from Ft. Hood, TX, to W. Germany in the fall of 1978. The second study attempted to validate the CMs under more realistic combat conditions by focusing on troops from Ft. Riley, KS, being airlifted on USAF C-141 aircraft to participate in REFORGER '79 during the winter. The latter study also examined the effects of deployment on cognitive performance without the CMs.

## EXPERIMENT 1

Subjects. The sample comprised 179 soldiers (18 to 44 years old) from the 2nd Armored Div. transferring as a unit from Ft. Hood, TX, to W. Germany. Eighty-four of the subjects flew on one aircraft and followed the CM procedures; the remaining 95 control subjects deployed on a second plane. Both airliners departed the U.S. midday and arrived in Germany early the next morning, a time advance of 6 hrs.

Procedure. Organismic adjustment following flight can be conceptualized into three response classes: physiological activity, subjective reports of well-being, and work performance. While the latter is of paramount concern for military planning, the first two are obviously important in determining ultimate levels of efficiency. Thus, all three areas must be considered in a comprehensive evaluation of circadian dyschronism. Oral temperature was the physiological parameter assessed on all subjects, while a sub-sample of 15 soldiers in each group was studied more intensively. The latter completed self-report scales consisting of a fatigue checklist (10) and a diary of all activities. Performance tests comprised a four-choice reaction time task and an arithmetic test of summing successive pairs of single-digit numbers in a column of 50. Subjects in these "intensive" subgroups were selected from troops living in the barracks and were tested every 4 hrs. around-the-clock for 4 days two weeks prior to departure. Pre-flight temperature measurements for the remaining subjects, who lived off-post, were taken only during their normal duty hours at 0800, 1200, and 1600 CDT.

The CM procedures for the experimental group were initiated on the morning of departure. Subjects were restricted to a light, low carbohydrate breakfast with fruit juice, milk, and decaffeinated coffee; in fact, the majority ate nothing. Napping was prohibited throughout the day. Upon boarding the aircraft, they were welcomed first in German and then instructed by the Sergeant Major to set their watches ahead 6 hrs. since henceforth the unit would function on German time. A light "supper" was announced and served at 1745 CET (1145 CDT). It consisted of a ham and cheese sandwich, a small salad, cheese, and fresh fruit. No caffeinated beverages or sweetened soft drinks were allowed. Instead, milk, unsweetened fruit juices, and "Gatorade" were available. At 2200 CET the subjects were given 100 mg of dimenhydrinate to induce drowsiness. At 2300, the cabin lights were turned off, and everyone was instructed to sleep until 0405 CET when the cabin lights were turned on. Hot washcloths were distributed, and subjects were told to stretch, interact, and move about. A high-protein breakfast, including a 6 oz. steak and a two-egg cheese omelet, was served at 0430, with second helpings available. Consumption of caffeinated beverages was encouraged since the subjects were now on the downslope of their U.S. time-referenced circadian temperature cycle. The flight landed at 0630, and the remainder of the day was largely spent unpacking at the training base following a 90 min. bus ride from the airport. Napping was prohibited throughout the first day until 1800 to prevent reversion to U.S. time.

Control subjects, in contrast, followed a normal airline

routine. They ate a hot lunch and dinner on the aircraft at the usual U.S. times, plus a breakfast snack at 0810 CET. No alcoholic beverages were permitted. Although the cabin lights were turned off from 0215 until 0550 CET, individual reading lights were available and no constraints were placed on the subjects' activities. These subjects were allowed to nap whenever duties permitted during departure preparations and following arrival at the training base.

For the next six days, all subjects were housed in barracks and tested every 4 hrs. around-the-clock. Only light duties were assigned with no physical training or heavy labor.

Results. The most readily apparent evidence for CM effectiveness was seen in the self-reports. Experimental subjects reported significantly less fatigue during the first 24 hrs. in Germany than the control subjects ($p < 0.05$, t-test), and in fact showed little change from baseline (Fig. 1). Greater fatigue for the control subjects is also indicated by the fact that they slept longer than the CM subjects for the first two days in Germany (Fig. 2). This difference was significant even when sleep before 1800 on day 1 was excluded from the analysis (4.4 vs. 8.1 hrs., $p < 0.005$, t-test) in order to remove the potential bias resulting from the CM subjects not being allowed to nap during the day. In-flight observations demonstrated that the average amount of sleep on the plane was 5.5 hrs. for each group. It was the timing of sleep that differentiated the two groups.

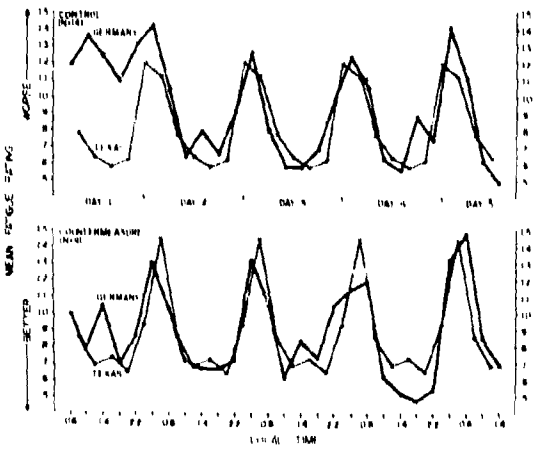


Fig. 1. Post-flight fatigue ratings compared to phase-shifted (+6 hrs.) 4-day mean ratings in Texas. Seven subjects omitted due to contradictory responses or loss of book.

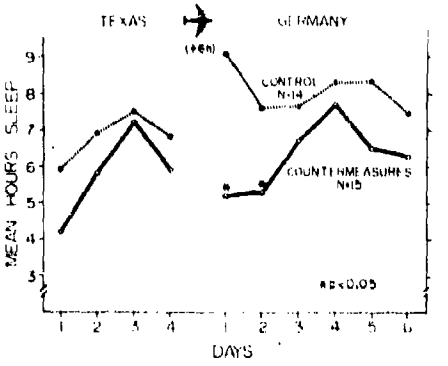



Fig. 2. Duration of daily sleep before and after deployment.

Support for the hypothesis of faster adaptation with the CMs is also provided by the oral temperature data. However, the interpretation of these results is limited by the lack of a 24-hr. baseline for the large groups and the relatively short five-day measurement period in Germany. As shown in Fig. 3, the curves of group mean temperature exhibited very rapid initial adaptation to the new time zone. Inspection of Fig. 3 suggests that, compared to the control group, the shape and amplitude of the CM function more closely approximates that of the intensive groups' baseline for the first two days in Germany. Both large groups appear rather similar thereafter. Whether there is a difference in the rate of final adaptation of the two groups can not be determined since data collection terminated after five days.

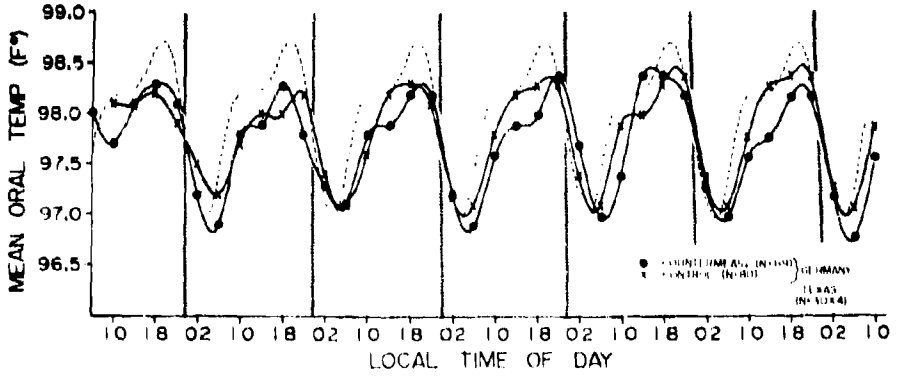



Fig. 3. Spline-fit functions of large groups' mean post-flight temperature compared to phase-shifted estimate of pre-flight rhythm based on combined small groups in Texas.

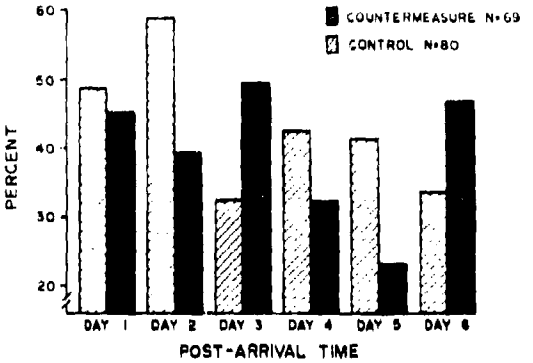



Fig. 4. Percent of subjects whose thermal acrophase indicated lack of phase adaptation (see text).

Group means tend to minimize the day-to-day variation of individuals and may obfuscate the oscillatory nature of the adjustment process by implying a smooth, gradual transition. Consequently, assessment of rhythmic structure was made by subjecting the data to a complex demodulation (CD) analysis (11) which outputs a pseudo-sinusoidal estimate of the times of the circadian maximum and minimum for each subject per day. Figure 4 shows the results of one compilation of these analyses. It depicts the daily percentage of subjects whose estimated acrophase (i.e., peak time) fell outside a one standard deviation range about the pre-flight mean acrophase (1713 $\pm$ 2.9 hrs) of the combined intensive groups. It is evident that phase adaptation proceeded irregularly and cyclically, with an approximate 3-day cycle. Unfortunately, the data terminate before any firm conclusions can be drawn regarding differential overall adaptation rates between the two large groups.

The group mean temperature curves for the intensive groups varied in a different fashion than those for the large groups, particularly during the first four days. The variability from day to day in both amplitude and phase may reflect the small number of subjects, the sleep-disrupting schedule, or chance. Adaptation of mean temperature appeared to be largely complete by day 6.

A CD analysis of each subject's temperature data provided individual estimates of the circadian and ultradian components. Due to the limitation of a 4-hr. sampling rate, the latter comprised frequencies of two and three cycles per day. Total energy may be represented by the sum of all frequencies, while shifts in energy may occur among the various components during periods of adaptation. Figure 5 indicates that the CM subjects maintained a relatively higher percen-

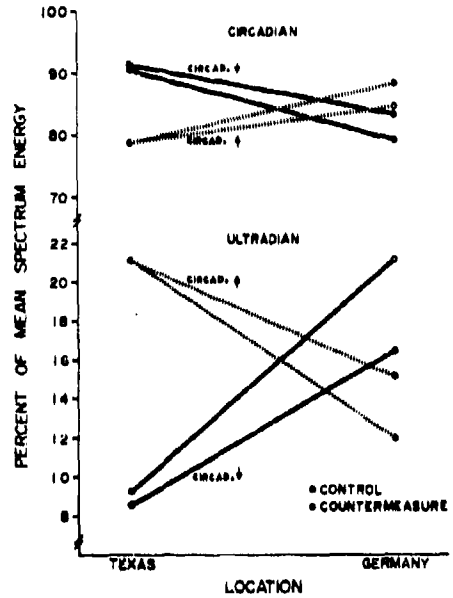



**Fig. 5.** Shifts in spectral energy of thermal rhythms after deployment. Subjects have been divided into those who increased (Control=7, CM=6) and those who decreased (Control=8, CM=9) circadian energy.

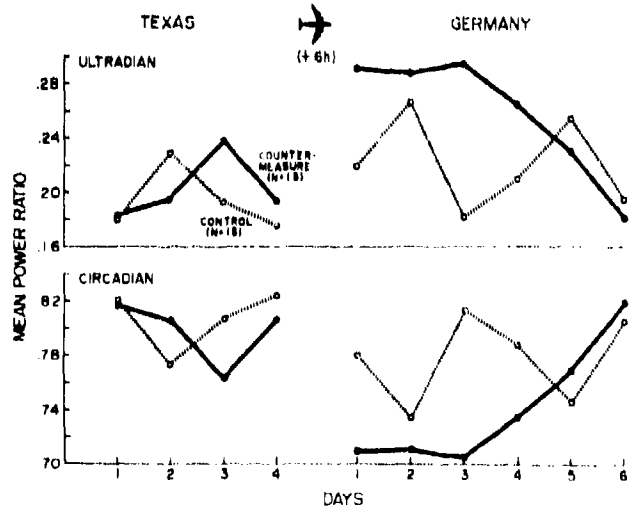



**Fig. 6.** Daily mean power ratios of thermal rhythms. Control group components differ from CM components on days 1, 3, and 4 in Germany and on day 3 in Texas ($p < 0.001$, t-test on arc sine transforms).

tage of ultradian energy post-flight regardless of whether they increased or decreased the spectral strength of the circadian component. This effect is dissected in Fig. 6, where the mean power ratios are plotted daily for each group. Both groups exhibited moderate day-to-day variability in Texas, and deviated even more following the flight. The CM subjects, however, changed smoothly and gradually back to the baseline from initial days of relatively high ultradian energy, while the control group varied erratically.

Because older individuals adapt with greater difficulty to altered work-sleep schedules (15), two subgroups of older subjects ($\geq$ 30 yrs.) were drawn for comparison. In Fig. 7 it is evident that the CMs were particularly efficacious in preserving the amplitude and phase of their circadian temperature rhythm. Obvious differences between the control and CM groups disappeared by day 3.

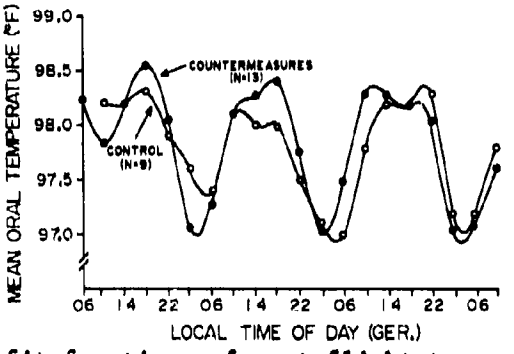



Fig. 7. Spline-fit functions of post-flight temperatures in older soldiers.

A final point concerns the mean daily temperature averaged over all times. As others have reported (13), this value was suppressed following the flight (Fig. 8). Although the group means were identical in Texas, the control group exhibited a consistently (but not significantly) greater decrease after day 1 in Germany. This finding is a further suggestion of the beneficial effects of the CMs.

Assessment of the third component of interest, test performance, was limited. The addition task exhibited the close covariation with body temperature which has been reported frequently for tasks of this type. However, no group differences were detected following the flight, performance levels in both groups remaining virtually unchanged. Data from the 4-choice reaction task, currently being transferred from tape to computer, are unavailable at this time.

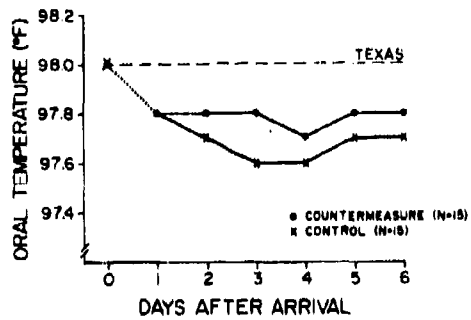



Fig. 8. Mean daily post-flight oral temperatures compared to each group's overall mean daily temperature in Texas.

## EXPERIMENT 2

The second experiment was designed to confirm the initial findings under more rigorous field conditions. Secondly, recognizing the critical importance of post-deployment cognitive functioning, we collected additional data to evaluate performance changes as a function of age. The study was carried out during winter REFORGER '79, with snow and extreme cold both in the U.S. and Germany.

Subjects. The CMs were tested on 120 subjects from an artillery battalion deployed on 4 aircraft: two each for the CM and control groups. Sixty subjects were selected from a maintenance battalion for the investigation of age effects. The "young" (N=29, mean age 21.0 yrs) and the "older" (N=29, mean age 34.2 yrs.) subjects deployed on several different aircraft with no CM treatment.

Procedure. Training and baseline testing were carried out for 4 days during the week immediately before departure from Ft. Riley, with three daily test periods corresponding to breakfast (0800), lunch (1200), and dinner (1630) times. No physiologic measures were taken. Oral temperature, the only feasible choice, would have been unreliable due to the weather. The self-report fatigue scale and the diary of the first study were employed along with scales for self-rating the abilities to concentrate, make decisions, reason clearly, and process information.

Direct assessment of performance was expanded considerably, although operational requirements limited testing of the CM group to one 3-min. task. This task, the "griddle", required subjects to encode and decode simulated map coordinates using an alphanumeric con-

version table. However, the test battery for the other 60 subjects included the griddle, as well as the trails test of visuospatial search, a logical reasoning task, letter cancellation, and short-term word recall. The battery was printed in a pocket-sized booklet and required about 20 min. to complete, at the end of which the subjects rated their overall performance. A technical specialist supervised the taking of the time-limited tests.

The troops deployed on USAF C-141 transports configured in four columns of webbed seats, a cramped arrangement which made sleeping difficult. The CMs mimicked the earlier procedures as much as possible, but were modified to conform to USAF schedules, standardized in-flight meals, seating, etc. All other subjects followed standard USAF cabin procedures.

Following deployment (+7hrs.), the troops were tested for 3 to 5 days in large tents which were poorly illuminated and heated. Four test sessions were held daily on the same schedule as in the U.S. except that a night test was added at 2100 hr.

Results. The self-report data provided the strongest support for CM effectiveness. Figure 9 shows that, while both groups experienced higher than normal fatigue after arrival, CM subjects were significantly lower than controls for the first two days in Germany ($p < 0.05$, t-test). Both groups failed to return to baseline levels following partial recovery on day 3. A very similar pattern of results appeared in the four self-rating scales of information processing, etc. The CM subjects exhibited significantly smaller decrements in self-rated effectiveness for the first two days, followed by a partial recovery (Fig. 10). The sleep data were essentially equivocal in Germany because the nature of the subjects' duties precluded ad libitum sleep.

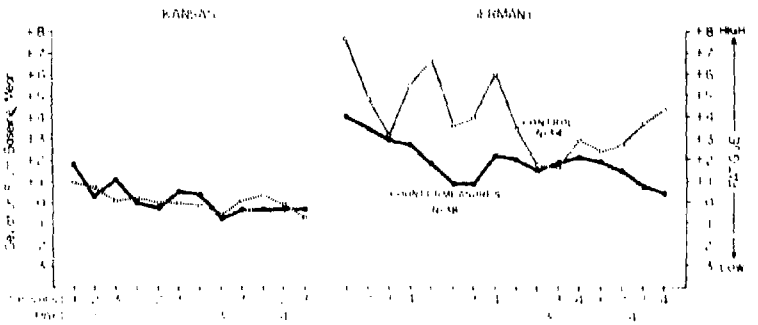


Fig. 9. Effect of CMs on self-rated fatigue in Exp. 2.

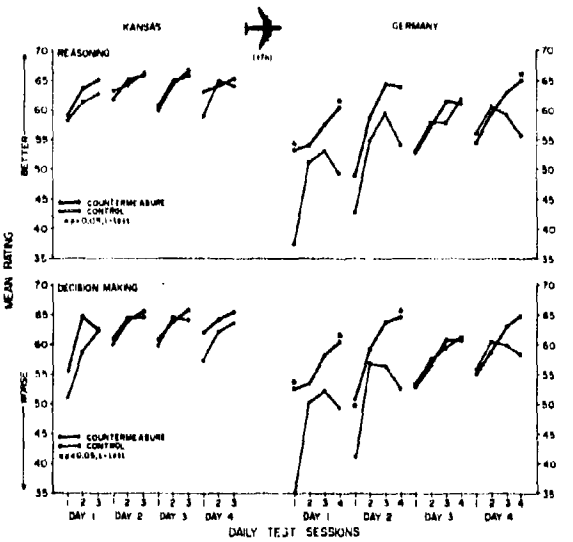



Fig. 10. Influence of CMs on self-rated cognitive abilities.

The number of items correct on the griddle test dropped 9.5% for the CM group and 12.8% for the control group on the first day following deployment, but this difference was not significant (Fig. 11). Gradual recovery occurred over the next three days. Stable accuracy levels were maintained only by the CM group.

In general, no consistent age differences were detected for self-report items or performance scores. Both old and young groups exhibited a post-flight increase in fatigue followed by partial recovery similar to the control group. Also, regardless of location, older subjects slept 20 min. less per day than younger soldiers, but the latter reported consistently lower scores on the cognitive self-rating scales throughout the entire study.

Performance on the cognitive test battery deteriorated 10 to 27% immediately after arrival, with gradual recovery over the next 1 to 4 days depending on the task. The ordering of the severity and length of task disruption corresponds to the estimated difficulty of the tests. Logical reasoning exhibited a decrease of 20% and 27% for young and old respectively, in the mean number of items correct on day 1 in Germany; baseline levels were not regained until day 4. Griddle output decreased 12% after arrival and recovered by day 3, a performance similar to the CM subjects. Accuracy during the first two days was highest in the morning and then declined. This diurnal vari-

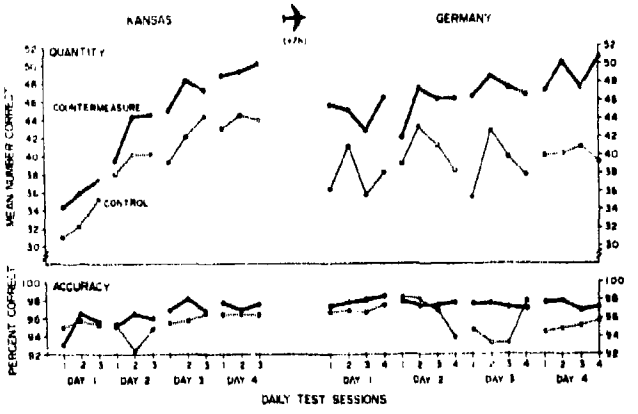



Fig. 11. Effect of CMs on encoding-decoding performance (Control N=34, CM N=38).

ability disappeared as response rate increased, so that accuracy rates displayed stable, pre-flight levels by day 4. Mean word recall dropped one word per test on day 1 and returned to baseline by day 2. No change was discerned for either group in letter cancellation speed or accuracy. The trails task, which requires the connecting of irregularly spaced targets in proper sequence, showed an unexpected post-flight improvement in performance.

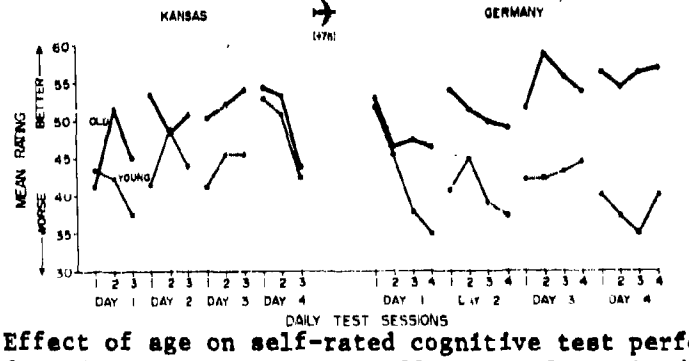



Fig. 12. Effect of age on self-rated cognitive test performance before and after deployment. Groups differ significantly ($p < 0.05$, t-test) on post-flight days (session 2) and 4 (sessions 1,2, and 3).

Older soldiers consistently rated their overall test performance higher than younger troops (Fig. 12). This difference increased in Germany: older subjects' ratings approximately paralleled actual performance recovery, while younger soldiers persistently rated their performance lower than it was. This finding has serious morale implications for young soldiers following deployment.

## DISCUSSION

On the basis of self-reports the CMs reduced jet lag in both studies. Countermeasure subjects slept less on the first two nights in Germany in Experiment 1, where some degree of individual control over sleep time was possible. The sleep duration of about 5.3 hrs. for CM subjects during the first 2 nights seems unusually low, possibly because the testing procedure required awakenings at 0200 and 0600. The lower fatigue scores for these subjects make it unlikely that the sleep result was due to difficulties in falling or staying asleep.

The oral temperature evidence supports the hypothesis that CMs promote more rapid physiological adaptation. An intriguing finding is the maintenance of greater ultradian spectral energy by CM subjects during the first 3 post-flight days. The presence of significant ultradian components strongly suggests an active transitional state wherein the underlying oscillator is readjusting itself to the phase requirements of a shift in the Zeitgeber schedule.

These two studies have been less successful in demonstrating substantial CM effects with actual performance tests. The addition test used in the first study was, in retrospect, insufficiently demanding to produce substantial deficits. As seen in Experiment 2, easier tasks may show little or no impairment following time zone shifts. While accuracy in the griddle task of Experiment 2 was better maintained for CM subjects, the effect was slight, and the major dependent variable, response speed, was unaffected. Interpretation of this result is complicated by the environmental factors, which may have introduced sufficient variance to overwhelm any effect of the CMs.

Despite the limitations of the performance results, the preponderance of evidence warrants the use of the CMs in future military deployments. Although further development is needed to determine which components are particularly efficacious, the current CMs are operationally feasible and potentially valuable in maintaining combat readiness.

The expanded cognitive test battery revealed palpable deficits following the flight, but did not distinguish between older and younger subjects. Several precautions affect any conclusion that age may be irrelevant in determining the effects of rapid transmeridian deployment: (a) The age of the "older" group may have been too low to produce the difficulties typically experienced by older travelers. (b) Test difficulty was targeted toward the high-level cognitive abilities of a HQ unit. The use of troops from a maintenance battalion

may have produced a "floor" effect which reduced the sensitivity of the tasks to flight-induced cognitive deficits. (c) Finally, the adverse environment may have caused excessive variance in the data and lower mean scores throughout the post-deployment observation span.

## REFERENCES

1. Aschoff, J., K. Hoffman, H. Pohl, and R. Wever. 1975. Reentrainment of circadian rhythms after phase-shifts of the zeitgeber. Chronobiologia. 2:23-78.
2. Christie, G. A., and M. Moore-Robinson. 1970. Project Pegasus: some physiological effects of travel across time zones. Clin. Trials J. 7:45-55.
3. Simpson, H. M., N. Bellamy, J. Bohlen, and F. Halberg. 1973. Double blind trial of a possible chronobiotic (Quiadon). Int. J. Chronobiology. 1:287-311.
4. Ehret, C. F., K. R. Groh, and J. C. Meinert. 1978. Circadian dyschronism and chronotypic ecophilia as factors in aging and longevity. Adv. Exp. Med. Biol. 108:185-213.
5. Graeber, R., R. Gatty, F. Halberg, and H. Levine. 1978. Human eating behavior: preferences, consumption patterns, and biorhythms. U.S. Army Natick Res. & Devel. Cmd., Tech. Rep. TR-78-022.
6. Fernstrom, J. D. 1976. The effect of nutritional factors on brain amino acid levels and monoamine synthesis. Fed. Proc. 35:1151-1156.
7. Wurtman, R. J. 1979. When -- and why -- should nutritional state control neurotransmitter synthesis? J. Neural. Transm., Suppl. 15:69-79.
8. Perlow, M. M. Ebert, E. Gordon, M. Ziegler, C. Lake, and T. Chase. 1978. The circadian variation of catecholamine metabolism in the subhuman primate. Brain Res. 139:101-113.
9. Klein, K. E., and H. M. Wegmann. 1974. The resynchronization of human circadian rhythms after transmeridian flights as a result of flight direction and mode of activity. pp. 564-570. In: L. E. Scheving, H. Halberg, and J. E. Pauly, Eds. Chronobiology. Igaku Shoin, Ltd., Tokyo.
10. Pearson, R. B., and G. E. Byars, Jr. 1956. The development and validation of a checklist for measuring subjective fatigue. School of Aviation Medicine, USAF, Randolph AFB, Texas, Report No. 56-115.
11. Walter, D. O. 1968. The method of complex demodulation. EEG Clin. Neurophysiol., Suppl. 27:53-57.

# Army GPS-Doppler Hybrid Navigation System

Jack Gray
US Army Avionics Research and
Development Activity
Fort Monmouth, N.J. 07703

Introduction - In accordance with the Department of the Army's approved Positioning and Navigation System (PANS) Materiel Need, the Army's stated objective is to field "a mix of externally referenced Positioning Subsystems, self-contained subsystems, and hybrid systems to allow for determination of heading, azimuth,and range for navigation and present position". The intent is to avoid losing completely Pos/Nav/orientation capability for critical missions through dependence by the Army upon any one Nav system in the field. More specifically, the fiscal 1980 Scientific and Technical Objectives Guide defines the need to provide superior accuracy and reliable navigation continuity over the battlefield under all visibility, terrain, and weather conditions for such missions as Aerial Scout Helicopter and Advanced Attack Helicopter. To satisfy these requirements, the Army is developing a family of advanced externally referenced and self-contained Pos/Nav systems and the techniques to hybrid these to satisfy mission requirements at lowest cost.

One of these, the Doppler navigator or, ASN-128, is a fully self-contained, EW Secure, sophisticated, dead-reckoning system. However due to residual heading/attitude and velocity errors, its position accuracy degrades as a function of distance traveled. Therefore, mission aircraft which require high accuracy at all times must use a form of position-updated doppler navigator.

The externally-referenced (Satellite-Based) Global Positioning System (GPS) provides high-accuracy, world-wide, position-fixing on a common coordinate system that can be used by all Army elements (ground, air,



PRECEDING PAGE BLANK-NOT FILMED

and marine) so that these units can locate themselves and coordinate tactics effectively. However, GPS User Equipment, as a stand-alone navigation system in a tactical environment, is susceptible to intentional and unintentional electronic jamming and/or interference and to terrain, forest, and buildings masking or attenuating satellite signals. GPS also depends on a fully-functioning, satellite constellation complex. Therefore, mission aircraft which require Pos/Nav continuity (as well as accuracy) on the battlefield require some type of augmentation for GPS.

Since the Doppler is effectively invulnerable to jamming in low-flying aircraft (because its highly directional, narrow-beam antenna will reject all signals not radiated from the small area directly below the aircraft), a hybrid of the GPS and Doppler Navigator can efficiently complement each other to satisfy the requirement of the PANS MN and of the STOG for the critical, continuous Pos/Nav of attack and surveillance aircraft. If the GPS is jammed, the Doppler system is still available and can help the GPS re-acquire its satellites. When the GPS is operative, it continuously keeps the Doppler system initialized to a high position accuracy.

PROGRAM INITIATION - Based on these considerations, an exploratory development program was established to derive the parameters that must be incorporated in a GPS/Doppler Hybrid and to assess feasibility experimentally. Prior to analyzing the GPS/Doppler Hybrid Navigation System, a brief description of each individual system is warranted.

NAVSTAR SYSTEM DESCRIPTION - NAVSTAR GPS is a space-based radio position-fixing and navigation system that has the potential for providing, on a global basis, highly accurate three-dimensional position, velocity, and system time to users equipped with suitable (passive) reveivers. As illustrated in Figure 1 NAVSTAR GPS consists of three major segments; namely, the space system, the control system, and the user system. These are bricfly discussed below.

- Space System · It is planned that the operational space segment will consist of three equi-spaced planes of satellites in circular, 12-hour (-10,000 nmi) orbits inclined approximately 63 degrees to each other. Each orbital plane is to contain eight suitably phased satellites, for a total of 24. Each satellite will transmit a composite waveform consisting of a Protected (P) Signal and a Clear/Acquisition (C/A) Signal in phase quadrature. The P Signal will be used by the "precision" military user and is being designed to resist jamming, spoofing, and multipath and also be deniable to unauthorized users by employing transmission security (TRANSEC) devices. The C/A

Signal will serve as an aid to the acquisition of the P Signal, and will also provide an uncoded (clear) navigation signal to both the military and civil user.

Both the P and C/A Signals are Pseudo-Noise Biphase Shift Keyed (PN/BPSK) continuous sinusoidal carriers, and both signals carry system data. System data will consist of such information as satellite ephemeris, ionospheric propagation corrections, and satellite clock biases. Each space vehicle will be assigned a unique set of pseudo-noise codes of seven days length for a P signal and one msec length for the C/A signal.

The navigation signals are transmitted on two channels; $L_1$ and $L_2$. Channel $L_1$, the Primary Navigation Channel will be 1575.4 MHz and carry the P and C/A Signal, but not simultaneously. System date will always be carried on both channels. The additional $L_2$ signal will permit the high accuracy user to more accurately determine the ionospheric group delay.

The signal waveform is specifically designed to allow system time to be conveniently and directly extracted in terms of standard units of days, hours, minutes, and integer multiples and submultiples of the second.

- Control System - Four widely separated Ground Monitor Stations will passively measure range and velocity time histories of all satellites in view. This information will be processed at the Master Control Station (possibly collated with a Monitor Station) to use in determining satellite ephemerises, clock drifts, electronic delays, etc.. An upload station located in CONUS will transmit the necessary system data corrections via a secure link to the satellites.

- User System - The user equipment consists of a receiver, antenna, data processor, and control and display unit. The receiver will process the signals from four suitably chosen satellites and will measure four independent pseudo-ranges and pseudo-range rates. The processor will then convert these eight independent measurements into three-dimensional position and velocity of the user, and phase and frequency corrections for the user's clock. The process of solving for position is to be carried out in an earth referenced coordinate frame, which would then be converted for display to geographic coordinates (Lat., Long.), UTM grid coordinates, or any other earth grid convenient for the user. The user equipment will also have the capability of accepting waypoint or destination coordinates in the geographic or UTM grids and providing the user with range, bearing, and cross-track error to any of these points.

Velocity accuracy characteristics of the fully developed system are:
Horizontal velocity accuracy referred to Lat/Long. = 0.1 knots (RMS)
Vertical velocity accuracy referred toCenter of Earth = 0.15 knots (RMS)

Position accuracy characteristics of the fully developed system are:
Horizontal position accuracy = 9 meters (RMS)
Vertical position accuracy = 14 meters (RMS)

AN/ASN-128 Doppler Navigation System Description - The AN/ASN-128 Doppler Navigation System is the Army's Standard Airborne Doppler Navigation Set and consists of a Receiver Transmitter Antenna (RTA), Signal Data Connector (SDC), and Computer Display Unit (CDU). Separately input Heading Reference and Attitude Reference are required.

The RTA and SDC constitute the Doppler Radar Velocity Sensor (DRVS) which continuously measures the velocity of the aircraft. The CDU provides control and display functions for the operator and contains the navigation computer. With inputs from external heading and vertical references, the ASN-128 provides accurate aircraft velocity and present position from ground level to altitudes well above 10,000 feet. It is completely self-contained and requires no ground-based aids.

The DRVS determines the three components of aircraft axis-referenced velocity from measurements of the Doppler frequency shift in radar energy transmitted toward and received back from the ground. As soon as prime power is applied to the DRVS, it transmits microwave energy towards the ground in four non-coplanar beams (See Figure 1) and measures the Doppler frequency shift in the back-scattered energy.

The four Doppler frequency shifts (in terms of components along the beam directions) are then sent to the computer. Using these inputs, together with pitch and roll, the CDU computes the three orthogonal components of velocity in aircraft axis-referenced coordinates. Aircraft velocity is then transformed through true heading to give velocities in north and east coordinates which are integrated to obtain changes in present position.

Velocity accuracy characteristics of the system are as follows:

Horizontal velocity accuracy (RMS) = .25% $V_t$ + .1 knots

Vertical velocity accuracy (RMS) =.15% $V_t$ + .1 knots

$G_{RAY}$

where $V_t$ = Vehicle's true velocity

Position CEP (for distances of not less than 10 nautical miles) shall not exceed 2 percent when inputs of attitude and heading meet the limits and accuracy of,

ROLL (RMS) accuracy = $3^{\circ}$ , ROLL Limits = $\pm 45^{\circ}$

PITCH (RMS) accuracy = $3^{\circ}$ , PITCH Limits = $\pm 30^{\circ}$

HEADING (RMS) accuracy = $1^{\circ}$
(MAGNETIC)

Project Plan - The basic approach for evaluating the effectiveness of a system like the GPS/Doppler Navigator must be empirical rather than analytical. The analytical approach is based on the construction of a mathematical model that includes prediction of system characteristics within the constraints imposed by the analyst. Whatever assumptions were made must be tested; predictions must be verified. Therefore, the empirical approach, collecting data and evaluating system effectiveness by observing performance characteristics in the field, must be the primary method.

In analyzing the GPS/Doppler Hybrid Navigation System, there was a wealth of data to show the performance of stand-alone Doppler and GPS navigation systems. Figure 2 shows the results of over-land navigation accuracy tests performed by the Doppler navigator indicating a Circular Probable Error (CEP) for radial position ranges from 1.3 to 2.0 percent of distance traveled.[1] Field test reports concerning GPS position accuracy indicate a Circular Probable Error (CEP of 12 meters.[2] Therefore, by using the improved navigation accuracy of GPS in combination with the Doppler Navigator, direct position updating of the Doppler system can be provided within the accuracy of the GPS.

[1] Research and Development Tech Report, ECOM 4512, Flight Evaluation of the Lightweight Doppler Navigation System, AN/ASN-128 (XE-1) Stanley J. Sokolowski, Avionics Lab., July 1977.

[2] Final User Field Test Report for the Navstar Global Positioning System, Phase I - Major Field Test Objective Report on navigation accuracy: position accuracy: General Dynamics, GPS-GD-025-C-US-7708, 25 June 1979.

Although this is not an optimized approach, it will demonstrate the actual performance gained by augmenting a fielded Doppler system with the currently developed GPS and provide valuable information for the next stage of hybrid development. There will be no filtering or data-smoothing in this first approach to integrating the GPS and Doppler system; in effect, the Doppler System is disregarded at the time of the update and the GPS is assumed to "know" the true position and velocity. Since hybrid system position will be referred throughout to known ground bench marks, the error contribution of both GPS and Doppler navigation to the hybrid will be derived.

The next step, based upon the results of the preceding tests, will be to apply Kalman or least squares filtering which will not only update position but extract system error growth rates. At this stage, the manner in which position/velocity/heading error propagates in the system will be known, and by properly modeling how each error propagates, it will be possible to attribute, after several position/velocity/heading fixes, proper proportions of the total position/velocity/heading error to each modeled error source. If the chosen models are correct, the position/velocity/heading error histories should converge, with time, to very small values. The final value they converge to is a function of the error noise amplitude and feequency content in the Doppler and GPS.

Empirical data for an actual GPS/Doppler Hybrid Navigation system is required, therefore, to verify the expected system performance and to provide a baseline. In order to obtain this data, a Texas Instruments GPS High Dynamics User Equipment was integrated with the AN/ASN-128 Doppler Navigator set through a common ROLM 1650 minicomputer, and tests were run in a mobile van in the Fort Monmouth, N.J. area. Figure 3 shows the GPS/Doppler Hybrid system which was installed in the mobile van.

The GPS user equipment consisted of the following:

Antenna/Preamplifier Assembly - Can receive RF signals from up to five satellites; filters, amplifies, and transmits the signals to the receiver/processor assembly.

Receiver - Consists of five single channel receivers connected to a matrix switch output, and a check module for system timing. Acquires, tracks, demodulates, and performs necessary processing to derive pseudo-range, pseudo-range rate, down-link data and system time from the satellite signals.

Processor - Provides overall GPS Subsystem control and performs navigation calculations.

Instrumentation Interface Unit - Provides intercommunications between the receiver/processor and data acquisition computer. This unit also loads the navigation programs into processor memory.

Control/Display Unit - Provides the human interface and operating mode control functions for overall receiver operation. The unit consists of a multifunction keyboard for receiver mode and navigation display control and alpha-numeric displays for monitoring of navigation parameters.

The Army's AN/ASN-128 Doppler Navigation System operates in conjunction with the Army's Standard AN/ASN-43 Heading Reference Set and MD-1 Attitude Reference Set. The ASN-128 Doppler Radar Navigation Set consists of the following:

Receiver-Transmitter Antenna (RTA) - Transmits RF energy toward the ground in four non-coplanar beams; measures the four Doppler frequency shifts (in terms of components along the beam directions) to the SDC.

Signal Data Converter (SDC) - Accepts heading and vertical reference synchro signals and, along with Doppler beam velocities, transmits serial digital output to the CDU computer.

Computer/Display Unit (CDU) - Accepts from the SDC beam velocities, Heading, Roll and Pitch; performs the Navigation Computations; provides intercommunications between the DRVS and Data Acquisition Equipments.

The Hybrid Navigation Computer consists of:

ROLM 1650 Minicomputer with 32K Core Memory.

GPS/Hybrid Computer Interface.

AN/ASN-128 Doppler Navigator ARINC/Hybrid Computer Interface.

Finally, the Data Acquisition system consists of:

Data Acquisition Computer - Provides overall data acquisition control and the interfaces between the GPS/Doppler and Data Acquisition Subsystems.

Magnetic Tape Unit - Records system test data/parameters and loads data acquisition software programs into the Data Acquisition Computer.

Teletype Unit - Provides for operator control of Data Acquisition Subsystem and, at operator's option, types out all or portions of the system test data/parameters.

Van testing of the GPS/Doppler system not only provides a formidable dynamic environment to ascertain the system's performance, but also is extremely cost effective in comparison to flight tests. The van tests effectively simulate the environment that would be present were the system undergoing a nap-of-the-earth (NOE) flying scenario. Whereas NOE flight involves flight as close to the earth's surface as vegetation or obstacles permit (while generally following the contours of the earth), van tests are, in effect, "flight" on the earth's surface.

As a preliminary to the van testing, a set of detailed 1:24,000 U.S. Geological Survey map for the New Jersey area were used in deriving the latitude/longitude and Universal Transverse Mercator Grid points of easily identifiable landmarks (bridge, intersections, etc.). "Closed loop" navigation courses were selected for these van tests. Checkpoints were chosen such that the landmarks were at least 10 nautical miles apart. During actual van testing, the GPS/Doppler position was recorded as each landmark was passed. Each van run lasted no more than two hours due to the limited GPS satellite constellation visibility over New Jersey. Following each test run, the recorded GPS/Doppler data was reduced, and radial position errors obtained. After the first few runs in the van, it was discovered that GPS positioning data (measured against local bench-mark coordinates) recorded radial position errors ranging from 1 meter to over 200 meters. No particular bias was evident, and re-check of the GPS User equipment in the van along with the Data recording equipment verified these were operating properly. Check with SAMSO revealed one of the satellites had a significant clock malfunction, and another had a clock that was variably questionable.

Additional runs in the van were then made, but, since a maximum of only three satellites could be counted on, local bench-mark altitude was inserted in our GPS User Equipment to overcome the unavailability of the fourth satellite's signals -- and thus allow the GPS set to calculate its best horizontal position. Radial positioning errors were still recorded in a range up to 450 meters. Our GPS set and data recording system were re-re-checked and found operating well. However, further discussions with SAMSO indicated that satellite

drift rates varied also during the time of their transit from the point of ground up-date in the Western United States to the Fort Monmouth area. Also variable was the time elapsed between any satellite's up-date and its observation on the ground at Fort Monmouth:- thus, when the constellation of three useful satellites is over the Fort Monmouth area, one may have been up-dated only 2 hours ago, but another may have been up-dated as much as 4 hours prior to the time of observation.

Still another factor found to influence the GPS Satellite Signal's positioning accuracy is the angular relation of the satellites with respect to the ground observation point. This relationship, referred to as Geometric Dilution of Positioning (GDOP), varies from orbital swing to orbital swing and, aside from the clock situation, itself significantly influences radial position error. With a full schedule of satellites aloft (24), the GDOP factor may be minimized through the ability of a ground receiver to have access to a constellation with a favorable GDOP. But, with only three working satellites, GDOP variation has full impact on positioning capability.

This situation regarding satellite data transmittal to the GPS User Equipment in our van at Fort Monmouth has significantly delayed the rate at which data could be cumulated for this project and has, therefore, prevented the incorporation in this paper of data and results from a set of satisfactorily completed runs. The recent launching of a new GPS satellite containing four re-designed clocks should improve GPS navigation performance and permit more rapid GPS/Doppler hybrid system testing, data collection, and processing.

SYNERGISTIC BENEFIT - In the course of executing the basic project plan to assemble, test, and evaluate GPS/Doppler Navigation Hybrid Techniques aimed at continuous, accurate positioning over the battlefield, an unexpected synergistic bonus was discovered. This bonus was in the form of an idea that occurred that, in addition to improved positioning, the independently measured GPS and Doppler velocity vectors could be used to derive vehicle heading and attitude as well.

GPS/Doppler Hybrid Velocity Heading Reference - It is well known that the overall navigation accuracy of a Doppler Navigation System is limited by the accuracy of the associated heading reference. In fact, with the high accuracy velocity characteristic of the new Army Doppler, the heading reference has become the major contributor of Doppler navigation error.

Two classes of heading references are used: magnetic and inertial. The former sensor is subject to errors due to variations in the earth's magnetic field, especially in the polar regions, as well as local distortions of the magnetic field where it is installed. The latter sensor, since it imploys a gyrocompass to maintain alignment, is subject to drift error.

In attempting to deal with the heading error source of the Doppler Navigator, it was suddenly realized that, by the addition of the GPS system to the Doppler Navigator, a totally unexpected capability was realized. By judiciously combining GPS and Doppler velocities, a true heading reference system can be generated.[3] The coordinate system shown in Figure 4 depicts the applicable geometry needed to derive the GPS/Doppler velocity-derived heading equation, namely,

$$\text{Heading} = \text{TAN}^{-1}\left[\frac{V_H V_E - V_N V_D}{V_N V_H + V_D V_E}\right]$$

where $V_N$, $V_E$ = GPS derived velocities

$V_N$, $V_D$ = Doppler derived velocities

With this GPS/Doppler Heading Reference, a non-magnetic, non-gyro-compass means of providing true heading has been found. It remains to be seen whether the accuracies ultimately possible with this system would justify using it as the primary heading reference for mission aircraft. However, it can be used to improve magnetic-and gyro-compasses via Kalman filtering to help overcome magnetic anomalies and to reduce gyro-compass errors through dynamic calibration/alignments. Also, where primary heading references may be a casualty, there now exists a "fall-back" heading reference.

CONCLUSIONS:

1. GPS-Doppler Navigator field testing in the New Jersey area must continue. The shape and range of the GPS position and velocity errors in this geographic area must be defined (as well as what has influenced these errors). Based on results of such testing, it may be that, from a world-wide deployment point of view under operational conditions, simple position update of a Doppler Navigator by GPS is of insufficient accuracy for attack and other high performance

aircraft. In this event, some form of multi-state Kalman filtering would have to be employed.

2. The experimental exploration of the Hybrid Velocity Derivation of Heading and Attitude, which arose as a synergistic bonus under this program, must continue. Such experiments will help characterize the error budget that applies to and governs the accuracy of the heading, roll, and pitch that are so sensed.

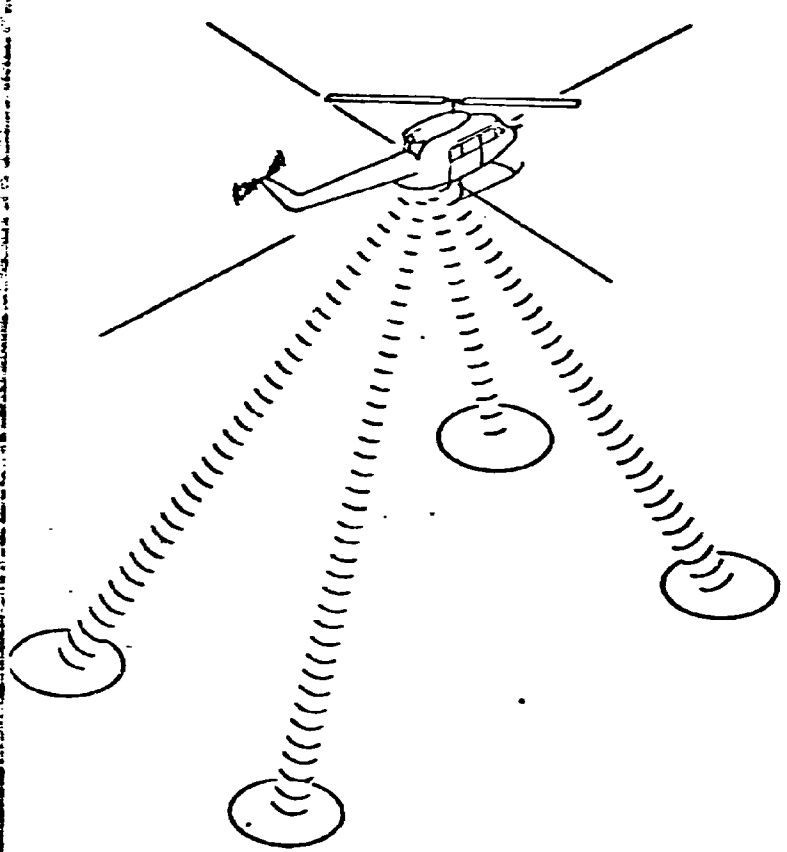
3. Any GPS-Doppler Hybrid Kalman filter should address not only position and velocity improvement but derivation of heading and attitude as well. These derivations then may be used with the outputs of the independent aircraft heading and attitude sensors to improve overall heading and attitude information without paying for more expensive heading and attitude sensors.

4. The principle of graceful degradation for high performance mission aircraft will be enhanced. Not only does this project offer "fall-back" capability in positioning under adverse conditions but, now, "fall-back" in heading and attitude as well.

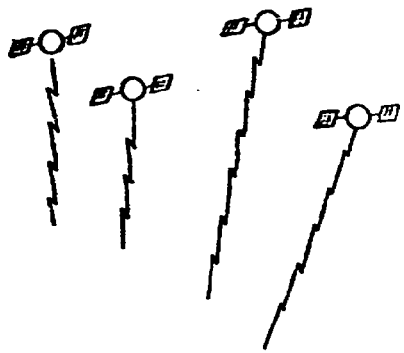
[3]Patent Application, Docket #D-2071, Title: GPS/Doppler Velocity Derived Heading Reference System, Jack Gray

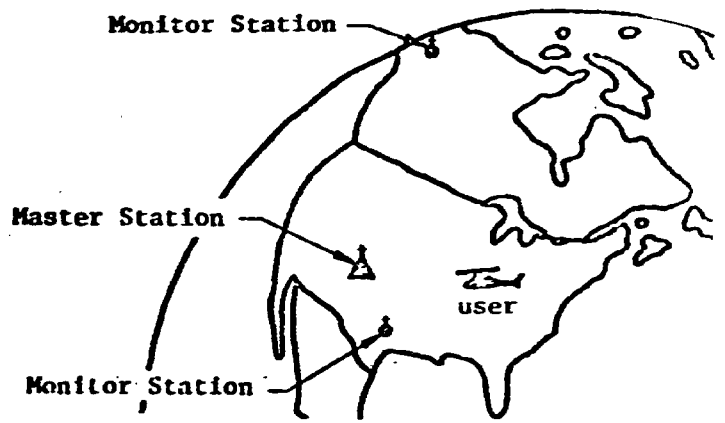
FIGURE 1 STAND ALONE NAVIGATION SYSTEMS

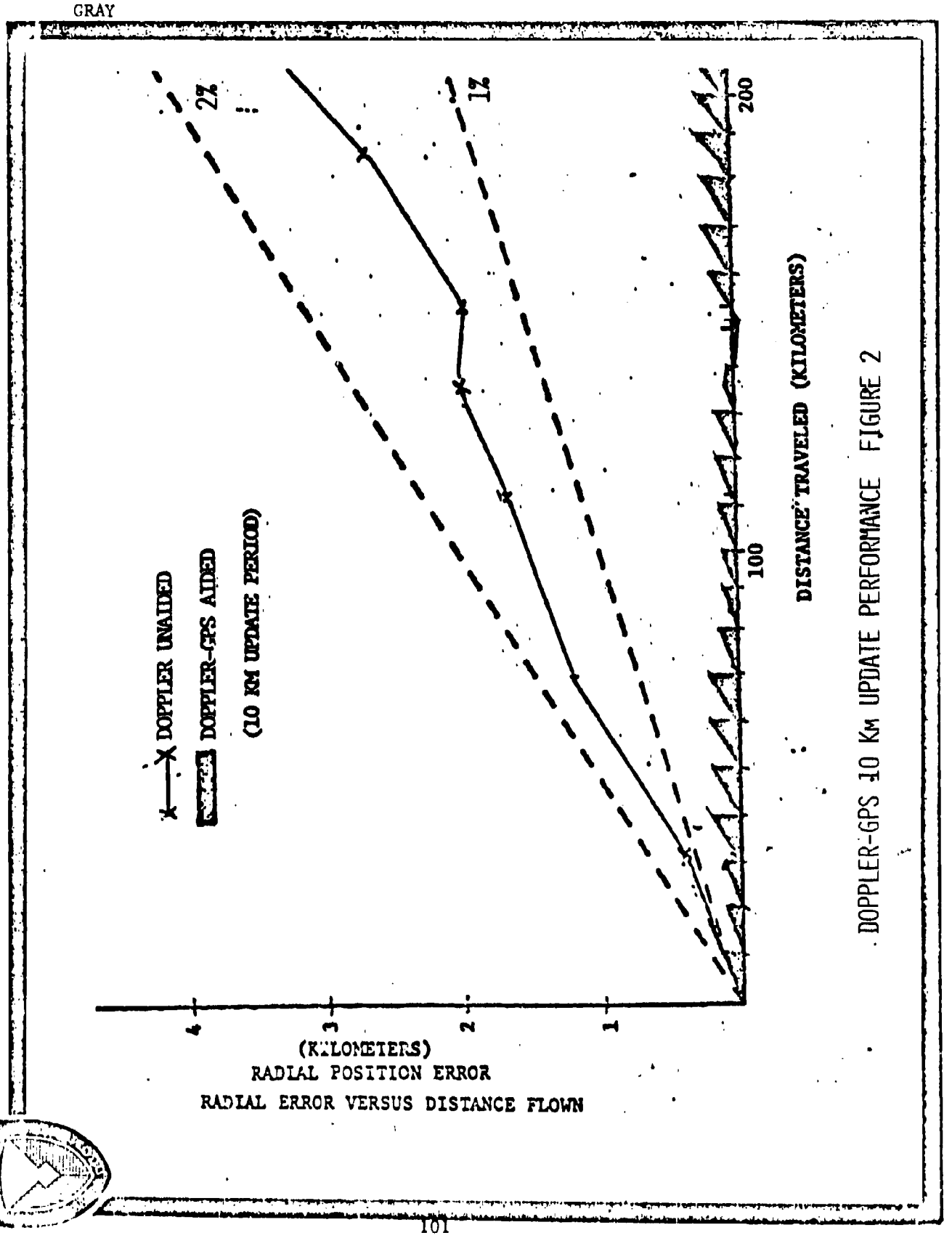
DOPPLER



GPS








DOPPLER-GPS 10 Km UPDATE PERFORMANCE FIGURE 2

# GPS DOPPLER HYBRID NAVIGATION SYSTEM

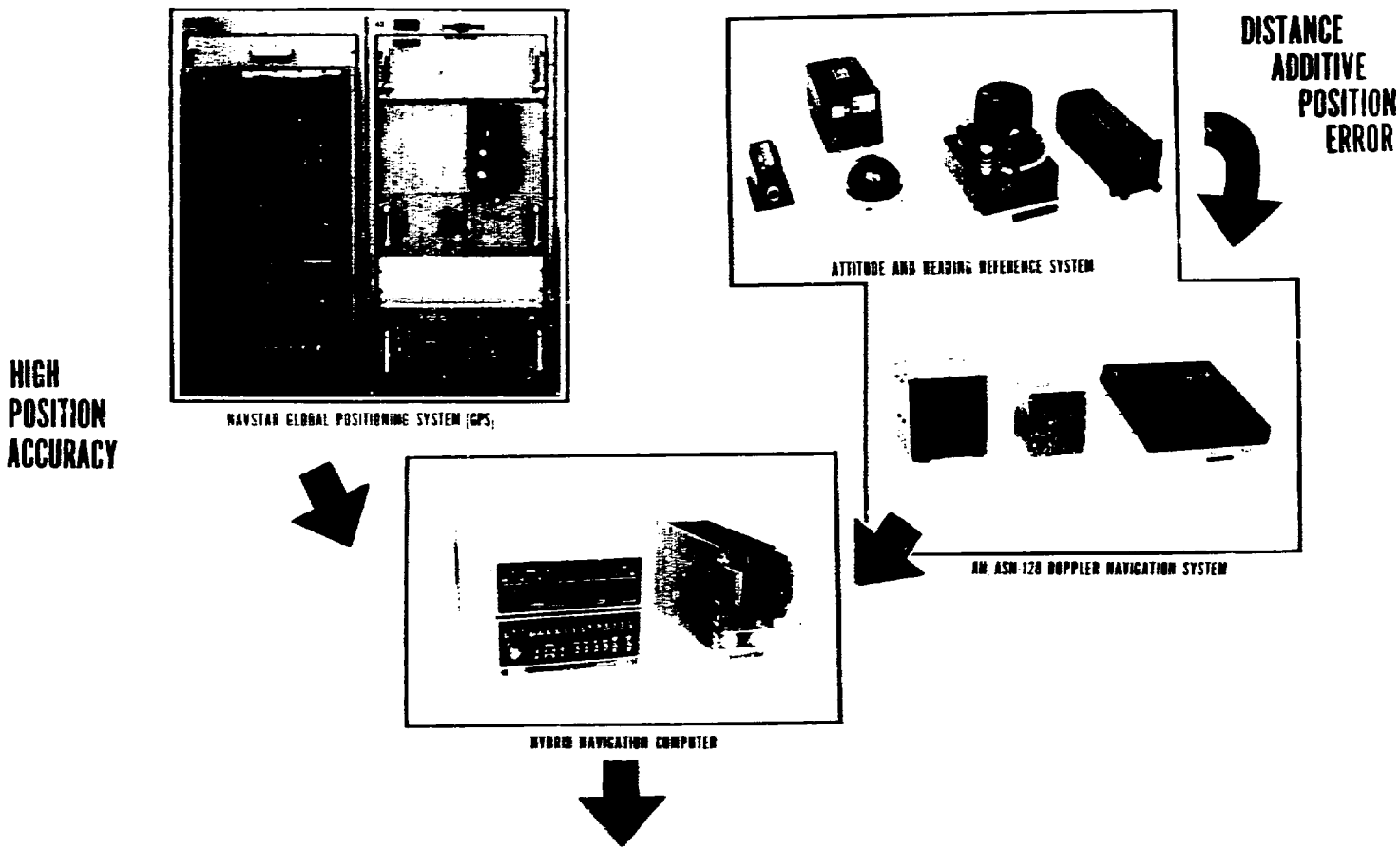



FIGURE 3

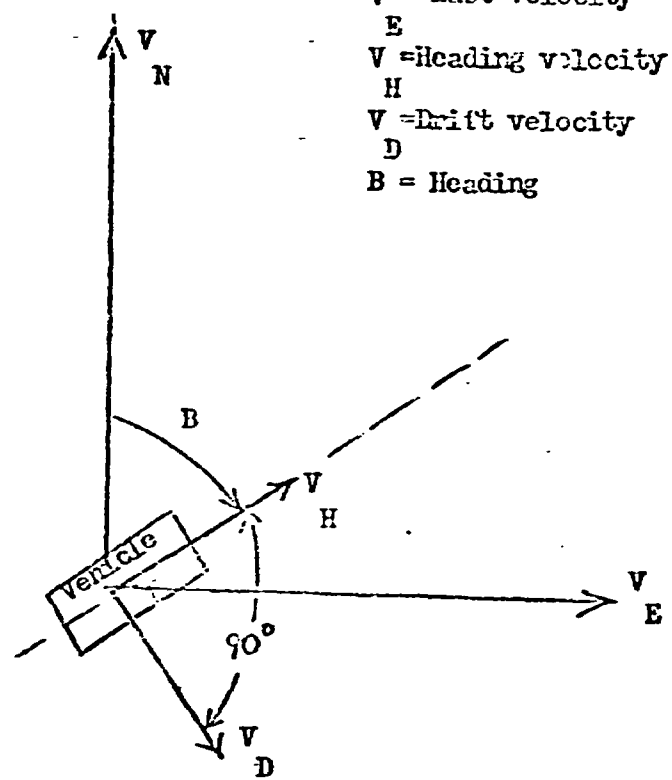
$V_N$ = North velocity

$V_E$ = East velocity

$V_H$ = Heading velocity

$V_D$ = Drift velocity

B = Heading




GPS/DOPPLER HEADING REFERENCE GEOMETRY FIGURE 4

*GROWER, CUTRIGHT and RUSSELL

# REGENERATION OF SURGICALLY EXCISED SEGMENTS OF DOG (U) ESOPHAGUS USING BIODEGRADABLE PLA HOLLOW ORGAN GRAFTS

*MARVIN F. GROWER, LTC, DC
DUANE E. CUTRIGHT, COL, DC and EMERY A. RUSSELL, JR., COL, DC
U. S. ARMY INSTITUTE OF DENTAL RESEARCH
WASHINGTON, D. C. 20012

The widespread multiple site tissue damage to hollow organs which may be produced by the high speed projectiles found in today's battlefield, with the resulting disruption to the vascular supply of the viscera, as well as the need for immediate treatment of the resulting damage, have produced a need for a simple and reliable method for repair of the resulting defects which will yield effective long-term functional results.

The current therapy for repair and replacement of the diseased or avulsed esophagus is by the use of autografts of viscus such as the stomach,(12) the colon,(1) jejunal loops,(3) isolated jejunal segments, (17) or split-thickness skin grafts.(10) None of these procedures produce totally satisfactory results and complications of reconstructive esophageal surgery may include:(14) necrosis of the graft; infection; inadequate blood supply; difficulties in suture retention; leakage at the anastomatic sites; stenosis of the anastomosis between the esophagus and the graft; gastric stasis; reflux; and eating disorders.

The object of this study was to test the feasibility of using a biodegradable polymeric implant constructed from the polymers and copolymers of polylactic acid (PLA) and polyglycolic acid (PGA) to replace an excised segment of the dog esophagus. On a conceptual basis, the use of a biodegradable polymer to fabricate a successful hollow organ graft holds promise, in that, if successful, it would obviate the need for multiple operations and the concern for the vascular integrity of the graft segment. Also, such off-the-shelf grafts would be readily available for immediate repair of a traumatic defect and could be cut to various lengths, depending on the defect to be repaired. Polylactic and polyglycolic acid polymers have been



PRECEDING PAGE BLANK-NOT FILMED

used as implant materials in the orbital floor,(6) as fixation devices for fractures,(13) and as suture material(4) and in these uses have been shown to exhibit minimal inflammatory response and satisfactory healing response at the implant or suture site. (2,4,5, 6,7,8,9,15) The degradation rates of polymers and copolymers of polylactic and polyglycolic acids have been shown to be a function of the type of polymer used as well as different copolymer ratios. (9,15,16)

## MATERIALS AND METHODS

The materials used and the working concentration of the solutions employed to fabricate the esophageal grafts were as follows: (1) 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP), (Eastman Kodak Co.). (2) Methylene chloride, analytical grade (Fisher Scientific and Baker Chemical). (3) Polyglactin 910 (90% polyglycolic acid: 10% polylactic acid copolymer) (Ethicon Corp.), [4 g polyglactin 910/50 ml HFIP]. (4) Polylactic acid (DL-PLA), d,l-polylactide (Southern Research Institute), [10 g DL-PLA/100 ml methylene chloride]. (5) Polylactic acid (60% L: 40% DL-PLA), 60% L-polylactide: 40% d,l-polylactide (Ethicon Corp.), [8 g 60% L: 40% DL-PLA/100 ml methylene chloride]. (6) 50% polylactic acid: 50% polyglycolic acid copolymer (50% PLA: 50% PGA) (Batelle Corp.), [8 g 50% PLA: 50% PGA/100 ml methylene chloride]. (7) The templates for fabrication of the esophageal grafts were Teflon cylinders. The Teflon rods used in the fabrication of Group I implants were 150 mm long x 20 mm o.d., while those for Group II were 145 mm long x 26.2 mm o.d. The Teflon templates were constructed from either solid Teflon rods machined to the desired dimensions or from hollow Teflon tubing which had plugs inserted at both ends to facilitate rotation of the tubes.

The esophageal grafts were prepared by building up a hollow cylinder of polylactic acid polymer fibers and films which were reinforced with polyglactin 910 rings by a process of sequential dipping and spraying PLA solutions on a slowly rotating Teflon template. Table 1 gives an example of the steps used in making the Group I grafts listed in Table 2. The PLA solutions were sprayed with a #152 dental atomizer (DeVilbiss Corp.) using 20-30 psi of nitrogen as a propellant at a distance of 10-18 inches from the rotating rod in a fume hood with an air flow of 150 CFM. Spraying of the polymer solutions listed in the materials section produced fibers of 10-25 microns in diameter and 3-10 cm in length which oriented themselves in a circular manner as they attached to each other around the rotating Teflon mandril. Sterilization of the esophageal implants was done using ethylene oxide for six hours at 55C to 60C followed by aeration for eight hours.

The surgical procedures used in this study were done on twelve mongrel dogs under oroendotracheal nitrous oxide, oxygen, and fluo-

Table 1.

Preparation of Series I Esophageal Grafts

1. Dipped 21 mm o.d. x 150 mm length Teflon mandril in DL-PLA solution and allowed to air dry.

2. One end of the coated rod was placed horizontally in the chuck of a slow speed motor, and the other end was put in a glass supporting tube which allowed for rotation of the rod. The motor was then set to turn at a low speed.

3. Sprayed rotating rod with 110 ml PLA solution (90 ml DL-PLA and 20 ml 60% L:50% DL-PLA).

4. Placed 9 polyglactin 910 rings on the PLA-coated mandril. (Polyglactin 910 rings were prepared by spraying polyglactin 910 dissolved in HFIP on a 22.5 mm o.d. Teflon mandril.) The rings were 5 mm in width and were placed 10 mm from each other in groups of 3 to make segments of 45 mm each.

5. Sprayed rod with 10 ml PLA solution (60% L:40% DL-PLA).

6. Dipped rod in 50% PLA/PGA solution and let air dry.

7. Sprayed rod with 40 ml DL-PLA solution.

8. Dipped rod in 50% PLA-PGA solution and let air dry.

9. Sprayed rod with 55 ml DL-PLA solution and placed in sealed jar which was placed in desiccator overnight.

10. Sprayed rod with 72 ml PLA solution (36 ml DL-PLA and 36 ml 60% L:40% DL-PLA).

11. Polymer coated rod was air dried in hood for 2 hours.

12. Placed polymer coated rod in lyophilizer for 48 hours to remove residual methylene chloride solvent.

13. After lyophilization the polymer implant was cut into 3 sections, removed from the Teflon mandril, and stored in a sealed desiccator prior to use.

thane general anesthesia. By a combination of sharp and blunt dissection the cervical esophagus was exposed, mobilized, and retracted from beneath the trachea. A segment of esophagus measuring approximately five centimeters was removed by sharp dissection after placement of umbilical tape slip ties to control secretions from the proximal and distal ends of the transected esophagus. The PLA graft was then anastomosed to the distal end of the esophagus by use of interrupted 000 Dexon sutures (Ethicon Corp.). The proximal end of the esophagus was then anastomosed to the PLA graft in such a manner as to prevent the esophagus from being twisted upon itself. The umbilical tape was removed from the esophagus and the anastomosis inspected for leakage. The surgical wound was closed in the usual layered manner and the suture line sprayed with Topazone (Eaton Veterinary Labs.).

The animals were given Bicillin 1.2 million units intramuscularly for three days post-operatively followed by Keflex, 250 mg three times a day for a period of one week. To prevent dehydration each animal received twice daily subcutaneous feedings of 750 cc of 5% dextrose and water for three days following surgery. On the fourth post-operative day the dogs were fed a liquid diet of canned dog food emulsified in water and fortified with fat (Pig Kalorie Supplement, Haver-Lockhart Labs.).

One of each of the twelve animals was sacrificed at three and nine days and two each at 14, 21, 30, and 56 days with an overdose of barbiturates. Two of the animals were retained for long-term study and as of the writing of this article, one is 10 months post-operative, while the other is 16 months post-operative.

At sacrifice, the graft sites were immediately removed in a cervical block to include the surrounding tissue and at least 2 cm of normal esophagus at either end of the graft. Excess tissue was trimmed off the specimens and they were placed in buffered 10% Formalin. After fixation, the graft sites were grossed serially into 5 to 8 mm transverse segments and photographed. Tissue sections were then prepared at 6 microns thickness and stained with Hematoylin and Eosin for histology.

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 Committee on the Revision of the Guide for Laboratory Animal Facilities and Care of the Institute of Laboratory Animal Resource, National Research Council.

## RESULTS

Graft Fabrication Table 2 shows the physical dimensions of the biodegradable esophageal grafts that were produced by the sequential spraying and dipping of the Teflon templates with the various forms of polymeric polylactic and polyglycolic acid. Examples of the

Table 2.

Physical Characterization of Biodegradable Esophageal Grafts

| Sample | o.d. mm[a] | Wall thickness mm[b] | Length mm | Wt g |
|---|---|---|---|---|
| | | GROUP I | | |
| 1 | 24.5 | 1.4 | 44 | 3.4 |
| 2 | 24.7 | 1.5 | 48 | 3.9 |
| 3 | 24.7 | 1.6 | 45 | 3.4 |
| 4 | 25.4 | 1.7 | 52 | 4.9 |
| 5 | 25.1 | 1.7 | 38 | 3.4 |
| 6 | 25.1 | 1.7 | 41 | 3.7 |
| | | GROUP II | | |
| 1 | 33.9 | 3.7 | 43 | 8.6 |
| 2 | 33.4 | 3.6 | 39 | 7.2 |
| 3 | 33.7 | 3.9 | 40 | 8.0 |
| 4 | 31.5 | 3.0 | 45 | 6.8 |
| 5 | 32.0 | 3.0 | 45 | 6.8 |
| 6 | 32.4 | 3.2 | 45 | 6.9 |

[a] Outside diameter of samples was measured around the polyglactin reinforcing rings. Internal diameter of Group I implants was 20 mm; Group II implants was 26.2 mm.

[b] Wall thickness of polymer between polyglactin 910 reinforcing rings.

actual grafts produced are presented in Figure 1, A and B. The fabricated grafts were rigid and showed very little tendency to flex. The inner surface of the graft was smooth and tended to be somewhat more solid in nature than the peripheria due to filling up of interfibrillar networks by the 50% PLA/PGA dips used in fabrication of the core portion of the implant. This solid core tended to provide a certain rigidity to the implant. The outer half of the implant was composed of a circular network of PLA fibers 10-25 microns in diameter which could be observed as a series of laminations around the inner core as seen in Figure 1A. This fibrillar coating allowed for the rapid infiltration of fibrin and fibrovascular tissue into the implant which resulted in a water-tight seal. These implants exhibited resistance to flex and collapse; however, the consistency of the wall material was fibrillar enough to allow the needle from a 3-0 Dexon suture (Ethicon Corp.) to be placed completely through its wall (Figure 1B). The sutures placed through the graft wall were retained in position and the wall of the graft did not show any tendency to tear after placement of the sutures. Comparison of the physical properties of Group I implants and Group II implants (Table 2) shows that the Group II implants had an inner diameter 6.2 mm greater than Group I and a wall thickness at least two times as thick. These larger implants were constructed to provide more resistance to lumen collapse which was noted in some two-week specimens, and to produce a larger diameter esophageal replacement which would be more resistant to esophageal stricture during the repair phase of healing. Ethylene oxide sterilization of the implants caused an average 4% decrease in length and a 6% decrease in diameter from the dimensions listed in Table 2. This dimensional shrinkage was also accompanied by a slight increase in flex resistance.

Clinical Findings The post-operative periods progressed uneventfully. There was very little swelling at the surgical site. Endoscopic examination of the grafts *in situ* at three days and eight days showed an unobstructed esophageal graft which was in continuity with the rest of the esophagus. Oral administration of slurried canned dog food was thus begun on day four after initial surgery. By the fourth week after surgery, the dogs showed some difficulty in oral feeding and endoscopic examination showed contraction of the repair tissues present in the graft site. Dilation with a series of metal bougies starting at a French #29 (9.5 mm) and ending at a French #45 (14.8 mm) was begun at this time and the graft sites were dilated biweekly until two weeks prior to sacrifice.

The dog presently surviving 16 months (Group I, #5) was dilated biweekly until six months post-surgery. Some esophageal constriction was seen at this time, but was readily relieved by bougienage therapy and a size #45 French dilator freely passed the length of the esopha-

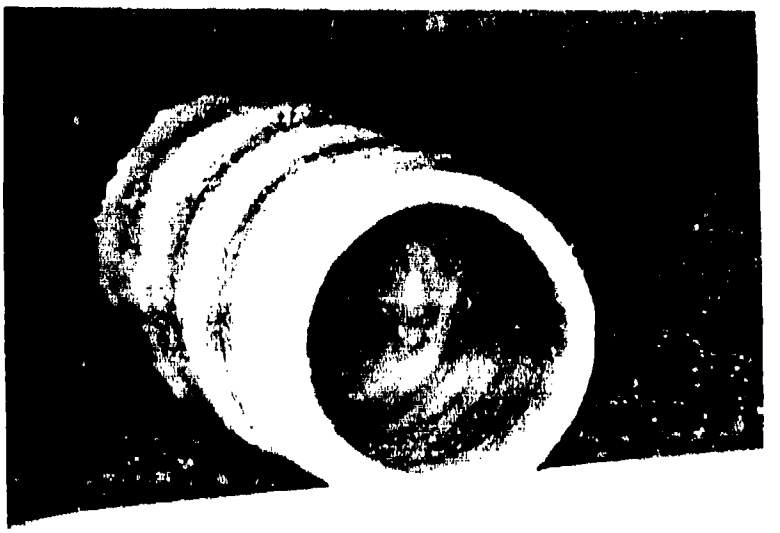


FIGURE 1A. End view of a Group II biodegradable polymer esophageal graft. x1.5



FIGURE 1B. Group I biodegradable polymer graft with 000 Dexon suture placed through wall. x3.0

gus after dilation. This dog was then dilated monthly for the next three months and then one more time two months later. This animal has now gone five months without any further dilation. The dog eats well and has shown an increase in body weight from 45 pounds prior to surgery, to 48 pounds at the present time.

The second surviving dog which had surgery done 10 months ago (Group II, #6) was not dilated until 40 days after surgery and then was dilated monthly for two months up to a size #45 French dilator, after which time it has received no further dilations (over seven months). Endoscopic examination showed that no esophageal constrictions were present and that the epithelium of the graft area was continuous and similar to that of the original esophagus. After an initial weight loss after the esophageal surgery, the dog eventually regained and has maintained his preoperative weight of 40 pounds.

Gross and Histologic Findings The gross morphology of the esophageal grafts after three days of implantation is shown in Figure 2. Gross examination of the implant site did not show any evidence of leakage around the anastomosis between the polymer tube and the esophageal tissue. All the original sutures were present and they showed only minimal inflammatory reaction around them. The implant did not show any evidence of collapse or evidence of loss of structural integrity at this time.

Histologic examination of the original connective tissue wall of the esophagus and the graft itself showed a layer of very early fibroblastic tissue demonstrating hemorrhage with early vascular proliferation. At the interface between the graft and the connective tissue there was evidence of platelet and fibrin accumulation within the interstices of the graft but only very little evidence of organization. The inflammatory response to the graft and sutures was minimal at this time. Although there was no proliferation of the early fibroblastic tissue into the interstices of the graft, there was a definite increase in mitotic activity in the surrounding connective tissues. This was evidenced by an increased number of fibroblasts and small vascular channels proliferating around the edges of the graft.

Nine days after surgery the interface between the connective tissue and the graft showed a marked maturation of the fibrous connective tissue and ingrowth of the vasofibroblastic response into the interstices of the graft material. The thickness of the connective tissue wall measured from the outer surface of the tracheal cartilage rings was 2 mm and the tracheal wall did not show any evidence of erosion due to presence of the graft. Histologically there could be seen a beginning orientation of the connective tissue fibers in a circular manner around the graft site. There was very little tendency to form giant cells surrounding the polymer fibers.




FIGURE 3. Gross section of a 21 day esophageal implant (cross section). (w) regenerated esophageal wall, (l) lumen of esophagus, (i) residual implant material, (t) trachea.




FIGURE 2. Three day gross specimen of the anterior anastomosis of the esophagus with the biodegradable polymer graft (longitudional section). (e) anterior esophagus, (a) anastomosis site, (g) main body of polymer implant.

The interface between the graft and the original connective tissue showed an ingrowth of 1.5 mm of fibrous connective tissue into the interstices of the graft itself with minimal inflammation present. At the margins or at the interface between the graft and the esophageal epithelium there was a proliferation of new epithelium at least 3 mm down over the graft site.

Examination of the 14 day samples revealed a vasoblastic connective tissue wall of approximately 3 mm which surrounded the esophageal implant. In certain areas of the connective tissue graft interface there was some mild multinucleated giant cell formation where the polymer had been incorporated into the tissue and this extended back about 1.5 mm from the apparent edge of the graft itself, indicating that the vasofibroblastic ingrowth had penetrated at least 1.5 to 2 mm into the graft. In addition, there was minimal evidence of collagen organization into both circular and longitudinal bands in the implant area as well as epithelial ingrowth down the edges of the graft itself.

Gross examination of the 21-day samples, as seen in Figure 3, showed that there was a vasofibroblastic wall of about 5 mm (measured from the outer tracheal ring) which had formed both as a result of tissue growth around the implant and growth into the implant itself of about 2.5 mm. Histologically, the advancing front of granulation tissue into the graft itself was characterized by occasional multinucleated giant cells which were shown to be phagocytosing the polymer after it had been hydrolyzed. The more solid inner portions of the implants containing the polyglactin 910 supporting rings appeared to have become delaminated from the outer portions which were penetrated by granulation tissue ingrowth. Portions of this residual graft material could be seen in some sections as seen in Figure 3, although in other areas this material was absent. The collagen fibers present in the new esophageal wall showed a variable orientation and there was variable epithelial migration up to 5 mm from the anastomosis along the inner aspect of the graft site.

By eight weeks the graft appeared to be almost completely resorbed and a collagenous tube with a wall thickness up to 5 mm measured from the tracheal rings was present, as seen in Figure 4. In addition, it was evident that the esophageal implants had not affected or caused any erosion of the trachea. The lumen of the esophagus was open although some contraction of the repair tissue was evident as indicated by the narrowing of the esophageal lumen and there was no macroscopic evidence of graft material. Although on microscopic examination (Figure 5) some evidence of residual polymer between the collagen fibers appeared to be present. The Dexon sutures, on the other hand, appeared to have become completely degraded. Only minimal inflammation and slight edema were present at the graft site and the collagen showed a variable orientation with a

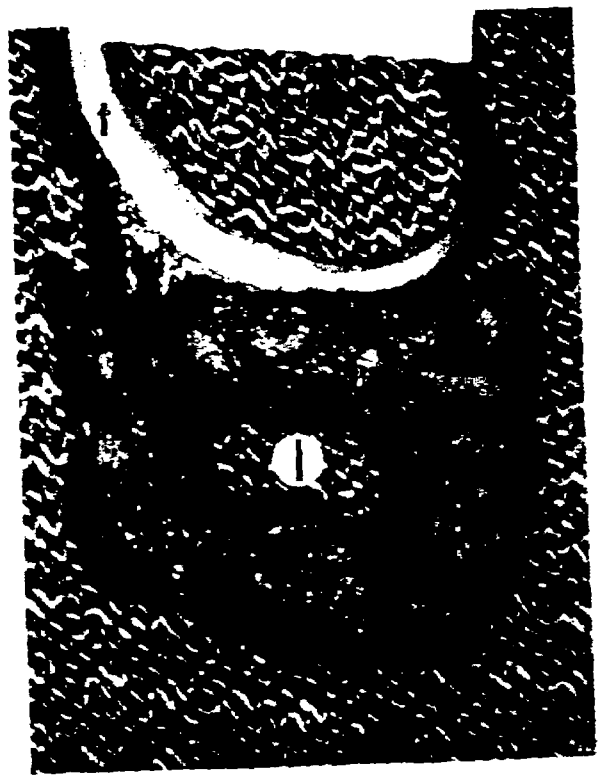



FIGURE 4. Gross section of 8 week esophageal implant (cross section). (w) regenerated esophageal wall, (l) lumen of esophagus, (t) trachea.




FIGURE 5. Histologic section of the 8 week esophageal graft at the graft-esophagus interface.
(e) epithelium of esophagus, (eg) epithelium growing on luminal aspect of graft, (c) collagenous repair tissue, (p) residual polymer, (r) giant cell reaction, (l) lumen of esophagus. x48

small amount of circular orientation. There were islands and sheets of epithelium growing over the surface of the implant throughout much of the graft lumen as seen in Figure 5, although the lumen was not completely epithelialized.

## DISCUSSION

This study demonstrated the feasibility of using a biodegradable hollow organ implant fabricated from polylactic acid and polyglycolic acid polymers to regenerate an excised segment of the dog esophagus. The use of an inner core fabricated from polyglactin 910 rings, 100% PLA fibers and laminations of 50% PLA/50% PGA provided for rigidity of the implant. The outer covering of PLA fibers provided for strength and allowed for a vasofibroblastic ingrowth into the graft as well as promoting growth of a collagenous sheath of tissue around the peripheria of the graft. The fabrication of 100% PLA into 10-25 microfibers, as was done in this study, allowed degradation of some polymer in as little as 14 to 21 days (Figure 3) by hydrolysis of tissue fluids and cellular action; although some polymer appeared to be retained for up to 8 weeks (Figure 5). This resorption time is in contrast to the time required for solid PLA plates. Plates are partially resorbed in six weeks and total resorption occurs in six months.(13,16)

The fibrillar esophageal grafts which were constructed in this study were strong, easily handled, and could be trimmed to any required size. They were resistant to leakage, retained sutures well, and showed minimal inflammatory reaction at the anastomotic sites. The ability of these implants to undergo sterilization by ethylene oxide would enable them to be routinely stocked and sterilized in a hospital setting. The minimal shrinkage seen in the implants with ethylene oxide sterilization at 55-60C was probably due to some dimensional change exhibited by the d,l-polylactide fibers (15); however, the use of L-polylactide fibers and polyglactin 910 supporting rings tends to inhibit these dimensional changes.

The PLA polymeric material appeared to have a favorable response on the healing reactions seen as complete repair of the graft site by a collagenous sheath connecting the two ends of the esophagus was achieved by two weeks. Histologically, this repair was by a dense hyalinized type of connective tissue in sheath form.

The connective tissue sheath allowed the animal to eat freely with minimal discomfort until the fourth week after surgery, when some contraction of the new segments was noted. This constriction was readily relieved by dilation of the esophagus with metal bougies. Similar problems may also be seen with autogenous viscus replacements (14) and the course of bougienage therapy used on the long-term surviving dogs is similar to the therapeutic regimens used on human patients with constrictive esophagitis.(11)

The ability of the two surviving dogs to go without dilation for over 5 to 7 months suggests that maturation of the collagen fibers making up the repair may have occurred. The greater resistance to esophageal contraction shown by the dog receiving Graft II, #6 (Table 2) suggests that the wider and thicker grafts used in Series II may provide more optimal healing stimuli than the narrower and thinner grafts of Series I. Dilation of the grafts caused only minimal trauma to the lining epithelium, which rapidly healed. The esophageal grafts in these two long-term animals have also apparently resulted in a functionally effective repair which allowed adequate nutritional support of the dogs, as evidenced by the maintenance and/or increases seen in their initial body weights. This compares favorably with results achieved using viscus grafts in which patients report difficulties in eating and slow weight gain for 12 to 18 months or longer. (14)

In conclusion, it has been shown in this study, for the first time, that: (1) It is possible to construct a completely biodegradable off-the-shelf graft which can replace lost segments of hollow organs; (2) Regeneration of the hollow organ occurred by a new tube of connective tissue lined by epithelium utilizing the technique of neogenesis within a biodegradable polymer-copolymer framework; (3) The replacement supported the dogs' nutritional intake requirements; (4) The replacement showed adequate strength and allowed for maintenance of esophageal diameter by bougienage therapy; (5) It did not appear to exhibit problems such as the need for multiple operations, leakage at anastomosis sites, lack of blood supply, and postoperative infections seen with other therapeutic procedures in current use; and (6) Grafts made of spun biodegradable PLA and PGA polymers meet the requirements for an effective and easy-to-use replacement for traumatized or lost hollow organs which are encountered in combat military surgery, and should be studied further.

## REFERENCES

1. Belsey, R. Reconstruction of the Esophagus with Left Colon. J. Thorac. Cardiovasc. Surg. 49:33-55, 1965.

2. Brady, J.M., Cutright, D.E., Miller, R.A., Battistone, G.C., and Hunsuck, E.E. Resorption Rate, Route of Elimination, and Ultrastructure of the Implant Site of Polylactic Acid in the Abdominal Wall of the Rat. J. Biomed. Mater. Res. 7:153, 1973.

3. Brain, R.H.F. Steatorrhea in Oesophagastric Surgical Practice. Proc. R. Soc. Med. 46:438, 1953.

4. Conn, J.C., Oyasu, R. Welsh, M., and Beal, J.M. Vicryl (Polyglactin 910) Synthetic Absorbable Sutures. Am. J. Surg. 128:19-23, 1974.

5. Cutright, D.E. and Hunsuck, E.E. Tissue Reactions to the Biodegradable Polylactic Acid Suture. Oral Surg. 31:134-139, 1971.

6. Cutright, D.E. and Hunsuck, E.E. The Repair of Fractures of the Orbital Floor Using Biodegradable Polylactic Acid. Oral Surg. 33:28-34, 1972.
7. Cutright, D.E., Beaseley, J.D., and Perez, B. Histologic Comparison of Polylactic and Polyglycolic Acid Sutures. Oral Surg. 32: 165-173, 1971.
8. Cutright, D.E., Hunsuck, E.E., and Brady, J.M. Fracture Reduction Using a Biodegradable Material. Oral Surg. 29:393-397, 1979.
9. Cutright, D.E., Perez, B., Beasley, J., Larson, W.J., and Posey, W.R. Degradation Rates of Polymers and Copolymers of Polylactic and Polyglycolic Acids. Oral Surg. 37:142-152, 1974.
10. Edgerton, M.T. One Stage Reconstruction of Cervical Esophagus or Trachea. Surgery 31:239, 1952.
11. Flood, C.A. Bougienage Therapy for Constrictive Esophagitis. Gastrointest. Endosc. 25:130-132, 1979.
12. Gavriliu, D. The Long-Term Clinical State After Resection with Jejunal Replacement. IN Surgery of the Oesophagus, Ed. Smith, R.A. and Smith, R.E. London. Butterworth. pp 29-33, 1975.
13. Getter, L., Cutright, D.E., Bhaskar, S.N., and Augsburg, J.K. A Biodegradable Intraosseous Appliance in the Treatment of Mandibular Fractures. J. Oral Surg. 30:344-348, 1972.
14. Gunning, A.J. and Marshall, R. The Oesophagus. Part II: Replacement of the Oesophagus. Clin. Gastroenterol. 8:292-304, 1979.
15. Kulkarni, R.K., Moore, E.G., Hegyeli, A.F., and Leonard, F. Biodegradable Poly(lactic Acid) Polymers. J. Biomed. Mater. Res. 5: 169-181, 1971.
16. Miller, R.A., Brady, J.M., and Cutright, D.E. Degradation Rates of Oral Resorbable Implants (Polylactates and Polyglycolates): Rate Modification and Changes in PLA/PGA Copolymer Ratios. J. Biomed. Mater. Res. 11:711-719, 1977.
17. Seidenberg, B., Rosenak, S.S., Hurwitt, E.S., and Som, M.L. Immediate Reconstruction of the Cervical Esophagus by a Revascularized Isolated Jejunal Segment. Ann. Surg. 149:162-171, 1959.

# THREE-DIMENSIONAL FINITE ELEMENT COMBUSTION INSTABILITY ANALYSIS

ROBERT M. HACKETT, PhD
PROPULSION DIRECTORATE, US ARMY MISSILE LABORATORY
REDSTONE ARSENAL, ALABAMA 35809

## INTRODUCTION

The phenomenon of oscillatory combustion instability in solid rocket motors results from the responsiveness of the combustion process to oscillations in the flow environment. Because of the high probability of combustion instability in low signature motors and the attendant jeopardy to successful motor and system performance, the capability of instability prediction is of unquestioned importance. The state-of-the-art in the field of linear analysis of combustion instability is based on a perturbation of the acoustic field in the burning propellant and an evaluation of the growth/decay coefficient associated with the acoustic pressure waves; a positive coefficient indicates an amplification of the waves and, therefore, instability, and a negative value implies attenuation of the waves and stability.

In early 1975 the development of a Standardized Stability Prediction Method for Solid Rocket Motors [1] was begun by Aerojet Solid Propulsion Company under contract with the Air Force. It was decided that this method would employ the NASTRAN finite element fluid analysis option which was developed for rocket acoustic cavity analysis [2,3,4,5]. The NASTRAN finite element analysis is axisymmetric and approximates the effect of radial slots (Figure 1) on the cavity acoustics. As long as the slot width is relatively narrow, the approximation provides an accurate model of the cavity acoustic response, but as the slot width increases, the accuracy diminishes. The NASTRAN option as used in the standardized method does not provide for the coupling of the vibratory response of the solid propellant grain to that of the acoustic cavity and hence does not provide a means of evaluating the damping of the acoustic oscillations by the propellant grain response. An additional limiting feature of the

standardized method is that it employs a post processor, separate from the other analyses, for evaluating the growth/decay coefficient.

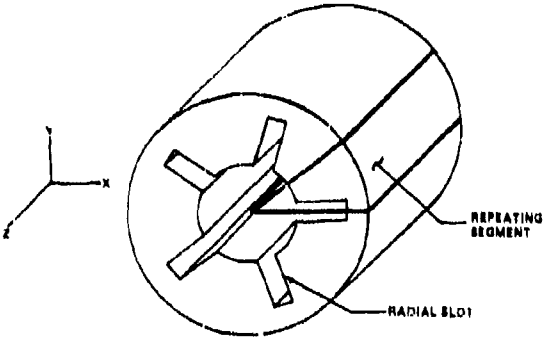



Figure 1. Three-dimensional cavity-solid propellant model.

The three-dimensional analysis package presented herein was developed primarily to provide 1) more generality, and therefore more accuracy in modeling complex cavity geometries, 2) a means of predicting the damping of acoustic oscillations by the solid propellant grain, 3) an integrated program designed solely for the purpose of combustion instability prediction with ease of use, and 4) a means of coupling the combustion instability analysis to an internal ballistics analysis. The developed package possesses all of these positive features as well as certain inherent, somewhat unappealing features associated with "bigness" which will be pointed out later in the paper.

The main aspects of the developed three-dimensional code are listed as follows and will be elaborated upon subsequently in the paper:

1) It utilizes a three-dimensional finite element mesh generator especially adapted to provide input to the program.

2) It utilizes the principle of dihedral symmetry which enables a consideration of only the smallest repeating geometrical segment.

3) It couples the response of the gas cavity with that of the solid propellant grain to enable the calculation of the frequency of the coupled system and the damping provided by the propellant grain.

4) It provides for modeling the propellant grain as a nearly incompressible material (which differs from the common minimum potential energy formulation).

5) It utilizes the principle of condensation, wherein, in this case, the fluid pressure degrees-of-freedom are designated "master" and the solid propellant displacement and mean pressure parameter degrees-of-freedom are designated "slave." This enables a major reduction in the size of the problem; the number of equations is reduced from the total number of degrees-of-freedom of the coupled system to the number of fluid pressure degrees-of-freedom.

6) It provides the option of considering the response of the gas cavity alone (which models the cavity boundaries as being "acoustically hard"). This option might be utilized in certain cases where a savings in computer costs or storage is a dominant consideration. In this case the condensation approach previously described obviously would not be taken.

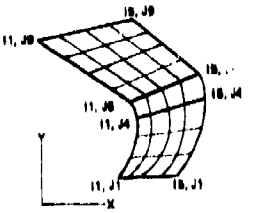
7) It calculates the three-dimensional potential flow field.

8) It evaluates the stability integrals for the calculation of the net driving/damping coefficient for each mode.

## FINITE ELEMENT MESH GENERATION

It is only necessary to develop a finite element mesh for one repeating segment (Figure 1) of the total cavity-grain rocket geometry. This is true because of the employment of the principle of dihedral symmetry in the program. Although the three-dimensional element used in the program for both the cavity region and the propellant grain is a tetrahedron [6], the mesh is that of bricks, each brick, or quasi-hexahedron, being comprised of five basic tetrahedra. The breakdown of the quasi-hexahedra into tetrahedra is performed internally. The mesh generator is an efficient routine which automatically creates the complete finite element mesh, for both cavity and solid propellant regions, from a minimal amount of input. Each repeating segment is sectioned in the longitudinal or z-direction, with each section comprised of a number of quadrilateral parts which are identified by a counterclockwise listing of their part boundary curves.

Part boundary curves may be ellipses as well as straight lines and their points of intersection are designated by I,J indices as shown in Figure 2.




PART 1 (CAVITY)

I1, J1 TO I5, J1 (CURVE 1)
I5, J1 TO I5, J4 (CURVE 2)
I5, J4 TO I1, J4 (CURVE 3)
I1, J4 TO I1, J1 (CURVE 4)

PART 2 (CAVITY-SOLID PROPELLANT INTERFACE)

I1, J4 TO I5, J4 (CURVE 3)
I5, J4 TO I5, J5 (CURVE 2)
I5, J5 TO I1, J5 (CURVE 5)
I1, J5 TO I1, J4 (CURVE 4)

PART 3 (SOLID PROPELLANT)

I1, J5 TO I5, J5 (CURVE 5)
I5, J5 TO I5, J9 (CURVE 6)
I5, J9 TO I1, J9 (CURVE 7)
I1, J9 TO I1, J5 (CURVE 8)

Figure 2. Finite element mesh definition for cavity-solid propellant.

By coupling the output of a standard internal ballistics analysis computer code with the mesh generator one can obtain a combustion instability analysis at any stage of performance, i.e., at any instant from ignition of the propellant to burnout. This has been done through the definition of additional part boundary curves, which conform to the geometry of the different zones of burning for the differing cavity-grain configurations. With the transfer of a small number of geometric parameters from the output of the internal ballistics code to the mesh generator, a complete combustion instability analysis can be initiated and automatically performed. The designation of these parameters and a detailed description of the development of the coupling program are found in References [7] and [8]. Presently the coupling of an internal ballistics analysis with the combustion instability analysis is limited to the consideration of axisymmetric and star cross-section geometries, but increased part boundary curve definition is a matter of expansion of the program, not of additional development.

## DIHEDRAL SYMMETRY

If a geometrically defined body is made up of identical segments symmetrically arranged with respect to an axis, the degrees-of-freedom for a finite element analysis can be transformed into uncoupled symmetrical components, thereby greatly reducing the number of equations which must be solved simultaneously [9]. A further reduction can be effected if each segment has a plane of reflective symmetry. Dihedral symmetry is the term applied to this latter condition. It can be seen from Figure 1 that a typical solid rocket geometry meets this requirement and, therefore, the principle of dihedral symmetry can be employed in a three-dimensional (cylindrical coordinate) analysis. The application of the principle to this problem is explained in detail in Reference [10] and will not be reproduced here, but the resulting analysis will be discussed.

Employing the principle of dihedral symmetry, the program computes three distinct types of acoustic harmonics: the zero harmonic, the K harmonics, and the N/2 harmonic. The zero harmonic exists for all cases. The number of possible K harmonics is given by

$$K = 1, \ldots, L \tag{1}$$

where the harmonic index L is given by

$$L = \frac{N - 1}{2} \text{ (if N is an odd number)} \tag{2a}$$

$$= \frac{N - 2}{2} \text{ (if N is an even number)} \tag{2b}$$

where N is the number of radial slots or lobes. The N/2 harmonic exists only when N is even. Referring to the geometry of Figure 1, then, one could calculate the zero, first, and second harmonics, the latter two being K harmonics. The longitudinal modes associated with each harmonic are calculated as requested by the user. The result of this calculation (eigensolution) is the natural circular frequency associated with each acoustic mode and the corresponding acoustic pressure distribution (normalized acoustic pressure at each finite element nodal point) for the smallest repeating segment. The pressure distribution for each of the other segments is then simply calculated automatically through the dihedral transformation matrix. The acoustic velocity components (constant for the region occupied by each cavity tetrahedral element) are computed from the acoustic pressure nodal point values for those cavity elements which are adjacent to the propellant grain.

The theoretical formulation of the complete three-dimensional finite element analysis, in which the natural circular frequency, the acoustic pressure distribution, and the element acoustic velocities are calculated, is given in Reference [11] and will not be repeated here. Reference [12] also presents the theoretical finite element formulation which was used in the development of this program.

## COUPLED RESPONSE

The presence of the solid propellant grain can significantly shift the acoustic system frequency from that of the gas phase alone, a portion of the acoustic energy being dissipated by the deformable solid material. This effect can be one of the more significant sinks for acoustic energy in both large and small rocket motors, the amount of damping depending on the grain geometry and mechanical properties and on the acoustic mode shape and natural frequency.

In order to evaluate the coupled cavity-solid propellant grain response it is necessary to model, by the finite element method, both cavity and grain. This, of course, greatly increases the size of the problem to be solved, from the standpoint of number of initial degrees-of-freedom. The coupled finite element formulation of Reference [12], which is coded in the three-dimensional program, is expressed in matrix form by:

$$\left( \begin{bmatrix} F & 0 \\ -U^T & K \end{bmatrix} - \lambda^2 \begin{bmatrix} C & U \\ 0 & \begin{matrix} M & 0 \\ 0 & 0 \end{matrix} \end{bmatrix} \right) \begin{Bmatrix} p \\ \Delta \\ H \end{Bmatrix} = 0 \qquad (3)$$

where [F] is the fluid inertia matrix, [C] is the fluid compressibility matrix, [K] is the solid stiffness matrix, [M] is the solid consistent mass matrix, [U] is the matrix which couples acoustic pressure degrees-of-freedom to solid displacement degrees-of-freedom, {p} is the acoustic pressure vector, {Δ} is the solid displacement vector, {H} is the mean pressure parameter vector (to be elaborated upon later), and $\lambda^2$ is the eigenvalue of the coupled system.

The structural damping can be attributed to the out-of-phase response of the solid propellant grain, which is measured in terms of the complex shearing modulus of the grain, which, in turn, results in a complex eigenvalue for the coupled system. The imaginary part of the complex eigenvalue obtained from the eigensolution is the natural circular frequency of the coupled system while the real part is the structural damping rate.

Since the complex shearing modulus is frequency dependent, a series of iterations may be necessary before the accurate value of complex modulus for input into the program is determined.

## PROPELLANT GRAIN MODELING

Since the propellant grain is only accurately modeled as a nearly incompressible material, the well-known standard Navier displacement formulation, in conjunction with the Ritz procedure, would lead to inaccuracies in the finite element modeling of the grain. In order to avoid this situation the solid finite element formulation utilized in this program is that of a linear displacement-linear mean pressure tetrahedron [13]. It is similar to the Herrmann variational formulation [14], which employs a linear displacement function and a constant mean pressure function. The finite element modeling of the propellant grain used in this program is outlined in detail in Reference [12] and will not be repeated here.

## EIGENVALUE ECONOMIZER - CONDENSATION

The extraction of eigenvalues and eigenvectors is a much more expensive operation than is the solution of simultaneous linear equations. It requires roughly twice as much time to extract a single eigenvalue as to do a single "static" analysis. In order to reduce or "condense" the number of degrees-of-freedom in the eigensolution, the following technique is utilized in the three-dimensional program. Further details of the method are found in Reference [15].

The original formulation of the coupled system is given by Equation 3, where the number of degrees-of-freedom is equal to the number of cavity nodal point pressures, plus the number of solid propellant nodal point displacement components, plus the number of solid propellant nodal point mean pressure parameter values (one at each propellant grain nodal point), for the analyzed repeating segment. The condensed formulation is given by:

$$\left( \left[ F_r \right] - \lambda^2 \left[ C_r \right] \right) \left\{ p \right\} = 0 \tag{4}$$

where

$$\left[ F_r \right] = \left[ \begin{array}{c|c} I & \\ \hline K^{-1}U^T & \end{array} \right]^T \left[ \begin{array}{c|c} F & 0 \\ \hline -U^T & K \end{array} \right] \left[ \begin{array}{c|c} I & \\ \hline K^{-1}U^T & \end{array} \right] \tag{5a}$$

and
$$\left[C_r\right] = \left[\begin{array}{c} I \\ \hline K^{-1}U^T \end{array}\right]^T \left[\begin{array}{c|c} C & U \\ \hline O & \begin{array}{c|c} M & O \\ \hline O & O \end{array} \end{array}\right] \left[\begin{array}{c} I \\ \hline K^{-1}U^T \end{array}\right] \tag{5b}$$

where I is the identity matrix. The relationship between initial and reduced degrees-of-freedom is

$$\left\{\begin{array}{c} p \\ \hline \Delta \\ \hline H \end{array}\right\} = \left[\begin{array}{c} I \\ \hline K^{-1}U^T \end{array}\right] \left\{p\right\} . \tag{6}$$

## UNCOUPLED RESPONSE

Although one of the most important features of the three-dimensional program is the coupled cavity-propellant response, the option of a cavity analysis alone is available to the user. In this case the cavity-solid propellant interface is modeled as an "acoustically hard" boundary. The cavity only option would greatly reduce the solution time and cost and would, in some cases, perhaps suffice. The user need make only two simple adjustments to the input data, and these adjustments are described in the user's guide to the three-dimensional code [16].

## POTENTIAL FLOW CALCULATION

A separate operation is carried out in the sub-routine which performs a potential flow analysis for the purpose of determining the mean flow field in the rocket cavity. As in the case of the eigensolution, only the smallest repeating geometrical segment need be considered, and, for this calculation, only the cavity portion of the segment with the proper boundary conditions. The same general formulation of the finite element model equations of motion is utilized except that, in this case, the fluid is considered to be incompressible. The mass flow into the cavity from the burning propellant surface is modeled as a cavity-solid propellant interface nodal point quantity. It is calculated by summing the interface surface areas associated with each nodal point lying on the cavity-solid propellant interface. The solution of the resulting set of linear equations for the mean flow velocity components (constant for each cavity tetrahedron) is explained in Reference [11] and will not be discussed here.

## EVALUATION OF STABILITY INTEGRALS

The final operation consists primarily of the evaluation of the stability integrals associated with the various driving/damping coupling mechanisms which occur in the cavity chamber in the presence of combustion and flow. The stability integrals presently incorporated into the program are those derived in Reference [17] for the three-dimensional case, along with the flow-turning formulation also found in Reference [17]. The use of linear pressure (and therefore constant velocity) tetrahedral elements to represent the cavity region enables an explicit evaluation of the stability integrals, given the acoustic nodal point pressures and element velocities and the mean flow element velocities from the finite element solutions. At present the code does not contain a routine for evaluating nozzle damping or particulate damping. These additional calculations will be added.

The evaluation of stability integrals is for the purpose of calculating the driving/damping coefficient, $\alpha$, a fact well-known to the combustion community. A positive $\alpha$ indicates a stable mode. The net value of $\alpha$ computed by the three-dimensional code is a summation of the computed values of $\alpha_{PC}$ (pressure coupling), $\alpha_{VC}$ (velocity coupling), $\alpha_{FT}$ (flow-turning), and $\alpha_{SD}$ (structural damping). The value of $\alpha_{SD}$ is obtained from the complex eigensolution described in an earlier section; the other three $\alpha$-values are obtained from the evaluation of the stability integrals. It is known that the pressure coupling mechanism always drives the acoustic oscillations, that the velocity coupling mechanism may either drive or damp the oscillations, and that the flow-turning mechanism always damps the oscillations. Response factors [18,19] are input into the program as multiples of the stability integrals for the calculation of the $\alpha$'s obtained from the different coupling mechanisms. These propellant grain-dependent response functions are obtained from other analyses and utilized as direct input into the three-dimensional program.

## COMPARISON WITH ANALYTICAL SOLUTIONS AND EXPERIMENTAL OBSERVATIONS

In order to affirm the accuracy and usefulness of any numerical analysis package it is necessary to make comparisons of results obtained numerically with available analytical closed-form solutions and experimentally obtained results. In the case of a complex analysis procedure such as that of combustion instability, available bases of comparison are limited in both regards. The following results are presented to support the contention of accuracy and utility of the developed package.

Analytical Comparisons. A right circular cylindrical cavity 254mm long and 254mm in diameter was modeled using an 18° repeating segment consisting of 32 quasi-hexahedral elements, 4 in the radial direction and 8 in the longitudinal direction. Acoustic frequencies corresponding to the first and second logitudinal modes (1L and 2L), the first tangential mode (1T), the first tangential-first longitudinal mode (1T-1L), and the first radial mode (1R) were solved for and are compared with the corresponding analytically obtained frequencies of the cavity. The values are shown in Table 1.

TABLE 1. NUMERICALLY AND ANALYTICALLY CALCULATED ACOUSTIC FREQUENCIES FOR CYLINDRICAL CAVITY

| Mode | Frequency (Hz) | |
|---|---|---|
| | Analytical Solution | Finite Element Solution |
| 1L | 2000 | 1988 |
| 2L | 4000 | 4042 |
| 1T | 2400 | 2378 |
| 1T-1L | 3124 | 3151 |
| 1R | 4994 | 5033 |

As described earlier, a separate potential flow calculation is made in the combustion instability analysis package for the purpose of evaluating the stability integrals. A comparison of the finite element potential flow solution with the analytical solution for the cylindrical cavity is shown in Figure 3.

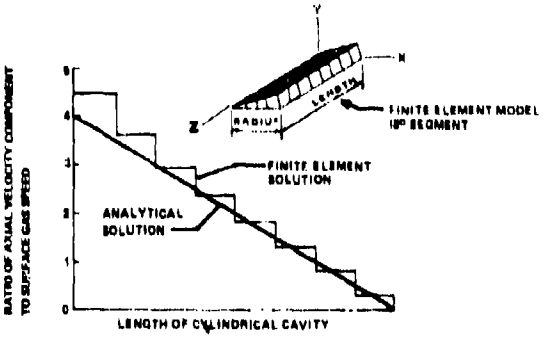



Figure 3. Comparison of finite element and analytical potential flow field solutions for a circular cylinder.

The accuracy of the finite element frequency calculations is seen in Table 1. These results are highly favorable, especially when considering the relatively small number of elements used to model the cavity. The somewhat less satisfying agreement in the case of the finite element potential flow solution can be attributed to the relatively small number of elements, especially near the exhaust end of the cavity where the axial velocity component has its largest value. Additional comparisons with closed-form solutions of the same type have yielded results similar to those cited here.

Experimental Comparisons. The experimental data used for comparison purposes were obtained from Hercules, Inc., Allegany Ballistics Laboratory (ABL) in Cumberland Maryland and result from testing of an early experimental model of the Smokeless CHAPARRAL Motor Assembly, conducted in early 1977. The finocyl geometry of the pulsed motor and the finite element model consisted of a small diameter-to-length ratio cylindrical cavity having four radial slots over approximately one-third of the cavity length. Experimental longitudinal mode frequency and stability data can be compared to corresponding finite element analysis results in Table 2. The test data and numerical results found in Table 2 correspond to a burn distance of 8.03mm. The finite element analyses were performed on the CDC 6600 computer at Redstone Arsenal. The lack of close agreement observed when comparing corresponding net $\alpha$-values can to some extent be attributed to the use of somewhat inaccurate values of response factors, which are presently considered in the program as properties of the propellant, but which are also flow-dependent. However, the qualitatively good agreement between corresponding values can be observed.

TABLE 2. SMOKELESS CHAPARRAL DATA AND FINITE ELEMENT ANALYSIS RESULTS

| Test Data | | Finite Element Analysis | |
|---|---|---|---|
| Frequency (Hz) | $\alpha$ ($sec^{-1}$) | Frequency (Hz) | $\alpha$ ($sec^{-1}$) |
| 323 | -94 | | |
| 290 | -41 | 307 | -198 |
| 644 | -266 | | |
| 640 | -290 | 593 | -220 |
| 909 | -192 | | |
| 927 | -145 | 909 | -148 |

The results from additional finite element analyses of the early experimental model of the Smokeless CHAPARRAL Motor Assembly are shown in Table 3. The most notable aspect of this set of finite element analysis results is the prediction of instability in some of the tangential modes (positive values of $\alpha$). This compares quite favorably with ABL tests of the same assembly, which exhibited clear instabilities in tangential modes. Also, ABL test results showed 5-6 separate tangential modes evident in the frequency range of 9000-20,000 Hz.

TABLE 3. HIGHER ACOUSTIC MODE FINITE ELEMENT ANALYSIS RESULTS FOR SMOKELESS CHAPARRAL MOTOR

| Harmonic | Mode | Frequency (Hz) | $\alpha$ ($sec^{-1}$) |
|---|---|---|---|
| K (=1) | 1 | 7992 | -147 |
| | 2 | 13,476 | + 33 |
| | 3 | 15,303 | - 55 |
| | 4 | 15,915 | + 46 |
| | 5 | 17,907 | - 92 |
| N/2 (=2) | 1 | 8981 | -232 |
| | 2 | 15,948 | +162 |
| | 3 | 20,516 | -119 |
| | 4 | 23,693 | - 62 |
| | 5 | 25,644 | - 85 |

The comparisons between numerical and experimental results, which can be made from Tables 2 and 3, would appear to support the contention that the developed finite element package can be of valuable use in predicting the instability of solid rocket systems.

## CONCLUSIONS

The developed computer package performs a linear analysis of the irrotational motions of an inviscid, compressible fluid coupled to the motions of a nearly incompressible, linearly viscoelastic solid, performs a linear potential flow analysis of the irrotational motions of an inviscid, incompressible fluid, and then determines the effect of the flow field and of combustion on the calculated acoustic oscillations. There are obvious limitations attached to any code which is as basic as the restrictions listed above dictate, but it is felt that the developed code presented herein is probably as sophisticated as the present state-of-the-art warrants. It is viewed as having much potential as both a design and a research tool. As the state-of-the-art in combustion technology advances, it is felt that

the code can be relatively easily revised and updated to include the new technology; at least it was designed with that provision in mind.

One of the most attractive features of the code is the ease of use. Other extremely important attributes are the fact that it is three-dimensional, that it performs a coupled cavity-solid propellant analysis (or, alternatively, a cavity-only analysis), and that all analyses are automatically linked. Features of the code which do not enhance its reputation also exist, and they too should be pointed out. It is a large program requiring a large amount of storage and it may require long run times, as is the case with any three-dimensional finite element program. Presently, the entire program is in-core computation, but this will, in all probability, be modified. The eigenvalue routine used in the analysis may not be the most efficient one available. This possibility is currently being investigated, and if a more efficient routine can be found, it will be used in place of the one presently employed. In certain instances a two-dimensional (axisymmetric) uncoupled analysis provides sufficient accuracy, and for such cases use of the three-dimensional code might not have merit.

It is felt that the demonstrated attributes of the three-dimensional code far outweigh any foreseen disadvantages, and that it can provide the means of performing important, and heretofore impossible, analyses.

## REFERENCES


1. Standardized Stability Prediction Method for Solid Rocket Motors, Vol. II (Multidimensional Computer Program Users Manual, by D. P. Dudley and R. D. Waugh) and Vol. III (Axial Mode Computer Program Users Manual, by R. L. Lovine and R. C. Waugh), Aerojet Propulsion Company, Sacramento, California, February 1976.
2. Herting, D. N., "Application Manual for NASTRAN Acoustic Cavity Analysis," The MacNeal-Schwendler Corporation, Report MS 116-2, May 1971.
3. Herting, D. N., J. A. Joseph, L. R. Kuusinen and R. H. MacNeal, "Acoustic Analysis of Solid Rocket Motor Cavities by a Finite Element Method," The MacNeal-Schwendler Corporation, Report MS 116-3.
4. NASTRAN Cyclic Symmetry Capability User's Guide, The MacNeal-Schwendler Corporation, EC-180, July 1972.
5. Analysis of a Solid Rocket Motor Cavity, The MacNeal-Schwendler Corporation, Report MS 200, November 22, 1972.
6. Zienkiewicz, O. C., The Finite Element Method in Engineering Science, McGraw-Hill, London, 1971.

7. Hackett, R. M., "A Coupled Interior Ballistics Finite Element Combustion Instability Analysis Procedure," Technical Report T-78-72, US Army Missile Research and Development Command, Redstone Arsenal, Alabama, July 1978.
8. Hackett, R. M., "A Coupled Interior Ballistics Finite Element Combustion Instability Analysis Procedure - Part II," Technical Report T-79-68, US Army Missile Command, Redstone Arsenal, Alabama, June 1979.
9. MacNeal, R. H., R. H. Harder and J. B. Mason, "NASTRAN Cyclic Symmetry Capability," TM X-2893 NASTRAN: Users' Experiences, NASA, September 1973, pp. 395-421.
10. Hackett, R. M., "Three-Dimensional Accustic Analysis of Solid Rocket Motor Cavities," Technical Report RK-76-7, US Army Missile Command, Redstone Arsenal, Alabama, November 1975.
11. Hackett, R. M., "Three-Dimensional Finite Element Acoustic Analysis of Solid Rocket Motor Cavities," Journal of Spacecraft and Rockets, v 13, n 10, October 1976, pp. 585-588.
12. Hackett, R. M., and R. S. Juruf, "Three-Dimensional Finite Element Analysis of Acoustic Instability of Solid Propellant Rocket Motors," Proceedings of the 13th Annual Meeting of the Society of Engineering Science, Inc., Hampton, Virginia, November 1976, pp. 1641-1651.
13. Hughes, T. J. R. and H. Allik, "Finite Elements for Compressible and Incompressible Continua," Proceedings of the Symposium on Application of Finite Element Methods in Civil Engineering, Vanderbilt University, Nashville, Tennessee, November 1969, pp. 27-62.
14. Herrmann, L. R., "Elasticity Equations for Incompressible and Nearly Incompressible Materials by a Variational Theorem," Journal of the American Institute of Aeronautics and Astronautics, v 3, n 10, October 1965, pp. 1896-1900.
15. Cook, R. D., Concepts and Applications of Finite Element Analysis, John Wiley & Sons, Inc., 1974.
16. Hackett, R. M., "User's Manual for FLAP3," Technical Report TK-77-4, US Army Missile Research and Development Command, Redstone Arsenal, Alabama, July 1977.
17. Final Report - T-Burner Testing of Metallized Solid Propellants, Report No. AFRPL-TR-74-28 (F.E.C. Culick - Editor) Air Force Rocket Propulsion Laboratory, Edwards, California, October 1974.
18. Hart, R. W. and F. T. McClure, "Combustion Instability: Acoustic Interaction with a Burning Propellant Surface," Journal of Chemical Physics, v 30, June 1959, pp. 1501-1514.
19. Price, E. W., "Experimental Solid Rocket Combustion Instability," Tenth Symposium (International) on Combustion, The Combustion Institute, 1965, pp. 1067-1082.

# EXPERIMENTAL CHEMOTHERAPY: A RAPID AND SIMPLE SCREENING METHOD FOR DRUG BINDING TO DNA

FRED E. HAHN, PH. D.
WALTER REED ARMY INSTITUTE OF RESEARCH
WASHINGTON, D.C. 20012

## I. BACKGROUND

Histories of wars and of infectious diseases have been interwoven since the time of antiquity. In World War II, some 500,000 American servicemen acquired malaria with an attending loss of 6.6 millions of man days. During 1965, the number of American soldiers evacuated from Vietnam because of chloroquine-resistant malaria, equalled the number evacuated because of wounds (1). The invasion of Taiwan from mainland China, planned in 1949, had to be abandoned because of a catastrophic outbreak of schistosomiasis which the assembled troops acquired while practicing landing maneuvers on inland lakes in Fukien province. Earlier, the campaign of Napoleon in Egypt faltered because of schistosomiasis and trachoma in the expeditionary force.

Drugs for the treatment of those communicable diseases against which there exists no effective immunoprophylaxis are a military necessity when troops must be deployed in unsanitary parts of the world. The Russian Civil War (1917-1924) was accompanied by 25,000,000 cases of epidemic typhus. Today, such patients would be treated successfully with chloramphenicol or tetracyclines.

In the preparation for warfare, the development of chemotherapeutic drugs is an absolute necessity. Search for, and development of, such drugs still employs empirical methods. The discovery of lead compounds and their successful molecular modification would be facilitated by the introduction of exact scientific pre-screens. This article describes such a screening method.

## II. SCIENTIFIC INTRODUCTION

Numerous drugs against parasitic diseases, such as malaria, schistosomiasis, trypanosomiasis, and leishmaniasis form complexes with DNA (Table 1). DNA-binding compounds are among inhibitors of nucleic acid polymerase reactions of DNA viruses and retroviruses (Table 2) and furnish leads to the development of virus chemotherapy (2). The largest number of DNA-complexing drugs is found among anticancer compounds (Table 3). DNA-complexing compounds also eliminate drug-resistance plasmids from cultures of multiresistant bacteria and, hence, restore sensitivity of such organisms to antibiotic and synthetic drugs (3,4). The agents in Tables 1 - 3 act as DNA template poisons that inhibit the replication of DNA and/or the transcription of RNA from DNA.

Table 1. DNA-Complexing Antiparasitic Drugs

| | |
|---|---|
| Acriflavine | Hydroxystilbamidine |
| Berberine | Miracil D |
| Berenil | Pentamidine |
| Chloroquine | Propamidine |
| Congocidine | Quinine |
| Ethidium Bromide | Quinacrine |
| Hycanthone | Stilbamidine |

Table 2. DNA-Complexing Compounds with Antiviral Action

| | |
|---|---|
| Distamycin A | Nitroacridine C-283 |
| Ethidium Bromide | Tilorone |

Table 3. DNA-Complexing Antitumor Compounds

| | |
|---|---|
| Adriamycin | Echinomycin |
| Anthramycin | Ellipticine |
| Bleomycin | Mithramycin |
| Chromomycin | Nitroacridine C-283 |
| Daunomycin | Nogalamycin |
| Dichloro-diammino platinum | Olivomycin |
| Distamycin A | Sibiromycin |

DNA-binding compounds can be designed (5). They are also found among antibiotics (Table 3) as well as among empirically discovered synthetic drugs (Table 1). Since the ability of a compound to form a molecular complex with DNA signals potential chemotherapeutic

activity, it follows that an <u>in vitro</u> pre-screen for DNA binding will be a useful step in experimental chemotherapy programs.

The subject of this article is a simple, rapid, and inexpensive screening procedure for drug binding to DNA. It yields numerical results which are proportional to biochemical and biological activities of the tested compounds. The method is based upon the displacement of methyl green from its complex with duplex DNA. It requires only a simple spectrophotometer and the (commercially available) methyl green-DNA reagent.

## III. METHYL GREEN AND ITS COMPLEX WITH DNA

Methyl green, MG, is a basic triphenylmethane dye (Fig. 1). It is used as a histochemical stain for DNA. The DNA-MG reagent is an experimental substrate for the determination of deoxyribonucleases (6, 7). The first qualitative observation of a displacement of MG from DNA by drugs was made for chloroquine and quinacrine (8).

The use of the DNA-MG complex in displacement analysis is based upon the fact that free MG in polar solutions of pH >5 undergoes spontaneous molecular rearrangement to its colorless carbinol base (9) so that the liberation of the dye from DNA can be followed spectrophotometrically as an exponential decrease in absorbance at 640 nm with a first order rate constant of 0.65 x $hr^{-1}$ (10). In its complex with DNA, MG is stabilized against molecular rearrangement and, hence, maintains its color.

In order to understand the molecular mechanism of the displacement of MG, its complex with DNA has been studied (10). MG binds to double-stranded (calf thymus) DNA with a stoichiometry of one dye mole-

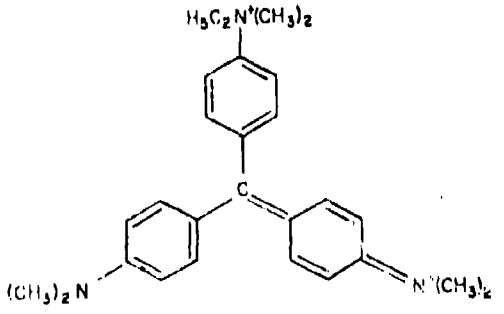
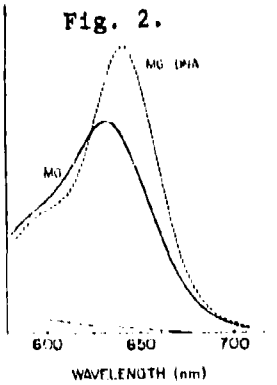



Fig. 1. METHYL GREEN

Fig. 2.



cule per 13 nucleotides (9). This is attended by a bathochromic shift of 10 nm in the absorption maximum of MG and a hyperchromic increase of 40 per cent in this maximum (Fig. 2). Bathochromic shifts indicate the binding of single dye molecules rather than of dimers or aggregates (11). Hyperchromicity suggests peripheral binding to DNA rather than insertion between base pairs (intercalation) in the interior of the double helix: intercalation places chromophores into a less polar environment and, thus, produces hypochromic changes in their absorption spectra.

The naturally single-stranded DNA of ϕX174 also causes the bathochromic shift in the absorption spectrum of MG but does not produce hyperchromicity. After ϕX174 DNA and MG are combined in solution, the absorption of the dye decreases with a time course similar to that of free MG. A small portion of MG remains stably bound to this single-stranded DNA (10).

The synthetic duplex DNA-like polymer, poly[d(A-T)], causes the same manifestations as DNA in the absorption spectrum of MG and also stabilizes the dye. In contrast, poly(dG-dC) produces a slight hypochromic change at 640 nm, followed by slow decolorization with an absolute endpoint of 33 per cent of MG remaining stably bound to the polymer (10).

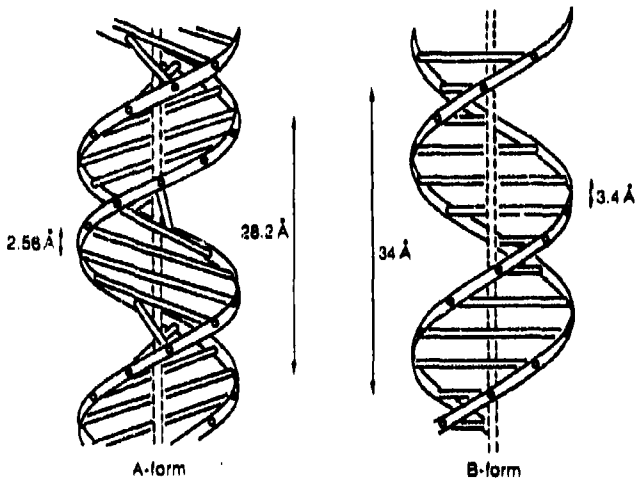
Besides by base composition and the homopolymeric nature of its component strands, poly(dG-dC) differs from DNA and poly[d(A-T)] by its preference for the A-conformation, although variable portions coexist in the B-form (12). Fig. 3 depicts the B-conformation of DNA in aqueous, and the A-conformation in dehydrated, environments.

Stable binding of MG requires double-stranded DNA in the B-conformation. This is proved by observations that ethanol at graded concentrations releases MG from DNA. Effects of ethanol on DNA in solution have been extensively studied*.

---

*From 20 vol per cent, ethanol begins to denature DNA as indicated by a sharp drop in viscosity which decreases further with increasing alcohol concentrations (13). Up to 40 vol per cent, ethanol decreases the median denaturation temperature, $T_m$, of DNA, indicating progressive denaturation; but above 40 vol per cent ethanol, $T_m$ increases again owing to the beginning of the B → A transition (14). At 65-78 vol per cent ethanol, DNA exhibits (by circular dichroism)the conformational change from the B- to the A-conformation (15). In 100 per cent ethanol, DNA is completely denatured (16).

FIG. 3. Schemes of the A and B forms of DNA.



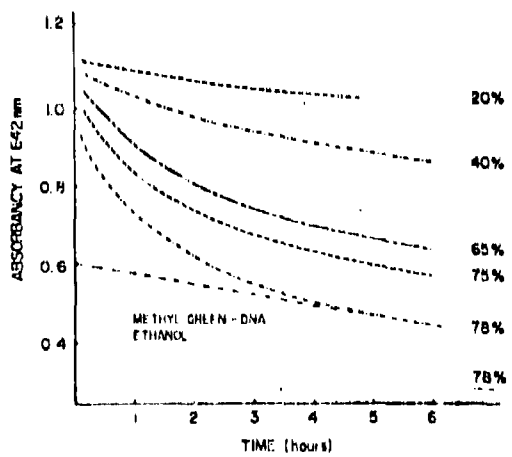




FIGURE 4 Effect of ethanol on the Methyl Green-DNA complex. Time courses for the release of Methyl Green from its complex with DNA were measured at different volume percents of ethanol; for 78% alcohol the extrapolation to zero time of the linear portion of the release kinetics is shown. Concentrations: 106 μg/ml of MG-DNA complex.

The effects of ethanol from 20 to 78 vol per cent on the DNA-MG complex are shown in Fig. 4. Increasing ethanol concentrations liberate MG at increasing rates whose initial curvatures show transition to linearity beyond 4.5 hrs (i.e. the time required to decolorize free MG). When the curved portions of the release kinetics were corrected by subtracting the contribution of the linear process, it was discovered that the course of MG liberation was a composite of the linear rate and of an initial burst of release of the dye, followed by a first order decay such as is seen for free MG. Apparently, the slow linear liberation of the dye was the result of progressive denaturation of DNA, while the sudden release of the bulk of MG was caused by the B → A transition of DNA. MG was not extracted from DNA by ethanol: placing a sample of the solid DNA-MG compound into 100 per cent ethanol did not decolorize it upon prolonged exposure.

The forces that bind MG to DNA are predominantly electrostatic: $Mg^{2+}$ at concentrations from $10^{-3}$ to 5 $x10^{-2}$ displaces MG from DNA at rates and to endpoints shown in Fig. 5. In contrast, urea at 6 M (having a pH of 7.2) displaces only 38 per cent of MG from DNA.

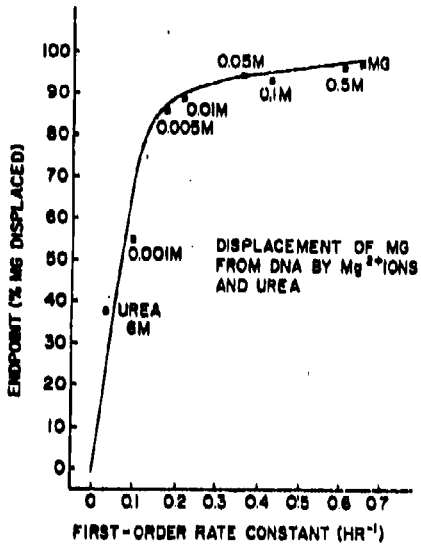



FIGURE 5: Displacement of MG from DNA by magnesium ions and urea. $Mg^{2+}$ and urea were used at the concentrations indicated in the diagram; $1.88 \times 10^{-5}$ *M* MG was considered bound to $2.45 \times 10^{-4}$ *M* DNA phosphorus present.

Urea is known to break non-polar bonds. The bimodality of the curve in Fig. 5 reveals that ∿90 per cent of MG are displaced by a process whose endpoints are strongly dependent upon the rate constant of the displacement reaction while at $Mg^{2+}$ concentrations of $>10^{-2}$ M, the displacement of the remaining MG is much less dependent upon the 1st order rate constants. Apparently, the binding of MG to DNA is bimodal, involving perhaps binding to both strands of the double helix across the minor groove as well as binding to individual strands. Binding to single strands is, indeed, indicated by the reactivity of MG with the DNA of ϕX174 (10).

The electrostatic nature of the interaction of MG with DNA is also illustrated by the effect of the dye on the thermal denaturation profile ("melting curve") of DNA, Fig. 6. MG, acting as a counterion to the DNA phosphates, shifts this profile to higher temperatures ($\Delta T_m$ = ∿12° C for the stoichiometric complex). Crystal violet which differs from MG only be the absence of a quarternary ammonium nitrogen has no significant effect on the thermal denaturation of DNA (Fig. 6). Evidently, the quarternary amino group of MG is prominently involved in the binding of the dye to DNA.

## IV. KINETICS OF DISPLACEMENT OF MG BY DRUGS

Before undertaking the survey of the displacement of MG from DNA by drugs, it was necessary to study the kinetics of the displacement reaction. The test compound was the antimalarial quinacrine (Fig. 7). This drug binds to DNA by intercalation (17) with a stoichiometry of one quinacrine molecule per two base pairs (18). Intercalation unwinds the double helix and causes drastic changes in the molecular architecture of the B-conformation of DNA, foremost a doubling of the intervals between the base pairs at the intercalation

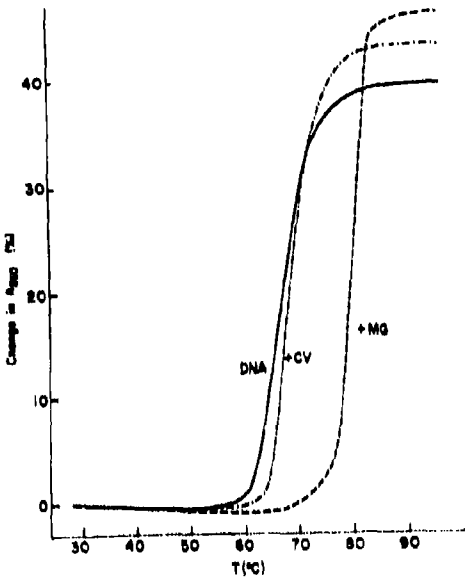



FIGURE 6: Thermal denaturation of 30 μg/ml of calf thymus DNA in the presence of Methyl Green or Crystal Violet and in their absence. Increases in absorbances at 260 nm were calculated with respect to DNA's absorbancy at room temperature. Concentrations of MG and CV, $7.5 \times 10^{-6}$ *M*.




Fig. 7. Quinacrine

sites. This may distort the binding sites for MG. Also: the $pK_1$ (of the tertiary aliphatic amino group, Fig. 7) is ∿10 and the $pK_2$ (of the ring-substituted secondary amino group, Fig. 7) is ∿8. Hence, quinacrine exists at physiological pH as a di-cation and may, like $MG^{2+}$ (Fig. 5), displace MG from DNA by ionic competition. In fact, in addition to intercalation, one quinacrine molecule is bound per 3 nucleotides by electrostatic attraction to the periphery of the double helix (18).

Fig. 8 shows that at a molar ratio of free quinacrine/DNA-bound MG of 5, the displacement reaction is of first order with time, while at a ratio of 1.25 it is of second order. At a "standard" molar ratio of 2.5, the reaction is of first order for the initial 2 hrs of the displacement reaction and then changes to second order kinetics.

These observations can be explained stoichiometrically. Binding of MG to DNA occurs with 0.077 dye molecules per nucleotide, while quinacrine binds to the extent of 0.58 molecules per nucleotide. Since 7.5 more quinacrine binds to DNA than the number of MG molecules which it displaces, it follows that with initially low or critically decreasing concentrations of free quinacrine, these concentrations become one of the rate-limiting factors of the displacement reaction, the other being the concentrations of remaining bound MG. For the screening method reported here, the molar ratio of free drugs to DNA-bound MG was standardized at 2.65 which affords a 3 hr period for determining the first order rate constant of the displacement reaction.

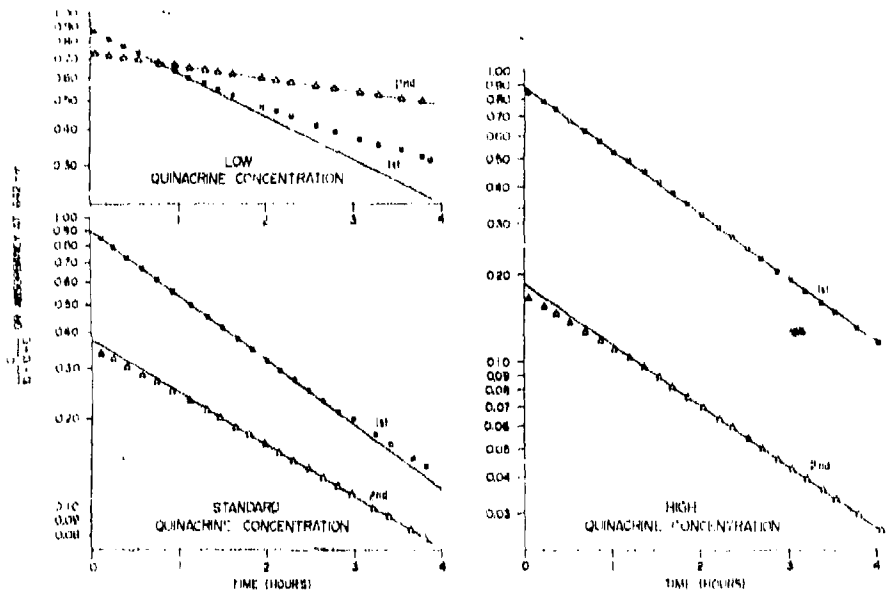



FIG. 8.—*First-order (1st) and second-order (2nd) kinetics for displacement of methyl green from its complex with DNA by quinacrine*

*a* = initial concentration of DNA-bound methyl green (18.8 μM); *b* = initial concentrations of quinacrine (50 μM, standard concentration of quinacrine; 100 μM, high; 25 μM, low); *c* = concentration of the amount of methyl green which remained bound to DNA. On the ordinate absorbance is plotted for first-order kinetics, and $c/(b - a + c)$, for second-order kinetics.

The "absolute endpoint" of MG displacement is determined after keeping the reaction mixtures in the dark for one or two days, ascertaining periodically that no more MG is displaced from DNA.

The first order rate constant of the decolorization of free MG is 0.65 x $hr^{-1}$ and has not been attained in any displacement by drugs. Hence, the rearrangement of liberated MG is not a rate-limiting factor in displacement analysis and all measured displacement rates are characteristics of the drugs studied.

## V. DISPLACEMENT OF METHYL GREEN FROM DNA BY DRUGS

Results of displacement analyses of MG from DNA by drugs are summarized in Fig. 9. The bimodal correlation between first order reaction constants and absolute endpoints of displacement can be virtually superimposed on the same correlation (Fig. 5), resulting from the displacement of MG by graded concentrations of $Mg^{2+}$. As in Fig. 5, there is an inflection point at ~90 per cent displacement. Up to this value, end points are strongly dependent upon the rate constants of the dis-

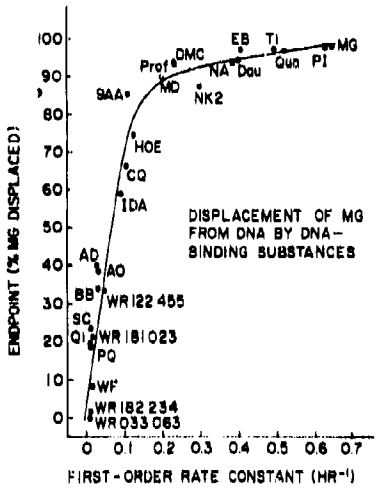



FIGURE 9. Displacement of MG from DNA by drugs and dyes. Displacing compounds: propidium iodide (PI), quinacrine (Qua), tilorone (TI), ethidium bromide (EB), daunomycin (Dau), Nitroakridin 3582 (NA), distamycin A (DMC), proflavine (Prof), nitroakridin 2 (NK 2), Hoechst 33258 (HOE), chloroquine (CQ), irehdiamine (IDA), actinomycin D (AD), Acridine Orange (AO), berberine (BB), side chain of quinacrine and chloroquine (SC), quinine (QI), primaquine (PQ), warfarin (WF), α-(2-piperidyl)-3,6-bis(trifluoromethyl)-9-phenanthrenemethanol hydrochloride (WR 122455), 6-methoxy-8-(4-amino-1-methylbutylamino)lepidine·4HCl·2C$_2$H$_5$OH·H$_2$O (WR 181023), 6-methoxy-8-(4-amino-1-methylbutylamino)quinaldine dihydrochloride (WR 182234), and α-(di-*n*-heptylaminomethyl)-6-bromo-9-phenanthrenemethanol hydrochloride (WR 033063); the displacing compounds were used at a concentration of $5 \times 10^{-5}$ *M*; $1.88 \times 10^{-5}$ *M* MG was considered bound to $2.45 \times 10^{-4}$ *M* DNA phosphorus present; 1-cm light path. Miracil D (MD) and 6-chloro-2-methoxy-9-methylaminoacridine were used at a concentration of $5 \times 10^{-6}$ *M*; $1.88 \times 10^{-6}$ *M* MG was considered bound to $2.45 \times 10^{-5}$ *M* DNA phosphorus present; 10-cm light path.

placement reactions, while for the displacement of the remaining 10 per of MG, the dependence is less marked. The bimodality of displacement again suggests the existence of two discrete binding modes of MG to DNA in addition to differences in the abilities of the tested compounds to interfere with these modes either by ionic competition or by structural distortion of the B-configuration of DNA

The most active displacing drugs in the upper branch of the curve (Fig. 9) are the strong and stoichiometric intercalators, propidium iodide quinacrine, tilorone, ethidium bromide, daunomycin, two nitroacridines, proflavine, and miracil D. The non-intercalating antibiotic, distamycin A, which binds to DNA pseudo-irreversibly (19) is also found in this part of the diagram.

The left branch of the curve lists some non-intercalating substances such as Hoechst 33258, or the aliphatic side chain of quinacrine (and chloroquine) as well as those intercalative agents which bind more weakly to DNA [chloroquine, primaquine (17)] or bind with a lower stoichiometry, owing to steric hindrance (actinomycin D, berberine, quinine, irehdiamine).

The phenanthrene derivative WR 122 455 will on structural grounds, be able to intercalate into DNA, but it lacks a ring nitrogen or a substituted amino group for electrostatic attraction to DNA phosphates; the 2-piperidyl substituent is weakly basic (pK 2.8) and is not coplanar with the phenanthrene ring. Intercalation, therefore,

will be relatively weak, owing to lack of electrostatic anchoring and possible steric hindrance. This should be compared to the strong binding of propidium iodide and ethidium bromide which are phenanthridines. Deletion of the two amino groups from ethidium is known to decrease the DNA binding constant by factors of 10 to 20.

WR 181 023 is a structural analog of primaquine. Fig. 9 shows that the MG displacement parameters are similar to those of primaquine. Both compounds have a low $pK_2$ of the 8-amino group and, hence, will exist at physiological pH as monoprotonated cations which anchor poorly to DNA. The $pK_2$ of primaquine is 3.1.

The nature of the low endpoints of MG displacements (<10 per cent) by warfarin and by the experimental substances WR 182 234 and WR 033 063 is not well understood. Warfarin has failed all other tests for interaction with DNA. For practical purposes, endpoints of <10 per cent MG displacement and rates close to 0 should be regarded as non-significant.

## VI. CORRELATION OF MG DISPLACEMENT DATA WITH BIOLOGICAL ACTIVITIES

A study of the ability of DNA intercalators to eliminate drug-resistance genes from the R-plasmid R1 (in Salmonella typhimurium) has yielded activities against the kanamycin resistance gene (3) that are directly proportional to the MG displacements of the active drugs, screened by the method reported here. This is documented in Fig. 10 (10).

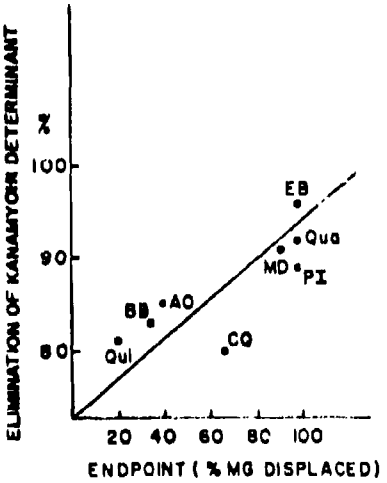



Fig. 10

The results of Fig. 10 are explained by the fact that intercalation binding into plasmid DNA blocks selectively the replication of this DNA so that in growing bacterial cultures plasmid$^+$ cells are continuously being diluted by the plasmid$^-$ progeny (4).

The ability of selected intercalants to facilitate the disassembly of ribosomes in vitro has been measured in terms of the relative rates of this process (20). The diagram, Fig 11, shows these relative rates as a function of the absolute endpoints of MG displacement (20). The curve exhibits the same bimodality as thatshown in Figs. 5 and 9, suggesting that the

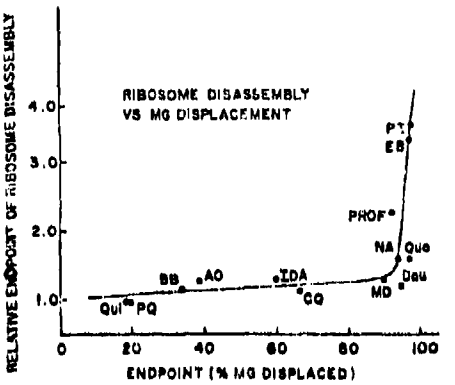



FIGURE 11Effect of test compounds on disassembly of ribosomes (Wolfe et al., 1972) as a function of Methyl Green displacing end points.

effects on ribosomes are caused by drug interactions with double-stranded runs of ribosomal RNA (20).

By a differently designed method it was found for a series of planar phthalanilides that concentrations that displaced 25 per cent of DNA-bound MG were directly proportional to antibacterial concentrations of the same compounds (21). This has been the first contribution of a form of MG displacement analysis to a program in experimental chemotherapy.

A comparison of arbitrary endpoints (at 18 hrs) of MG displacements by graded concentrations of proflavine, actinomycin D, and anthramycin has shown anthramycin to be the least active compound, although it was an effective inhibitor of DNA and RNA (but not protein) biosyntheses (22). Anthramycin binds to DNA slowly without intercalation (23).

Semi-quantitative estimation of MG displacement has been used to select 19 substituted N-heterocyclic compounds as potential immunosuppressive agents in the graft vs. host reaction that results from the development of cellular immunity (24). Seven of these preselected compounds, i.e. 37 per cent, showed significant immunosuppressive activity (24). Since they were tested at different concentrations, the numerical results do not lend themselves to regression analysis. It is noteworthy, however, that a MG displacement prescreen predicted correctly pharmacological activity of 37 per cent of a screened population of compounds.

After the present screening method had been applied to the study of the binding of the antibiotic, distamycin A, to DNA (25), the method was used to pilot the development of distamycin congeners with graded numbers of N-methylpyrrole rings in the molecule (26). Increasing numbers of N-methylpyrrole constituents correlated with increasing displacement of MG (26) and with increasing inhibitions of the DNA polymerase reactions of Escherichia coli and of Rous sarcoma virus (rev. in 27).

Finally, MG displacement has been used to discover the presence of DNA-complexing compounds in crude fermentation mixtures of an antibiotics search program. Aliquots of such unresolved mixtures were added directly to solutions of the MG-DNA reagent, and several DNA-binding antibiotics were isolated from those mixtures which had displaced MG from DNA (28).

## VII. DISCUSSION AND SUMMARY

The principal result, reported in this article, is that spectrophotometric analysis of the displacement of MG from its complex with double-stranded DNA by DNA-binding clinical and experimental drugs yields reaction rates and absolute endpoints which are correlated to biochemical and chemotherapeutic activities of the tested compounds, when these activities are caused by DNA template toxicity. Tables 1 and 3 document that this mechanism of action is prevalent among antiparasitic and anticancer drugs.

Structural rules for the design or the recognition of intercalating DNA complexers are well developed (5) but MG displacement analysis yields numerical indications of relative affinities and/or stoichiometries of DNA binding, as discussed in Section V.

Structural rules for peripheral (i.e. non-intercalative) binding to DNA are not so well developed, and MG displacement can serve as a facile indication that such binding occurs (25).

Finally, the ability of fermentation beers from antibiotics search programs to displace MG will signal the presence of DNA-binding compounds of completely unknown structures and facilitate their isolation (28).

Structural variation of DNA binding drugs can be piloted by MG displacement analysis (21, 26) in the expectation that enhanced MG displacement will signal enhanced chemotherapeutic activity.

## VIII. REFERENCES


1. Tigertt, W.D.: Military Med. 131 (Suppl.) 131, 853-856 (1966)
2. Hahn, F.E.: Antibiot. & Chemother. 27, in press, Karger, Basel, (1980)
3. Hahn, F.E.: ibid. 20, 196-226 (1976)
4. Hahn, F.E.: Naturwissenschaften 66, 555-562 (1979)
5. Hahn, F.E.: ibid. 62, 449-458 (1975)
6. Kurnick, N.B.: Arch. Biochem. Biophys. 29, 41-53 (1950)
7. Kurnick, N.B.: Methods Biochem. Anal. 9, 1-38 (1962)
8. Kurnick, N.B. and Radcliffe, I.E.: J. Lab. Clin. Med. 60, 669-688 (1962)
9. Kurnick, N.B. and Foster, M.: J. Gen. Physiol. 34, 147-159 (1950)
10. Krey, A.K. and Hahn, F.E.: Biochemistry 14, 5061-5067 (1975)
11. Michaelis, L.: Cold Spring. Harb. Symp. Quant. Biol. 12, 131-142 (1947)
12. Arnott, S. and Selsing, E.: J. Mol. Biol. 88, 551-552 (1947)
13. Frisman, E.V., Veselkov, A.N., Slonitsky, S.V., Karavaev, L.S. and Vorob'ev, V.I.: Biopolymers 13, 2169-2178 (1974)
14. Usatyi, A.F. and Shlyakhtenko, L.S.: Biopolymers 13, 2434-2446 (1974)
15. Ivanov, V.I., Minchenkova, L.E., Minyat, E.E., Frank-Kamenetskii, M.D. and Schyolkina, A.K.: J. Mol. Biol. 87, 817-833 (1974)
16. Herskovits,T.T., Singer, S.J. and Geiduschek, E.P.: Arch. Biochem. Biophys. 94, 99-114 (1961)
17. Allison, R.G. and Hahn, F.E.: Antimicrob. Agents & Chemother. 11, 251-257 (1977)
18. Krey, A.K. and Hahn, F.E.: Mol. Pharmacol. 10, 686-695 (1974)
19. Krey, A.K., Allison, R.G. and Hahn, F.E.: FEBS Lett. 29, 58-62 (1973)
20. Wolfe, A.D., Allison, R.G. and Hahn, F.E.: Biochemistry 11, 1569-1572 (1969)
21. Rauen, H.M., Norpoth, K., Untersberg, W. and Haar, H.: Experientia 21, 300-304 (1965)
22. Bates, H.M., Kuenzig, W., and Watson, W.B.: Cancer Res. 29, 2195-2205 (1969)
23. Kohn, K.W.: In Antibiotics III, 3-11, Springer, Heidelberg (1975)
24. Zeleznick, L.D., Crim, J.A. and Gray, G.D.: Biochem. Pharmacol. 18, 1823-1827 (1969)
25. Krey, A.K. and Hahn, F.E.: FEBS Lett. 10, 175-178 (1970)
26. Zunino, F. and DiMarco, A.: Biochem. Pharmacol. 21, 867-873 (1972)
27. Hahn, F.E.: Pharmac. Ther. A. 1, 475-485 (1977)
28. Zeleznick, L.D.: personal communication

# ON THE USE OF ACTIVE HIGHER HARMONIC BLADE PITCH CONTROL FOR HELICOPTER VIBRATION REDUCTION

*CHARLES E. HAMMOND, PhD
JOHN H. CLINE
STRUCTURES LABORATORY
UNITED STATES ARMY RESEARCH AND TECHNOLOGY LABORATORIES
(AVRADCOM)
NASA LANGLEY RESEARCH CENTER
HAMPTON, VIRGINIA 23665

Vibration levels have been a problem in helicopters since their inception. The reason for this lies in the method whereby the helicopter generates its lift, namely, the rotor system. As the rotor blades rotate they encounter a continuously changing aerodynamic environment which results in a continuously changing environment which results in a continuously changing aerodynamic loading on the blades. This changing environment is repeated on each revolution of the rotor. Hence, the rotor develops aerodynamic loads which are oscillatory in nature. These oscillatory loads are transferred directly to the helicopter airframe through the mechanical connection of the rotor to the airframe, i.e., the rotor-shaft/transmission attachment. Oscillatory loads are also transmitted to the airframe by impingement of the rotor wake on the upper portion of the airframe, but the mechanically transferred loads are in most cases much more significant than the aerodynamically transferred loads.

Because of the symmetrical placement of blades in a rotor, the oscillatory loads felt by the airframe occur at frequencies which are multiples of the number of blades times the rotational frequency of the rotor. For example, if $\Omega$ is the rotational speed for a four-bladed rotor, the oscillatory loads would occur at frequencies of $4\Omega$, $8\Omega$, $12\Omega$, etc. Conventionally, these frequencies are denoted as 4P, 8P, 12P, etc. The oscillatory loads which occur at the first harmonic of the blade passage frequency, i.e., 4P for



PRECEDING PAGE BLANK-NOT FILMED

a four bladed rotor, are the loads of primary concern in any helicopter vibration reduction program. The reason for this is that these loads are generally significantly higher than the higher frequency loads, and as a result tend to mask the influence of the higher frequency loads. There have been cases reported, however, where the higher frequency loads became more significant when the first harmonic loads were reduced, e.g., reference 1.

Past efforts to reduce the vibration levels in helicopters have employed methods such as airframe tuning to avoid resonance of the structural frequencies with the frequencies of the oscillatory loads (ref. 2) and application of vibration control devices such as tuned vibration absorbers (ref. 3, 4), and vibration isolators (ref. 5). These approaches have been successful in the sense that the vibration levels on current generation helicopters are about one-half the levels which were achievable on previous generation machines. These lowered vibration levels are a result of a considerable amount of dedicated research and trial and error development of vibration control devices conducted primarily by the helicopter industry.

The vibration control devices, while successful in reducing vibration levels, add a significant amount of weight to the aircraft. The weight penalty on current helicopters which can be attributed directly to vibration control devices amounts to about two percent of the aircraft gross weight. Although this is the generally accepted industry method of presenting the weight penalty, it is more significant from a user point of view to note that this weight penalty amounts to about fifteen percent of the aircraft payload. Thus, if vibration reduction can be achieved at lower weight penalties, significant increases in the aircraft useful payload can result.

It is the purpose of this paper to discuss an approach to vibration control different from the vibration control devices mentioned above, which depend to a large degree on the addition of mass for their effectiveness. This approach, which will be referred to as higher harmonic control (HHC), is aimed at altering the aerodynamic loads on the rotor before they are transferred to the airframe. This is in contrast to the vibration control devices discussed earlier which attempt to deal with the oscillatory loads after they have been generated. Higher harmonic control, as will be discussed later, is a method whereby the aerodynamic loading on the blade is tailored in such a way that the vibratory loads transferred to the airframe are minimized. Preliminary design studies by Hughes

Helicopters (ref. 6) have indicated that weight penalties on the order of 0.5 percent of gross weight are achievable with the higher harmonic control concept.

## SYMBOLS

| | |
|---|---|
| J | Magnitude of optimal control penalty function |
| T | Transfer matrix relating higher harmonic inputs to vibratory responses |
| $W_Z$ | Matrix of response weights |
| $W_\theta$ | Matrix of control weights |
| Z | Column of vibratory responses |
| $Z_o$ | Column of baseline vibratory responses (without higher harmonic control) |
| $\theta$ | Column of higher harmonic inputs |

Superscripts

| | |
|---|---|
| T | Transpose of a matrix |
| ^ | Estimated value from Kalman filter |
| * | Optimum higher harmonic inputs |

## HIGHER HARMONIC CONTROL CONCEPT

Higher harmonic control is achieved by superimposing non-rotating swashplate motions at the blade passage frequency (4P for a 4 bladed rotor) upon the basic collective and cyclic flight control inputs. The frequency of the inputs is picked at the blade passage frequency because this is the frequency of the loads which are to be suppressed. The amplitude and phase of the higher harmonic inputs are chosen so as to achieve minimization of the responses being controlled.

This approach to control vibratory loads has been the subject of a number of recent wind tunnel investigations, e.g., references 7, 8, and 9. These investigations, which were each conducted on significantly different types of rotor systems, all

showed that higher harmonic control was successful in reducing the vibratory loads transmitted by the rotor to the airframe. These tests further indicated that the amplitude of higher harmonic blade pitch inputs required to achieve the desired reductions was small; on the order of one degree for the conventional helicopter flight envelope.

The primary parameters which determine the success of the higher harmonic inputs in reducing the vibratory loads are the amplitudes and phases of the various inputs. In the references 7, 8, and 9 these inputs were determined through trial and error testing. This trial and error approach is satisfactory if one is using a single input to control a single response. However, when three controls are used to control one or more responses, then the number of possible combinations of inputs becomes too numerous for the trial and error approach to be successful. Furthermore, if the higher harmonic control technique is to be applied to production helicopters then some systematic means must be available to determine, automatically, the required inputs. The means for automatically determining the higher harmonic inputs constitutes a closed loop active control system.

## Active Control System

The active control system to be discussed here is the approach which has been taken at the Structures Laboratory, U.S. Army Research and Technology Laboratories (AVRADCOM). Other researchers in the field, notably the Boeing Vertol researchers (reference 8) are pursuing somewhat different approaches. A schematic of the active control system employed in obtaining the results reported herein is shown in figure 1.

In this case a four-bladed rotor wind tunnel model (to be discussed later) was used and the 4P higher harmonic inputs were used to control the 4P vibratory responses in vertical force, pitching moment, and rolling moment. In figure 1, the vibratory responses from the model (containing all the harmonics) are input to an electronic control unit (ECU). The ECU actually performs two separate functions, the first of which is to extract from the total vibratory response signals the amplitude and phase of the 4P contribution, since it is this contribution which is to be minimized. The ECU contains an analog implementation of a demodulation scheme which provides the sine and cosine components (from which the amplitude and phase may be determined) of the 4P responses in real time.

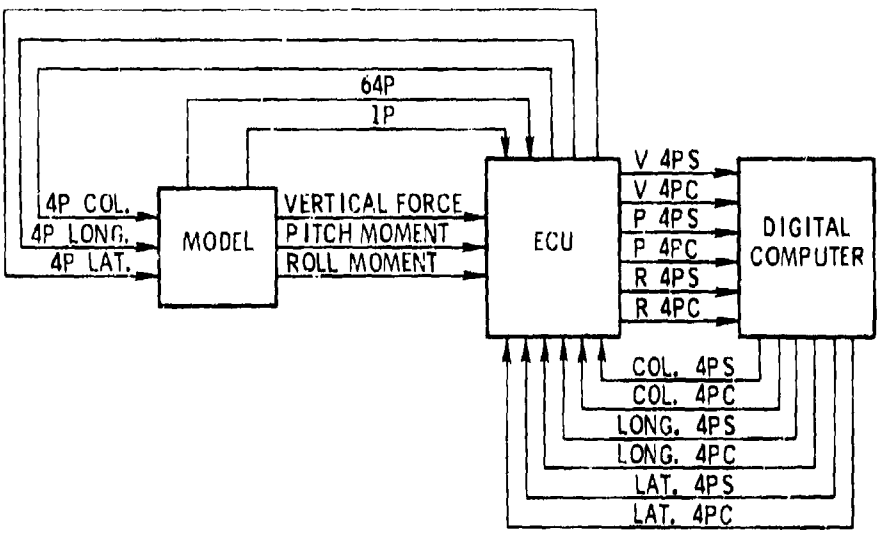



Figure 1. Block diagram of closed loop higher harmonic control system.

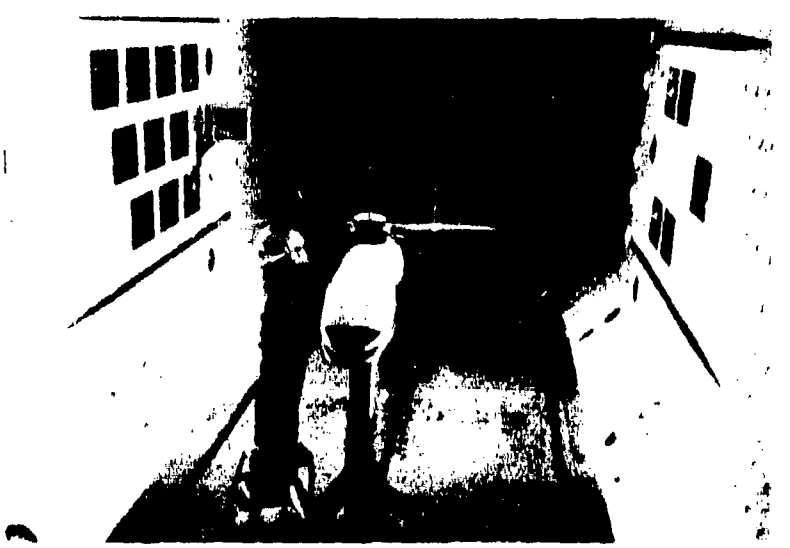


Figure 2. Aeroelastic Rotor Experimental System (ARES) installed in the Langley Transonic Dynamics Tunnel.

The sine and cosine components of the 4P responses are passed from the ECU to a digital computer which contains the software for the control algorithms. The nature of the control algorithms will be discussed in a subsequent section. The control software makes use of the measured responses to previous 4P higher harmonic inputs to determine the "optimum" higher harmonic inputs. The sine and cosine components of these "optimum" inputs are output from the computer as d.c. voltages which are passed to the ECU. The ECU then performs its second function which is to convert the d.c. voltages from the computer to 4P oscillatory analog signals having the correct amplitude and phase to drive the control system servos. The model then responds to these inputs and the control loop begins again.

The 1P and 64P signals shown on figure 1 are timing signals used by the ECU in extracting the 4P components of the responses.

## Control Algorithms

The control algorithms employed in the program make use of digital optimal control theory (ref. 10). In implementing the theory, it is assumed that the 4P system response may be described by the following equations

$$\{z\} = \{z_o\} + [T]\{\theta\} \qquad (1)$$

Note that these equations constitue a static linear representation. The equations state that the system 4P response is made up of a baseline response plus a response which is related to the 4P inputs by a transfer matrix. Thus, if the number of responses is the same as the number of inputs and if the baseline responses and transfer matrix are known, then a set of 4P inputs could be found which would null the 4P responses.

The first portion of the control strategy is thus to determine the baseline response and the transfer matrix. Since it is undesirable to turn the control system off to measure the baseline response, and since information about the system is available from past HHC inputs and the resulting responses, an identification algorithm is used to determine $Z_o$ and T. The identification algorithm used is the Kalman filter (ref. 11). This algorithm may be thought of as a generalized form of a least-squares algorithm which accounts for the fact that the measured responses may be contaminated by noise and the transfer matrix may be changing with time.

Once the baseline responses and the transfer matrix are known, the "optimum" outputs are determined as those inputs which will minimize the performance index

$$J = Z^T W_Z Z + \theta^T W_\theta \theta \tag{2}$$

If it is assumed that the transfer matrix is known without error, then the inputs which minimize the above performance index are given by

$$\theta^* = -[\hat{T}^T W_Z \hat{T} + W_\theta]^{-1} [\hat{T}^T W_Z \hat{Z}_o] \tag{3}$$

Note from equation (3) that if the response weighting matrix, $W_Z$, is the identity matrix and the control weighting matrix, $W_\theta$, is zero, then the result from equation (3) is the same as solving equation (1) directly for the inputs which will give zero responses. The weighting matrix on the responses allows one to place more emphasis on reduction of some of the responses than others. The control weightings allow one to limit the amplitude of controls allowed.

The Kalman filter used in estimating the baseline responses and transfer matrix is a recursive algorithm and thus each new measurement of the responses leads to an updated estimate of the baseline responses and transfer matrix. With each update of these parameters, updated "optimum" inputs are calculated and applied to the rotor control system, and the cycle begins again. This control system is adaptive in that the estimates of the parameters used in the model, equation (1), are continuously updated and the updated parameter estimates are used to determine the optimal inputs.

The control algorithms are executed very quickly by the computer and permit updating the optimal control solution every revolution of the rotor. The algorithms would actually permit more rapid updating of the solutions, but it is felt that once-per-revolution updating is sufficient to accommodate the most rapid changes in flight conditions which might be experienced by a helicopter.

## DESCRIPTION OF MODEL AND TESTS

The basic wind tunnel model used in this investigation was the Structures Laboratory Aeroelastic Rotor Experimental System (ARES) shown in figure 2. This model is the successor to the model described

in reference 12, and it is used for aeroelastic investigations of model scale rotor systems. These investigations are conducted in the Langley Transonic Dynamics Tunnel (TDT).

The TDT is a continuous-flow tunnel with a slotted test section and is capable of operation over a Mach number range up to 1.20 at stagnation pressures from .01 to 1 atmosphere. The tunnel test section is 4.9 m square with cropped corners and has a cross sectional area of 23 $m^2$. Either air or Freon-12 may be used as a test medium in the TDT. For this investigation, Freon-12 at a nominal density of 3.09 $Kg/m^3$ was used as a test medium. The advantages of using Freon-12 as a test medium for aeroelastic model testing have been discussed in references 13 and 14.

The ARES is powered by a 35 kw variable frequency synchronous electric motor connected to the rotor shaft through a belt-driven, two-stage reduction system. The model pitch attitude is changed using a remotely controlled hydraulic actuator and electric servo system. The rotor control system is a conventional swashplate system which is remotely controlled through the use of three electronic servos and hydraulic actuators. The high frequency response characteristics of this control system are necessary for the higher harmonic inputs.

Instrumentation provisions on the ARES allow continuous measurement of model control settings, rotor forces and moments, blade loads, and pitch link loads. Model pitch attitude is measured by an accelerometer, and rotor control positions are measured by linear potentiometers connected to the swashplate. Rotor blade flapping and lagging are measured by rotary potentiometers mounted on the rotor hub and geared to the blade cuff. The rotating blade data are transferred to the fixed system through a 60-channel, horizontal disk slip-ring assembly. Rotor forces and moments are measured by using a six-component strain-gage balance mounted below the drive system. The balance is fixed with respect to the rotor shaft and pitches with the fuselage. Fuselage forces and moments are not sensed by the balance.

The vibratory forces and moments used as response inputs to the higher harmonic control algorithms were taken from the balance. This means that the moment responses used by the control algorithms were made up of the rotor hub moments plus the rotor inplane shears times the offset distance between the rotor hub and the balance center. This offset distance was 51.44 cm.

The rotor system used in this investigation was a four-bladed articulared rotor system. The blades were dynamically scaled to be representative of a current generation rotor system. The blades had swept tips consistent with their full-scale counterpart, but the swept tips were not significant with respect to the higher harmonic control program.

The rotor was tested over a range of advance ratios (tunnel speed/rotor tip speed) consistent with the full-scale flight envelope. Because of tunnel limitations, advance ratios below .2 were not possible. The rotor rotational speed was set so as to achieve a full-scale tip Mach number. At each advance ratio the rotor was trimmed to a condition which represented a 1-g flight condition for the full-scale aircraft. Blade flapping was trimmed with respect to the shaft.

## DISCUSSION OF RESULTS

The results to be discussed in this section were obtained using the closed loop active control system discussed earlier. In obtaining these results, the model was trimmed at a given advance ratio, and data were recorded to establish the vibratory responses without higher harmonic control. The automatic control system was then turned on and allowed to stabilize. With the controller still on at its stabilized condition, data were recorded to establish the vibratory responses with higher harmonic control. The following results present a comparison with and without higher harmonic control of the vibratory responses, blade loads, and control loads.

The success of the higher harmonic control in reducing the vibratory responses is shown in figures 3, 4, and 5, where the variation of the responses with advance ratio are shown both with and without higher harmonic control. Figure 3 shows the variation of the vibratory vertical force. As may be seen from this figure, the higher harmonic control was quite successful in reducing this vibratory response. Reductions of from 70 to 90 percent were obtained over the range of advance ratios tested. The vibratory pitching moment shown in figure 4 indicates reductions of from 33 to 68 percent and the vibratory rolling moment shown in figure 5 indicates reductions of from 0 to 46 percent.

The fact that the order of the reductions which could be obtained in the vibratory pitching and rolling moments was much less than the reductions obtained in the vertical force is a result for which no explanation has been established. Mathematically, since

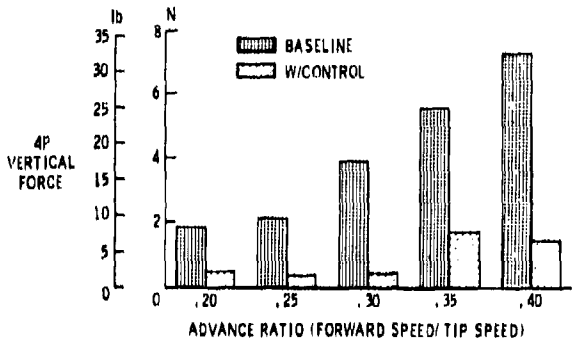



Figure 3. Variation of vibratory vertical force with advance ratio.

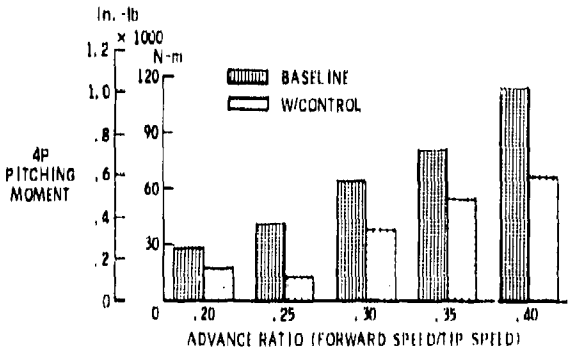



Figure 4. Variation of vibratory pitching moment with advance ratio.

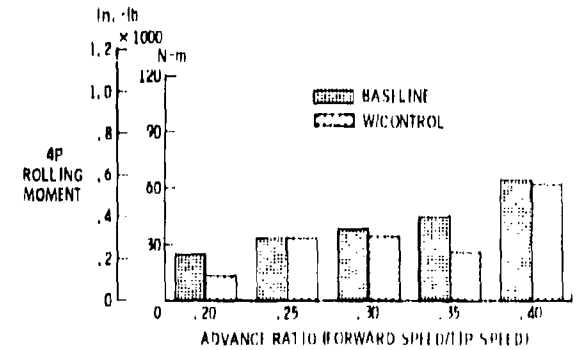



Figure 5. Variation of vibratory rolling moment with advance ratio.

three inputs were used to control three responses, it should have been possible to drive each of the responses to near zero values. A considerable amount of testing was done to explore this apparent anomaly, but a satisfactory explanation was not found during the wind tunnel test program.

It should be pointed out that the results presented in figures 3-5 were obtained by weighting the vertical force response more heavily than the moment responses (equations (2), (3)). Numerous combinations of the weightings were explored during the test, and it was found that the weightings play a significant role in the levels of vibration reduction which can be obtained. It was found, for example, that with the proper combination of weights, the moments could be reduced more than is shown in figures 4 and 5, but at the expense of less reduction in vertical force.

Efforts to understand why moment response reductions greater than those shown in figures 4 and 5 could not be obtained in conjunction with large reductions in vertical force response are continuing. Indications are that the problem lies in the sensor location, i.e., the moments being sensed by the balance contained hub moment as well as hub shear contributions. Further tests are being performed on the model in a hover facility to reconcile this issue.

It is imperative when evaluating a system which appears to promise high payoff for low investment, e.g., significant vibration reduction with a low weight penalty, that all avenues of possible side effects be explored. In the case of higher harmonic control, since the concept is based on tailoring the blade aerodynamic loads to achieve reductions in the vibratory responses, an examination of the higher harmonic inputs is appropriate. The results to be shown are from the same test points at an advance ratio of .3 as the vibratory responses shown earlier. The results at other advance ratios were similar.

The radial distribution of blade alternating flapwise bending moment (½ peak-to-peak values) is shown in figure 6. Similar distributions for the edgewise moment and torsion are shown in figures 7 and 8, respectively. As may be seen, there is a small reduction in the flapwise bending moment, a significant increase in the edgewise bending moment, and a moderage increase in the torsional moment. With the exception of the edgewise moment, these results are consistent with the open loop results obtained previously (ref. 9).

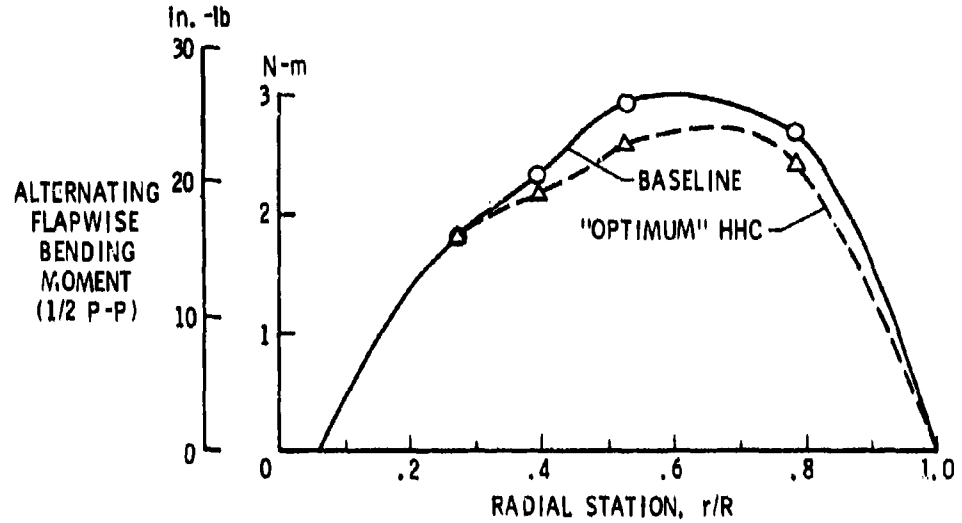



Figure 6. Radial distribution of blade alternating flapwise bending moment (½ peak-to-peak values) at an advance ratio of 0.3.

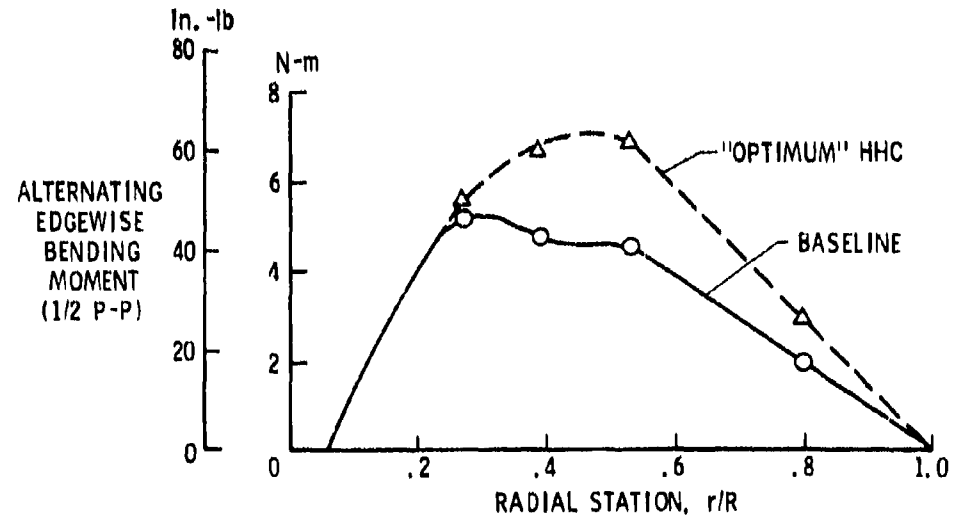



Figure 7. Radial distribution of blade alternating edgewise bending moment (½ peak-to-peak values) at an advance ratio of 0.3.

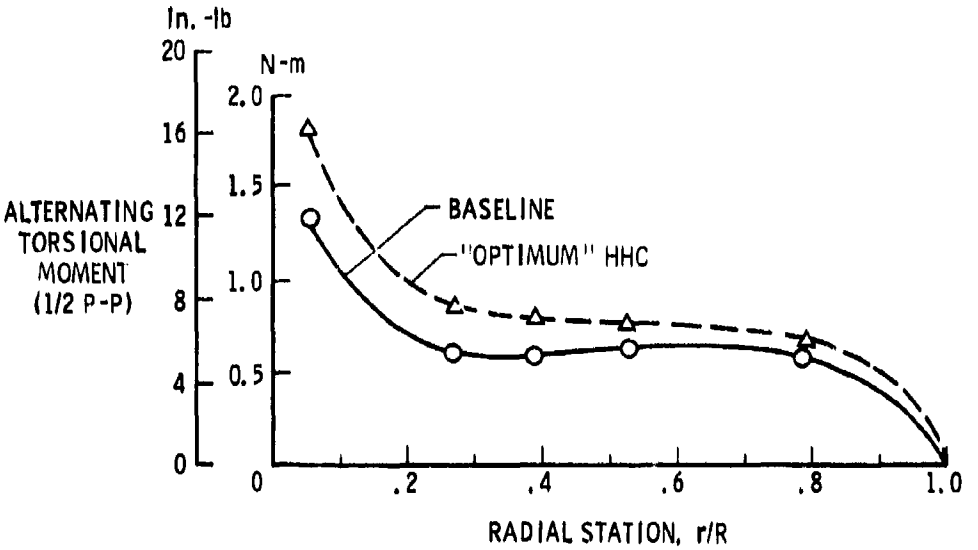



Figure 8. Radial distribution of blade alternating torsional moment (½ peak-to-peak values) at an advance ratio of 0.3.

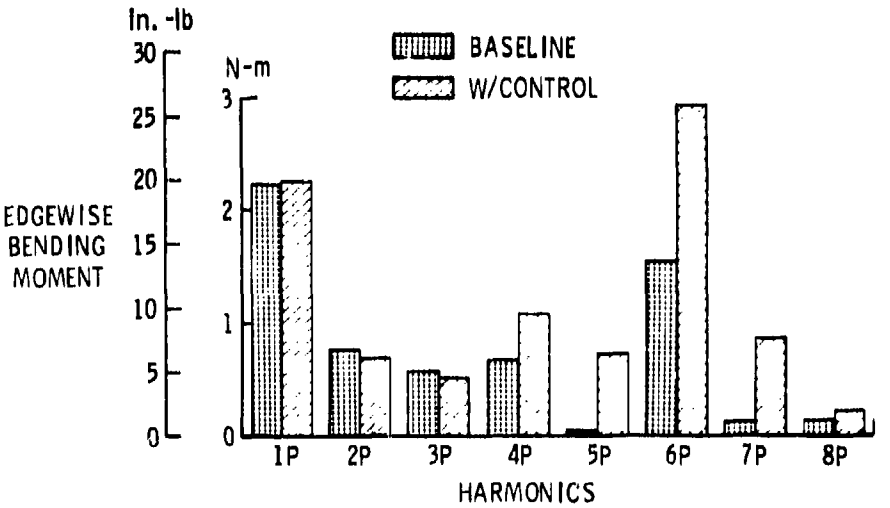



Figure 9. Harmonic decomposition of edgewise bending moment at 53 percent span, and advance ratio of 0.3.

The cause of the increase in the edgewise moments appears to be associated with placement of the blade's natural frequencies relative to the rotor harmonics. Figure 9 presents a harmonic decomposition of the edgewise bending moment at 53 percent span. As may be seen, there is a strong contribution at 6P without higher harmonic control, and this contribution is aggravated when higher harmonic control is applied. The strong contribution at 6P without higher harmonic control is indicative of a blade natural frequency near 6P. Excitation of this mode by the higher harmonic control comes from the fact that 4P cyclic motion of the non-rotating swashplate results in 3P and 5P motions of blade pitch in the rotating system, whereas 4P collective motion of the swashplate results in 4P blade pitch changes. Any impurity of the 3P blade pitch motions could excite the 6P natural blade mode since it is a second harmonic of the 3P input.

The indication from the edgewise moments is that if a a new rotor is designed to incorporate higher harmonic control, blade frequency placements subject to constraints imposed by the higher harmonic control must be a design consideration. Further, for flight testing of higher harmonic control on existing aircraft, the blade loads must be carefully monitored to avoid any excessive stresses. It should be noted that the edgewise loads with higher harmonic control shown in figure 7 are well within the design load envelope for these blades, but the fact that higher harmonic control can produce a significant increase in the loads must be recognized, particularly in flight test programs.

Figure 10 presents the pitch link loads with and without higher harmonic control as a function of advance ratio. As may be seen, and as was expected, there is an increase in the control loads when the higher harmonic control is applied. The source of the increase may be attributed directly to the higher harmonic inputs as may be seen from figure 11. This figure presents a harmonic decomposition of the pitch link load at an advance ratio of .3. Note that the increase in load with higher harmonic input occurs at frequencies of 3P, 4P, and 5P which are the excitation frequencies in the rotating system. These increases in control system loads are consistent with previous findings (ref. 9) and the magnitude of the increases have not caused significant concern among designers. Again, however, these increases must be considered in any flight test program.

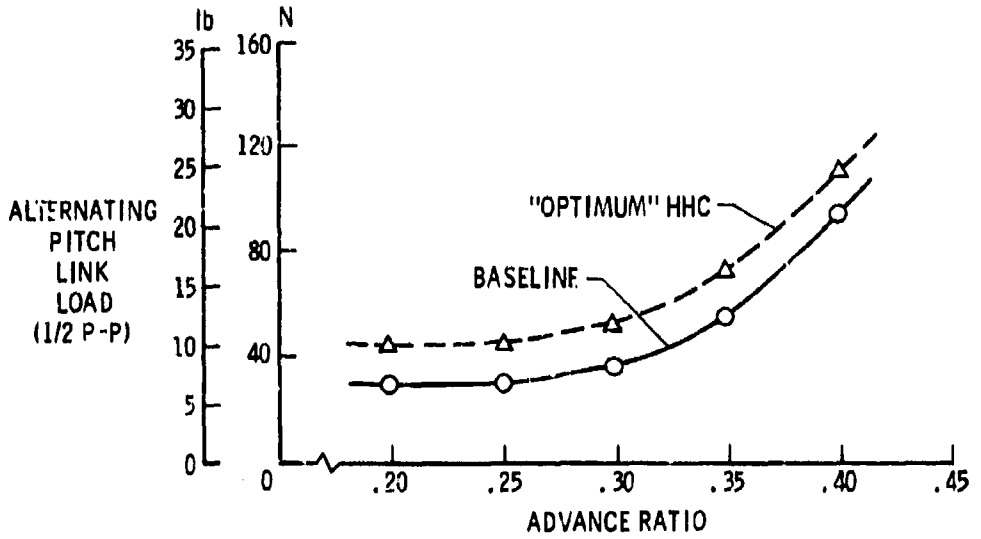



Figure 10. Variation of alternating pitch link load (½ peak-to-peak values) with advance ratio.

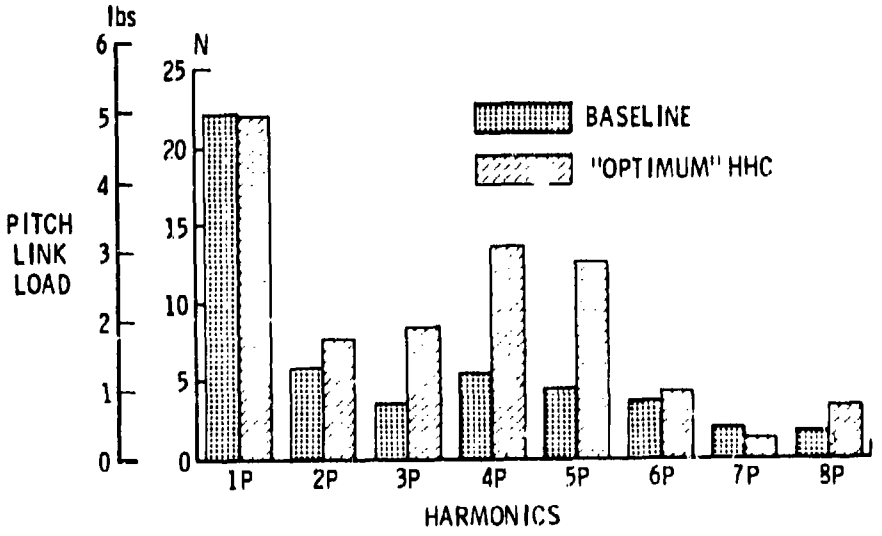



Figure 11. Harmonic decomposition of pitch link load at an advance ratio of 0.3.

## CONCLUDING REMARKS

Results have been presented from a wind tunnel test of a dynamically-scaled helicopter rotor model in which an active control system employing higher harmonic blade pitch was used for helicopter vibration reduction. This test was the first time that an adaptive control system employing optimal control theory has been used for this purpose. The test was successful in that the control algorithms functioned flawlessly and significant reductions in vibratory responses were achieved. An open issue remains, however, as to why even greater reductions in the vibratory responses were not obtained. Further testing is being conducted with the model to resolve this question.

The test results indicate that higher harmonic control can lead to increases in blade and control system loads. For the model tested, increases were evident in the edgewise bending and torsional moments, as well as the pitch link loads. Although the increased loads were considerably below the design limits for the model tested, the fact that blade and control system loads can increase must be considered in any flight test demonstration of the higher harmonic control concept.

Further wind tunnel testing of the active control concept presented in this paper will be conducted in August 1980. Preparations are also underway for a flight test demonstration of the wind-tunnel-developed system. The flight tests will be conducted under contract by Hughes Helicopters using an OH-6A helicopter early in 1981.

## REFERENCES

1. Desjardins, R. A., and Hooper, W. E.: Rotor Isolation of the Hingeless Rotor BO-105 and YUH-61 Helicopters. Second European Rotorcraft and Powered Lift Aircraft Forum, Paper No. 13, Sept. 1976.

2. Ricks, R. G., and Gabel, R.: Vibration Optimization of the CH-47C Helicopter Using NASTRAN. Symposium on Mathematical Modeling in Structural Engineering, Langley Research Center, Oct. 1979.

3. Amer, K. B., and Neff, J. R.: Vertical Plane Pendulum Absorbers for Minimizing Helicopter Vibratory Loads. J. American Helicopter Society, Vol. 19, No. 4, Oct. 1974, pp. 44-58.

4. Paul, W. F.: Development and Evaluation of the Main Rotor Bifilar Absorber. Proceedings of the 25th Annual National Forum, American Helicopter Society, May 1969.

5. Flannelly, W. G.: The Dynamic Antiresonant Vibration Isolator. Proceedings of the 22nd Annual National Forum, American Helicopter Society, May 1966.

6. Wood, E. R., and Powers, R. W.: Practical Design Considerations for a Flightworthy Higher Harmonic Control System. Proceedings of AIAA/ASME/ASCE/AHS 21st Structures, Structural Dynamics, and Materials Conference, May 1980.

7. Sissingh, G. J., and Donham, R. E.: Hingeless Rotor Theory and Experiment on Vibration Reduction by Periodic Variation of Conventional Controls. Rotorcraft Dynamics, NASA SP-352, February 1974.

8. McHugh, F. J., and Shaw, J., Jr.: Benefits of Higher Harmonic Blade Pitch: Vibration Reduction, Blade-Load Reduction and Performance Improvement. Proceedings of the Aermican Helicopter Society Mideast Region Symposium on Rotor Technology, August 1976.

9. Hammond, C. E.: Helicopter Vibration Reduction Via Higher Harmonic Control. Proceedings of the Rotorcraft Vibration Workshop, NASA Ames Research Center, February 22-23, 1978.

10. Bryson, A. E., Jr., and Ho, Y. C.: Applied Optimal Control. John Wiley & Sons, 1975.

11. Goodwin, G. C., and Payne, R. L.: Dynamic System Identification: Experiment Design and Data Analysis. Academic Press, 1977.

12. Hammond, C. E., and Weller, W. H.: Wind-Tunnel Testing of Aeroelastically Scaled Helicopter Rotor Models (U). 1976 Army Science Conference, West Point, NY, June 22-25, 1976.

13. vonDoenhoff, A. E., Braslow, A. L., and Swartzberg, M. A.: Studies of the Use of Freon-12 as a Wind Tunnel Testing Medium. NACA TN-3000, 1958.

14. Hammond, C. E., and Weller, W. H.: Recent Experience in the Testing of a Generalized Rotor Aeroelastic Model at Langley Research Center. Second European Rotorcraft and Powered Lift Aircraft Forum, Paper No. 35, Sept. 1976.

# DELIBERATE AIR POLLUTION: THE ART OF SMOKE SCREENING (U)

FRANK V. HANSEN, MR.
RICARDO PENA, MR.
*ROBERT K. UMSTEAD, MR.

US Army Atmospheric Sciences Laboratory
White Sands Missile Range, New Mexico 88002

## INTRODUCTION

Smoke screening as a battlefield countermeasure is highly dependent upon a number of atmospheric parameters, not a function of munition expenditures alone. Consideration must be given to ambient atmospheric conditions, forecasts, and the aptly named "fog of war" and its affects upon the optical characteristics of the atmosphere. Determination of the optimum smoke density to render the atmosphere opaque to energy from the visible through the far infrared band of the electromagnetic spectrum can be accomplished by an algorithm based upon atmospheric optics and turbulent diffusion hypotheses. The algorithm KWIK was initially conceived as a munitions expenditures model but rendered versatile enough to be used for large area screening and the "seeability" on a battlefield.

The philosophy behind the design and implementation of the algorithm was to keep it simple by using conventional, well-proven formulae and calculating variables within the program from regression fits to easily obtained or measured atmospheric parameters. The model was formatted in a modular sense such that as better experimental results became available changes could be made with minimum effort.



PRECEDING PAGE BLANK-NOT FILMED

## BACKGROUND

The decision to develop a multispectral obscuration model immediately led to a straw man outline with secondary and tertiary options for the structure of the algorithm to assure flexibility. The advantages and disadvantages of atmospheric optic and diffusion postulates that could be used were weighed against the meteorological observations that would be available on a battlefield. A determination was made that eight available surface and sensible weather observations were basic to the development of a workable obscuration scheme. As a consequence, the atmospheric optics model selected was one proposed by Downs [1], being compatible with the required meteorological data. The optics model was then complemented by a diffusion approach advocated by Gifford [2]. The complementary models require a methodology for classifying atmospheric stability. The six category Pasquill [3] scheme was selected on the basis of compatibility with the input parameters. The relationships between the Pasquill categories and observational data and the eight meteorological inputs to the model are given in tables 1a. and 1b.

TABLE 1a. RELATION OF PASQUILL CATEGORIES TO WEATHER CONDITIONS

A - Extremely unstable
B - Moderately unstable
C - Slightly unstable
D - Neutral
E - Slightly stable
F - Moderately stable

| Surface Windspeed (m/s) | Daytime Insolation | | | Nighttime Conditions | |
|---|---|---|---|---|---|
| | Strong | Moderate | Slight | Thin Overcast or $\geq$ 4/8 Cloudiness | $\leq$ 3/8 Cloudiness |
| < 2 | A | A-B | B | | |
| 2 | A-B | B | C | E | F |
| 4 | B | B-C | C | D | E |
| 6 | C | C-D | D | D | D |
| 6 | C | D | D | D | D |

TABLE 1b. METEOROLOGICAL INPUT PARAMETERS FOR STABILITY, OPTICS, AND DIFFUSION CALCULATIONS

| | |
|---|---|
| Ceiling height (feet) | Temperature (degrees F) |
| Cloud cover (percent) | Dew point (degrees F) |
| Visibility (miles) | Wind direction (degrees) |
| Precipitation (yes or no) | Windspeed (knots) |

The Gaussian plume and puff hypothesis chosen to model semicontinuous and quasi-instantaneous sources, uses longitudinal, lateral, and vertical dispersion parameters that are stability dependent and, in the case of vertical diffusion, a function of the aerodynamic roughness of the earth's surface. The dispersion parameters are calculated using power laws attributed to Pasquill [3]. Aerodynamic roughness is evaluated using the approach of Kung [4], based upon the height of the roughness elements, i.e., trees, bushes, grasses, etc.

The straw man approach highlighted additional factors that would influence obscuration on the battlefield including (1) relative humidity effects upon the hygroscopic characteristics of the smoke aerosols, (2) the need for optimum impact separations of artillery-delivered smoke projectiles, (3) rate of fire calculations for efficient dissemination of chemical smokes, and (4) munition expenditure estimates used for planning future operations.

## THE FOG OF WAR

The visual range in the atmosphere is directly affected by natural aerosol concentrations and particle size distributions occurring over any optical path. Visibility may be reduced by dry haze, wet or relative humidity haze, fog, or air pollution. Battlefield visibility can be compromised by the additive effects of the fog of war, i.e., pollution induced by deliberate smoke screening, dust thrown up by the mass movement of heavy vehicles and intense artillery barrages, or smokes from burning vehicles. The reduction of visibility by natural aerosols plus the fog of war had a direct effect upon countermeasure obscuration used to deny target acquisition.

Battlefield pollution coupled with the liquid water content of the atmosphere control the aerosol scattering and water vapor absorption of visible light and infrared radiation in the

atmosphere. The attenuation of an optical path by scattering and absorption results in a higher threshold level of detection. The net effect is a need for less smoke to render an optical path opaque. The KWIK algorithm was designed to take these phenomena into account, i.e., aerosol mass concentrations with respect to battlefield visiblity and attenuation by water vapor as a function of relative humidity.

The inclusion of visibility in the determination of extinction coefficients for finite optical paths is a major factor in the reduction of munitions expenditures necessary to establish and maintain obscuring screens on a battlefield. Conversely, as discussed below, the KWIK algorithm, with some modifications, can be used to estimate the degradation of visibility by the products of a dirty battlefield.

## SMOKE SCREENING AND OBSCURATION

Countermeasure obscuration may be considered from two viewpoints, the first being the attenuation of near horizontal optical paths where the crosswind integrated concentration of smokes such as white phosphorous or zinc chloride is of prime importance. The second is large area screening whereby detection of ground targets by airbone observers or remotely piloted vehicles is denied. Here the vertical or slant path integrated concentration is the prime parameter. The determination of the horizontal or vertical integrated concentrations necessary to obscure an optical path to the threshold level is almost wholly dependent upon the basic optical calculations. The optics portion of KWIK is adapted from an approach to atmospheric transmission suggested by Downs [1]. Transmittance of light at various wavelengths through a path is determined by calculating the attenuation due to absorption by water vapor, scattering by haze or fog, and precipitation. When the attenuation due to atmospheric conditions is known, the attenuation due to smoke that is required to lower tansmittance to a threshold contrast can be computed.

Absorption is attributable to the amount of precipitable water in a path, assuming the water vapor concentration in the atmosphere is well behaved and exhibits a scale height of about 2 km. The water vapor concentration expressed in centimeters per kilometer of path length may then be given as:

$$W = W_o e^{-(L \sin \theta)/2} , \qquad (1)$$

where $W_o$ is the precipitable water along a path L and $\theta$ is the angle between the horizontal and the height of a target above or below an observer. $W_o$ is computed from a regression equation relating precipitable water and dew point temperature ($T_d$):

$$W_o = 0.4477 + 0.0328T_d + 1.2(10)^{-3}\, T_d^2 + 1.84(10)^{-5}\, T_d^3 \qquad (2)$$

Equation (2) was fit to data extracted from Downs [1] and is considered valid for any meteorological condition.

The amount of water vapor in the path, W, is given by:

$$W = W_o \int_o^L e^{-(L \sin \theta)/2}\, dL \,. \qquad (3)$$

Transmission through the absorbing component of the atmosphere is calculated by using an error function absorption law developed by Elsasser [5]

$$T = 1 - \mathrm{erf}\,(z) \,, \qquad (4)$$

$$\mathrm{erf}\,(z) = \frac{2}{\sqrt{\pi}} \int_o^z e^{-z^2}\, dz \,, \qquad (5)$$

where $z = 0.5\ \beta\ \sqrt{\pi W}$, and $\beta$ is the error function absorption coefficient as a function of wavelength.

Downs [1] states that the Elsasser approach is unable to correctly address long wavelengths and suggests using an approach described by Fisher [6] for the far infrared wavelengths, with the computation of transmission due to absorption by water vapor given by:

$$T = e^{-0.0681W} \,. \qquad (6)$$

Reduction in transmittance due to attenuation by haze and fog can be calculated by using Mie theory. Downs [1] indicates that the Mie scattering coefficient decreases with altitude such that its

behavior can only be estimated. The following expressions for $\sigma_M$ (Mie scattering coefficient) are, at best, an approximation to the behavior of the $\sigma_M$ versus altitude relationship

$$\sigma_M = \sigma_{hf} e^{-L \sin \theta / 4.1} ; \; V_R \geq G(\lambda) , \tag{7}$$

$$\sigma_{M_1} = \sigma_{hf} e^{L \sin \theta \ln (0.1/\sigma_{hf})} ; \; V_R < G(\lambda) , \; 0 < L \sin \theta \leq 1 \text{ km}$$

$$\sigma_{M_2} = 0.128 e^{-L \sin \theta / 4.1} ; \; V_R < G(\lambda) , \; 1 \text{ km} \leq L \sin \theta < \infty$$

$$\sigma_M = \sigma_{M_1} \left( \sigma_{M_2} \right) , \tag{8}$$

where $\sigma_{hf}$ is the extinction coefficient determined from a linear regression as a function of visibility and wavelength, based upon Downs' evaluation. $V_R$ is visibility, and $G(\lambda)$ is the scale height of $\sigma_M$. $G(\lambda)$ is not constant; rather it is a function of altitude, visibility, and wavelength. Transmission along a path with attenuation $\sigma_M$ can be determined by the equation

$$T = e^{- \int_0^L \sigma_M(L) \, dL} , \tag{9}$$

after substituting a value for $\sigma_M$ according to equations (7) and (8). If precipitation is indicated (by input parameter), then the value for transmittance in equation (8) is set to one and a calculation is made for attenuation by precipitation instead.

Reduced transmittance owing to attenuation by precipitation can be obtained from

$$T = e^{- \int_0^L \sigma_r(L) \, dL} , \tag{10}$$

where L is the path length and $\sigma_r$ is an attenuation coefficient determined from a regression equation as a function of visibility and

wavelength. The total transmittance along an optical path is the product of the partial transmittances

$$T_{total} = T_a T_{hf} T_p T_s \,, \tag{11}$$

where

$T_a$ = Transmittance due to attenuation by absorption

$T_{hf}$ = Transmittance due to attenuation by haze and fog

$T_p$ = Transmittance due to attenuation by precipitation

$T_s$ = Transmittance due to attenuation by smoke

$T_s$ is then calculated from equation (11), and the desired threshold contrast of transmittance for a particular wavelength can be determined from:

$$T_s = \frac{T_{tc}}{T_a T_{hf} T_p} \,, \tag{12}$$

where $T_{tc}$, the threshold contrast, is based upon the Koschmieder [7] theory but set equal to 0.10 for visible wavelengths and 0.05 for infrared.

The approach used to determine the line of sight integrated concentration, CL, of the smoke screen necessary to attenuate an optical path to a threshold level is based on the transmittance of equation (12) as a function of CL. The CL-value necessary to attenuate an optical path to deny target acquisition may be determined from the Bouguer-Beer law written as

$$T_s = \frac{I}{I_o} = e^{-\alpha CL} \tag{13}$$

where I is the illuminance at a target, $I_o$ is the illuminance at the light source, $\alpha$ the extinction coefficient, and $T_s$ the reciprocal of attenuation. Rearranging and solving equation (13) yields

$$CL = \frac{\ln T_s}{-\alpha} \tag{14}$$

where CL is the minimum integrated smoke concentration.

Equation (14) is applicable to both visible and infrared wavelengths, if values of $\alpha$ are known. Listed in table 2 are average extinction coefficients for the two screening chemical smokes considered by the KWIK algorithm, i.e., bulk white phosphorous (WP) and hexachloroethane (HC) (zinc chloride) for the indicated spectral bands.

TABLE 2. EXTINCTION COEFFICIENTS FOR WHITE PHOSPHOROUS AND HEXACHLOROETHANE GENERATED SMOKES

| Spectral Band | Wavelength ($\mu$m) | $\alpha$, $m^2 g^{-1}$ | |
|---|---|---|---|
| | | Zinc Chloride | White Phosphorous |
| Visible | 0.4 to 0.7 | 3.30 | 2.46 |
| Near IR | 0.75 to 1.2 | 1.50 | 1.50 |
| Mid IR | 3 to 5 | 0.12 | 0.21 |
| Far IR | 8 to 14 | 0.05 | 0.28 |

The development of diffusion formulae for predicting the obscuring power of chemical smokes starts with the assumption of an instantaneous point source of material diffusing in three dimensions. For a Gaussian distribution of diffusion taking place independently in the three coordinate directions, the equation can be stated as

$$\chi_{(x,y,z)} = \frac{(2\pi)^{-3/2}\, Q_T}{\sigma_x\, \sigma_y\, \sigma_z} \exp\left\{ -\frac{1}{2}\left[\left(\frac{x - \overline{V}t}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2 + \left(\frac{z}{\sigma_z}\right)^2\right]\right\} \qquad (15)$$

where $\chi$ is concentration in g $m^{-3}$, $Q_T$ the total release of material in g, x, y, z the coordinate directions, $\overline{V}$ the mean windspeed in m $s^{-1}$, t is time, and $\sigma_x$, $\sigma_y$, $\sigma_z$ the dispersion parameters.

Integration of equation (15) yields the continuous source equation for sources and receptors near the ground as

$$\chi = \frac{Q}{\pi\, \overline{V}\, \sigma_y\, \sigma_z} \exp\left[ -\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2 + \left(\frac{z}{\sigma_z}\right)^2\right] \qquad (16)$$

where Q is a time rate of release. If equation (16) is integrated, the result is the crosswind integrated concentration (CWIC) of the plume

$$\chi_{CWIC} = \left(\frac{2}{\pi}\right)^{1/2} \frac{Q}{\bar{V}\,\sigma_z} \exp\left[-\frac{1}{2}\left(\frac{z}{\sigma_z}\right)^2\right] \tag{17}$$

which is the basic form for estimating obscuration from sources generated by the HC smoke mix. The vertical dispersion parameter $\sigma_z$ is determined using the Pasquill [3] power law in the form

$$\sigma_z = cx^d \,. \tag{18}$$

The coefficient and index values will be discussed at the end of the section.

Screening and obscuration are not restricted to the along wind case of equation (16); consequently, a wind direction correction factor must be considered for head, tail, and quartering wind conditions. The correction factor may be derived from considerations concerned with finite line sources, cumulative effects of multiple sources, and discrete point line sources. The correction factor is eliptical and given by

$$\delta^2 = \frac{m^2 n^2}{m^2 \sin^2\theta + n^2 \cos^2\theta} \tag{19}$$

where m = 3.71 and n = 1, the eliptical semi-axes, and $\theta$ the angle between the mean wind direction and the optical path.

The exponential term on the right hand side of equation (17) is only partially sensitive to stability and downwind travel distances associated with smoke screening. Consequently, numerical evaluation for the six stability categories and along wind distances of 50 to 150 m shows that $\exp[-1/2(z/\sigma_z)^2]$ may be taken as constant and set equal to 0.916.

The assumption can be made that the line of sight integrated concentration CL of the screening aerosol calculated from equation (14) is equal to the value of $\chi_{CWIC}$ evaluated from equation (17). This allows equation (17) to be rearranged, after substitution of equation (18) and consideration of the ramifications of equation (19) and solved for the along wind travel distance x. Included in

the solution are the source efficiency term $\lambda$ and the relative humidity-related yield factor $\Omega$, which serve to modify the source strength Q. Thus

$$\chi = \delta^{-1}\left[\frac{0.731\ \lambda\ \Omega\ Q}{C\ \overline{V}\ CL}\right]^{d^{-1}} \tag{20}$$

after combining all the constants. The integrated concentration required for obscuration reaches a minimum at the distance x downwind from the source. Accordingly, x is the calculated impact separation of the smoke projectiles.

Obscuration calculations for quasi-instantaneous sources, i.e., as generated by bulk WP munitions requires that the integrated concentration equation be written in the form

$$\chi_{CWIC} = \frac{\lambda\ \Omega\ Q_T}{\pi\ \sigma_{x_I}\ \sigma_{z_I}}\ \exp\left\{-\frac{1}{2}\left[\left(\frac{x - \overline{V}t}{\sigma_{x_I}}\right)^2 + \left(\frac{\overline{Z} - z}{\sigma_{z_I}}\right)^2\right]\right\} \tag{21}$$

where the dispersion parameters $\sigma_{x_I}$ and $\sigma_{z_I}$ are not to be confused with those associated with a continuous source. The term $\overline{Z}$ represents the height of the puff centroid above the surface. Owing to the nature of the exponential on the right hand side of equation (21) which requires knowledge of two downwind travel distances and three heights for solution, the integrated concentrations were determined for a unit downwind distance of 100 m.

Bulk WP is an exothermal smoke source, with only a fraction of the total material available for screening. The major portion of the phosphorous smoke is transported vertically in the thermal plume. An analysis of available data yielded the efficiencies or percent of payload available shown in table 3. Also tabulated in table 3 are numerical evaluations of the exponential term of equation (21) for each stability category (K factors).

TABLE 3. EFFICIENCIES AND K FACTORS FOR BULK WHITE PHOSPHOROUS SMOKE MUNITIONS AS A FUNCTION OF PASQUILL STABILITY CATEGORY

| Pasquill Category | K Factor | Efficiency |
|---|---|---|
| A | 0.4633 | 0.07 |
| B | 0.3631 | 0.10 |
| C | 0.2056 | 0.14 |
| D | 0.0647 | 0.28 |
| E and F | 0.0725 | ≈ 0.30 |

This allows equation (21) to be restated as

$$\chi_{CWIC} = \frac{K \lambda \Omega Q_T}{\pi \sigma_{x_I} \sigma_{z_I}} . \tag{22}$$

Projectile impact separations can now be determined from

$$x = 100 \frac{\chi_{CWIC}}{CL} \tag{23}$$

where the constant is the unit screen length and CL is calculated from equation (14).

Large area screening can be treated by integration of equation (16) in the vertical yielding

$$\chi_{VIC} = \left(\frac{2}{\pi}\right)^{1/2} \frac{Q}{\bar{V} \sigma_y} \exp\left[-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2\right] \tag{24}$$

which is applicable to the use of fog oil generators. The rendering opaque of large areas such as air fields is dependent upon the inherent and apparent contrast between two objects on the ground with respect to the sky-ground ratio and the threshold contrast of the objects. The problem of vision looking downward from aircraft has

been discussed by Duntley [8] and Middleton [9]. Middleton suggests that the threshold contrast over a slant

range $\overline{R}$ is given by

$$\varepsilon = C_o \left[ 1 - \left( \frac{B_m}{B_o} \right) \left( 1 - \exp 3.912 \, \overline{R}/V_r \right) \right]^{-1} \tag{25}$$

where $C_o$ is the inherent contrast, $B_m/B_o$ the sky-ground ratio, and $V_r$ the visual range. Values of $\varepsilon$ so determined may be used in equation (14) to evaluate slant range integrated concentrations. Sherwood [10] found that $\chi_{VIC}$'s of 0.33 g $m^{-2}$ were necessary to screen a large area from aerial observation.

The obvious use of equations (24) and (25) is the determination of the number and separation of oil fog generators required to screen a large area. Equation (24) can be manipulated to yield the separation distance $y_s$ as

$$y_s = 2 \, \sigma_y \left[ \ln \left( \frac{2}{\pi} \right)^{1/2} \frac{Q}{\overline{V} \, \sigma_y \, \chi_{VIC}} \right]^{1/2} \tag{26}$$

and the number of generators $N_g$ required by dividing the screen width L by $y_s$ plus one, or

$$N_g = \frac{L}{y_s} + 1 \, . \tag{27}$$

The dispersion coefficients $\sigma_y$, $\sigma_z$, $\sigma_{x_I}$, and $\sigma_{z_I}$ are based upon the Pasquill [3] power laws and given by

$$\sigma_y = a x^b \tag{28}$$

$$\sigma_z = c x^d \tag{29}$$

$$\sigma_{x_I} = \sigma_{x_o} + 0.74 \, a x^b \tag{30}$$

$$\sigma_{z_I} = \sigma_{z_o} + 0.667\, cx^d \qquad (31)$$

where $\sigma_{x_o}$ and $\sigma_{z_o}$ are the initial dispersion or burst functions. Downwind dispersion parameters for the quasi-instantaneous sources are approximately two-thirds of those for continuous sources as shown by Pasquill [3], which is reflected by equations (30) and (31). The coefficients and indices for the dispersion parameters as a function of stability and aerodynamic roughness are tabulated in table 4 for three roughness lengths.

TABLE 4. COEFFICIENT AND INDEX VALUES FOR THE POWER LAW DISPERSION PARAMETERS

| Stability Category | a | b | c ($z_o$ = 1 cm) | d ($z_o$ = 1 cm) | c ($z_o$ = 10 cm) | d ($z_o$ = 10 cm) | c ($z_o$ = 100 cm) | d ($z_o$ = 100 cm) |
|---|---|---|---|---|---|---|---|---|
| A | 0.40 | 0.90 | 0.154 | 0.94 | 0.279 | 0.90 | 0.615 | 0.83 |
| B | 0.32 | 0.90 | 0.133 | 0.89 | 0.225 | 0.85 | 0.539 | 0.77 |
| C | 0.22 | 0.90 | 0.121 | 0.85 | 0.213 | 0.81 | 0.533 | 0.72 |
| D | 0.143 | 0.90 | 0.108 | 0.81 | 0.195 | 0.76 | 0.456 | 0.68 |
| E | 0.102 | 0.90 | 0.078 | 0.78 | 0.139 | 0.73 | 0.348 | 0.65 |
| F | 0.076 | 0.90 | 0.062 | 0.72 | 0.117 | 0.67 | 0.309 | 0.58 |

## SEEABILITY ON THE BATTLEFIELD

A battlefield may be considered as being mesometeorological in scale, i.e., areas ranging from hundreds to thousands of square kilometers. If the density of meteorological observations is large and timely, the optics portion of KWIK may be used to calculate the attenuation of optical paths for each weather observational site. The attenuations may be plotted and analyzed much like synoptic data to prepare "seeability" charts for a battlefield. Battle plans for future engagements with estimates of munition expenditures anticipated number of burning vehicles and vehicular dust conditions can be used to predict visibility conditions which may be used for attenuation forecasts. Seeability, prognostications can be used for planning purposes, i.e., what weapons system will be effective on the next day's predicted dirty battlefield.

## DISCUSSION

The complete KWIK smoke obscuration model has been discussed by Umstead, Peña, and Hansen [11], including the development of the scheme to establish rates of fire, impact separations in adverse wind conditions, and the special considerations for munitions expenditures in the mid and far infrared regions of the spectrum. Owing to space limitations, these subjects are beyond the scope of this paper.

The basic scheme developed for artillery delivered smoke projectiles assumes that smoke obscuration operations will be conducted with batteries or battalions firing in an open sheaf rather than parallel or normal sheaf patterns. This concept, coupled with relative humidity dependent yield factors in the smoke model, will result in a more efficient use of smoke on a battlefield. Preliminary studies suggest savings in munitions up to 20 to 30 percent over current methods.

The use of the KWIK algorithm for large area screening operations and predicting attenuation degradation on the dirty battlefield does not detract from the original intent of the model, but enhances its capabilities. The outputs of these offshoots of the primary model can be utilized to improve the munition expenditure estimates generated for countermeasure obscuration purposes. Large area screening systems utilizing oil fogs are only usable in the visible portion of the spectrum. Consequently, the portion of KWIK expressed by equations (24), (25), and (26) only apply to the 0.4- to 0.7-μm band.

Owing to the extinction characteristics of HC smoke, the approach personified by equation (20) is valid only in the visual and near infrared portions of the electromagnetic spectrum. The superior characteristics of WP allow it to be utilized to countermeasure devices operating in the mid and far infrared regions.

## CONCLUSIONS

The KWIK obscuration model is highly versatile and has been programmed to operate on a variety of machines ranging from programmable desk calculators to digital computers. Three versions are in existence: the Fortran, the real-time with three options, and a deferred-time version for generating munition expenditure estimate tables based upon climatological input data. The algorithm may be used in threat analysis studies or as a subroutine in force on force

scenarios. The modular concept used to develop KWIK allows greater flexibility than found in predecessor smoke models.

## REFERENCES

1. Downs, A. R., 1976, "A Review of Atmospheric Transmission Information in the Optical and Microwave Spectral Regions," Ballistics Research Laboratories Report 2710.
2. Gifford, F. A. 1968, "An Outline of Theories of Diffusion in the Lower Layers of the Atmosphere," Meteorology and Atomic Energy (D. Slade, ed.) US Atomic Energy Commission, Washington, DC.
3. Pasquill, F., 1974, Atmospheric Diffusion, 2nd ed., Halsted Press, div. of John Wiley & Sons, Inc., NY.
4. Kung, E. C., 1963, "Climatology of Aerodynamic Roughness Parameter and Energy Dissipation in the Planetary Boundary Layer of the Northern Hemisphere," Annual Report, 1963, Studies of Variations in Boundary Conditions on the Atmospheric Boundary Layer, Contract DA-36-039-AMC-00878, Department of Meteorology, University of Wisconsin, Madison, WI.
5. Elsasser, W. M., 1942, "Heat Transfer by Infrared Radiation in the Atmosphere," Harvard Meteorological Series 6, Harvard University Press, Cambridge, MA.
6. Fisher, D. F. et al., 1963, "Transmissometry and Atmospheric Transmission Studies Final Report," University of Michigan, Institute of Science and Technology.
7. Koschmieder, H., 1924, "Theorie der Horizontalen Sichtweite," Beitr Phys Frein Atmos, 12:33-53, 171-181.
8. Duntley, S. Q., 1948, "The Visibility of Distant Objects" JOSA, 38:237-249
9. Middleton, W. E. K., 1952, Vision Through the Atmosphere, University of Toronto Press, Toronto, Canada.
10. Sherwood, T. K., 1949, "The Geometry of Smoke Screens," J Meteorol, 6:416-419.
11. Umstead, Robert K., Ricardo Pena, and Frank V. Hansen, 1979, "KWIK: An Algorithm for Calculating Munition Expenditures for Smoke Screening/Obscuration in Tactical Situations," ASL-TR-0030, US Army Atmospheric Sciences Laboratory, White Sands Missile Range, NM.

# ANALYSIS OF MULTIPASS LASER AMPLIFIER SYSTEMS FOR STORAGE LASER MEDIA (U)

JAMES F. HARVEY, LTC, SC
MILITARY RESEARCH ASSOCIATE GROUP
LAWRENCE LIVERMORE NATIONAL LABORATORY
LIVERMORE, CA. 94550

## I. Introduction

Highly efficient short pulse high power lasers have many potential applications, including laser fusion drivers. One technique for achieving high powers in short pulses is to use a storage laser amplifier. A storage laser amplifier uses a laser medium with a long lived upper laser level. The upper laser level can accumulate energy from a pumping source over a relatively long time. This stored energy is then extracted by stimulated emission over a relatively short time. Examples of such storage laser media are Nd:YAG, Nd:Glass, V:$MgF_2$, Tm:Glass, $CO_2$, and Group VI media (e.g. Sulfur and Selenium).

The single pass amplifier system depicted in figure 1 (top) is the simplest approach to amplifying a laser light pulse. The laser beam is passed once through the laser medium. The beam is amplified as it extracts energy stored in the medium's upper laser level. The single pass amplifier performance is limited in that it cannot simultaneously provide high energy gain and high efficiency. Under certain conditions these limitations can be overcome by using a multipass system such as the one depicted in the lower part of figure 1.

In this paper a single pass amplifier is investigated first using the Frantz-Nodvik theory of short pulse laser amplification.[1-5] The multipass system is then treated by sequentially applying the single pass extraction equations for each extraction pass. In order to find the gain coefficient for each extraction, the changes occurring in the laser medium and in the laser beam fluence between extraction passes are determined using a simple three level laser kinetics model. The single pass and



PRECEDING PAGE BLANK-NOT FILMED

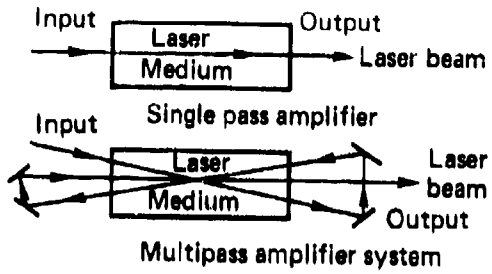



Fig. 1

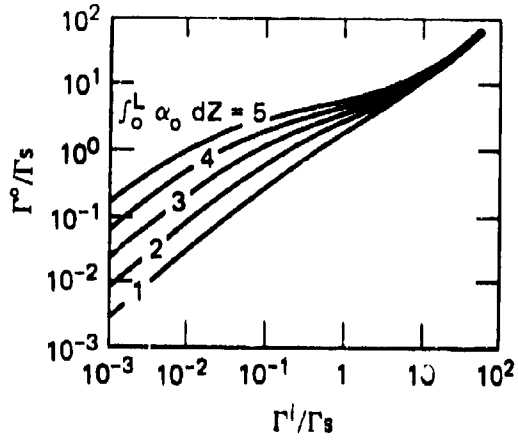



Fig. 2

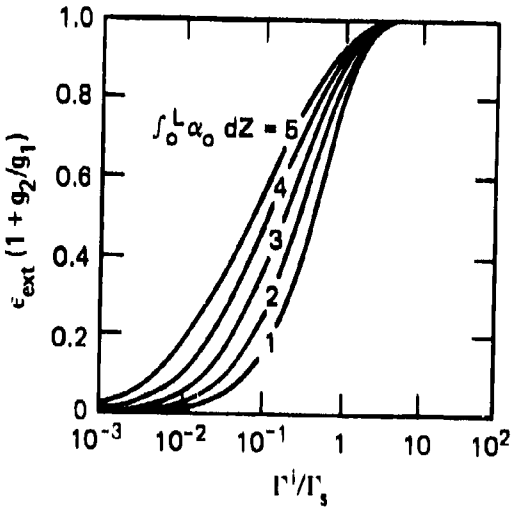



Fig. 3

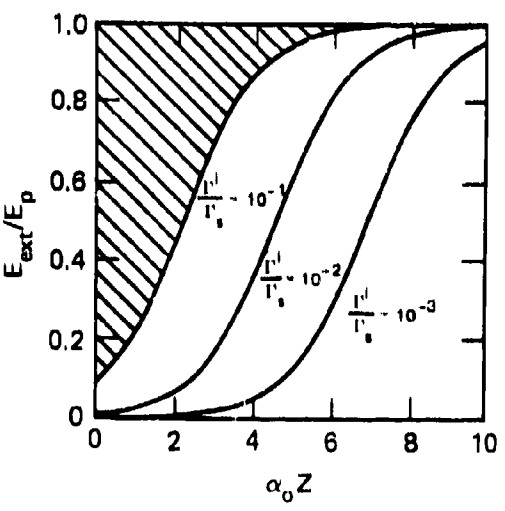



Fig. 4

multipass results are compared to determine when a multipass system has an advantage over the single pass amplifier The effects of the variation of different medium and system parameters on amplifier performance are then discussed. Finally several specific storage amplifier systems which are candidates for laser fusion drivers are discussed.

## II. Single Pass Extraction

In this section the extraction characteristics of a single pass amplifier are considered. First the equations required to calculate the behavior of a single pass amplifier are developed. These equations are then used to study the extraction characteristics of these amplifiers.

For practical systems the laser pulse length and extraction period are generally short compared with the upper and lower level lifetimes. This condition on the lower level lifetime is called "bottlenecked" extraction. Then during the extraction period the only changes in the upper and lower laser level populations, $n_2$ and $n_1$, are due to the stimulated transitions induced by the photon field of the laser beam. The rate equations for the populations of the two laser levels and the photon field become

$$(1/C)\partial I/\partial t + \partial I/\partial z = dI/dz = \alpha I \tag{1}$$

and

$$\partial \alpha/\partial t = -\alpha I/\mathcal{F}_s, \tag{2}$$

where the laser beam propagation is in the z direction and I is the intensity of the laser light. The gain coefficient, $\alpha$, is defined by

$$\alpha \equiv \sigma[n_2 - (g_2/g_1)\, n_1], \tag{3}$$

where $\sigma$ is the cross section for stimulated emission and $g_2/g_1$ is the upper to lower level degeneracy ratio. The amplifier medium loss is generally negligible. The saturation fluence is defined as

$$\mathcal{F}_s \equiv h\nu/[\sigma(1 + g_2/g_1)], \tag{4}$$

where $h\nu$ is the energy of the laser photon. Eqns. (1) and (2) can be solved to give

$$\mathcal{F}^o/\mathcal{F}_s = \mathrm{Ln}[\exp(\int_0^L \alpha_o(t_e)dz)\{\exp(\mathcal{F}^i/\mathcal{F}_s)-1\}+1]. \tag{5}$$

In this equation $\Gamma$ is the laser pulse energy fluence and is defined as

$$\Gamma \equiv \int_{-\infty}^{\infty} I(z,t)dt. \tag{6}$$

$\Gamma^i$ and $\Gamma^o$ are the input and output fluences respectively. The small signal gain coefficient, $\alpha_o$, in the gain-length integral is the gain coefficient immediately prior to extraction and is evaluated at the time $t_e$ of extraction.[1-4]

The output fluence thus depends only on the input fluence normalized by the saturation fluence and on the integral of the gain coefficient along the amplifier axis. It does not depend on the specific distribution of values of the gain coefficient along the amplifier axis.

When the signal being amplified remains much less than a saturation fluence then from eqn. (5)

$$\Gamma^o/\Gamma_s \approx (\Gamma^i/\Gamma_s)\exp(\int_0^L \alpha_o dz). \tag{7}$$

The output fluence depends exponentially on the gain-length integral. Conversely when the signal being amplified becomes greater than $\Gamma_s$ then

$$\Gamma^o/\Gamma_s \approx (\Gamma^i/\Gamma_s) + \int_0^L \alpha_o dz. \tag{8}$$

In this case the output fluence depends linearly on the gain-length integral. This behavior can be seen in figure 2, which is a graph of equation (5) for several possible values of the gain-length integral.

The laser beam energy extraction efficiency is defined as

$$\varepsilon_{ext} \equiv (\Gamma^o - \Gamma^i)/(E_p L) \tag{9}$$

where L is the amplifier length and $E_p$ is the pump energy density deposited in the upper laser level. Eqn. (9) can be rewritten as

$$\varepsilon_{ext} = (\Gamma^o/\Gamma_s - \Gamma^i/\Gamma_s)/[(1 + g_2/g_1)(\alpha_o L)_M], \tag{10}$$

where $(\alpha_o L)_M$ is the maximum possible gain-length integral for a given pump energy density, namely,

$$(\alpha_o L)_M = (\sigma/h\nu)E_p L = E_p L/[\Gamma_s(1 + g_2/g_1)]. \tag{11}$$

Efficient extraction only occurs at input fluences comparable to or greater than the saturation fluence. Under these circumstances

$$\varepsilon_{ext} \approx \int_0^L \alpha_o dz/[(\alpha_o L)_M(1 + g_2/g_1)] \qquad (12)$$

The factor $\int_0^L \alpha_o dz/(\alpha_o L)_M$ represents the loss due to the depletion in the gain prior to extraction. This depletion is caused by decay from the upper level population during the finite time period in which the pump energy is deposited in the upper laser level. This effect will be investigated during the discussion of the multipass system. To focus attention on the extraction period itself for now, $\int_0^L \alpha_o dz$ will be considered equal to $(\alpha_o L)_M$. Then there is no decay from the upper level and no population in the lower level. The extraction efficiency for such a single pass is plotted in figure 3 as a function of input fluence. For all values of input fluence a larger value of the gain-length integral will result in a larger efficiency for a given $\Gamma^i/\Gamma_s$. This effect of the gain-length on efficiency becomes less pronounced at higher input fluences. At large input fluence, as shown in eqn. (12), the efficiency is limited to the value $1/(1 + g_2/g_1)$ because the net gain becomes zero when $n_2/g_2 = n_1/g_1$. The efficiency is thus very sensitive to the degeneracy ratio. If the extraction period is long compared with the lower level lifetime the extraction is often said to be "unbottlenecked". In this case the above analysis is still valid, but a zero degeneracy ratio should be used for extraction since the lower level population will not live long enough to influence the extraction. The extraction efficiency then will not be so sensitive to the real degeneracy ratio.

Figure 3 shows that the extraction efficiency for a single pass amplifier falls rapidly from unity with decreasing $\Gamma^i/\Gamma_s$ ($\Gamma^i/\Gamma_s \ll 1$). The reason for this reduced efficiency is demonstrated in figure 4, which shows the general behavior of the energy extracted per unit volume, $E_{ext}$, along the length of a laser amplifier for a small input laser light pulse. Initially the light pulse is in the small signal regime and experiences exponential growth. Since the amplified pulse is still small compared with the saturation fluence, it has not extracted a significant portion of the energy stored in the upper laser level. As the pulse is amplified it becomes large enough that it is extracting most of the stored upper level population. The curve of extracted energy density now approaches the available stored energy density. The energy remaining in the form of an upper laser level population after the passage of the extracting laser pulse represents unextracted energy and therefore produces a decrease in extraction efficiency. In fig. 4 the area between the horizontal line $E_{ext}/E_p = 1$ and the extracted energy curve represents the energy left in the amplifier after the extraction process. This area is shaded for the $\Gamma^i/\Gamma_s = 0.1$ curve. In the small signal

regime the pulse amplification is high, but the extraction efficiency is low. In the saturated regime the pulse amplification is low, but the extraction efficiency high. Considerations such as the control of parasitics limit the gain-length for practical amplifiers. For practical amplifiers the saturated extraction regime is only reached after the laser pulse has travelled a significant fraction of the amplifier length, unless the input fluence is comparable to a saturation fluence. So the practical single pass amplifier can provide high energy gain at low extraction efficiency in the exponential growth regime. Alternatively it can provide high extraction efficiency with low energy gain in the saturated extraction regime. But it cannot obtain both high energy gain and high extraction efficiency simultaneously. Efficient extraction begins to occur for input fluence values comparable to a saturation fluence.

## III. Multipass Extraction

These limitations of the single pass amplifier can be overcome by using multipass extraction as illustrated in fig. 1. In such an approach a small input fluence is amplified in the exponential gain regime, and the resulting output pulse returned into the same medium. The returned pulse has sufficient fluence to extract that part of the stored energy remaining after the first extraction pass. High energy gain can then be realized with a higher extraction efficiency than is possible from a single pass amplifier.

In order to analyze the multipass amplifier system it is only necessary to repeat the extraction calculation for $\Gamma^o/\Gamma_s$ for each extraction pass using the appropriate values for $\Gamma^i/\Gamma_s$ and the gain-length integral. The appropriate input fluence value for the nth extraction pass is the output fluence of the previous extraction pass reduced by the optical loss during the turnaround time, namely,

$$\Gamma_n^i/\Gamma_s = (1-T)\Gamma^o_{(n-1)}/\Gamma_s, \quad n \neq 1, \tag{13}$$

where T is the optical transmission coefficient. In this notation the input fluence $\Gamma^i$ to the amplifier system is denoted by $\Gamma^i 1$. The subscript n on a quantity indicates its value prior to or during the nth extraction pass.The value of the gain-length integral to be used in the nth extraction calculation will depend on the prior changes that have taken place in the two laser level populations, $n_2$ and $n_1$, and is given by

$$\int_0^L \alpha_{on}(t_e)dz = \sigma\left[\int_0^L n_{2n}(t_e)dz - (g_2/g_1)\int_0^L n_{1n}(t_e)dz\right]. \tag{14}$$

These changes have occurred during the pump period, during previous extraction passes, and during the optical turnaround time between previous extraction passes. The changes in the populations during an extraction pass are proportional to the energy extracted

$$\Delta \int_0^L n_{1n} dz = -\Delta \int_0^L n_{2n} dz = (r_n^o - r_n^i)/(h\nu). \tag{15}$$

Figure 5 illustrates the model used in analyzing the laser medium. The cross section for stimulated emission between the upper and lower excited states is $\sigma$. The population in the upper state, $n_2$, is also depleted by nonstimulated transitions which may be collisional or radiative. The total rate of depletion of the upper laser level, $n_2/\tau_2$, is characterized by a lifetime, $\tau_2$. A fraction, $f_{21}$, of this upper state decay adds to the lower laser level population, $n_1$, and the remainder decays to unrelated states. The lower state is also characterized by a lifetime $\tau_1$, and the total rate of depletion of the lower laser level is $n_1/\tau_1$.

During the pumping time, $\tau p$, the upper laser level is populated by some form of energy deposition, e.g. flashlamp or laser light pumping in the case of most solid state lasers, photolytic bleaching-wave pumping in the case of the group VI lasers, and collisional pumping in the case of the $CO_2$ laser. The volumetric pump rate $R_p$ is $E_p/\tau_p$ and is assumed constant for time $\tau_p$. The rate of change in the upper level population due to decay and pumping is governed by

$$dn_2/dt = R_p - n_2/\tau_2. \tag{16}$$

The rate of change of the lower laser level population is determined by

$$dn_1/dt = f_{21} n_2/\tau_2 - n_1/\tau_1. \tag{17}$$

Equations (16) and (17) are subject to the initial conditions that the two levels are unpopulated at t = 0 in the pumping period. Under these conditions equations (16) and (17) have the solutions

$$n_2(\tau_p) = (E_p/h\nu)(\tau_2/\tau_p)[1 - \exp(-\tau_p/\tau_2)] \tag{18}$$

$$n_1(\tau_p) = (f_{21} E_p/h\nu)(\tau_1/\tau_p)[1 - \exp(-\tau_p/\tau_2)\tau_2/(\tau_2 - \tau_1) + \exp(-\tau_p/\tau_1)\tau_1/(\tau_2 - \tau_1) \tag{19}$$

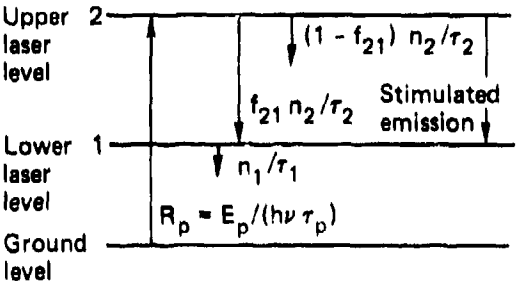



Fig. 5

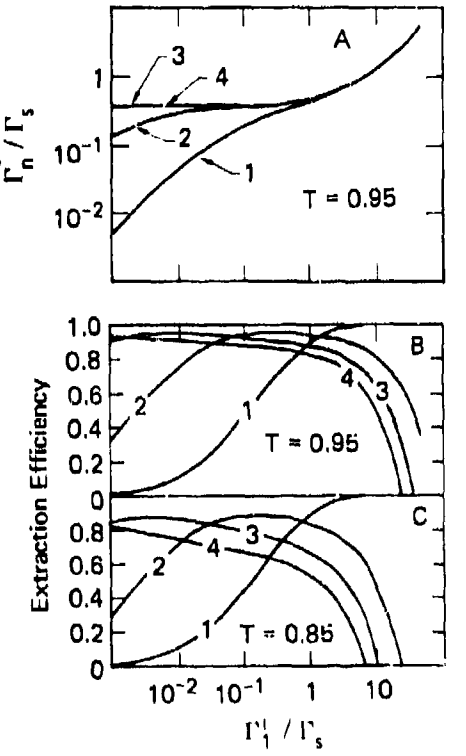



Fig. 6

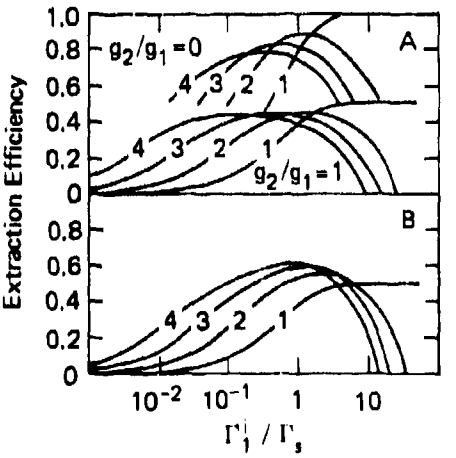



Fig. 7

During the optical turnaround time $\tau_T$ the decay of the two laser levels has the same form as during the pumping time, but there is now no pumping of the upper level. Equations (16) and (17) still describe $n_1$ and $n_2$ if $R_p$ is set equal to zero. Now the initial conditions are determined by past changes in the laser level populations. The initial values of $n_1$ and $n_2$ are constants determined by the values of $n_1$ and $n_2$ after the previous extraction. Under these conditions equations (16) and (17) have the solutions

$$\int_0^L n_{2n}(\tau_T)dz = \int_0^L n_{2n}(\tau_T = 0)dz \exp(-\tau_T/\tau_2) \tag{20}$$

and

$$\int_0^L n_{1n}(\tau_T)dz = f_{21} \int_0^L n_{1n}(\tau_T = 0)dz[\exp(-\tau_T/\tau_2) - \exp(-\tau_T/\tau_1)] \tau_1/(\tau_2 - \tau_1) + \int_0^L n_{1n}(\tau_T = 0)dz \exp(-\tau_T/\tau_1). \tag{21}$$

Eqn. (15) is used to calculate the values of $\int_0^L n_{2n}(\tau_T=0)dZ$ and $\int_0^L n_{1n}(\tau_T=0)dz$ in terms of the known value before the previous extraction. Eqns. (20) and (21) are used in eqn. (14) to calculate the gain-length integral prior to extraction. The gain-length integral value is then used in eqn. (5) to calculate the output fluence.

## IV. Results of Multipass Extraction Analysis

The multipass analysis described in section III was applied to several hypothetical systems. Different combinations of parameter values were used in order to study their effects on the performance characteristics. Figs. 6 to 11 show the results of this analysis.

### A. Multipass Extraction Characteristics

In this section the dependence of the multipass amplifier performance on the amplifier gain-length integral, the medium degeneracy, the laser level relaxation characteristics, and the number of extraction passes will be specifically examined. To illustrate consider fig. 6, consisting of 3 sets of curves labeled "A", "B", and "C" respectively. The set of curves labeled "A" are plots of normalized output fluence for each of a total of 4 extraction passes. The sets labeled "B" and "C" are plots of the cumulative extraction efficiency, for each of the 4 passes. The cumulative extraction efficiency for the nth pass is

$$\varepsilon_{ext} = (\Gamma_n^o - \Gamma_1^i)/(E_p L). \tag{22}$$

All the curves in fig. 7 are plots of cumulative extraction efficiency. All of the curves in figs. 6 and 7 are plotted against input fluence normalized to the saturation fluence. The parameters for pump time, optical turnaround time, $f_{21}$, and degeneracy ratio were made equal to zero in the figure 6 calculations in order to isolate the effects of optical loss.

The curves for the first extraction pass in figure 6 correspond to the $\int^{L}\alpha_0 dz = 4$ curves for the single pass amplifier in figures 2 and 3. Figure 6A illustrates the increase in extracted energy possible by utilizing more than one extraction pass. This advantage is most dramatic for the lower input fluences. At inputs higher than a saturation fluence the output curves for different numbers of passes become almost indistinguishable on this scale. This is because an input fluence greater than a saturation fluence will extract a significant portion of the stored energy on the first pass. In addition, the input fluence itself is a major part of the resulting output fluence. Differences in performance characteristics for different numbers of passes are more apparent from the extraction efficiency curves in figure 6B. Figure 6B shows that there are clearly defined ranges where a specific number of extraction passes is optimum in terms of efficiency and output fluence. These results show that for input fluences less than a saturation fluence two or more extraction passes are favored over a single pass. However, operating a multipass system with more extraction passes than the optimum number can be heavily penalized in efficiency due to the increased optical losses to the high fluence beam between passes. This can be seen by comparing the efficiencies for different numbers of extraction passes in figs. 6B or 6C. The penalty for operating with too many extraction passes is especially true for $r^1{}_1/r_s \approx 1$, where for these conditions the single pass amplifier is highly favored. Too few passes will not extract the stored energy efficiently. Too many passes will incur heavy optical losses. Of course any specific system will be designed considering a tradeoff between the greater efficiency of operating with the optimum number of passes against the added costs of the optical and system elements to provide each additional extraction pass. For this purpose a separate calculation must be made for the specific system being considered.

As seen by comparing fig. 6B with fig. 6C, increasing the optical losses between passes by 10 percent decreases the maximum extraction efficiency for the high gain multipass system by roughly 10 percent. This is understandable because the higher optical loss puts greater penalties on each additional pass. Since in a high gain amplifier system the pulse is being continually amplified, the greatest optical loss will occur immediately before the final extraction pass. The loss at this stage of the amplification

process will generally be much greater than the other optical losses. The total optical loss for all turnaround periods will be primarily the loss at this maximum loss stage. The total optical loss will then depend approximately linearly on the optical loss ratio, (1-T), for a single stage. For the same reason, the penalty for operating with too many passes increases substantially.

The purpose of the efficiency curves in figure 7 is to demonstrate the effects of the degeneracy ratio and gain recovery on the multipass system efficiency. To isolate these effects, the pump time and $f_{21}$ are set equal to zero. The efficiency curves in figure 7A illustrate the dramatic effect of the degeneracy ratio in reducing efficiency for bottlenecked extraction. When $\tau_1/\tau_2 < 1$ the lower level will relax faster than the upper level. In this case the gain coefficient will actually increase or "recover" during the turnaround time. The efficiency curves in figure 7B are for an example of such gain recovery, where some of the efficiency which would be lost due to the non-zero value of degeneracy ratio is regained during an appropriate optical turnaround time. Fig. 7B shows that a multipass system with gain recovery is favored over a single pass system for inputs up to nearly 10 saturation fluences. This is considerably higher than for the systems without gain recovery. It certainly includes the entire input range of practical interest. This recovery does not continue indefinitely. Eventually the lower level will be effectively relaxed and additional time will only deplete the upper level and thus the gain. When gain recovery is possible there is an optimum value of the turnaround time which will maximize the gain coefficient. It can be calculated from eqns. (14), (20), and (21) by setting the time derivative of the gain coefficient equal to zero. In eqn. (3) for the gain coefficient the degeneracy ratio multiplies the lower laser level population density. For this reason increasing the degeneracy ratio exaggerates any effects which depend on the behavior of the lower level population. Such processes include gain recovery during the turnaround time or gain degradation during the turnaround time and the pump time.

The efficiency curves in figure 7 are for an $(\alpha_0 L)_M$ value of 1.5 while those in figure 6 are for a value of 4. A comparison then of the curves in figure 6B with the top curves in figure 7A illustrates the effects of $(\alpha_0 L)_M$ variation. The higher value of $(\alpha_0 L)_M$, due to a greater pump energy deposition, will result in higher values of the gain coefficient throughout the multipass extraction operation. The higher gain coefficient will result in smaller input fluences being amplified efficiently into the saturated extraction regime. As shown by comparing figs. 6B and 7A the efficiency curve for any extraction pass will then be broader for the higher value of $(\alpha_0 L)_M$. When the fluence amplification

is comparable to the optical loss between passes the signal cannot grow significantly with subsequent extraction passes. In such a situation a multipass amplifier system has no advantage over a single pass amplifier.

During the pump time the upper level population increases from zero until the loss rate, $n_2/\tau_2$, equals the pumping rate $R_p$. The lower level population also approaches a steady value when the loss rate, $n_1/\tau_1$, equals the rate of increase from upper level decay, $f_{21}n_2/\tau_2$. Depending on the specific values of $g_2/g_1$, $\tau_1/\tau_2$, and $f_{21}$ the gain may rise monotonically to a steady value or it may have a maximum. Using a pumping time which produces as high a gain coefficient as possible will produce the highest output fluence from the subsequent extraction. However, the upper level is decaying during the entire pumping time. Therefore a tradeoff must be made between efficiency and output fluence in selecting a pumping time. The pumping rate will depend on the capacity of the pump technology and on any limits to the pumping power due to damage considerations. The total energy density deposited is then this limiting pump rate multiplied by the optimized pump time. For some systems it may be necessary to limit the total deposited energy density to a smaller value determined by other damage criteria. In this case the shortest possible pump time to deposit this energy density will produce the greatest efficiency and the greatest output fluence. In general the shorter the lifetime of the lower state compared with the lifetime of the upper state, the less lower state population will accumulate to degrade the gain. Where the degeneracy ratio is zero, the lifetime ratio will not matter. But for nonzero $g_2/g_1$, smaller values of the lifetime ratio will lead to better amplifier performance.

B. Optimum Extraction Efficiency Profiles

As discussed in the previous section there are a large number of parameters which affect the multipass amplifier system performance characteristics. These parameters interact in a complex manner, and each may cause significant effects on the performance characteristics in different ranges of parameter values. For this reason, one can't identify only one or two parameters which dominate the system performance. It is possible however to form some qualitative and semi-quantitative conclusions regarding the tradeoff of the parameters. The form of the extraction equation and the expressions governing the level populations lend themselves to a reduction of the parameter space into a minimum set of key ratios and nondimensional quantities: $\Gamma/\Gamma_s$, $\tau_T/\tau_2$, $\tau_p/\tau_2$, $\tau_1/\tau_2$, $g_2/g_1$, $(\alpha_{0L})_M$, $f_{21}$, and T. In figures which follow, $\tau_1/\tau_2$ vs $g_2/g_1$ and $(\alpha_{0L})_M$ vs $g_2/g_1$ were chosen as pairs of key parameters to form 2 dimensional parameter spaces in which to plot efficiency profiles.

A set of optimistic but reasonable values for the remaining parameters were chosen, with some variation within a figure and between figures to illustrate important parameter effects. The parameter $\Gamma_1/\Gamma_s$ was limited to values of $10^{-2}$ since multipass configurations are most useful at lower initial input fluences. The maximum number of extraction passes which can be made is a system limitation that will be different for each specific multipass amplifier system. The highest efficiencies for up to 6 passes were plotted in these profiles.

In considering a multipass amplifier the designer is frequently faced with selecting a candidate laser medium. At this point in the system design process attention is focused on the medium characteristics, $g_2/g_1$, $\tau_1/\tau_2$, and $f_{21}$, rather than on system parameters. Figures 8 and 9 are plots of efficiency profiles in $\tau_1/\tau_2$ - $g_2/g_1$ space to facilitate medium selection considerations. These figures show that efficiency requirements divide the $\tau_1/\tau_2$ - $g_2/g_1$ plane into fairly restrictive regions. As an example consider the system depicted by the solid curves in fig. 8. To achieve an extraction efficiency greater than 0.5 either the lifetime ratio must be less than 0.15 or the degeneracy ratio must be less than 0.6. For less optimistic system parameter values or for nonzero $f_{21}$, as illustrated in the other profiles in figs. 8 and 9, this bounding value of $\tau_1/\tau_2$ or $g_2/g$ is even more tightly constrained. As depicted in fig. 8 extraction efficiencies greater than 0.5 became fairly insensitive to the degeneracy ratio for values greater than 0.6 and highly insensitive to the lifetime ratio for values above 0.15. For larger lifetime ratios the lower level lifetime has become greater than the time during which there is a significant upper level population. Then there is insignificant decay of the lower level. Further increases in lower level lifetime can have no effect on the gain and therefore none on the efficiency. Conversely the insensitivity to the degeneracy ratio is physically due to gain recovery counteracting the degeneracy ratio's constraint on extraction. A small decrease in lifetime ratio will have a large influence on the gain recovery. It will thus counterbalance large increases in the degeneracy ratio, and the extraction efficiency will not be sensitive to $g_2/g_1$.

In optimizing a system design the various systems parameters must be traded off against each other. In figures 10 and 11 the dependence of the extraction efficiency on $(\alpha_0 L)_M$ and $g_2/g_1$ is examined. The parameter $(\alpha_0 L)_M$ is determined by the pump energy deposited and the cross section for stimulated emission of the medium. For a multipass system limited to 6 extraction passes and with reasonable optical loss between passes $(\alpha_0 L)_M$ must be above 1 for the efficiency to be above 0.5. For $(\alpha_0 L)_M$ between

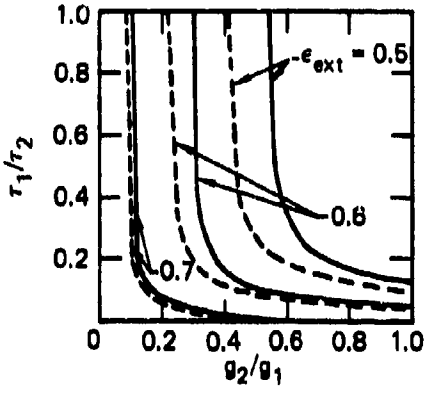



Fig. 8

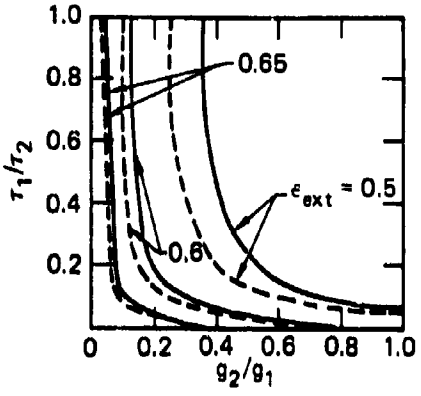



Fig. 9

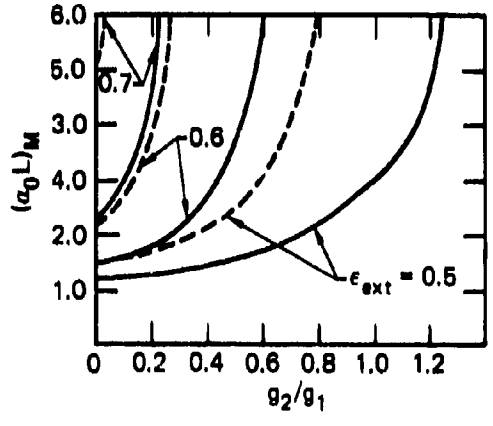



Fig. 10

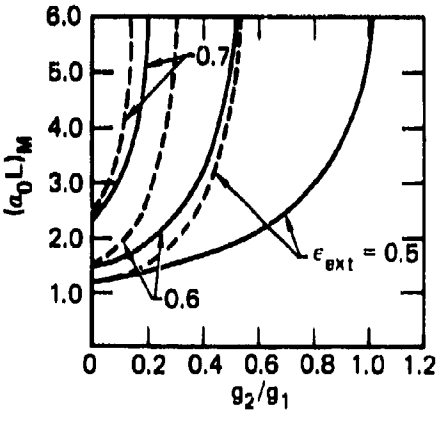



Fig. 11

1 and 2 the extraction efficiency is relatively insensitive to variation in $g_2/g_1$ compared with variation in $(\alpha_0 L)_M$. For $(\alpha_0 L)_M$ above 2 the extraction efficiency becomes more sensitive to variation in $g_2/g_1$ than to the variation in $(\alpha_0 L)_M$. The same general conclusions are reached if $\tau_p/\tau_2$ is paired with $(\alpha_0 L)_M$ to produce a set of profiles.

The characteristic shape of these curves is due to the rapidly diminishing effect of $(\alpha_0 L)_M$ on the extraction efficiency. For example consider the situation when the degeneracy ratio is zero. Then the lower level population has no effect on the extraction characteristics. The energy stored in the upper laser level will be depleted by the laser beam extraction and by the upper level decay during each turnaround time period. With each additional extraction pass optical loss depletes the energy from the laser beam. A higher value of $(\alpha_0 L)_M$ will result in a higher amplification of the laser beam, as discussed in Section. II. This will reduce the number of passes required to amplify the beam into the saturated extraction regime of efficient extraction. Fewer extraction passes result in smaller optical losses and less upper level decay. Thus a given portion of the stored energy can be extracted with higher efficiency for a higher $(\alpha_0 L)_M$. The number of extraction passes required to achieve a given amplification is approximately inversely proportional to the gain-length. So the incremental increase in efficiency with increased $(\alpha_0 L)_M$ diminishes quickly as $(\alpha_0 L)_M$ becomes large. This is especially true when the total number of extraction passes becomes very small.

If the degeneracy ratio increases, the $1/(1 + g_2/g_1)$ factor in the extraction efficiency causes a significant decrease in efficiency. To maintain the same efficiency along a constant efficiency contour $(\alpha_0 L)_M$ must increase at a nonlinear rate as the degeneracy increases, as shown in figures 10 and 11. As $(\alpha_0 L)_M$ becomes large it must increase at an extreme rate to maintin a given value of efficiency.

If the optical loss and losses due to upper level decay were negligible, the efficiency profiles in figures 10 and 11 would asymptotically approach limits of constant values of $g_2/g_1$ as $(\alpha_0 L)_M$ becomes very large. These limits can be calculated from eqn. (12), $(g_2/g_1)_{limit} = (1 - \varepsilon)/\varepsilon$, where $\varepsilon$ is the efficiency of a given profile. For example these limits are $g_2/g_1 = 1$ for the 0.5 efficiency profile and $g_2/g_1 = 0.429$ for the 0.7 profile. When the optical and decay losses are nonzero, these limiting values of $g_2/g_1$ will decrease. On the other hand gain recovery can cause them to increase. For the conditions represented by the solid lines in figure 10 the 0.5 profile approaches a limiting value of $g_2/g_1$ of greater than 1 due to gain recovery. But since gain recovery diminishes for smaller degeneracy ratio, the 0.7 profile limit is

less than 0.429, because losses have decreased it.

## V. Performance Characteristics of Specific Multipass Amplifier Systems

In order to evaluate any specific multipass system a specific calculation of output fluence and extraction efficiency, as discussed in Section III, must be performed. Figures 12 through 15 are examples of such calculations made for specific systems which have been considered as laser fusion candidates.

In analyzing real systems the extraction may not be uniform across the amplifier cross section as assumed in the analysis in sections II and III. This nonuniformity is typically due to the extracting laser beam having a nonuniform cross sectional profile or to the energy deposition being nonuniform during the pumping process. These nonuniformities can be handled by dividing the amplifier cross section into small elements, each of which has an approximately uniform cross sectional profile. The final output fluences are then averaged over all the cross sectional elements. Any pumping nonuniformity along the amplifier axis will be integrated out in the gain-length integral. The example calculations in figures 12-15 have taken these nonuniformities into account when appropirate.

Figures 12 and 13 plot extraction efficiency and output fluence against input fluence for a $V:MgF_2$ medium with a nonuniform extracting pulse profile.[6-7] At a moderate level of pump energy deposition, producing an $(\alpha_0 L)_M = 2$, the $V:MgF_2$ system can deliver significant output fluences for input fluences greater than 1 J/cm$^2$ with 0.7 extraction efficiency and with 4 or fewer extraction passes. The Tm:Glass system in figure 14,[6,8-9] on the other hand is limited to less than 0.5 extraction efficiency for similar input fluences and the Selenium (Group VI) system in figure 15[6,10-12] produces only slightly greater than 0.5 efficiency even though it has a larger gain-length product, $(\alpha_0 L)_M = 3$. The disparity in these system performances can be attributed primarily to differences in their medium characteristics. The $V:MgF_2$ system is phonon terminated, therefore the lower level lifetime is essentially zero and the degeneracy ratio is zero to account for the completely unbottlenecked extraction. The Tm:Glass system has a degeneracy ratio of approximately 0.56 and a laser level lifetime ratio of 25. For the Selenium system the degeneracy ratio is 0.75 and the lifetime ratio is approximately 0.2. Locating these points in the parameter space of figure 8 and 9 shows that the $V:MgF_2$ system is capable of greater than 0.7 extraction efficiency (for a uniform beam profile) while the Selenium and Tm:Glass systems are

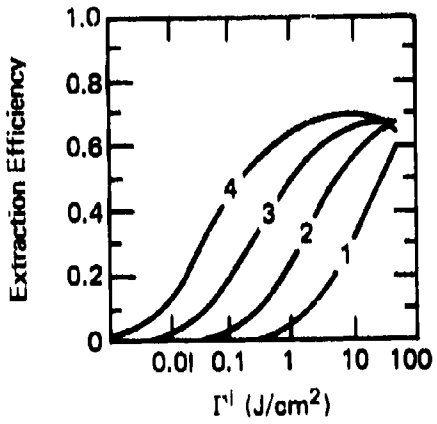



Fig. 12

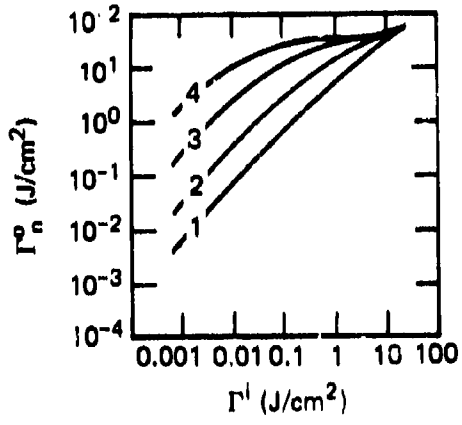



Fig. 13

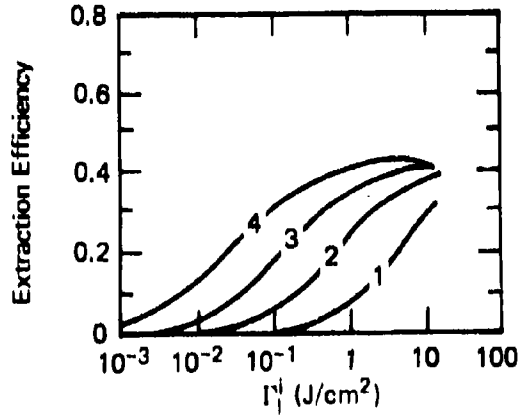



Fig. 14

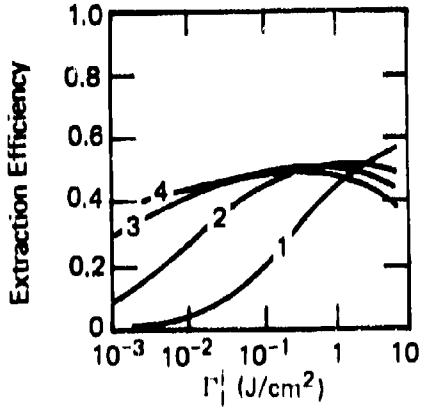



Fig. 15

near the 0.5 efficiency profile only for the more optimistic system parameters.

## VI. Summary

This paper has examined laser amplifier extraction from a storage medium. Two general energy extraction configurations were considered. First a single pass amplification of the laser beam was considered. Then a multipass amplifier system where the laser beam is returned through the medium for additional extraction one or more times following the initial pass was examined. The multipass technique provides the opportunity to extract energy from upper laser level states missed during the first pass. For this reason it has the significant advantage over the single pass amplifier of providing high energy gain and high extraction efficiency simultaneously for a wide range of parameter values. When there is no gain recovery the multipass amplifier configuration is generally favored when the input fluence is less than a saturation fluence. Multipass amplifier systems with significant gain recovery can be superior to a single pass amplifier for all values of input fluence of practical interest. Optical losses between passes reduce efficiency and can heavily penalize systems using too many extraction passes. During the pump time or between passes decay of the energy stored in the upper laser level will decrease the gain-length integral and therefore the efficiency of extraction.

For bottlenecked extraction without gain recovery a nonzero degeneracy ratio will limit the extraction process and hence the extraction efficiency. When the lower laser level population decays faster than the upper level population, gain recovery can remove some of this limitation on extraction. Under these conditions there is an optimum value of optical turnaround time which will produce the maximum improvement in efficiency.

An increase in $(\alpha_0 L)_M$ due to greater pump energy deposition will improve the efficiency. This improvement diminishes rapidly with larger values of $(\alpha_0 L)_M$.

A given acceptable value of extraction efficiency places severe restrictions on the possible values of either the degeneracy ratio or the laser level lifetime ratio. On the other hand the extraction efficiency is insensitive to a large range of values of either the degeneracy ratio or the lifetime ratio.

## Acknowledgement

I would like to acknowledge the extensive work done by Charles V. Duncan in initiating the extraction analysis described in this paper. I also want to thank Roger A. Haas for conceiving the original idea of this project, for his guidance during the study, and for helpful discussions to clarify the physics issues involved.

## References


1. L. M. Frantz and J. S. Nodvik, J. Appl. Phys. 34, 2346 (1963).

2. J. B. Trenholme and K. R. Manes, "A Simple Approach to Laser Amplifiers," Lawrence Livermore National Laboratory, UCRL-51413 (1972).

3. W. H. Lowdermilk and J. E. Murray, "The Regenerative Amplifier: I - Theory and Numerical Analysis," to be published in J. of Appl. Phys., manuscript no. 8829R.

4. R. A. Haas, "Laser System Architectures - Energy Extraction," Laser Program Annual Report - 1977, Lawrence Livermore National Laboratory, Report UCRL-50021-77 (1978), p. 7-13.

5. A. J. Glass, "Multipass Amplification," Sec. 2.4.1 in NOVA CPD Final Report, Laser Fusion Program, LLL-Misc. 111, ed. by T. J. Gilmartin and R. O. Godwin, Mar. 1978.

6. W. F. Krupke, Lawrence Livermore National Laboratory, Livermore, CA, private communication.

7. L. F. Johnson, H. J. Guggenheim, and R. A. Thomas, Phys. Rev. 149, 179 (1966).

8. W. F. Krupke, E. V. George, R. A. Haas, "Advanced Lasers for Fusion," in Laser Handbook vol. 3, M. L. Stitch, Ed. (North-Holland, Amsterdam, 1979).

9. R. R. Jacobs and W. F. Krupke, "Rare-Earth: Solid State Hybrid Laser Systems, Physics Issues," Laser Program Annual Report - 1977, Lawrence Livermore National Laboratory, Report, UCRL-50021-77 (1978), p. 7-69.

10. J. R. Murray and C. K. Rhodes, J. of Appl. Phys. 47, 5041 (1976).

11. H. T. Powell and J. J. Ewing, Appl. Phys. Lett. 33, 165 (1978).

12. R. A. Haas, "Group VI Laser System Analysis," Laser Program Annual Report - 1977, Lawrence Livermore National Laboratory, Report, UCRL-50021-77 (1978), p. 7-40.

Figure Captions

1. Laser Amplifier Energy Extraction Configurations.

2. Single Pass Amplifier Performance Characteristics. Normalized output fluence is plotted as a function of normalized input fluence for several values of the gain length integral.

3. Single Pass Amplifier Performance Characteristics. Normalized extraction efficiency is plotted as a function of normalized input fluence for several values of the gain-length integral.

4. Single Pass Amplifier Performance Characteristics. The fraction of the pump energy density extracted by the laser beam is plotted as a function of the normalized position along the amplifier axis for several values of the normalized input fluence. The position is normalized as the gain-length.

5. Three Level Laser Model for Multipass Amplifier Extraction Analysis.

6. Performance Characteristics of a Multipass Amplifier when $(\alpha_0 L)_M = 4$, $\tau_1/\tau_2 = 0.1$, and $g_2/g_1 = \tau_T/\tau_2 = \tau_p/\tau_2 = f_{21} = 0$. This figure is divided into three sets of curves. Set A are normalized output fluence curves for T = 0.95. Sets B and C are extraction efficiency curves for T = 0.95 and 0.85 respectively. Each curve is identified with its extraction pass number.

7. Performance Characteristics of a Multipass Amplifier when $(\alpha_0 L)_M = 4$, $\tau_1/\tau_2 = 0.1$, $\tau_p/\tau_2 = 0$, and $f_{21} = 0$. Extraction efficiency is plotted as a function of normalized output fluence. Each curve is identified with its extraction pass number. The upper set of curves in section A of this figures has $\tau_T/\tau_2 = 0.1$ and $g_2/g_1 = 0$, while the lower set has $\tau_T/\tau_2 = 0$ and $g_2/g_1 = 1$. The set of curves in section B of this figure has $\tau_T/\tau_2 = 0.1$ and $g_2/g_1 = 1$.

8. Multipass Amplifier Extraction Efficiency Profiles when $(\alpha_0 L)M = 4$, $\tau_1/\tau_2 = 0.25$, $\tau_T/\tau_2 = 0.05$, $\Gamma_{i_1}/\Gamma_s = 10^{-2}$. Solid curves are for $f_{21} = 0$. Dashed curves are for $f_{21} = 1$. Each curve is identified with its efficiency. Each profile represents the maximum extraction efficiency achieved in a total of 6 passes.

9. Multipass Amplifier Extraction Efficiency Profiles when $(\alpha_o L)_M = 2$, $\tau_p/\tau_2 = 0.25$, $\tau_T/\tau_2 = 0.05$, $\Gamma^i{}_1/\Gamma_s = 10^{-2}$. Solid curves are for $f_{21} = 0$. Dashed curves are for $f_{21} = 1$. Each curve is identified with its efficiency. Each profile represents the maximum extraction efficiency achieved in a total of 6 passes.

10. Multipass Amplifier Extraction Efficiency Profiles when $\tau_T/\tau_2 = 0.05$, $\tau_1/\tau_2 = 0.1$, $f_{21} = 0$, $\Gamma^i{}_1/\Gamma_s = 10^{-2}$. Solid curves are for $\tau_p/\tau_2 = 0.25$. Dashed curves are for $\tau_p/\tau_2 = 0.50$. Each curve is identified with its efficiency. Each profile represents the maximum extraction efficiency achieved in a total of 6 passes.

11. Multipass Amplifier Extraction Efficiency Profiles when $\tau_T/\tau_2 = 0.05$, $\tau_p/\tau_2 = 0.25$, $f_{21} = 1$, $\Gamma^i{}_1/\Gamma_s = 10^{-2}$. Solid curves are for $\tau_1/\tau_2 = 0.1$. Dashed curves are for $\tau_1/\tau_2 = 0.3$. Each curve is identified with its efficiency. Each profile represents the maximum extraction efficiency achieved in a total of 6 passes.

12. Performance Characteristics of a V:$MgF_2$ Multipass Amplifier System. Extraction efficiency is plotted as a function of input fluence when $\tau_p/\tau_2 = 0.174$, $\tau_T/\tau_2 = 0$, $\tau_1/\tau_2 = 0$, $g_2/g_1 = 0$, $(\alpha_o L)_M = 2.077$, $T = 0.95$ and $f_{21} = 0$. Each curve is identified with its extraction pass number.

13. Performance Characteristics of a V:$MgF_2$ Multipass Amplifier System. Output fluence is plotted as a function of input fluence. Each curve is identified with its extraction pass number.

14. Performance Characteristics of a Tm:Glass Multipass Amplifier System. Extraction efficiency is plotted as a function of input fluence when $\tau_p/\tau_2 = 0.125$, $\tau_T/\tau_2 = .025$, $\tau_1/\tau_2 = 25$, $g_2/g_1 = 0.56$, $(\alpha_o L)_M = 1.88$, $T = 0.95$, and $f_{21} = 0$. Each curve is identified with its extraction pass numbers.

15. Performance Characteristics of a Selenium ($^1S_o \rightarrow {}^3P_1$). Multipass Amplifier System. Extraction efficiency is plotted as a function of input fluence when $\tau_p/\tau_2 = 0.1$, $\tau_T/\tau_2 = 0.05$, $\tau_1/\tau_2 = 0.2$, $g_2/g_1 = 0.75$, $(\alpha_o L)_M = 3.0$, $T = 0.95$, and $f_{21} = 0$. Each curve is identified with its extraction pass number.

# REDEFINITION OF THE FOUR FUNDAMENTAL FORCES

ROBERT J. HEASTON, PHD
OFFICE OF THE DEPUTY CHIEF OF STAFF FOR
RESEARCH, DEVELOPMENT AND ACQUISITION
HEADQUARTERS DEPARTMENT OF ARMY
WASHINGTON, D. C. 20310

## INTRODUCTION

Unification of the fundamental forces has been one of the great theoretical problems in physics in the twentieth century. Beginning in 1918 with Weyl and continuing through the last thirty-five years of Einstein's life, many different attempts were made to unify the electromagnetic and the gravitational forces. Moreover, since the four fundamental forces were first defined in the early forties, extensive efforts by numerous investigators have gone into measuring and attempting to unify two or more of these forces. This paper defines the four forces and discusses some of the difficulties in unifying the forces. A new approach to unification will be presented with a discussion of the consequences and predictions of this approach.

The four fundamental forces are defined in Table 1. These four forces are all that are necessary to characterize all phenomena. From an Army perspective, the strong force is only of interest in the basic structure of matter and in nuclear weapons effects. The electromagnetic force is involved in the structure of matter, all electronic devices, all chemical reactions, explosives, and propellants. The weak force occurs in nuclear weapons effects. The gravitational force becomes involved in every load carrying device and in the motion of aircraft, projectiles and missiles. Frequently, such as in a fuze, more than one force is involved.



PRECEDING PAGE BLANK-NOT FILMED

Table 1. Four fundamental forces. Adapted from 1, p. 52 and 2, p. 453. One fermi (fm) is $10^{-13}$ centimeter.

| Force | Natural Occurrences | Relative Strength | Range |
|---|---|---|---|
| Strong | Nuclear Forces | 1 | 0.5 to 1 fm |
| Electromagnetic | Atomic Forces | $10^{-2}$ | 0.01 fm to Indefinite |
| Weak | Radioactive Decay | $10^{-5}$ - $10^{-13}$ | Singularity to 0.01 fm |
| Gravitational | Astronomical Forces | $10^{-39}$ | Point Source to Indefinite |

The four forces are generally listed in the order of increasing strength as represented by their coupling constants. These coupling constants are essentially the ratios of the strengths of the indicated force to the strength of the strong force. Thus, the strong force, which is the strongest, has a coupling constant of unity when compared to itself. The strong interaction is 137 times greater in strength than the electromagnetic force and about $10^{39}$ times greater than the gravitational force, so that the coupling constants for these two forces are $10^{-2}$ and $10^{-39}$, respectively. The weak force coupling constant is quoted at various magnitudes: $10^{-13}$, $10^{-7}$, or the Fermi theory value of $1.02 \times 10^{-5}$. These coupling constants have become the universally accepted way of fingerprinting the four forces.

Any effort to try to understand the four forces and to achieve unification must cope with a vast array of variables. Not all of these variables can be explained here. There are different forces, forms of interaction, effects on matter, relative strengths, spins, ranges, variations with range, mediators, saturation effects, charges, and selection rules. Even the interpretation of what unification means may take different paths. Hardly a single thread of continuity runs through the four forces.

A detailed review was made of various approaches to unification of the four forces and of the variables that were identified in the approaches. A side-by-side comparison of the different assumptions and steps in Weyl's, Einstein's, guage symmetry, and supergravity approaches to unification led to the following observation. It was observed that the only uniform step that everyone has adopted to date is the a priori assumption that two or more of the strong, electromagnetic, weak and gravitational forces are accepted without question. The almost completely independent way of theoretically defining and measuring the four forces, the limited success of indirect unification through families of mediating particles, the extraordinary complexity of the mathematics, the omission of general relativity from the theories, and the sheer intensity of unfulfilled efforts confirm the need to possibly question even the a priori assumption. Consequently, the primary basis of this paper is that, it may be possible that unification might be easier to achieve if the a priori statement of the four forces is redefined.

## REDEFINITION

The phenomena that led to the definition of the four forces in the first place is accepted at face value. Only a new explanation is needed. Two of the forces have a strong justification for being retained unchanged. These are the electromagnetic and the gravitational interactions. There are several similarities between these two, especially when compared in their simple forms of the Coulomb electrostatic force and the Newton inverse square law. Both are static, point-source, inverse square relationships. Each has a potential that is similarly defined. They have both been verified over ranges from less than one fermi to galactic distances. An alternate explanation of the four forces is that there are other forces or potentials very much like those of Coulomb and Newton which could explain the phenomena associated with the strong and the weak forces. This alternative, if applied rigorously, leads to a result which, at first glance, is counterintuitive but in the final analysis fulfills almost all predictions expected for unification. It is hypothesized that $E = mc^2$ and $E = h\nu$ are functions of potentials which represent corresponding forces in the structure of matter. One of these forces is called the Einstein force and the other is the Planck force, respectively. When combined with the Coulomb and Newton forces, as indicated in Table 2, the redefined fundamental forces are referred to as the unified interaction theory (first published as a paper in reference 3).

Table 2. Redefined forces and interactions of the unified interaction theory.

| Interaction | Force | Equation |
|---|---|---|
| Strong | Einstein | $\vec{F}_c = \frac{mc^2}{r}$ |
| Quantum | Planck | $\vec{F}_h = \frac{hc}{2\pi r^2}$ |
| Electromagnetic | Coulomb | $\vec{F}_e = \frac{e^2}{r^2}$ |
| Gravitational | Newton | $\vec{F}_G = \frac{m^2 G}{r^2}$ |

The redefined interactions consisting of the strong, quantum, electromagnetic and gravitational interactions replace the currently accepted strong, electromagnetic, weak and gravitational interactions. The new list of interactions are specifically defined by the forces (first published in abstract form in reference 4) listed in Table 2. Each force has a definite equation. Three of the forces are inverse square and one is inverse linear with distance. All forces are defined in terms of point-source, static interactions between two masses. The Einstein force is identified as the strong interaction. This force is essentially a rest mass energy gradient. The Planck force defines a new interaction, the quantum interaction, which will be explained in a later section. The Coulomb and Newton Forces are accepted without any new interpretation. The Coulomb force is related to the electromagnetic interaction only in terms of the electrostatic component.

## CORRELATIONS AND PREDICTIONS

Four criteria are used to assess the unified interaction theory; mathematical correlation of known relationships; physical correlation with measured phenomena; fulfillment of predictions expected of a unified theory; and, unique predictions of the new theory. These criteria will be applied in the following manner. First, relationships mathematically deduced from the four redefined

forces will be presented. Then, the criteria of physical measurements, unification predictions, and unique predictions will be explained for the Einstein and Planck forces. Finally, the criteria will be applied to the theory in general.

Based only upon the four redefined forces of the unified interaction theory, it is possible to deduce several functions and relationships that are either well-known or cited in the scientific literature. Only the consideration of equivalency conditions or ratios are needed to make these mathematical correlations. Figure 1, on the following page, has been especially structured to graphically aid in showing these mathematical correlations and in describing different phenomena. In fact, over thirty different concepts mentioned in this paper are keyed to Figure 1. The absolute magnitudes of the four redefined forces in Table 2 are displayed in Figure 1 as a plot of logarithm of force versus logarithm of distance. Values on the ordinate and abscissa correspond to calculations based upon an idealized point source interaction between two protons. Figure 1 was originally made to scale but has been condensed and somewhat distorted to fit a more convenient size. Points K, M, O, P, Q, R, S, and T will all be used to illustrate some phenomena or particular relationship. Letters for the points were generally chosen to provide some key to each illustration. The results to be described with the aid of Figure 1 could just as well be shown on a linear plot of force versus distance but it would not have been as convenient. The reader may verify any of the results indicated in the following statements. Generally, the last term in each sentence is the function or relationship obtained from performing the operation stated.

Equivalent conditions of force and distance are indicated at points Q, O, and S in Figure 1. The equivalence between the Einstein force and the Planck force magnitudes, when $F_c = F_h$ at point Q, occurs at a distance defined by the Compton wavelength. Similarly, the Einstein force is equal to the Coulomb force, $F_c = F_e$ for point O, at the classical radius. The Einstein force is equivalent to the Newton force, $F_c = F_g$, at the gravitational collapse limit, or one-half the Schwartzchild limit at point S. If the equation for the Newton force is stated at the gravitational collapse limit, the result is a constant force, $F_S$, which has previously gone unnoticed in the Einstein field equations.

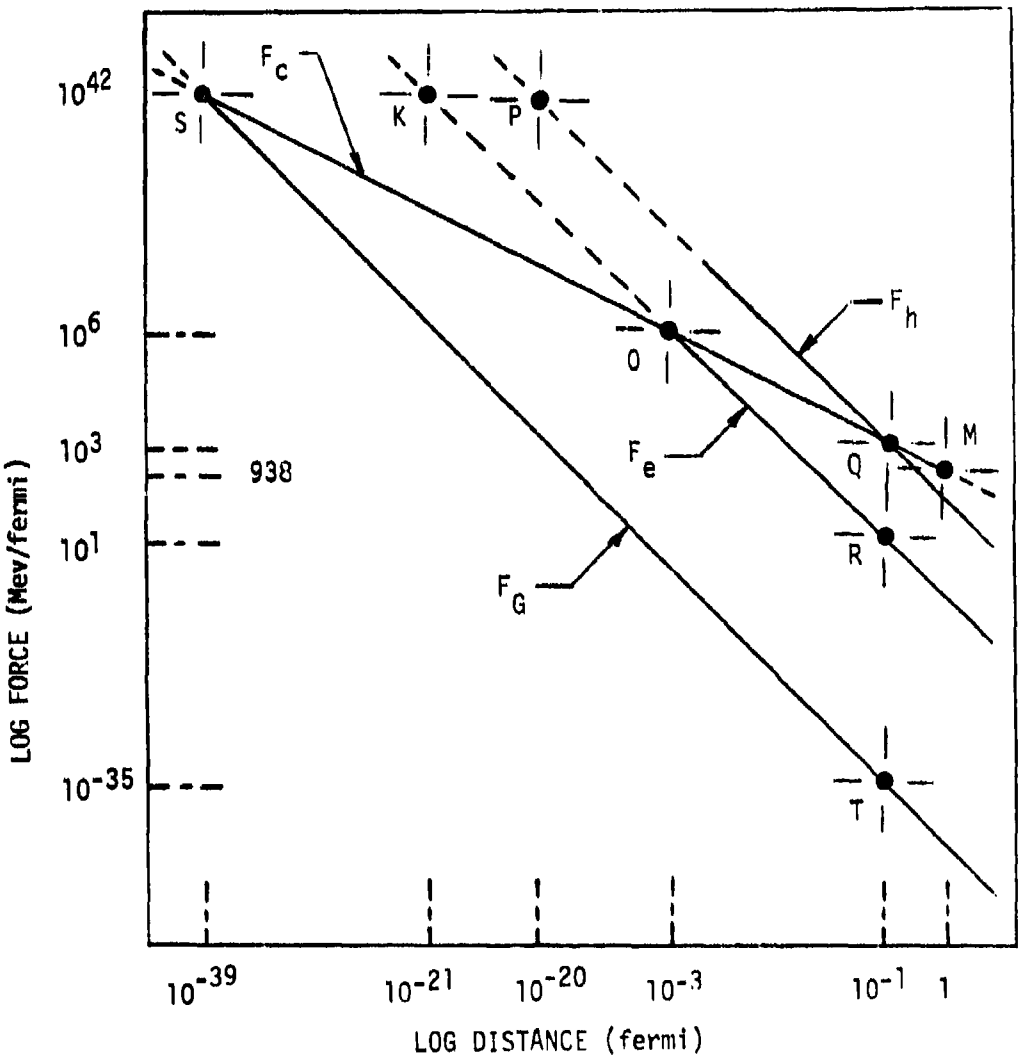



Figure 1. Plot of log force versus log distance for hypothetical proton-proton interaction based upon the redefined forces of the unified interaction theory. Solid lines are probable ranges of the forces. Not to scale but numbers approximately match points. One dyne is equivalent to $6.24 \times 10^{-8}$ Mev/fm. One newton is $6.24 \times 10^{-3}$ Mev/fm. Because of the units used, all forces at one fermi have the same numerical magnitude in Mev/fm as the rest mass energy in Mev (point M).

$$r_S = \frac{mG}{c^2}$$

$$F_G = \frac{m^2G}{r_S^2}$$

$$F_G = \frac{c^4}{G} = F_S$$

The force $F_S$ is a critical value of the Newton force at the gravitational collapse limit. The lengths at points K and P in Figure 1 can be explained in terms of equivalence of the constant force at K and P to this critical value. When the Coulomb force is equal to $F_S$ above, $F_e = F_S$, the resulting length is a function very similar to one derived by Kursunoglu (5, p. 1539). The equivalence at point P of the Planck force with $F_S$, $F_h = F_S$, is a function frequently used in astrophysics, the Planck length (6, p. 12). The mass required in the Einstein force to make $F_c = F_h = F_S$ at the Planck length is the Planck mass (6, p. 1215).

A number of significant ratios of forces, one to the other, are possible from Figure 1. The ratio of $F_h$ to $F_c$, which is unity at the Compton wavelength at point Q may be defined as the strong interaction coupling constant. The ratio of $F_e$ to $F_c$ at the Compton wavelength, $F_P/F_C$ is the electromagnetic interaction coupling constant. The ratio of $F_e$ to $F_h$ for any constant length is the fine structure constant. The ratio of $r_e$ to $r_h$ at any constant force is the square root of the fine structure constant. The ratio of $F_G$ to $F_c$ at the Compton wavelength, $F_T/F_Q$, is the gravitational interaction coupling constant. The ratio of $F_S$ to $F_T$, particularly for the mass of the pion, is the Eddington number. The relationship $(F_h/F_G)^2$ at any constant length is also the Eddington number. The two ratios, $F_e$ to $F_G$ and $F_h$ to $F_G$, at the Compton wavelength, or any constant length, are principal cosmic numbers. Everyone of the above explanations is a straight-forward mathematical manipulation of the four redefined forces. No other unified theory has been able to give such a mathematical correlation of so many different relationships.

The Einstein force is both defined and predicted by the unified interaction theory. From a physical phenomena perspective the Einstein force gives the right force magnitudes over the correct

range as measured for the strong interaction. The Einstein force fulfills theoretical and experimental expectations in being an inverse linear function of distance. The total energy relation

$$E^2 = (pc)^2 + (mc^2)^2$$

treats the rest mass energy as if it were a potential energy. In fact, all of the forces for points M, O, Q, and S can be shown to be some constant multiplied by $(mc^2)^2$. Figure 1 indicates that below the Compton wavelength, point Q, and the classical radius, point O, that the Einstein force may be exceeded by the Planck force and the Coulomb force, respectively. Both of these conditions, as well as the total energy equation, support the condition of asymptotic freedom where the electromagnetic and the weak interactions could possibly equal or exceed the strong interaction at higher system energies. The Einstein force is totally consistent with the principle of asymptotic freedom. A unique prediction of this theory is that all particles should be effected by the Einstein force. The masses of the leptons may be too small to isolate this effect from the other forces.

The Planck force has been experimentally measured (7, p. 161) on a macroscopic scale. It has been expressed in the form

$$F_h = \frac{hcA}{360d^4}$$

where d is the distance between two neutral plates of area A. This force has been attributed to quantum fluctuations. That is why the Planck force is associated with a quantum interaction. The unified interaction theory has essentially made the association, too, that the force attributed to quantum fluctuations on a macroscopic scale may be attributed to quantum fluctuations on a microscopic scale. The introduction of the Planck force accomplishes the same function already utilized in the modeling of the force coupling constants. According to Perkins (8, pp. 17-21), all of the coupling constants can be displayed as functions of $2\pi/hc$. In Figure 1, since the Einstein force is equal to the Planck force at point Q, the relative strengths of the various forces may be related to the Planck force rather than the Einstein force. In fact, this comparison may be made at larger distances. For example, the strong interaction coupling constant represented by the relationship of $2\pi g^2/hc = 10$ is the ratio of $F_c$ to $F_h$ at 2.1 fm, which compares to the experimental value of 2 fm reported by Meyerhof (9, p. 224). Calculation of this result is readily understood in terms of the unified interaction theory. The electromagnetic coupling constant, $\alpha = 2\pi e^2/hc$, is also easily

developed as explained in the mathematical correlations section of this paper. Similarly, the gravitational coupling constant may be expressed as $2\pi m^2 G/hc$ as required. The acceptance of the Planck force makes all of these coupling constants directly available as functions of $2\pi/hc$ without the arbitrary introduction of the Compton wavelength as a coincidental empirical length. If it is acceptable to state that the strength of the strong interaction is the relative magnitude between it and another force, such as the electromagnetic force, at the Compton wavelength for two protons or hadrons, then it should be rational to expect that the weak interaction is a similar relationship between a proton and an electron. Consequently, the relative strength of the Planck force at its Compton wavelength for an electron compared to the Planck force at its Compton wavelength for a proton is

$$\left(\frac{m_e}{m_p}\right)^2 = 2.97 \times 10^{-7}$$

This value of the weak interaction coupling constant is somewhat lower than the Fermi theory but is closer to actual measurements. If this result is used as the weak interaction coupling constant, and analogous reasoning is used as for the proton-electron coupling model in beta decay, the resulting mass of the electron's neutrino is predicted to be $2.78 \times 10^{-4}$ Mev and the mass of the muon's neutrino is $5.75 \times 10^{-2}$ Mev. Both of these masses would be different if the weak interaction coupling constant for neutrino generation were different from that for the proton-electron generation.

Since it has already been shown that the ratio of the Coulomb force to any Planck force at a constant distance is equal to the fine structure constant and that the consequence of $r_e$ divided by any $r_h$ at the same force is equal to the square root of the fine structure constant, then it follows that any Planck force magnitude is related to any other Planck force value as some function of the fine structure constant. The same is true of the relationship of any Coulomb force to another Coulomb force and of the point of equivalence of any Einstein force with the Planck force to another equivalence point. On a linear plot of force versus distance, these stair-step relationships as a function of the fine structure constant quickly diverge. These various relationships of the different forces as a function of the fine structure constant are completely analogous to the pertubation theory expansion of a series function of the fine structure constant in quantum electrodynamics.

The unified interaction theory predicts an unusual force, $c^4/G$, with an absolute magnitude of $7.55 \times 10^{41}$ Mev/fm. The combination $c^4/G$ is an integral part of many equations in the literature but no attention has been called to it as being a force *per se*. Because of its unique derivation from the unified interaction theory, $c^4/G$ has been referred to by others as the Heaston force. If the experimentally observed strong interaction is $10^{39}$ times the gravitational force between two protons and $c^4/G$ is $6.5 \times 10^{77}$ times the gravitational force (at 1 fm), then $c^4/G$ is $6.5 \times 10^{38}$ times the strong interaction. The Heaston force is obviously a super-strong force. Such a force has been qualitatively predicted and attributed to Kaufman by Feynman (10, p. 609) and to Kogut, Wilson and Susskind by Glashow (11, p. 45). It has been suggested in both cases as a binding-force for quarks. The Heaston force is strong enough to be the mechanism for the generation of black holes as well as a significant factor in cosmology. The reason for this statement is because $c^4/G$ appears in the Einstein field equation (6, pp. 431-434),

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu}$$

as well as in almost all solutions to this equation. Conditions may be defined such that the gravitational metric is zero, the Ricci tensor is $R_{oo}$, and the total energy is $T_{oo}$. Because of conditions imposed by the Christoffel indices, a factor of one-half is introduced so that

$$R_{oo} = \frac{4\pi G}{c^4} T_{oo}$$

This is precisely the result obtained by Weyl (12, p. 242) for the case of a stationary gravitational field where he referred to the inverse of $c^4/G$ as the greek letter kappa without indicating any awareness that kappa was an inverse force. The $4\pi$ is introduced to account for spatial density. Based upon the new awareness that $c^4/G$ is a superstrong force, the result in the last equation may be interpreted as stating that, at the beginning of the universe, the total mass/energy of the universe was confined in the dimension $R_{oo}$ and held together by the Heaston force.

Many physicists and astrophysicists make the assumption of physical units so that $c = h = G = 1$. Because of this assumption, $c^4/G$ is unity and loses visibility in theoretical derivations. On the other hand, the effect of such a choice of units is to normalize all forces so that $c^4/G$ is the maximum force possible and all other forces are relative. The unified interaction theory confirms this implication. The Heaston force is the critical value at the

gravitational collapse limit where the Newton force is equal to the Einstein force. The prediction of the Kursunoglu length at Point K in Figure 1 involves the Heaston force. Kursunoglu (5) attempted to unify the four fundamental forces based upon a time-independent spherically symmetric field solution to the field equations. He obtained a fundamental constant, $r^2 = 2Ge^2/c^4$, which he attributed to be a measure of deviation from the theory of relativity. This constant resulted from a roundabout unification of the electromagnetic and gravitational fields which Kursunoglu could not quite explain. The specific explanation is given by the unified interaction theory at point K. Accepted theories for black hole generation are based upon the Planck length which is the point where the Planck force is equal to the critical value represented by the Heaston force. The Planck mass, which is recognized as the maximum possible mass of any single particle occurs when $F_c = F_p = F_s$ in Figure 1. All of these observations support the conclusion that $c^4/G$ may be the maximum possible force in the universe. This means that gravitational collapse reaches an asymptote rather than continuing to a singularity. The possibility of such an asymptote is predicted (6, pp. 1196-1217) as an option for theoretical definition of the universe. In his attempts to derive a unified field theory, Einstein predicted (13) that gravity, which is the weakest of all forces, should have a role to play in particle structure. The Heaston force, which is based upon the limit of gravitational collapse, fulfills the prediction by Einstein.

The unified interaction theory can be extrapolated to create a picture of the structure of a single particle. The outer "boundary" of a particle is defined as the quantum surface, $r_Q$. It occurs at the Compton wavelength created by the equivalence of the Einstein force and the Planck force. The proton has a fuzzy boundary that averages to an experimental boundary equivalent to the quantum surface. Within the particle is a "charge surface" located at the classical radius, $r_0$. At the heart of a particle is a gravitational core where $c^4/G$ predominates. A three-layer structure like this has been constructed (14) for particles, particularly protons, from experimental observations. The radius of a particle is $h/2\pi mc$. The circumference is $h/mc$. If all of the charge is concentrated between the charge surface and the quantum surface, the rotation of a charged loop equal to the Compton wavelength in radius gives the Bohr magneton. The Einstein force is equivalent to a centripetal force at the quantum surface rotating at the constant speed of light, which correlates with one theoretical observation (15). The moment of inertia yielding such a force implies that particles have to be cylindrical discs rotating on an axis through the center of the disc. Seventeen years

of experimental observations on the magnetic moment of the electron indicates that the best model for all of the measurements is a thin right circular cylinder (16, p. 80).

The above conditions cannot be achieved without a further consequence which is the most speculative as well as the most unique prediction of the unified interaction theory. There would be no observed changes in intrinsic spin of a particle without the following hypothetical process. Whenever a particle is accelerated (or decelerated) in translation, it changes in size in all three dimensions from a fixed size at rest to zero volume at the speed of light. This process is called trilation, as compared to translation and rotation. Lorentz made this initial assumption in his derivation (17, p. 21). It can be shown that trilation can produce the same volume change as the Lorentz-Fitzgerald contraction. Moreover, the process of trilation is essentially stated by the Schrodinger wave equation. The most common interpretation of $|\psi|^2$ is that $|\psi|^2$ is the probability that a particle can be found in a box or phase space of arbitrary volume. Suppose that this volume were the rest mass volume. Consequently, when a particle is accelerated, the probability $|\psi|^2$ of the particle occupying its rest mass volume at rest is unity and the corresponding probability at the speed of light is zero. Nothing is changed but the interpretation. Volume is also a hidden variable in relativity theory. Forces are expressed as forces per unit volume (6, p. 159; 12, pp. 64 and 201; 17, p. 13) as well as for only a small volume of space. The trilation process is normalized into the theoretical results. Thus, the proposed model of particle behavior is as follows. A particle absorbs energy. It is accelerated as it changes size. In order to compensate for conservation of angular momentum and to maintain its constant quantum surface rotation at the speed of light, part of the energy is dumped overboard. Maxwell's equations are obeyed. Intrinsic spin is coupled to translation. Quantum numbers may be logically related. Rotation of the quantum surface in revolutions per second is the same as the frequency in $E = h\nu$. Deltas in spins, rather than absolute values, are explained. The particle has a built-in capability to recall its total energy state in any relativistic situation. Collision cross sections decrease with higher energy. Wave-particle duality has an integral physical justification. Spin speeds have a boundary. Quantized behavior has a physical interpretation. Based upon this model, a resonance particle is one that cannot hold together for more than one revolution. Its quantum surface does not develop.

Finally, an overall assessment of Figure 1 and the unified interaction theory indicates an extraordinary paradox. Points M, O, Q, R, and T in Figure 1 are concerned with forces and dimensions such as the coupling constants, fine structure constant, Compton wavelength and classical radius which are all relationships familiar to high energy physics. These phenomena are usually associated with dimensions of 0.001 to 1 fermi. On the other hand, points K, P, and S which contribute to defining the gravitational collapse limit, Planck length, Planck mass and $c^4/G$ are part of the language of astrophysics and cosmology. The dimensions associated with these points range from $10^{-39}$ to $10^{-20}$ fm. The paradox exists in that the world of galaxies and megaparsecs is dependent upon relationships derived from conditions dependent upon dimensions orders of magnitude less than the conditions associated with the microscopic world of particle physics which takes place at the relatively long range of one fermi. It is extraordinary that the equation for the gravitational collapse limit at point S and $10^{-39}$ fm for a proton is also extrapolated in astrophysics into the empirical relationship, $MG/Rc^2 \approx 1$, where M is the mass of the universe and R is the radius of the universe (18, p. 1162). The reason for this paradox is that the large magnitudes of $c^4/G$ and the inverses of the extremely small dimensions give the large numbers that are needed in cosmology. That is the very reason why every large cosmic number, except the Hubble constant (which is obtained from telescopic observations) can be constructed from various ratios of forces and distances in Figure 1. The major achievement of the unified interaction theory may very well be this explanation for the first time of the close association between the relationships of physics and astrophysics.

## CONCLUSIONS

After the initial hypothesis that $E = mc^2$ and $E = h\nu$ are functions of potentials that have corresponding forces in the structure of matter, all of the manipulations of the unified interaction theory are mathematically and dimensionally precise as well as internally consistent. No other theory has been able to give a coherent integrated explanation of so many different relationships of physics and astrophysics. A long desired goal of relating general relativity to quantum mechanics may have been achieved. Although the language is the same, the interpretations using the unified interaction theory are often different. The phenomena are accepted at face value but the cause and effect explanations are not the same. Experiments could possibly be devised to test the predictions of the Einstein and the Planck forces and their application to all particles, the masses of the neutrinos, the existence of the quantum surface, the

coupling of the quantum numbers, the internal structure of particles, the trilation process, and the role of the fine structure constant. Effects of the Heaston force may be implied only indirectly but its main value should be in theoretical developments in cosmology. The ultimate value of the redefined forces will probably be their use as a key step in explaining the different particle masses. A whole new world below one fermi has been opened up. The mathematical unity of the unified interaction theory is so overwhelming that serious attempts should be made to verify or disprove the predictions resulting from redefining the four fundamental forces.

## REFERENCES

1. STEVEN WEINBERG. Unified Theories of Elementary-Particle Interactions. Scientific American, 231, 1, 50-59 (July 1974)

2. GEOFFREY MANNING. EPIC as a Window on the Forces of Nature. Nature, 255, 453-458 (5 June 1975)

3. ROBERT J. HEASTON. Unified Interaction Model of the Four Fundamental Forces. Speculations in Science and Technology, 1, 1, 71-75 (April 1978)

4. ROBERT J. HEASTON. Speculations on a Unified Model of the Four Fundamental Forces. AAAS Publication 77-2, Contributed Papers Abstract No. 20, Denver: 143rd AAAS National Meeting (20-25 February 1977)

5. BEHRAM KURSUNOGLU. Consequences of Nonlinearity in the Generalized Theory of Gravitation. Physical Review D, 13, 6, 1538-1560 (March 1976)

6. CHARLES W. MISNER, KIP S. THORNE, AND JOHN ARCHIBALD WHEELER. Gravitation. San Francisco: W. H. Freeman and Co. (1973)

7. J. D. BJORKEN. "Quantum Electrodynamics." Vol. 11, Encyclopedia of Science and Technology, New York: McGraw-Hill, 160-162 (1971)

8. DONALD H. PERKINS. Introduction to High Energy Physics. Reading, Massachusetts: Addison-Wesley Publishing Co. (1972)

9. WALTER E. MEYERHOF. Elements of Nuclear Physics. New York: McGraw-Hill (1957)

10. RICHARD P. FEYNMAN. Structure of the Proton. Science, 183, 4125, 601-610 (15 February 1974)

11. SHELDON LEE GLASHOW. Quarks with Color and Flavor. Scientific American, 233, 4, 38-50 (October 1975)

12. HERMANN WEYL. Space-Time-Matter. New York: Dover Publications, Inc. (1922)

13. ALBERT EINSTEIN. "Do Gravitational Fields Play an Essential Part in the Structure of the Elementary Particles of Matter?" in The Principle of Relativity, New York: Dover Publications, Inc., 191-198 (1923)

14. MUHAMMAD M. ISLAM. The Optical Model at High Energies. Physics Today, 23-27 (May 1972)

15. N. S. AKULOV. Are Electrons and Positrons Quantum Rotators? Sov. Phys. Dokl., 23 (5), 326-328 (May 1978)

16. H. R. CRANE. The g Factor of the Electron. Scientific American, 1, 72-85 (January 1968)

17. H. A. LORENTZ. "Electromagnetic Phenomena in a System Moving with any Velocity less than that of Light." in The Principle of Relativity, New York: Dover Publications, Inc., 11-34 (1923)

18. KENNETH L. NORDTVEDT, JR. Gravitation Theory: Empirical Status from Solar System Experiments. Science, 178, 4066, 1157-1164 (15 December 1972)

*HENRY & LAIBLE

# IMPROVEMENTS IN CB PROTECTIVE CLOTHING

*MALCOLM C. HENRY, DR.
ROY C. LAIBLE, DR.
US ARMY NATICK RESEARCH & DEVELOPMENT COMMAND
NATICK, MASSACHUSETTS 01760

## INTRODUCTION

Chemical Protective Clothing, as it exists today in the US Army, has a number of unique features. It is a clothing system utilizing a variety of materials, including textiles, polymeric foam, unsupported films, elastomers and coated fabrics. These materials make up the current handwear, footwear, overgarment and headgear. The Chemical Biological Protective Mask is currently being studied and a significant effort is under way at the Chemical Systems Laboratory, ARRADCOM, to improve this item. For the purposes of this paper, the protective mask will not be discussed further except to recognize it as a component of the entire CB clothing system, its main role being to defeat chemical warfare agents that can enter through the auronasal route. Obviously, it also protects the face and eyes as well. The remainder of the CB protective clothing system is designed to protect against toxic or corrosive chemical warfare agents capable of affecting or penetrating through the skin. These latter agents are known as percutaneous agents and can cause casualties quickly at very low levels of concentration.

There are several guiding parameters in devising CB protective clothing that, in total, allow the scientist to focus on the problem rather quickly.

First of all, there are three basic mechanisms that can be utilized to design clothing that will intercept chemical agents and prevent them from coming in contact with the skin. A material can be selected that will provide a complete barrier to the agent, a



PRECEDING PAGE BLANK-NOT FILMED

material can be selected that will sorb the agent for some finite time preventing it from moving through the selected material and reaching the skin, and thirdly, a material can be selected which will react with the chemical agent and detoxify or convert it to a harmless state. There are no other alternative mechanisms.

The second guiding principle is the recognized necessity for clothing to breathe or, in pragmatic terms, to be comfortable. This is the reason our normal clothing uses textile materials in preference to all other material forms. Textiles provide comfort and subject the wearer to the least heat stress of any form of material while providing warmth, cooling, flexibility, protection against abrasion and all other factors required for clothing items. The clear conclusion from these considerations is that if possible a textile CB clothing system would be most desirable.

The third guiding principle is that any item of CB clothing that is developed will be stored in warehouses for an undetermined period of time, and issued to troops when needed. There is a long lead time between the time for development, the adoption of the item, procurement, storage and, finally, issue to the troops. It is imperative that all of the items of CB clothing that are put into the system be capable of defeating any and all of the agents in the enemy's arsenal, both now and those that may be developed in the future. The active ingredient in the overgarment, for example, is activated carbon and for good reason. Activated carbon is a universal adsorbent, i.e., it adsorbs on its surface, all polar molecules, which include all the percutaneous chemical agents known today. More important, activated carbon-based systems will work equally effectively for any new chemical agents yet to be developed. Activated carbon has advantages of strong adsorption and an extraordinarily high adsorptive surface area (approximately 1000 sq meters/gram of material.)

Finally, a "given" that continually must be considered is that our troops are expected to survive a first attack. US policy is to not be the first to use chemical agents; therefore, it follows we will be attacked first and we must be prepared to survive the attack as a number one priority.

In 1970, the US Army introduced a radically different and new system in Chemical Protective Clothing, which has been alluded to in the earlier part of this Introduction. (Figure 1) This system is comprised of a Chemical Biological Protective Mask, a butyl rubber-coated fabric hood, butyl rubber handwear and footwear, and a two-

piece overgarment composed of a nylon/cotton outer fabric and a polyurethane foam layer impregnated with activated carbon backed with nylon. This new system of clothing provides complete protection against all percutaneous and blistering chemical agents for long periods of time, when worn as directed.



Fig 1. COMPLETE CHEMICAL PROTECTIVE OUTFIT

Since adoption of these items, NARADCOM has embarked on a series of research and development efforts to improve the overall properties of the clothing components while the Chemical Systems Laboratory, ARRADCOM, has continuing efforts ongoing to develop an improved respirator. The NARADCOM effort discussed in this paper presents some of the results of this research and development effort.

## RESULTS AND DISCUSSION

### 1. Improved Flame/Thermal Resistant Overgarment

The new CB overgarment and the other components are all composed of organic materials that sustain combustion once ignited, and, indeed, even when removed from the flame environment, continue to burn to the point of total consumption. (Table I and Figure 2) It was recognized that certain military occupation specialties (MOS's) are exposed to flame/thermal flux intensities more often or more likely than others, particularly combat vehicle crewmen, helicopter pilots, tankers and the like.

TABLE I

FLAME TESTS ON CHEMICALLY PROTECTIVE OVERGARMENT MATERIALS

| | Flame Resistance Method 5903 | | | Flame Resistance Method 5905 | |
|---|---|---|---|---|---|
| Fabric System | AF(sec) | AG(sec) | CL(in) | AF(sec) | (% Consumed) |
| Standard Foam/ Nylon Tricot Laminate | 44 | * | 6.5 | ** | 100 |
| Ideal Garment (Goal) | 0 | 0 | 3.0 | 0 | 20 |

*Sparking occurs during afterflaming but glowing stops when flaming stops
**Consumed during test so no afterflame measured

The standard overgarment was placed on several mannikins and exposed to the fire pit test. The uniform caught fire almost immediately and continued to flame after removal from the flame source. The flames had to be extinguished or the garment would have continued to burn. When the garment was re-examined at that time, the uniform was mostly destroyed. In addition, the mannikins were equipped with sensors capable of measuring both the heat flux and the temperatures reached. These measurements showed that if the mannikin had been a soldier dressed in the overgarment and exposed to such a heat flux, second and third degree burns could have been expected. The experiment was repeated several times. Both the results mentioned and the flaming as shown in Figure 2 were typical in every case.

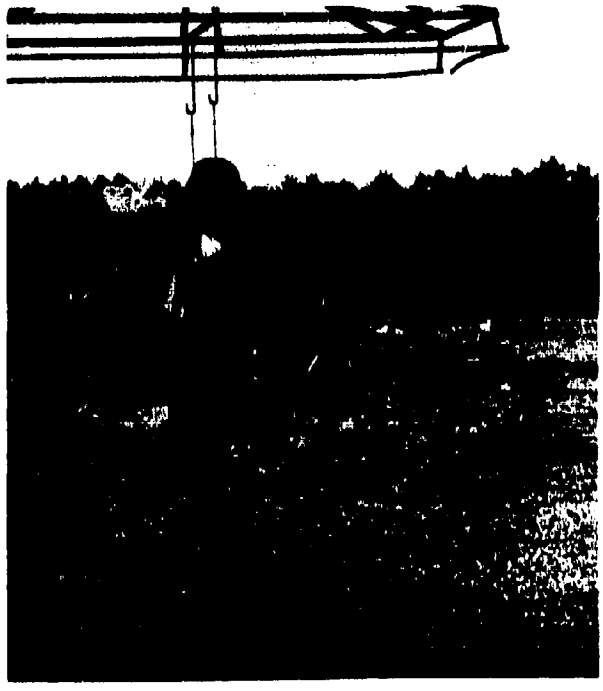


Fig 2. STANDARD OVERGARMENT 8 SECONDS AFTER FIRE PIT EXPOSURE

A product improvement effort was therefore initiated at NARADCOM to attempt to make the overgarment, at the very least, flame-resistant, and hopefully, flameproof. A number of approaches were attempted and the most successful approach selected as a modification to be considered for an improved item to be adopted in 1981.

The concept we have utilized is one never before successfully utilized in clothing. The principle is called "Intumescence." Our experiments have shown that utilizing this principle, slightly modified, a flame-resistant overgarment system can be developed which offers excellent flame/thermal protection.

The basic principle of "Intumescence" is the formation of a protective foam upon the application of a high thermal flux or flame source. The foam provides a thermal insulative barrier against the transfer of heat to the object. Intumescent paints have been used

for a number of years to protect surfaces of wood and metal in a variety of applications.

Basically, an intumescent formulation is composed of three descriptive components, a chemical that can be thermally decomposed to a strong dehydrating acid which once formed, dehydrates a second chemical quickly to copious quantities of carbon particulates. A third chemical is usually present that liberates gases at an appropriate temperature to foam the newly formed carbon particles.

In our case, with the overgarment foam, we were able to do away with the latter foaming component since we were already dealing with a foam item. Therefore, all that was needed was the dehydrating acid-forming chemical and a chemical readily dehydratable to carbon. Formulation experiments with this system finally resulted in the best combination of a commercial ammonium polyphosphate, tripentaerythritol and an acrylic binder to hold the chemicals in place in the foam.

In addition, the nylon-cotton outer fabric which could pose a flame and melt drip hazard was replaced with a flame-resistant Nomex III. Finally, the nylon supporting backing material which also was a hazard was replaced by a lightweight cotton fabric. During the processing of the intumescent foam, which contains the backing fabric for strength, the cotton fibers absorb sufficient intumescent chemicals rendering it also, flame/thermal resistant. Figure 3 shows the comparison between the standard and flame-resistant modified overgarment.

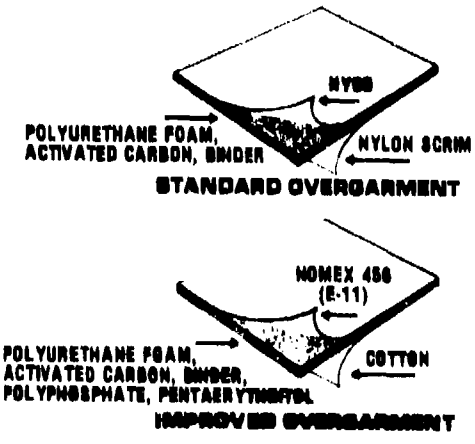



Fig 3. COMPARISON OF STANDARD AND FLAME RESISTANT OVERGARMENTS

The laboratory testing of the improved overgarment showed that the goals listed in Table I were met or exceeded. Uniforms were made of the new material, placed on instrumented mannikins and exposed to fire pit testing. In the fire pit, the overgarment survived a 1200°F fire completely intact. (Figure 4) The sensors on the mannikin indicated that the mannikin had received no second or third degree burns whatsoever.

Chemical agent protective qualities are not seriously degraded by this intumescent formulation. In fact, the same system exposed to the fire pit test provided over 24 hours protection when tested against both mustard gas and a nerve agent (GB).

Another approach which could be considered to improve the overgarment would be to make the foam thinner. In this way, the comfort factor could be enhanced and from a design point of view, alternate configurations could be considered such as CB protective liners instead of overgarments. The limitation of this approach is a mechanical one; commercial production equipment is not likely to be able to cut foam thickness less than .050 in. thick, at least with the reproducibility and consistency required for these kind of materials.




Fig. 4. FLAME HARDENED PROTECTIVE GARMENT AFTER FIRE PIT EXPOSURE

Foam thicknesses of .050 in. have been prepared, impregnated with activated carbon and are currently being evaluated. Another alternative currently being explored is the possibility of spraying the foam directly onto the fabric in which case very thin foam thicknesses can be achieved. A considerable amount of work has yet to be done before decisions can be made as to whether this approach has practical merit.

## 2. Chemical Agent Protective Textile Materials

If one were able to develop a chemical agent protective activated carbon-based textile system, it would have merit for many reasons. Such a system could provide increased drape, comfort and possibly could be laundered and regenerated. Because it would not have its activity lowered by the presence of binders or flame bonding, longer chemical agent protective lifetimes could be anticipated. Longer lifetime in storage could be realized resulting in considerable savings.

Four possibilities exist whereby one could transition activated carbon into textile fibrous systems. Figure 5 illustrates these possibilities.

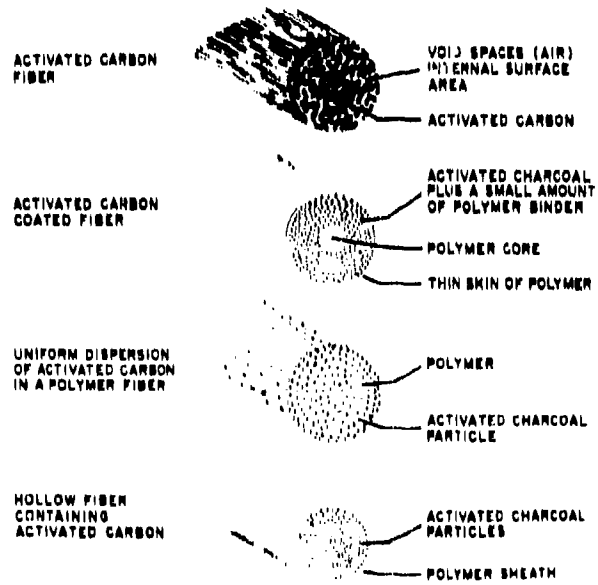



Fig 5. ACTIVE CARBON FIBER TYPES

a. Activated Carbon Fibrous Systems

Pure activated carbon fibers can be made by several different processes which will not be discussed in this paper. Probably the largest single disadvantage of activated carbon fibers is the brittle nature of these fibers and their lack of strength. One way to overcome these deficiencies is to wrap or braid the carbon yarns with another fiber to give the resulting composite fiber, flexibility and strength. Both fibers of this type and fabric have been made using this approach. Figure 6 shows a fabric made of a cotton-wrapped activated carbon yarn and then knitted into a fabric with Nomex.



FIG. 6. COTTON WRAPPED ACTIVE CARBON YARN IN A KNITTED STRUCTURE

Experimental uniforms of these fabrics have been fabricated and evaluated from a comfort point of view with results indicating improvements over the standard foam overgarment. These same uniforms have been worn in troop trials over obstacle courses and the report from these tests indicates a high degree of acceptance.

Another approach is the concept of activated carbon fabrics. In this case, we select a precursor already woven which by judicious selection of temperature and gaseous environments can be converted to a fabric with high sorption capacity and adequate strength. Because no binder is present, the activity of the active carbon fabric is all retained. The precursors which have been actively considered for the preparation of active carbon fabrics are acrylic, rayon and polyacrylonitrile (PAN) fabrics. Of these, the polyacrylonitrile has the advantage of providing a greater yield of end products than do the other two. At NARADCOM, experimentation is being conducted on carbon fabrics prepared from both rayon and PAN fabrics, but we have an active interest in a third precursor, pitch. Work to date on pitch fabrics has shown that very high yields are possible along with considerable activity and moderate strength. This work is continuing to optimize the process and to produce the fabric quickly and cheaply on existing equipment. This could result in a completely domestic source of active carbon fabric.

b. Hollow Core Activated Carbon-Filled Fibers

A second approach wherein one could enter the field of activated carbon-based textile systems is one where activated carbon particles are contained in a microporous hollow fiber. The potential advantage in this concept is that one would have basically a textile material or fabric that would serve as a chemical agent protective assembly when these chemical agents were present. At all other times, it would function as a standard fabric offering the wearer the advantages of any textile system.

For illustrative purposes, a scanning electron microscope photograph of a hollow fiber filled with particulate activated carbon is shown in Figure 7. This is a cross-section of the as-spun fiber 200 microns in diameter which after drawing will be reduced to 75 - 100 microns. This latter dimension is well in the range of practical textile fibers.

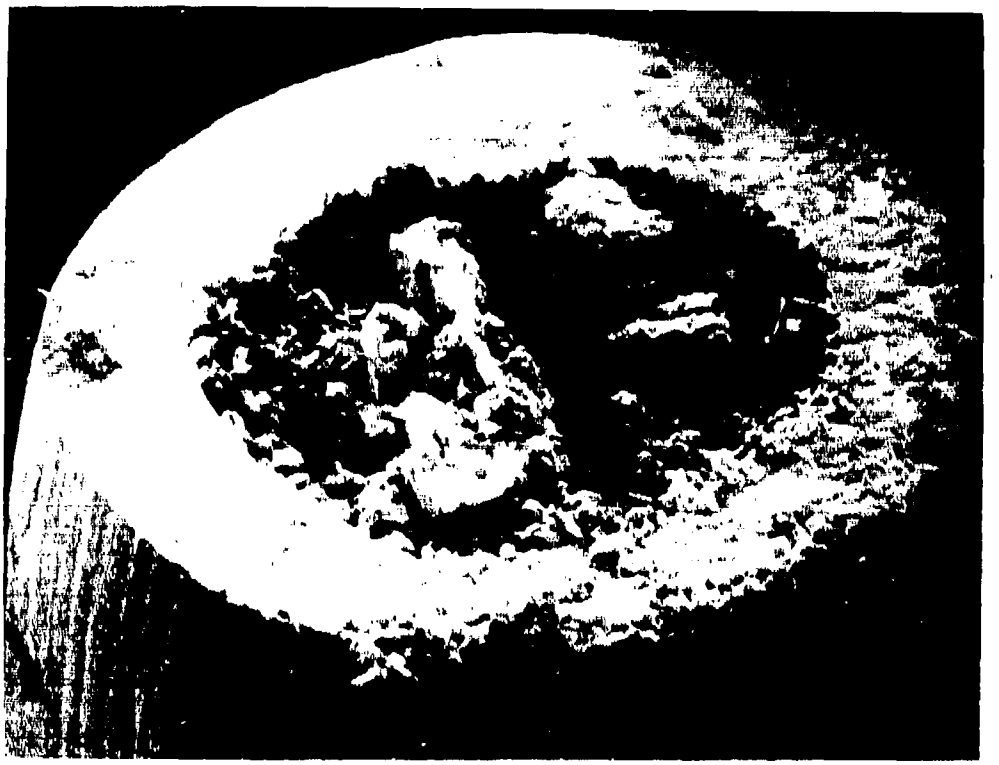


Fig 7. CROSS SECTION OF HOLLOW FIBER

One of the most interesting aspects of this work involved making the shell of the hollow fiber porous so that the agent would pass through that shell and be sorbed in the active carbon. In the initial studies, combinations of polypropylene and wax were spun into fibers. After the fibers filled with carbon were drawn, the wax was extracted to produce pores. These pores were too small for rapid permeation of vapors. Subsequent studies showed that a combination of polypropylene, a low molecular weight polyethylene and wax produced the proper crystalline morphology.

In Figure 8, the scanning electron micrograph shows a spherulitic crystal structure separated by extended chain crystals.



Fig 8. MICROSTRUCTURE OF HOLLOW FIBER WALL (3000X)

Pores capable of high permeation rates are available both between spherulites and as shown in the photograph within the spherulite itself. Fabrics have been woven of these fibers which for all practical purposes have the feel and appearance of a regular fabric.

In conclusion, a variety of approaches are being considered for the continued development of improved chemical agent protective clothing systems. The development of enhanced protection via improved activated carbon systems is leading the procession of new material configurations followed by a number of alternate routes any of which could alter the picture of what materials will be used in future generations of Chemical Protective Clothing.

As expected, laboratory measurements of the sorptive capabilities of the active carbon in these textile forms is high. Research and development efforts are continuing to ascertain the limitations, advantages and disadvantages of these approaches. Overall we expect the advantages will far outweigh the disadvantages.

### 3. CB Protection for the Extremities

Thus far, little has been said about the butyl rubber components of the current chemical protective clothing.

Experiments to develop improved elastomers are under way. The need for enhanced chemical agent protection, enhanced flexibility, and physical tactility are properties that are being emphasized. However, we must also consider the desirability of countering disadvantages such as the flammability and the marked tendency of butyl to swell in petroleum solvents. One avenue of research has been the exploitation of a novel rubber called phosphazene. This family of elastomers, based upon the backbone of phosphorous and nitrogen and containing side groups has unusual properties. These properties include a high degree of flame resistance, good low temperature flexibility and good resistance to swelling in petroleum solvents. The resistance to chemical agents is good and can be enhanced by compounding and curing variations. This work is currently under way at NARADCOM and has had an additional spinoff. Some of the compounding ingredients and methods which enhanced the chemical protective qualities of phosphazene rubber appear also to have beneficial results when applied to commercial elastomers.

## SUMMARY

Generally, three areas of active research have been identified in this paper. The first is the flame hardening of the overgarment which is proceeding with high success and an objective of implementation by 1981. The second area discussed involved the various methods of utilizing a sorptive material such as active carbon. These approaches such as active carbon fibers, active carbon fabrics, and

hollow fibers are still in an experimental stage. The third area, protection of the extremities, has produced some very interesting results but needs more work to reduce the concepts to reality in terms of availability and cost. These three areas were chosen because of the progress made in each area and because they hold the prospect of high success potential.

Other approaches have been and are being considered. They include the use of reactive groups alone or combined with sorption. The $XXCC_3$ impregnated garment long in the Army system utilized a chloramide to react with the mustard gas and V-type nerve agents, while chlorinated paraffin sorbed the G-agent. We would anticipate that future combinations would be based upon the insertion of appropriate functional groups on a sorbent material like carbon or a fibrous material. The problem is that a particular functional group may be fine for one agent and poor for another. Antagonistic functional groups on the same substrate might be unstable.

A second approach could utilize membranes in their broadest sense. The membranes would have to be barriers against the larger molecules typical of the chemical agents while allowing passage of air or better still, moisture. Some types of membranes may encourage the passage of water vapor while discouraging the passage of other molecules. We are considering these latter approaches but realize the difficulty in producing durable garments with universal chemical protection from these perhaps more novel techniques.

# THE ROLE OF ENDORPHINS IN THE PATHOPHYSIOLOGY OF SHOCK AND THE THERAPEUTIC BENEFIT OF OPIATE ANTAGONISTS (U)

*JOHN W. HOLADAY, Ph.D.
ALAN I. FADEN, MAJ, MC
DEPARTMENT OF MEDICAL NEUROSCIENCES, DIVISION OF NEUROPSYCHIATRY
WALTER REED ARMY INSTITUTE OF RESEARCH
WASHINGTON, D.C. 20012

## INTRODUCTION

The initial management of shock and trauma on the battlefield involves the rapid administration of intravenous fluids. Not only does this therapy require a skilled technician, but the problems of storage and availability of these fluids results in potentially fatal delays in treatment. In addition, the standard therapies employed in treating shock (e.g., fluids, steroids, and vasoactive agents) may not always reliably reverse the shock states which result from endotoxemia, hemorrhage, or spinal cord injury. We have sought a drug which would allow for the rapid stabilization of shock on the battlefield with minimal demands for technical skills and without the problems of storage and availability. We have shown that the opiate antagonist naloxone may be such a drug.

Within the past five years, it has been determined that opiate-like substances exist within the body and are involved in a variety of physiological and pathophysiological functions. Available evidence suggests that these endogenous opiates, collectively termed "endorphins", alter pain perception, body temperature, respiration, and may subserve other roles as well (1).

It is also known that endorphin systems are activated by stress (2,3) and that endorphins, like morphine, can produce marked hypotension following pharmacological administration (4). Since shock states are characterized by profound physiological stress, we investigated the possibility that the stress of shock would activate endorphin systems and thus contribute to the hypotension which

characterizes the shock syndrome. More importantly, if endorphins contribute to this pathophysiological effect of shock, then blockade of endorphins by opiate antagonists should reverse the hypotension of shock.

In this report, we present evidence that the pure opiate antagonist naloxone rapidly increases blood pressure and decreases mortality associated with shock caused by endotoxemia, hemorrhage, and spinal-cord transection. Shock studies conducted in rats, cats, and dogs demonstrate these therapeutic effects of naloxone across a variety of species. Additionally, naloxone treatment significantly improves neurologic recovery (paralysis) after spinal-cord injury in the cat. Results from additional experiments provide evidence for the site and mechanisms of these therapeutic effects of naloxone.

## MATERIALS AND METHODS†

Endotoxic and Hemorrhagic Shock in Rats. Since many anesthetic agents cause release of endorphins, we elected to study unanesthetized rats. Twenty-four hours prior to study, catheters were placed in the external jugular vein and tail artery; both cannulae were passed subcutaneously and threaded through a wire spring which was secured to the occipital area. This methodology, described in detail elsewhere (5), permitted evaluation of conscious, freely moving animals which remained in their home cages and which were not subjected to the stresses of handling immediately prior to study. Blood pressure (BP) and heart rate (HR) were continuously recorded using a microtransducer connected to a polygraph.

Escherichia coli lipopolysaccharide endotoxin was administered intravenously at a dose of 12 mg/Kg. When mean arterial pressure (MAP) had declined to a pre-established level of 65-70 mm Hg, animals received equal volumes of either saline or naloxone hydrochloride in varying doses.

For hemorrhagic shock, rats were prepared as described above. On the day following surgery, hemorrhagic shock was produced by withdrawing blood from the venous catheter. MAP was maintained at 40 mm Hg for a period of 20 minutes. This methodology resulted in the withdrawal of approximately 50% of the animals total blood volume by

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

the time of treatment and produced a 50% mortality in untreated rats. Animals were matched in pairs according to pre-treatment MAP, with 15 animals each assigned to either a saline or naloxone treatment group. A naloxone dose of 1 mg/Kg, given as a single iv bolus, was employed; saline controls received equal-volume iv injections.

Endotoxic and Hemorrhagic Shock in Dogs. Because of surgical procedures required for monitoring more complex cardiovascular parameters in dogs, studies were conducted on pentobarbital anesthetized animals. It was shown that naloxone has no significant effect on these parameters in the pentobarbital anesthetized, unshocked dogs used as controls. MAP was monitored through a femoral artery catheter. A pigtail catheter was placed in the left ventricle (LV) and the first derivative of LV pressure with respect to time (LV dp/dt max) was used as an index of LV contractility. A triple lumen Swan-Ganz catheter was passed into the pulmonary artery to measure pulmonary arterial wedge pressure ($PA_w$). This catheter was equipped with a thermistor tip to permit determination of cardiac output (CO) by a thermodilution method. Total peripheral vascular resistance (TPVR) and stroke volume were calculated from the CO, MAP and HR.

All animals received iv *E. coli* endotoxin at a dose (0.1 mg/kg) which produced a mortality of 80% at 24 hours in untreated animals. Following endotoxin administration, dogs were treated intravenously with either saline (14/group) or naloxone (2 mg/Kg; 6 dogs/group) in an equivalent volume. Survival was monitored at 24 hours.

For hemorrhagic shock studies, anesthetized dogs were prepared as above. These animals were bled down to 45 mm Hg and this MAP was maintained for 60 minutes. At this time the animals were treated with either naloxone (2 mg/Kg iv bolus, 2 mg/Kg/hr, n=5) or saline at equal volumes. One hour after treatment, shed blood was returned to surviving animals.

Spinal Shock and Spinal Trauma in Rats and Cats. Pentobarbital anesthetized rats and cats were studied. Rats were prepared with cannulae in the external jugular vein and tail artery; a guide tube was also affixed to the cranium for intraventricular (ivt) drug injections. Cats were anesthetized with pentobarbital, immobilized with gallamine triethiodide, and artificially ventilated. A femoral artery catheter was connected to a transducer to measure BP; heart rate (HR) was also monitored. A pigtail catheter was placed in the left ventricle (LV) through the right common carotid artery for measurement of cardiac contractility. Drugs were administered iv through a cannula in the femoral vein.

Both rats and cats were subjected to a dorsal laminectomy which exposed the spinal cord at the level of the seventh cervical vertebral (C7) segment. Following transection of the spinal cord, MAP fell to 20-30 mm Hg below baseline. In rats, the effects of iv as well as ivt injections of saline or naloxone were studied. Cats received only iv injections of these drugs. Doses and numbers of animals studied are depicted in figure legends.

A spinal-cord trauma technique was used to study the effects of naloxone on blood pressure and neurological outcome after spinal injury. After cannulating the femoral artery and vein, a calibrated weight (20 grams) was dropped 25 cm (a 500 gram-cm force) onto a 10 $mm^2$ plastic impact plate placed upon the spinal cord exposed at C7 as above. Forty-five minutes after injury, cats were treated intravenously with equal volume of saline or 2 mg/Kg naloxone. Cardiovascular parameters were monitored for four hours, after which catheters were removed and the laminectomy site was surgically closed. Animals were allowed to recover in home cages, and neurological examinations using a 5 point scale described elsewhere (6) were performed at 24 hours, 1, 2, and 3 weeks later.

## RESULTS

Figure 1 demonstrates the rapid improvement in blood pressure produced by 10 mg/Kg naloxone in an unasthetized rat subjected to endotoxic shock when compared to a saline-injected control animal. Following the precipitous drop in blood pressure produced by endotoxin, saline was without effect on this parameter. By contrast, naloxone produced a rapid return in blood pressure to control levels within seconds following administration (Fig 1).

A comparison of MAP and pulse pressure (PP; a crude index of cardiac performance) with and without naloxone for a group of rats is seen in Fig 2. After endotoxin hypotension, naloxone treatment resulted in a significant improvement in MAP and PP, whereas saline treatment was without effect. Additionally, naloxone administered to normotensive control rats not subjected to endotoxemia had no effect on these parameters, indicating a selective action in reversing shock hypotension instead of a direct effect of the drug by itself.

In other studies (data not shown), dose response effects of naloxone were determined (7,8). A dose of naloxone as low as 0.1 mg/Kg was shown to significantly improve MAP and PP, however maximum responses were obtained with 1.0 mg/Kg and 10.0 mg/Kg naloxone.

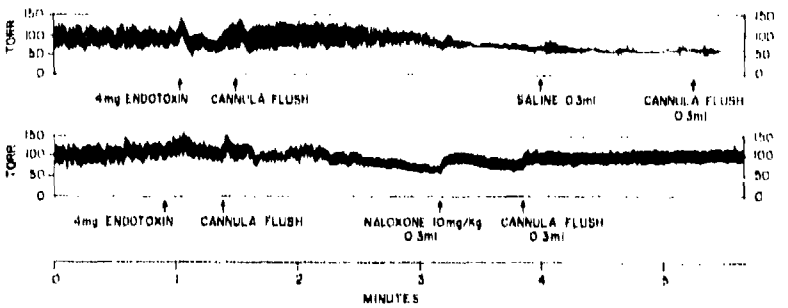



Fig. 1. Effects of intravenous saline (top) and naloxone (bottom) on the precipitious fall in blood pressure produced by 4 mg endotoxin in representative rats. Saline (0.3 ml) or 10 mg/Kg naloxone were injected after blood pressure fell to 65-70 mm Hg (torr).

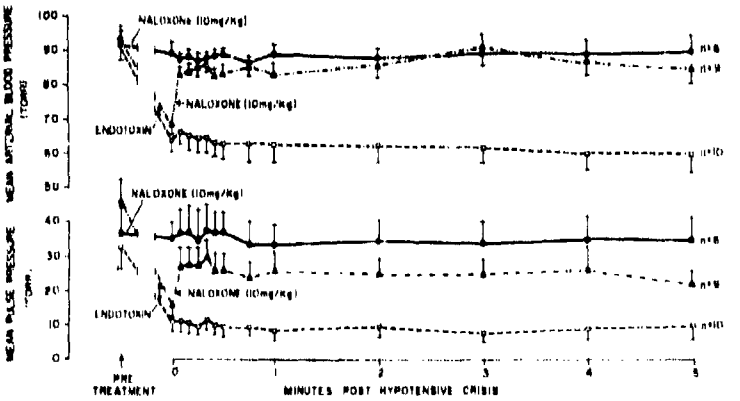



Fig. 2. The effects of naloxone or saline on mean arterial pressure (MAP, top) or pulse pressure (bottom) are compared with and without endotoxin-induced hypotension. Naloxone alone (●—●) did not affect MAP or PP. Endotoxin produced a significant drop in both MAP and PP which was reversed by naloxone (▲-▲) but unaffected by saline (o--o). Points represent averages ± SEM.

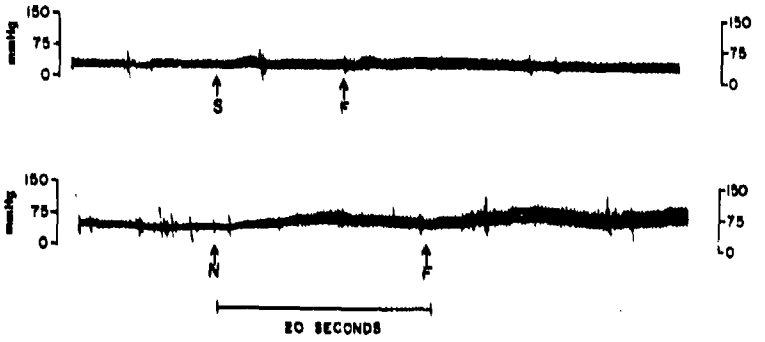



Fig. 3. Representative polygraph recordings showing the effect of saline (S, top) or 1 mg/Kg naloxone (N, bottom) on blood pressure following hemorrhagic hypotension. A 0.3 ml saline flush (F) was used to ensure drug delivery.

Figure 3 shows the effects of naloxone on blood pressure following hemorrhagic shock in representative rats. Naloxone, at a dose of 1 mg/Kg, significantly improved MAP and PP when compared to saline-injected, control rats (Fig. 4). More importantly, in this model of hemorrhagic shock, naloxone significantly improved survival with 13 of 15 naloxone-treated and 8 of 15 saline-treated rats surviving 24 hours (Fishers exact probability test, $p<.05$).

As seen in rat studies, the hypotension resulting from both endotoxic and hemorrhagic shock in dogs was rapidly reversed by naloxone at a dose of 2 mg/Kg iv (data not shown). Cardiac output (CO) was shown to be significantly improved (Figure 5), probably as a consequence of the increased cardiac contractility (CC) produced by naloxone in both shock models (data not shown). The fact that CO, and MAP all increased following naloxone injection, and venous return as well as total peripheral resistance were unchanged, suggests that naloxone was exerting its therapeutic effects directly or indirectly by improving cardiac contractility.

More critically, naloxone significantly improved survival following endotoxic or hemorrhagic shock in dogs. Twenty four hours following endotoxin treatment, 3 of 14 saline treated dogs (21%) were alive, whereas 5 of 6 naloxone treated animals (83%) survived. With hemorrhagic shock, results were even more striking. All of the naloxone-treated dogs survived 24 hours following hemorrhage and none of saline-treated animals remained alive.

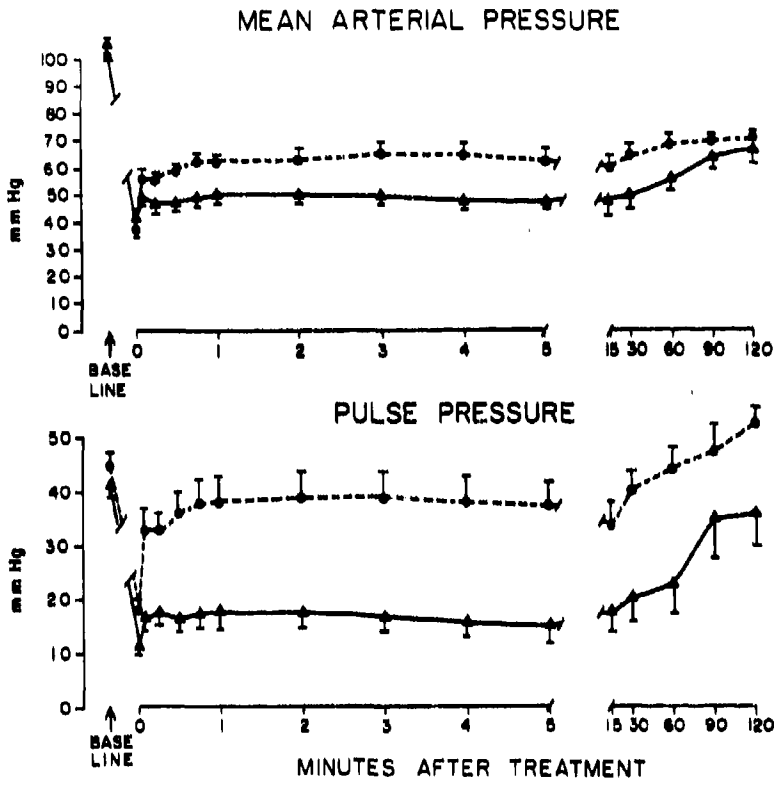



Fig. 4. The effects of 1.0 mg/Kg naloxone (●--●) and saline (▲—▲) treatment on MAP (top) and PP (bottom) following hemorrhagic shock. Naloxone treatment significantly improved these cardiovascular parameters, whereas saline was without a significant effect. Fifteen rats were studied in each group; vertical bars are ± SEM.

Spinal shock, which occurs following transection of the spinal cord, is also known to significantly reduce blood pressure. In rats, 48 µg of naloxone's active (-) or "levo-" isomer injected into the ventricles of the brain (ivt) produced an equivalent improvement in blood pressure as seen with 10 mg/Kg (-) naloxone injected parenterally. The (+) isomer of naloxone, which is chemically identical to (-) naloxone but biologically inactive at opiate receptors, had no effect on BP following ivt injection (Figure 6). These findings indicate that the effects of naloxone in this model are mediated by opiate

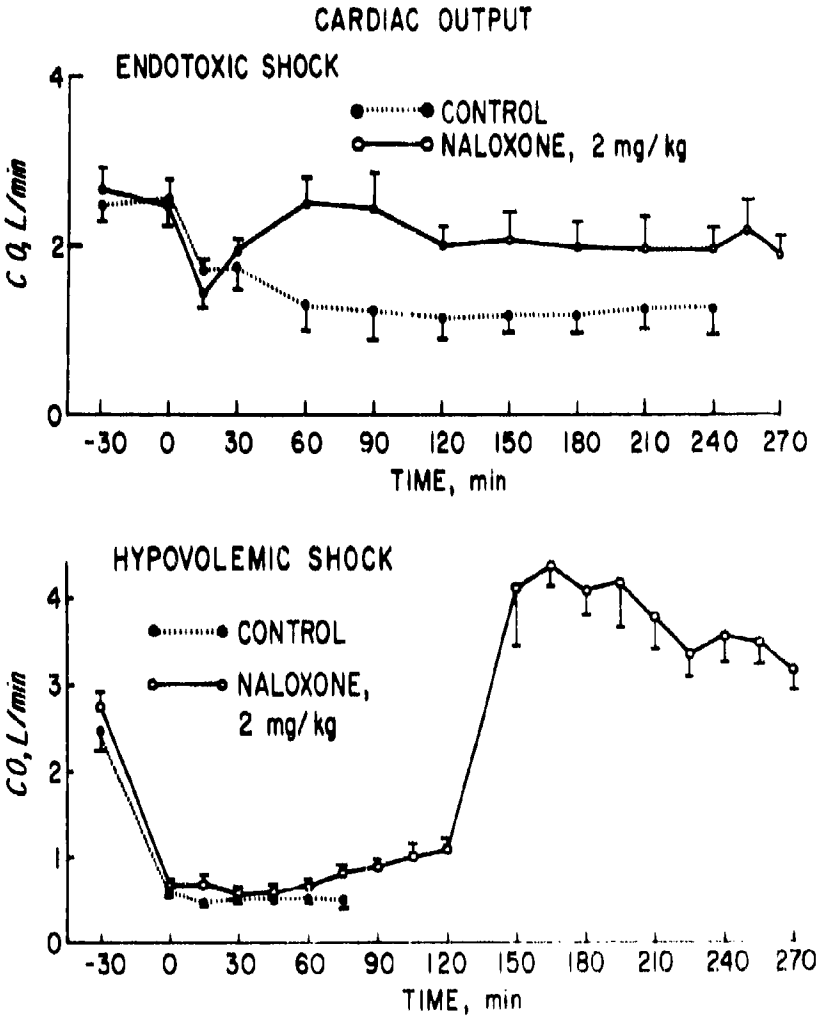



Fig. 5. Cardiac output (CO) was measured before and after treatment with naloxone or saline in canine models of shock. The upper figure demonstrates the decrease in CO produced by endotoxic shock; the lower figure represents the effects of hemorrhagic shock on this measure. Naloxone-treated dogs (o—o) experienced a sustained increase in CO, whereas CO in saline treated animals (●--●) continued to decline over time (injections at arrow). Data points for saline-treated dogs subjected to hemorrhagic shock end abruptly (bottom figure) since all 6 dogs died before shed blood could be reinfused. Vertical bars are + SEM.

receptors within the central nervous system (9,10). Injection of this very small ivt dose of (-) naloxone parenterally produced no effect. The fall in body temperature as well as respiratory depression following spinal shock were also reversed by naloxone (data not shown).

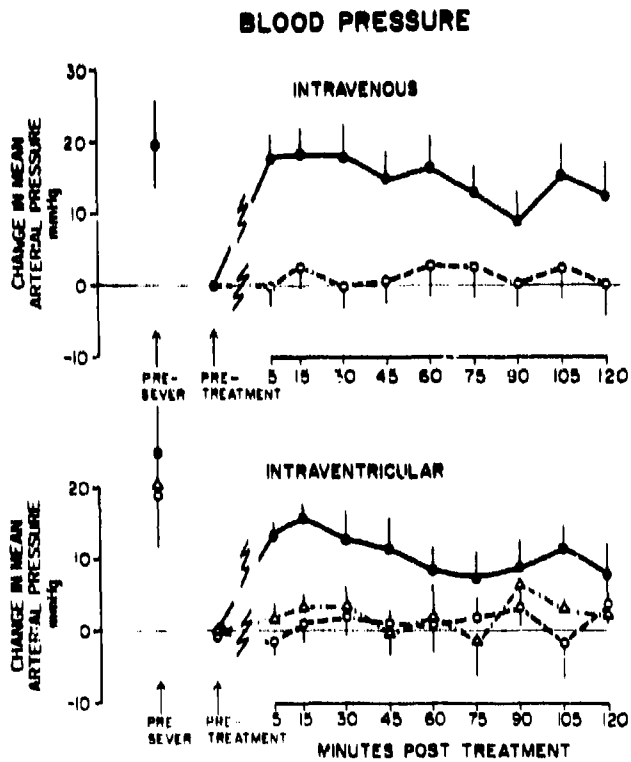



Fig. 6. Following rapid transection of the spinal cord in rats, MAP fell about 20 mm Hg. Rats receiving intravenous naloxone (top graph, 10 mg/Kg, 11 rats/group) had a sustained increase in MAP. In the bottom graph, the effects of 48 µg (-) naloxone (●—●) were compared to the effects of 48 µg (+) naloxone (△···△) or drug vehicle (○—○) following intraventricular injection in spinally transected, hypotense rats (6-7 rats/group). Vertical bars are ± SEM.

Spinal shock in cats also produced a significant fall in MAP and cardiac contractility (11,12). In these studies, four treatment groups were used. Cats were treated with naloxone, naloxone following vagotomy, naloxone following atropine, or saline. Since transection of the spinal cord leaves intact only parasympathetic innervation from the brain to the heart, the effect of naloxone on cardiovascular parameters would appear to be mediated by the vagus nerves. As in the rat, naloxone treatment significantly improved cardiovascular parameters. These effects of naloxone were completely blocked by vagotomy or atropine (Fig. 7). Thus, at least in spinal shock, naloxone acts upon parasympathetic centers in the CNS to improve cardiac performance via the vagus nerves which provide cholinergic innervation to the heart.

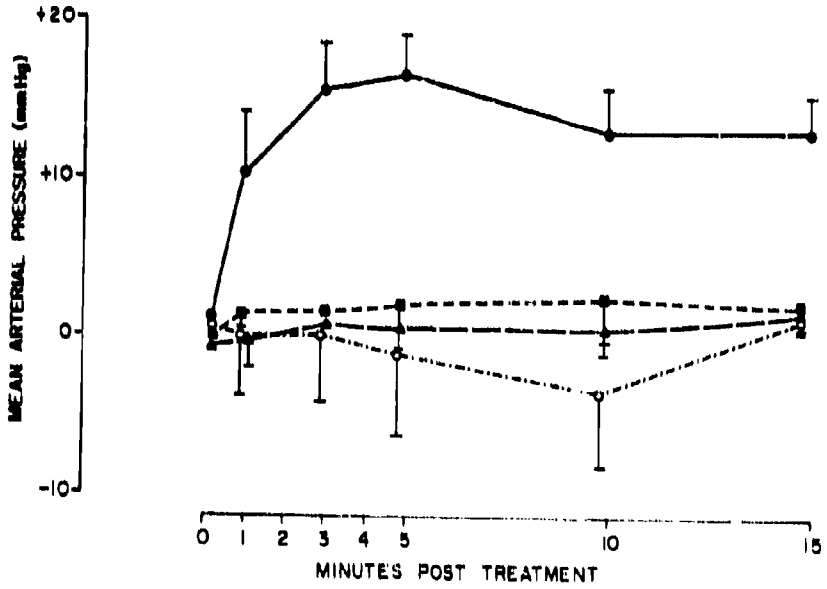



Fig. 7. Following hypotension caused by spinal cord transection in cats, animals received either naloxone (1 mg/Kg, ●—● ) or saline (▲--▲). Additionally, naloxone was injected into vagotomized (■--■) or atropinized (○--○) cats following the hypotension produced by spinal cord transection. Naloxone only improved MAP in cats <u>without</u> vagotomy or atropine; saline was without effect as well.

Following spinal cord trauma in cats, BP was again seen to drop approximately 15-20 mm Hg. Naloxone injections produced a significant increase in MAP, and saline was without effect on this measure. Neurologic examinations demonstrated that the naloxone-treated cats had significantly less paralysis than saline-treated animals at 24 hrs, 1, 2, and 3 weeks following spinal trauma (data not shown).

## DISCUSSION

The therapeutic efficacy of naloxone has been demonstrated following endotoxic shock in rats (1,7,8) and dogs (13), in hemorrhagic shock in rats (14,15,16) and dogs (17), and following spinal shock in rats (9,10) and cats (11,12). Since naloxone has been described as a pure opiate antagonist, its effects in altering physiological and behavioral parameters have been used to infer a blockade of endorphin-mediated events (1). From this perspective, our findings summarized above provide experimental evidence that endorphin systems significantly contribute to the pathophysiology of shock produced by a variety of means in different species. More importantly, the therapeutic efficacy of naloxone in all of these shock models points to its potential clinical utility in improving survival and recovery in humans.

Since we have shown that doses as low as 0.1 mg/Kg naloxone have therapeutic benefit in animals requiring greater than 100.0 mg/Kg for toxic effects, the large therapeutic ratio for this narcotic antagonist indicates a wide safety margin for its clinical use in treating shock.

Studies on the site and mechanism by which endorphins contribute to shock pathophysiology have recently indicated that pituitary endorphins gain access to the CNS where they ultimately produce the cardiovascular manifestations which characterize this syndrome (15,16). In these experiments, it was demonstrated that animals without pituitary glands, (hypophysectomized) do not respond to naloxone following hemorrhagic shock. By contrast, in control animals with intact pituitary glands, 10 µg of naloxone ivt effectively restored blood pressure, thus indicating a CNS site of action for these pituitary endorphins (16).

Once in the CNS, endorphins appear to act on specific opiate receptors since their effects are stereospecifically reversed by minute amounts of naloxone. Data obtained from spinally transected rats and cats indicate that parasympathetic centers, possibly in the brainstem, are involved. The action of endorphins in this region appears to indirectly depress cardiovascular function via the vagal-cholinergic input to the heart (12). Lastly, the end result of endorphin effects in

shock are to depress cardiac contractility and thereby decrease cardiac output and MAP concomitantly (12,13,17).

Recent data indicate that the improvement in blood pressure following injury to the spinal cord is associated with an improvement in neurological recovery (6). Collectively, these findings predict that the therapeutic effects of narcotic antagonists such as naloxone may significantly improve survival and functional recovery following battlefield injuries.

## SUMMARY

We have shown that the specific-opiate antagonist naloxone rapidly improved blood pressure and significantly decreased mortality associated with endotoxic, hemorrhagic, and spinal shock. These therapeutic effects of naloxone in the treatment of shock were demonstrated in rats, cats, and dogs. In addition, naloxone was shown to significantly decrease the paralysis resulting from spinal-cord injury in the cat. Available evidence indicates that naloxone produces those effects by antagonizing endogenous opiates (endorphins) secreted from the pituitary gland in response to the stresses of shock or spinal trauma. Moreover, these endorphins appeared to depress cardiovascular function by acting at opiate receptors in areas of the brain which regulate cardiac contractility. Our findings predict that the use of naloxone in the care and management of battlefield injuries may significantly improve survival and functional recovery.

## ACKNOWLEDGEMENTS

The authors are indebted to Thomas P. Jacobs and Clifton F. Johnson for their expert technical assistance as well as to Pat Conners for her preparation of this manuscript. We thank Endo Laboratories, Garden City, NY, for their generous gift of naloxone.

## REFERENCES

1. Holaday, J.W., Belenky, G.L., Faden, A.I., and Loh, H.H. Possible function of β endorphin. In: Neuro-Psychopharmacology (B. Saletu et al, Eds.), Pergamon Press, New York, pp.503-514, 1979.

2. Guillemin, R., Vargo, T., Rossier, J., Minick, S., Ling, N., Rivier, C., Vale, W., and Bloom, F. β endorphin and adrenocorticotropin are secreted concomitantly by the pituitary gland. Science 197:1367-1369, 1977.

3. Rossier, J., French, E.D., Rivier, C., Ling, N., Guillemin, R., and Bloom, F.E. Foot-shock induced stress increases β endorphin levels in blood but not brain. Nature 270:618-620, 1978.

4. Lemaire, I., Tseng, I., and Lemaire, S. Systemic administration of β endorphin: potent hypotensive effect involving a serotonergic pathway. Proc. Nat'l. Acad. Sci. U.S.A. 75:6240-6242, 1978.

5. Holaday, J.W. and Faden, A.I. Naloxone reversal of endotoxin hypotension suggests role of endorphins in shock. Nature 275:450-451, 1978.

6. Faden, A.I., Jacobs, T.P., Holaday, J.W., and Rigamonti, D. Opiate antagonist improves neurologic recovery after spinal injury. (Submitted for publication.)

7. Holaday, J.W. and Faden, A.I. Naloxone improvement of shock pathophysiology: evidence for opiate receptor involvement. In: Endogenous and Exogenous Opiate Agonists and Antagonists (E.L. Way, Ed.), Pergamon Press, New York, pp.479-482, 1979.

8. Faden, A.I. and Holaday, J.W. Naloxone treatment of endotoxin shock: stereospecificity of physiologic and pharmacologic effects in the rat. J. Pharmacol. Exp. Ther. (in press).

9. Faden, A.I. and Holaday, J.W. Naloxone reversal of hypotension caused by spinal transection. In: Endogenous and Exogenous Opiate Agonists and Antagonists (E.L. Way, Ed.), Pergamon Press, New York, pp.483-486, 1979.

10. Holaday, J.W. and Faden, A.I. Naloxone acts at central opiate receptors to reverse hypotension, hypothermia, and hypoventilation in spinal shock. Brain Res. (in press).

11. Faden, A.I., Jacobs, T.P., and Holaday, J.W. Naloxone alteration of physiologic parameters in spinally-transected animals. Trans. Am. Neurol. Assn. (in press).

12. Faden, A.I., Jacobs, T.P., and Holaday, J.W. Endorphin-parasympathetic interaction in spinal shock (submitted for publication).

13. Reynolds, D.G., Gurll, N.J., Vargish, T., Lechner, R., Faden, A.I., and Holaday, J.W. Blockade of opiate receptors with naloxone improves survival and cardiac performance in canine endotoxic shock. Circ. Shock (in press).

14. Faden, A.I., and Holaday, J.W. Opiate antagonists: a role in the treatment of hypovolemic shock. Science 205:317-318, 1979.

15. Holaday, J.W. and Faden, A.I. Hypophysectomy inhibits the therapeutic effects of naloxone in endotoxic and hypovolemic shock. Physiologist 22:57, 1979.

16. Holaday, J.W., O'Hara, M., and Faden, A.I. Pituitary endorphins are centrally involved in the pathophysiology of hypovolemic shock. (submitted for publication).

17. Vargish, T., Reynolds, D.G., Gurll, N.J., Lechner, R.B., Holaday, J.W., and Faden, A.I. Naloxone reversal of hypovolemic shock in dogs. Circ. Shock (in press).

FOOTNOTE

This material has been reviewed by the Walter Reed Army Institute of Research, and there is no objection to its presentation and/or publication. The opinions or assertions contained herein are the private views of the authors and are not be to construed as official or as reflecting the views of the Department of the Army of the Department of Defense.

# INFRARED TRANSMISSION MEASUREMENTS THROUGH SCREENING SMOKES: EXPERIMENTAL CONSIDERATIONS

GERALD C. HOLST, Ph.D.
CHEMICAL SYSTEMS LABORATORY, USAARRADCOM
ABERDEEN PROVING GROUND, MD 21010

## 1. INTRODUCTION

In order to evaluate the effect of tactical screening smokes on infrared transmission, it is necessary to understand the complex interactions among many variables. The transmission depends upon the bulk properties of the material (e.g., index of refraction), as well as the particle-size distribution, concentration, and pathlength. The measured or apparent transmission can be quite different from the true transmission because several simple basic facts are often overlooked. Therefore, it becomes quite difficult to compare data from different laboratories or from field tests because the experimental methodology is different at each location.

In principle, it is possible to calculate the transmission of the smoke if the particle-size distribution, concentration, pathlength, and the complex index of refraction are known. But these parameters are not always known precisely, and one resorts to experimentation to define them. The experiment becomes that of introducing a smoke with unknown infrared properties between the target and the detector. The ratio of the signal received with smoke to that without smoke is taken as the transmission.

In addition to transmissometers, the effects of screening smokes on thermal imaging systems is of interest. The experiment is similar to the transmissometer test. A smoke is introduced until the target can no longer be perceived by an observer. The amount of smoke required is a measure of the smoke's obscuring power. A second type experiment consists of introducing a "standard" smoke between several

different imaging systems and a target. The concentration is increased until the target has disappeared on several imaging systems. At this point, certain conclusions are drawn about the smoke's screening effectiveness, but the results may be a measure of the system's behavior and the observer's ability rather than the smoke's effectiveness.

Presented in sections II through V are the basic principles that the experimentalist should consider before designing the experiment.

## II. Laser Transmission and Scattering

Using Beer's law it is possible to calculate the transmission of an aerosol when the mass extinction coefficient $\alpha$, the concentration C, and the pathlength L are known:

$$T = \mathrm{EXP}\left[-\int_0^L \alpha C(\ell)\, d\ell\right] \tag{1}$$

Assuming spherical particles, $\alpha$ can be calculated from the Mie scattering theory:

$$\alpha = \frac{3}{4\rho}\int_r Q(m,\ r/\lambda)\,\frac{N(r)}{r}\,dr \tag{2}$$

where Q is the Mie scattering factor, $\rho$ is the mass density, m is the complex index of refraction, r is the radius of aerosol particle, $\lambda$ is the wavelength, and N(r) is the number density size distribution. The reason that $\alpha$ is usually obtained experimentally is that often neither m nor N(r) is known. From equation (2), it follows that, if the size distribution changes, than $\alpha$ will change also. Any dynamic process, such as coagulation, sedimentation, or introduction of a new aerosol with a different size distribution, will change N(r). For hygroscopic smokes, such as phosphorus and HC, the size distribution will depend upon the relative humidity and hence, $\alpha$ will be a function of relative humidity.

In recent experiments, fog oil smoke was produced by a vaporization condensation method. A laser transmissometer operating at $\lambda$ = 0.6328 μm measured the transmission as a function of time as shown in figure 1. The concentration was simultaneously obtained as shown in figure 2. The size distribution was obtained with cascade impactors at time t = 0.0 min and t = 51 min. The mode diameter increased and the total number of particles greater than 2.0 μm actually increased, suggesting that coagulation has occurred. Thus N(r) is changing with time. Using the data in figures 1 and 2, the extinction coefficient was calculated and is plotted in figure 3.

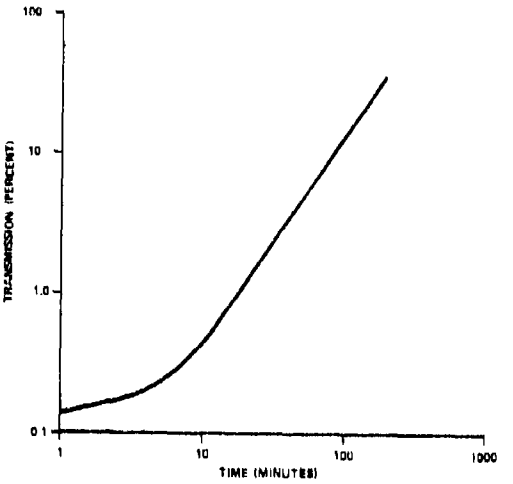



FIGURE 1. Transmission of a fog oil smoke as a function of time. Data obtained with a HeNe laser transmissometer ($\lambda$ = 0.6328 $\mu$m). Laser pathlength was 1.2 m. Since the transmission is changing with time, some type of dynamic process is occuring.

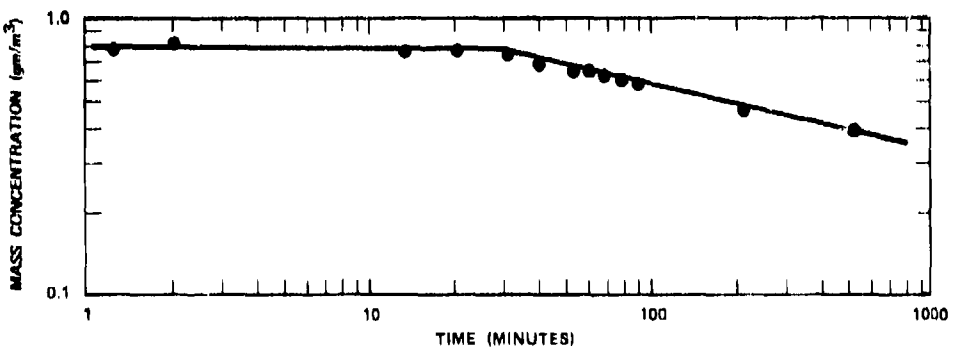



FIGURE 2. Mass concentration of a fog oil smoke as a function of time. Data obtained by gravimetric analysis of glass fiber filters. The aerosol was aspirated for 1 minute for each sample. The aerosol is stable for the first 20 minutes.

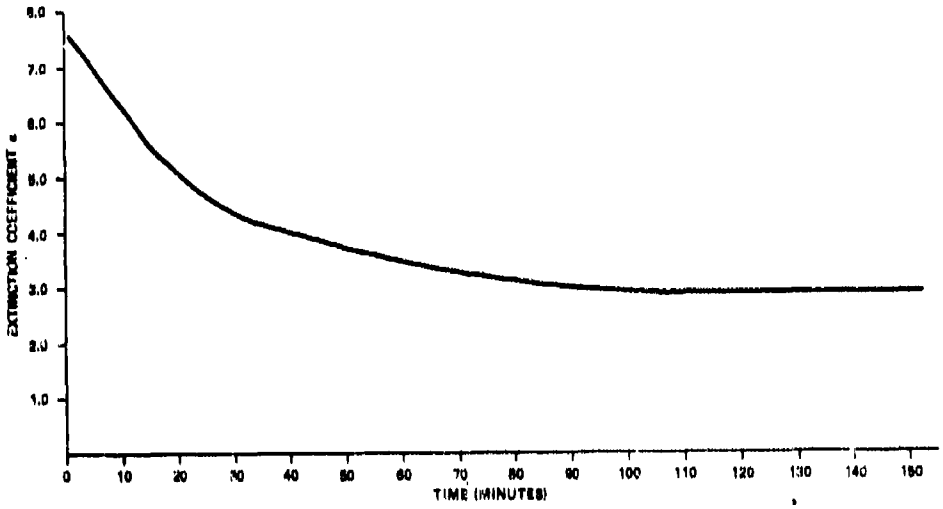



FIGURE 3. Calculated extinction coefficient as a function of time for a fog oil smoke. The tranmission data shown in Figure 1, mass concentration data shown in Figure 2 and the pathlength L = 1.2 m was used in Beer's law. Since α is a function of time, the age of the cloud is significant when specifying α

As a general rule, coagulation becomes significant when the number density exceeds $10^6$ particles/cm $^3$ and the diameter is less than 1 μm. Sedimentation, on the other hand, dominates when D > 1 μm in a chamber.

The extinction coefficient derived from Mie theory calculations assumes that any light scattered out of the incident beam contributes to total extinction. These calculations generally assume that the incident beam is a plane parallel wave of infinite extent and that the detector field of view (FOV) is essentially zero. In reality, the signal detected will depend upon how much scattering has taken place. Middleton[1] has shown that even for the single scattering case the amount of detected scattered light increases significantly as the detector FOV increases.

For dense aerosols, multiple scattering can occur. Second-order forward scattering has been treated theoretically[2]. The results indicate that, for typical polydisperse naturally occurring aerosols, a correction factor can be applied to the measured transmission to obtain the true extinction. The factor R which is a function of FOV,

size distribution and index of refraction appears in Beer's law as

$$T \text{ measured} = e^{-\alpha_m CI} = e^{-R\alpha_T CL} \tag{3}$$

where

$$R = \frac{\alpha \text{ Measured}}{\alpha \text{ True}} = \frac{\alpha_m}{\alpha_T} \tag{4}$$

Hooradian et. al[3] have shown that the detected signal is a function of receiver FOV for naturally occurring fogs. (See figure 4).

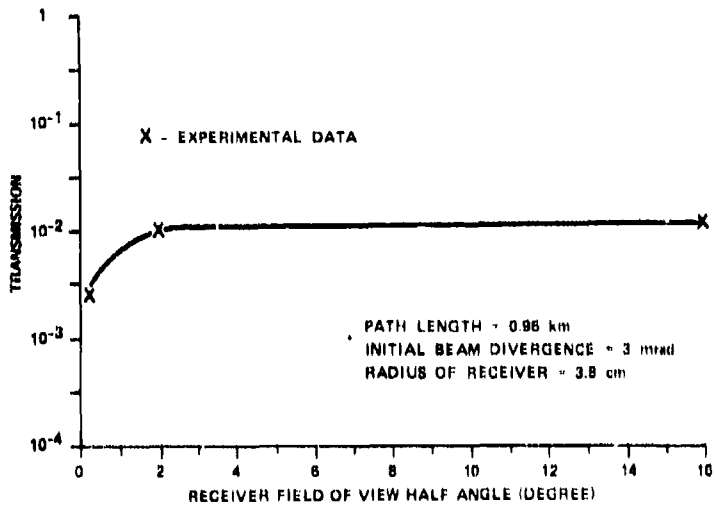



FIGURE 4. Measured transmission as a function of receiver field of view for naturally occuring water fogs (from Reference 3). The optical depth $\tau = 11.8$. The fog was not homogeneous. The laser transmissometer was a frequency doubled Q-switched Nd: YAG laser ($\lambda = 0.53$ μm). The solid line is the theoretical prediction which includes multiple scattering for dense aerosols.

## III. Broadband Detectors: Spectral Response Consideration

The extinction coefficient, as shown in equation 1, is, in general a function of wavelength. For broadband detectors, the measured extinction is a function of the source temperature and the detector spectral response, as well as the spectral transmission of the smoke. The total flux reaching a detector from a target with emissivity $\varepsilon_T$ at a single wavelength is given by

$$F_{OT} = T_a \varepsilon_T R_T + L_a \tag{5}$$

where $T_a$ is the transmission of the intervening atmosphere, $R_T$ is the blackbody power emitted by the target, and $L_a$ is the radiance of the atmosphere. When an emissive smoke is introduced, the flux reaching the detector becomes

$$F_{ST} = T_S T_a \varepsilon_T R_T + L_a{}^1 + T_a{}^1 R_s \tag{6}$$

where $T_s$ is the transmission of the smoke, $L_a{}^1$ is the radiance modified by the smoke, and $T_a{}^1$ is the transmission of the atmosphere between the smoke and detector. $R_s$ is a combination of the flux emitted by the smoke and the flux reflected off the cloud from external sources.

The current signal generated by these fluxes is given by I=SF, where S is the sensitivity of the detector. For broadband detectors the detector integrates over the wavelength of interest so that

$$I = \int_\lambda SF d\lambda \tag{7}$$

With synchronous detection, the source is modulated at some frequency and the detector electronics are sensitive to this frequency only. The path radiance and smoke emission are DC components and therefore are not measured. The measured ratio of smoke to no-smoke currents becomes

$$\langle T \rangle = \frac{I_{ST}}{I_{OT}} = \frac{\int S T_S T_a \varepsilon_T R_T d\lambda}{\int S T_a \varepsilon_T R_T d\lambda} \tag{8}$$

<T> is the spectral averaged transmission and depends upon the spectral character of S, $T_a$, $R_T$, and $T_s$.

The effect of spectral mismatch between the smoke and the detector only is shown in the following exaggerated example. Consider four different hypothetical systems, all of which are nominally classified as 8 to 14 μm radiometer-transmissometer systems. The spectral sensitivity of each is shown in figure 5. Consider a screening agent whose transmission is 80% from 8 to 11 μm and from 100% from 11 to 14 μm for unit pathlength and unit concentration. As the smoke concentration increases, the transmission in the 11 to 14 μm stays at 100%, and the transmission from 8 to 11 μm approaches zero. Thus, as far as detector 2 is concerned, the smoke is transparent. Since detector 1 integrates over the entire region, the transmission approaches 50% and with detector 4, the transmission goes to zero. Detector 3 shows some intermediate value. In figure

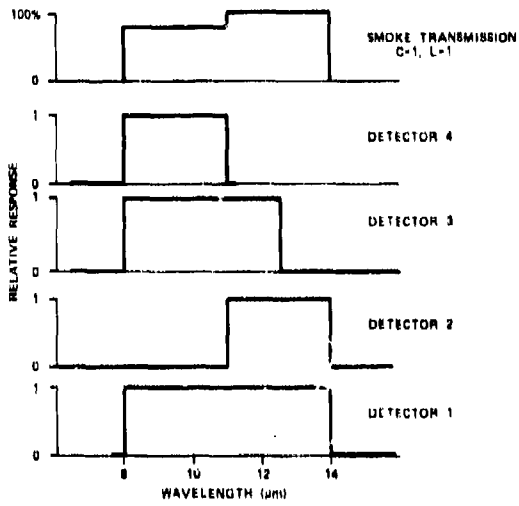



FIGURE 5. Spectral response of four hypothetical detectors. The relationship between these detectors and a hypothetical smoke is shown.

6, the expected transmission for each of these detectors is plotted as a function of concentration. The values were calculated with equation 8 and by letting $R_T = \epsilon_T = T_a = 1$. Although this example was exaggerated, the shape of the hypothetical smoke transmission is somewhat similar to phosphorus smoke[4,5]

The effect is not limited to the spectral mismatch of the smoke and detector. The spectral emission of the target (source) will also affect the measured transmission or extinction coefficient. Assuming that the source is a blackbody, calculations indicate that if a broadband HgCdTe detector is used, the measured transmission will decrease from 45% to 39% as the blackbody temperature is increased from 300°K to 1000°K.

Thus it is easy to see that there is no simple way of obtaining the smoke transmission with a broadband detector because of the measured transmission depends upon the target temperature, spectral response of the detector, and the atmospheric transmission. The atmospheric transmission is a function of pathlength, relative humidity, temperature, and local naturally occurring aerosols.

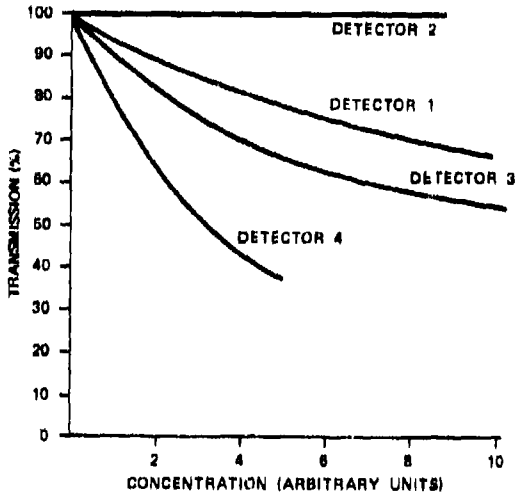



FIGURE 6. Calculated transmission as a function of concentration. The extremely large difference in transmission between detector 2 and detector 4 illustrates why spectral responses must be considered before collecting data.

## IV. Imaging Systems: Minimum Resolvable Temperature Considerations

The imaging systems, in addition to having the same limitations as discussed in the preceding sections, also rely upon the observer's evaluation of the image. As mentioned when evaluating the effectiveness of a smoke, the smoke concentration is increased until an observer can no longer detect the target. At this point, the target signature is below the internal noise of the entire detector-electronics-human observer system.

The noise level is referred back to the input as an equivalent temperature and is specified as the minimum resolvable temperature (MRT) above ambient for a "standard" observer and is expressed as a temperature differential $\Delta T$ above ambient. The MRT is a function of the angular subtense $\alpha$ of the target and is plotted for two systems in figure 7. These systems are identical in the sense that for all targets larger than $\alpha_c$ the exhibit the same MRT. At another target angular subtense $\alpha_1$, the MRT of the two systems are given by $\Delta T_{N1}$ and $\Delta T_{N2}$ respectively. Assume that a target is presented before these two systems and it is $\Delta T$ above the ambient. If an absorbing smoke is placed between the detectors and the target, then, since $\Delta T_1$ is larger than $\Delta T_2$, more smoke is needed to bring the target signature below the

MRT for system 1 than for system 2.

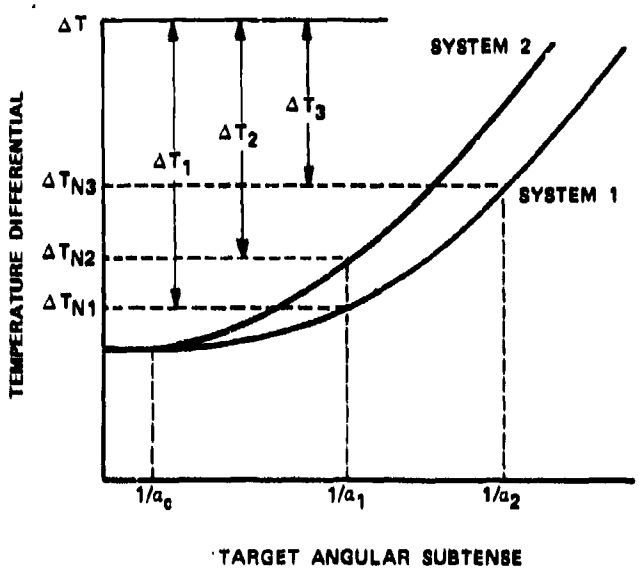



FIGURE 7. Minimum resolvable temperature of two hypothetical imaging systems as a function of target angular subtense.

Consider instead that, during two different tests, only system 1 is used and the target sizes are slightly different ($\alpha_1$ and $\alpha_2$). Since $\Delta T_1$ is greater than $\Delta T_3$, more smoke is needed to obscure target size $\alpha_1$. During field smoke trials, actual military targets (tanks, trucks, etc.) are used. The equivalent $\Delta T$ of the target depends upon the emissivity, target temperature, target size, and the spatial distribution of the heat (e.g., the engine area will be the hottest area). An equivalent $\Delta T$ is calculated so that the MRT can be used to describe system performance. Since the spatial distribution of the heat and the temperature can change from test to test, it is easy to see that the equivalent $\Delta T$ will change. Therefore, it becomes exceedingly difficult to compare field data. Conclusions reached about the effectiveness of smoke may result from differences in the detector MRT or target equivalent $\Delta T$ rather than the optical properties of the smoke.

In many field tests, trained observers are not used. The relationship between a trained observer and a casual observer is not clear. This is extremely important because the systems are characterized by how well a trained observer can perceive targets.

## V. Target Obscuration

Consider, now, an imaging device which processes the detected flux into various gray levels to produce an image on a TV screen. Let us assume that the device is adjusted so that, before the smoke is introduced, the hot target will appear as pure white and the background will appear as black. Let us assume that the device will insert 10 gray levels between these points so that each gray level is proportional $(I_{OT}-I_{OB})/10$ where $I_{OB}$ is the background signal without smoke.

Before any further analysis is possible, the exact mode of operation of the thermal imaging system must be known. There are two basic modes. In the first, the instrument is adjusted for optimum display and the controls are not further adjusted, even after the smoke is introduced. In the second mode, the internal automatic gain control (AGC) automatically adjusts the gain so that an optimum display is always present. The first mode is typical of laboratory-type systems and the second is typical of military imaging systems.

Let us first consider a device which will not be readjusted. There are three possible conditions which will obscure the target as shown in figure 8. If the flux differential between target and background is very small, then a relatively high transmission will obscure a target in A. In B very little emission/reflection energy is necessary to make the entire screen white.

In C, the relationship between emission/reflection and attenuation causes the image to disappear into a gray level and then the entire screen will be gray. Then

$$I_{ST} - I_{SB} \leq \frac{I_{OT} - I_{OB}}{10} \tag{9}$$

This results in an averaged transmission $(\bar{T}) \leq 1/10$. It indicates that the observer will be unable to see his target everytime the spectral averaged transmission drops below 10% for a 10 gray level system provided that he does not readjust his thermal imaging system. If the system is not optimized, i.e., both the background and target are in the gray levels, then the relationship presented above is relaxed in the same sense that a higher transmission smoke will produce the same effect. Thus, the three cases presented are "worst" cases.

Consider, now, the device which has an AGC or one that is readjusted for optimum display after the smoke is present. The only inequality that exists is when the attenuation and emission-reflection combined produces a signal differential between the target and back-

ground that is below the equivalent noise, current, $I_N$:

$$I_{ST} - I_{SB} \leq I_N \tag{10}$$

We can represent this current as an equivalent blackbody $R_N$ with emissivity $\epsilon = 1$, and a temperature differential $\Delta T_N$ above the ambient. Full analysis indicates that

$$\langle T^1 \rangle \leq \frac{\Delta T_N}{\Delta T} \tag{11}$$

which is the well-known equation that indicates the spectral average smoke transmission must be sufficiently small to reduce the target-background differential below the MRT of the imaging system. This average $\langle T^1 \rangle$ is different from that defined by equation 8.

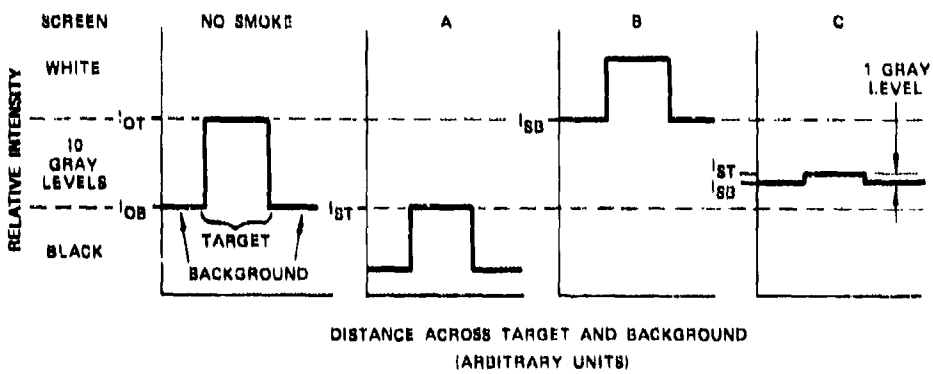



FIGURE 8. Three methods of obscuring a target detected with a thermal imaging system. With no smoke the target is pure white and the background is pure black. In A, smoke reduced the target signature below the black level. In B, the smoke emission/reflection raised the background into the white level. In C, emission/reflection combined with attenuation put the target-background intensity differential into a gray level.

For both systems (e.g., with ACG and without) we have assumed that the emissivities were equal to one and that the blackbody flux was wavelength independent and proportional only to the temperature. In practice, however, none of the assumptions are truly valid, and, therefore, we must consider the conclusions as guidelines rather than as facts.

Because of all the problems associated with thermal imaging systems, they should not be used to quantify the effectiveness of smoke. Rather they should be used solely for qualitative purposes.

## VI. Conclusions

The various difficulties encountered in measuring transmission of aerosols have been discussed. Specifically, scattered light, whether single or multiple, can enter a large FOV detector and give an artificially high transmission. Any dynamic process, such as sedimentation and coagulation, can alter the size distribution and thereby alter the results of transmission measurements. For aerosol measurements to be reproducible, it is necessary to generate exactly the same size distribution and concentration each time. A flow chart indicating possible errors in laser transmissometer measurements is shown in figure 9.

To intelligently analyze data from broadband detectors, the spectral components of the target, smoke, intervening atmosphere, and the detector must be known. To obtain consistent results, in addition to the aerosol reproducibility requirements, the target temperature and intervening atmosphere must always be the same. For small pathlengths, the atmospheric transmission is near 100%; however, for long pathlengths, transmission depends upon relative humidity, temperature, and atmospheric constituents, all of which can vary on an hourly basis. If two different detectors are used, the spectral responses must be identical. The flow chart illustrating these problems is shown in figure 10.

Field measurements are extremely difficult because the local meteorological conditions can disperse the cloud and create inhomogeneities in concentration. Futhermore, recent evidence suggests that fog oil droplets may partially evaporate. In that case since r is decreasing, even the extinction coefficient will change.

Finally, for target obscuration with a thermal imaging system with an AGC, sufficient smoke must be present to reduce the target-background temperature differential below the MRT of the system. The MRT depends upon the angular subtense of the target. The problems inherent to broadband detectors and transmissometers also apply. The flow chart is shown in figure 11. Note that only trained observers should be used if the only information available is the display screen.

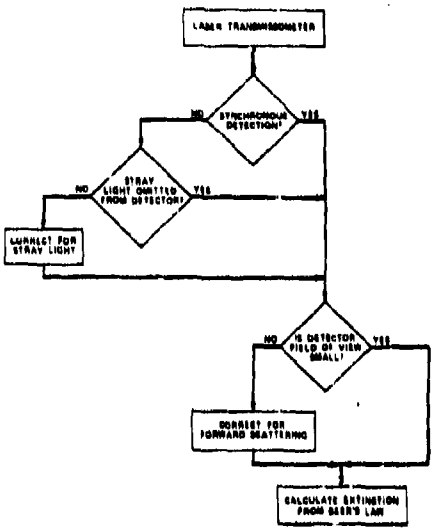



FIGURE 9. Flow chart of possible experimental errors encountered with laser transmissometers.

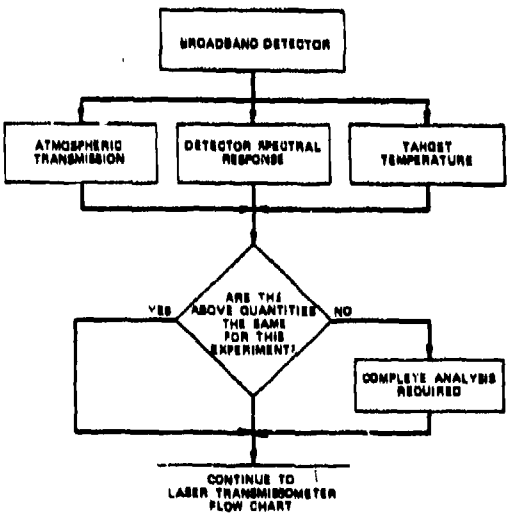



FIGURE 10. Flow chart of possible experimental errors encountered with broadband detectors.

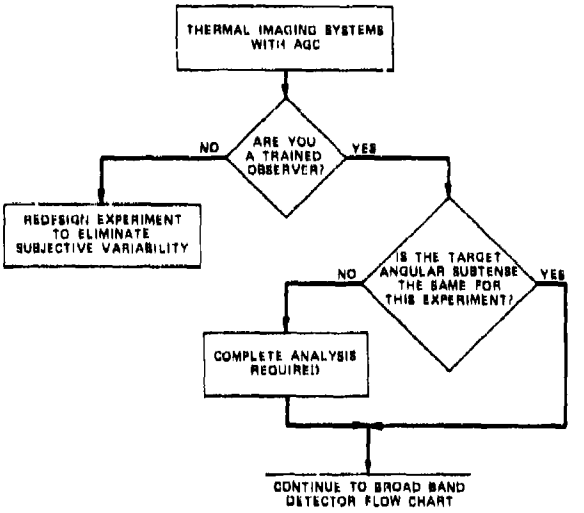



FIGURE 11. Flow chart of possible experimental errors encountered with thermal imaging systems.

References

1. W. E. K. Middleton. Vision Through the Atmosphere, University of Toronto Press, 1952, pg 177.

2. A. Deepak and M. A. Box, "Forward Scattering Corrections for Optical Extinction Measurements in Homogeneous Polydisperse Aerosols", Applied Optics, 17, 3169, 1978.

3. G. C. Mooradian, M. Geller, L. B. Stotts, D. H. Stephens, and Krantwald, "Blue-Green Pulsed Propagation Through Fog", Applied Optics, 18, 429, 1979.

4. M. Milham, "A Catalog of Optical Extinction Data for Various Aerosol/Smokes", ED-SP-77002, June 1976.

5. G. C. Holst and M. E. Milham, "Examination of the Correlation between Laboratory and Field Smoke Extinction Data", in Proceedings of the Smoke Symposium, HDL, April 1979.

# VEHICLE MOBILITY OR FIRING STABILITY
# A DELICATE BALANCE

*FRANCIS B. HOOGTERP, MR.
RONALD R. BECK, DR.
TANK-AUTOMOTIVE RESEARCH AND DEVELOPMENT COMMAND
WARREN, MI 48090

## INTRODUCTION AND MOTIVATION

The Army's current interest in mounting a large caliber cannon on a light weight tracked armoured combat vehicle has precipitated the evaluation of numerous vehicle-weapon system configurations. Yet the same question has continually resurfaced: "Is it feasible to mount a large cannon on a lightweight vehicle chassis?" Questions such as this can seldom be answered uniquely. The answer in this case depends on whether there is a strong enough need for such a system and on its mission, if and when it is fielded. The design community's job is to come up with the most workable weapon system concept within the given constraints and to attempt to relate the performance of this concept to that of some known combat vehicle. If this task is executed properly, then and only then, can the question be answered in a satisfactory manner.

The mission profile of such a vehicle dictates that it must be capable of destroying enemy tanks. Against the heavily armoured tanks of modern armies, this requires a relatively large cannon. This vehicle would also be expected to perform a reconnaissance role. In order to accomplish this successfully, the vehicle must possess superior mobility . A further requirement is that it shouldn't cost as much as a conventional tank. These restrictions are most easily satisfied by keeping the vehicle weight as low as possible. This tends to keep the cost per vehicle down and allows a fairly high horsepower per ton ratio. Any tactical gain obtained from increased vehicle acceleration performance, given by the high horsepower per ton ratio, will be



PRECEDING PAGE BLANK-NOT FILMED

quickly eroded unless a correspondingly good cross country ride performance is designed into the vehicle.

The overall vehicle survivability can be further enhanced by allowing the cannon to be operated in burst fire mode. This can increase the kill probability significantly in a surprise attack on an enemy tank. This increased kill probability can only be achieved, however, if the vehicle's stabilization system is capable of keeping the gun pointed at the target in the aftermath of the initial firing of the cannon. The desirability of burst fire capability is the major motivation for studying the platform firing stability problem. If the firing platform (i.e. the vehicle) remains more nearly stationary while the gun is recoiling from a previous firing, the stabilization system will be able to keep the gun on target more easily. A second reason for investigating the transient motions of the vehicle due to firing the cannon is the possibility that this motion might have an adverse effect on the crew of the vehicle. This is a problem that is somewhat unique to this type of vehicle-weapon system. The fact that this is a concern is testified to by gunners who have fired the M551 Sheridan gun (17.5 Ton with 152mm gun).

The design of a lightweight vehicle to carry a large cannon is not a new venture for the U.S. Army. Development of the M551 Armoured Reconnaissance/Airborne Assault Vehicle (also known as the Sheridan) was initiated in 1959. Approximately 1700 Sheridans were produced at the Cleveland Tank Automotive Plant by General Motors Corporation for the U.S. Army from 1966 through 1970 (1). The Sheridan is armed with a 152 mm main gun and the final design weighed in at approximately 35,000 lbs. The main gun was perhaps larger than desired in some respects. This came about because of the design constraint to allow the launching of the Shileleaugh Missile. The 152 mm gun could also fire conventional ammunition but the vehicle's violent reaction to such firings, gave many a gunner a severe headache.

The Sheridan had the additional constraint placed upon its design that it be air dropable. This requirement, along with the necessity to carry (and fire) such a large cannon, caused certain compromises to be made in the vehicle design, adversely affecting the vehicle's cross country ride performance.

The purpose of this study was to perform a sensitivity analysis, using available simulation models, and to determine the most influential vehicle parameters with respect to a vehicle's cross country ride performance and to the firing stability of an initially stationary vehicle. It is anticipated that the results produced will be of assistance to the designer of future vehicle-weapon systems.

Before discussing the results of the analysis, we will discuss the models and methodologies used. The model descriptions will be of necessity, somewhat brief, but will attempt to give an overall appreciation for the composition and complexity of each model. The discussion will also touch upon relevant model verification and validation efforts that have been accomplished and thereby try to convey some level of confidence in the results. In addition to discussing the models, a short dissertation is included on the measures of effectiveness employed for evaluating the ride comfort and firing stability of the various concepts.

## RIDE DYNAMICS MODEL

The vehicle dynamics model employed in this study is a two dimensional, 7 degree of freedom (for a vehicle with 5 roadwheels on a side) model. This model is implemented in FORTRAN and is combined with interactive data entry and postprocessor programs to provide a complete ride analysis package (4). Briefly the model consists of one second order differential equation describing the vertical motion of each roadwheel and two additional second order differential equations to describe the hull pitch motion about its CG and the vertical motion of the CG. The various spring and damper elements are treated in a piecewise linear fashion. Each of these suspension system elements is allowed to be represented by as many as nine separate (but connected) linear segments.

The track effects built into the model are rather simplistic but recent attempts to more completey validate the model have led to the inclusion of additional track effects by simply modifying the vehicle input characteristics slightly. A geometrical presmoothing of the profile of the terrain to be traversed is performed within the model and this "smoothed" profile is then used as the forcing function for each roadwheel. This presmoothing produces the approximate path that a slowly moving roadwheel would follow across the nondeflecting terrain. This results in very little modification of the milder profiles and will produce the most significant change to the sharpest obstacles. The effective vertical spring and damping characteristics of the combined roadwheel-track system are entered as piecewise linear functions. These spring and damper elements are attached between the modified terrain forcing function and the center of the roadwheel and are allowed to react only in the vertical direction.

One additional track effect was included for certain ride simulations performed. The rebound travels for the front and rear springs were restricted to 2.5 inches. The rationale for doing this is that the track prevents these roadwheels from dropping as far as

the intermediate wheels are allowed to. The 2.5 inch value was somewhat arbitrary but was selected after scrutinizing photographs of tracked vehicles traversing rough terrain. This effect can be pictured by imagining a vehicle being lifted completely clear of the ground. In this situation the front and rear roadwheels would actually be pulled upwards from their static positions by the tension in the track. In other situations, such as when the front wheel begins to cross a ditch, a limited amount of rebound travel does occur but never as much as would for an intermediate roadwheel.

For purposes of this paper, the evaluation of ride comfort will be considered only at the driver's station of the vehicle. In the concepts evaluated the driver sits nearly directly above the front roadwheel. The ride comfort of the driver for a particular velocity over a specified terrain profile is quantified in terms of the driver's vertical absorbed power. Absorbed power can be considered as being a frequency weighted rms acceleration, or more precisely as the power absorbed by the driver's body (2).

The frequency weighting is based on the average man's tolerance to acceleration at the various frequencies and was developed from extensive testing performed by the U.S. Army Mobility Systems Laboratory in the early 60's. A vehicle's limiting velocity for a given terrain is considered to be the lowest velocity at which the driver absorbs six watts of power (3).

Terrain roughness for the purpose of ride comfort evaluations is characterized by the rms (root mean square) of the profile measured in inches. A concept's ride limiting velocity, as defined above, is determined as a function of the terrain roughness. This relationship is considered to be the measure of the concept's ride comfort.

Ride comfort curves for two standard Army vehicles are shown in Figure 1. The solid curves represent the results predicted with this model for the M60A1 tank and for the MICV armored personnel carrier. The individual points, marked with an "x", for the M60A1

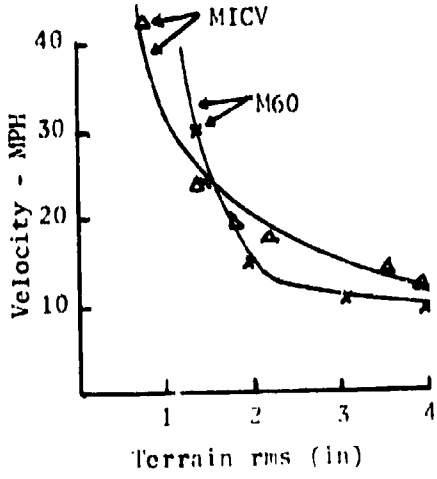



Figure 1
Ride Validation

were obtained from field tests conducted at Fort Knox in May 1979. The field tests from which the MICV data (denoted with a ▲) were obtained, were performed at Aberdeen Proving Ground in Nov. 1976.

## PLATFORM FIRING STABILITY MODEL

The purpose of the platform firing stability evaluations was to assess the relative hull pitch and roll motions for various vehicle configurations. Two distinct models were employed to obtain the results presented here.

The first effort was performed using a very large, general purpose, mechanical system formulation and simulation program. This program termed DADS, which stands for Dynamic Analysis and Design System, had been implemented by personnel at the University of Iowa and was modified, under a contract with the US Army, to support this particular application (5). Input to the program consists of the inertial and geometric properties of each body to be modelled and the locations and types of the various joints that interconnect the bodies. Each joint may also have a spring and damper associated with it and the attachment points for these elements are described relative to the appropriate body CGs.

The DADS program uses the physical description of the system (in this case the vehicle and weapon data) and generates an appropriate set of equations to represent the system. This set of differential and algebraic equations is then numerically solved for the user.

It is impossible to do justice to the power and utility of DADS in the span of a few lines. This program is an exceptional exploratory modelling tool because of the sophistication and completeness of the automatically formulated models. DADS also routinely generates detailed recordings of every force, acceleration, velocity, and displacement in the system. The DADS formulation, for pitch plane motion only, for a vehicle with 6 roadwheels per side resulted in a system of 86 equations in 86 unknowns. This detailed formulation has been used successfully to evaluate the firing stability of numerous vehicle concepts.

Since this study required the simulation of numerous additional concept variations, it was decided to develop a less detailed, but more efficient to use, model. The experience gained with the more sophisticated DADS model, provided guidance for the new implementation as well as baseline results against which the results of the more efficient but simpler model could be verified.

The model developed (dubbed PFIRS) was made available as part of a completely interactive, user oriented simulation package (6). PFIRS includes equations for hull pitch, roll, vertical, and side to side motions. The vehicle data representation employed is identical to that used by the ride dynamics model described in the previous section, with the addition of the necessary roll related parameters, trunnion position data, and the firing reaction characteristics.

This model accomodates firing from an oblique position and allows the vehicle to be sitting on a sideslope for the direct side firing situation. Due to the original program requirement for sideslope firing analysis, small angle approximations were not made for the roll equations. The pitch angle was assumed to remain small, however. This assumption was justified by the actual simulations. The largest pitch angle encountered for any concept was 11 degrees and in the majority of the cases it was below 8 degrees.

It was also assumed that the tracks would not slide relative to the ground. This seems to be a reasonable approximation since actual tests of the M551 firing straight ahead from level, hard packed gravel, indicated a sliding of only 1/4" (7).

The roadwheel motions are not described by differential equations. Instead the roadwheels are assumed to remain in contact with the ground (through the track) unless the associated suspension system member develops sufficient force to lift it. The justification offered for this approach is twofold. First the deflection of the roadwheel-track combination is very small compared to the hull motions of interest and therefore the effect of the roadwheel deflections on the suspension system forces should be minimal. Second, this may actually be a more realistic representation since we assume that the ground is nondeformable, and certainly the roadwheel-track is at best marginally deformable. This would tend to suggest a position constraint on the roadwheel rather than the more traditional spring-damper connection between the roadwheel-track and the ground. This implementation has the added advantage that it reduces the number of first order differential equations, for a vehicle with 10 roadwheels, from 26 to 6.

The performance measure used here to study platform firing stability is the maximum change in the pitch displacement (or roll displacement for firing off the side) of the vehicle. This measure should give a good indication of stability for evaluating the possible application of a burst fire mode of operation to a vehicle weapon system. Obviously the horizontal acceleration at the gunner's eyepiece

also has an important influence on the gunner's reaction to firing the gun.

If the horizontal acceleration at the CG is also desired, it can be estimated quite accurately. This fact is illustrated in Table 1 and is due to the extremely short duration of the recoil force. The estimated column, labeled EST, was computed by simply taking the maximum recoil force divided by the gross vehicle mass. The column labeled DADS was simulated with the DADS firing stability model.

| Concept | GVW tons | Gun Bore mm | Maximum Recoil ksi | Impulse Length msec | DADS g's | EST g's |
|---|---|---|---|---|---|---|
| I | 16 | 75 | 38 | 54 | 1.2 | 1.19 |
| II | 16 | 90 | 67 | 46 | 2.2 | 2.09 |
| III | 42.5 | 90 | 67 | 46 | 0.81 | 0.79 |
| M551 | 17.5 | 152 | 180 | 22 | 4.8 | 5.14 |

Table 1 Horizonal CG accelerations

It should be noted that the roll results throughout this paper are for firing directly off the side of the vehicle. The pitch results, on the other hand, are reported only for the vehicle firing straight ahead.

| Vehicle | simulated with DADS degrees | PFIRS degrees | Test Results degrees |
|---|---|---|---|
| I | 3.84 | 3.78 | -- |
| II | 7.0 | 7.34 | -- |
| M551 | 4.2 | 4.13 | 3.8 |

Table 2 Comparison of Maximum Pitch by Model

The firing data available for validating these models were very limited. The only useable data found were for the M551 Sheridan firing conventional ammunition from its 152mm gun. The maximum pitch

angle recorded for firing straight ahead, while resting on hard packed gravel, was 3.8 degrees (7). This compares quite favorably with the results obtained both from DADS and from PFIRS which are shown in Table 2. Also presented in Table 2 are two other concepts (the same concepts as in Table 1) which illustrate the level of agreement achieved between the two models. These comparisons are offered simply to verify the implementation of the models. This should not be misconstrued to imply that the model has been validated. Certainly there is a need to add to the single validation point offered (i.e. the M551 firing).

## RIDE COMFORT RESULTS

The Army's previous experience at mounting an oversize gun on a lightweight chassis resulted in a vehicle (the M551) with a vertical natural frequency of about 120 cpm. This is a very stiff suspension system when compared to the Army's other tracked vehicles. The natural frequencies and gross vehicle weights for several such vehicles are shown in Table 3.

| VEHICLE | GVW ksi | FREQUENCY cpm |
|---|---|---|
| M113 | 24 | 92 |
| M551 | 35 | 120 |
| MICV | 40 | 94 |
| M60 | 112 | 74 |
| XM1 | 120 | 72 |

Table 3 Typical Vertical Natural Frequencies

The effect of varying a vehicle's natural frequency on the vehicle's ride comfort is illustrated in Figure 2. Each of these concepts weighed 16 tons and had five roadwheels per side. Shock absorbers were positioned on the first, second, and the last roadwheels and a jounce damping ratio of 0.6 (the rebound damping was 2/3 of the jounce damping) was maintained for each concept. The only differences between concepts are in the force-deflection curves which describe each spring. The amount of jounce and rebound travel permitted each spring, however, also remained fixed at 14" and 5.5" respectively.

This portrayal supports the notion that if the suspension is made too soft the ride deteriorates on the harsher terrains. It is interesting to note also, that the 120 cpm version has the worst ride at all terrain roughness levels studied, yet this is precisely the design natural frequency of the M551.

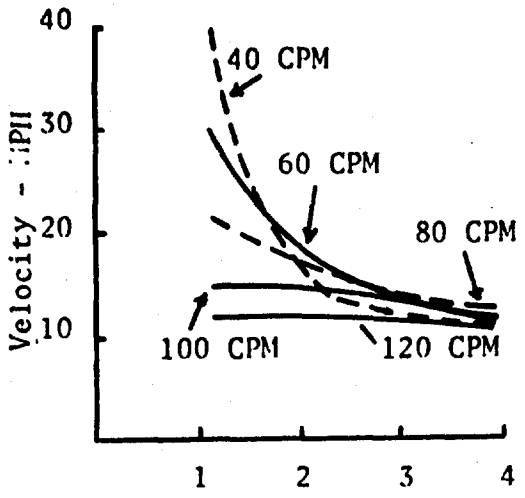



Figure 2
Natural Frequency vs. Ride

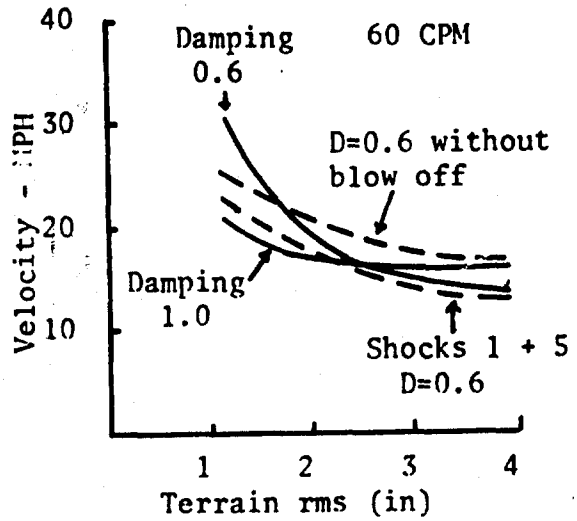



Figure 3
Damping vs. Ride

Figure 3 compares the baseline 60 cpm version of Figure 2 with three, otherwise identical, concepts with different damping characteristics. One version had a jounce damping ratio of 1.0 and blowoff at the same shock absorber velocity as for the baseline concept (resulting in a proportionately higher force). The second modification presented, maintained the damping ratio of 0.6 but removed the blowoff constraint. This permits the force produced by the shock absorber to increase without bound. The final concept shown in Figure 3 attains a jounce damping ratio of 0.6 with shocks only on the front and rear roadwheels. Blowoff is again provided and occurs at the same shock absorber velocity as before. This implementation provides an identical damping response, to a purely vertical motion, as does the baseline concept. This latter concept is of particular interest in view of the fact that the M551 has shocks only on the front and rear roadwheels.

The ride performance for various longitudinal CG positions is shown in Figure 4. It would appear that moving the CG forward can improve the ride comfort significantly over the lower rms terrains.

This result should be regarded with a word of caution. The ride comfort, as reported here, is measured only at the driver's location which is near the front of the vehicle. This implies that as the CG is moved forward, it is moved closer to the driver. Past studies have shown driver distance from the CG can greatly influence the magnitude of vertical acceleration that the driver would experience, and therefore also the power he would absorb. These results do not imply that the overall vehicle motion is reduced by moving the CG forward, but rather that the motion the driver feels is less severe.

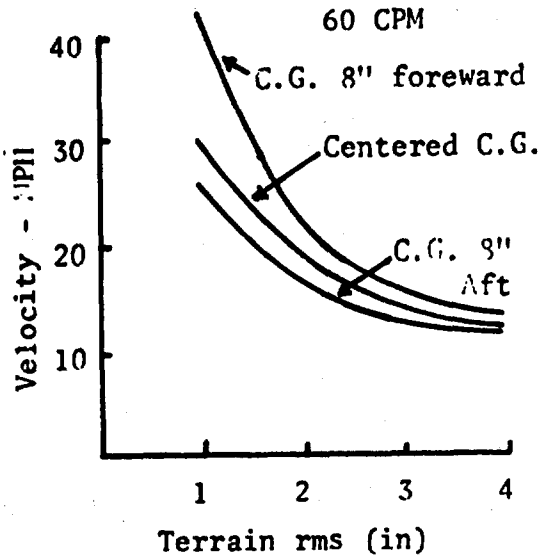



Figure 4
CG Position vs. Ride

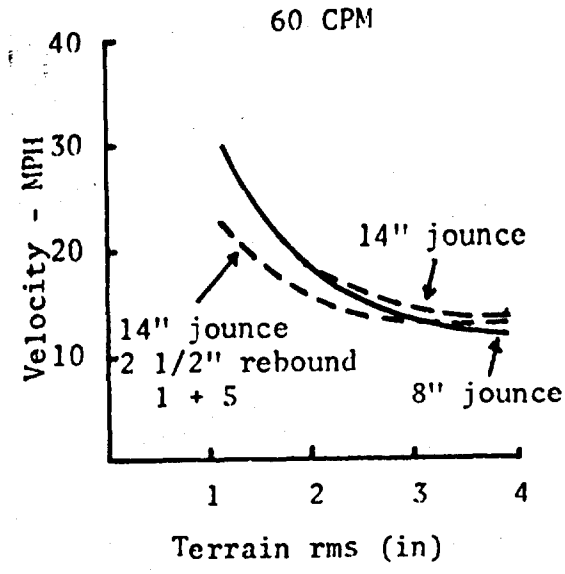



Figure 5
Wheel Travel vs. Ride

Figure 5 presents two final concept variations. The first comparison illustrates the rather expected impact of reducing the amount of jounce wheel travel. The baseline 60 cpm vehicle of Figure 2 is compared here to an otherwise identical vehicle with only 8" of jounce travel allowed for each roadwheel. The reduced wheel travel does not affect the ride on the milder terrains, but as the terrain roughness increases, the advantage of additional wheel travel becomes evident.

The effect that track tension can have on the front and rear roadwheels was discussed briefly in the model description section. It was stated that track tension greatly restricts the rebound travel of the front and rear roadwheels. The amount of rebound travel allowed would vary with different track-ground interface conditions but would, in general, decrease as track tension is increased. The final concept of Figure 5 shows the effect on ride quality of allowing only 2.5" of

rebound travel for the first and last roadwheels. All other vehicle parameters were identical to those of the baseline vehicle. The resulting deterioration in ride comfort could be simply attributed to model "sensitivities" but, more likely, it is an indication of how increasing track tension might affect a vehicle's ride quality.

## PLATFORM FIRING STABILITY RESULTS

Many of the parameters that affect the ride comfort of a vehicle, also affect the stability of the vehicle when used as a firing platform. The vertical natural frequency of the vehicle, for example, significantly influences both performance measures, but in opposite directions. The softer suspension systems tend to ride better but provide a less stable platform and vice versa. Other vehicle parameters, on the other hand, may affect one performance criteria while having little or no impact on the other. A case in point is the longitudinal CG position which can affect the ride comfort to a considerable degree as attested to by Figure 4, but has virtually no effect on the firing response.

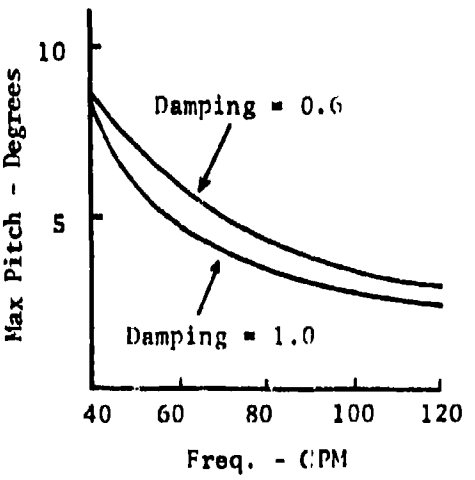



Figure 6
Natural Freq. vs. Firing

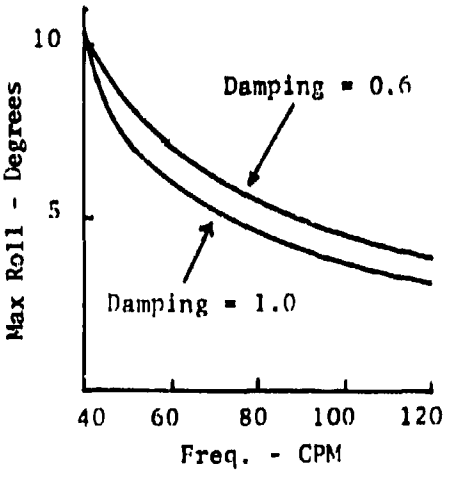



Figure 7
Natural Freq. vs. Firing

Figures 6 and 7 give the variation in platform response due to firing, as a function of the vertical natural frequency of the vehicle. Figure 6 shows the maximum pitch angle due to firing straight ahead. The concepts simulated to produce Figure 6 were also used to obtain Figure 7 which gives the maximum roll angle achieved when firing a single round straight off the side of the vehicle. These results are shown for two different damping levels with shock absorbers on the first, second, and the last roadwheels on each side.

One set of parameters that does not affect the ride comfort (at least as modelled here), but significantly impacts in the firing results, are the static CG height and the trunnion height above the CG. Figures 8 and 9 show the firing response variations when these two parameters are varied simultaneously. These results were for the 60 cpm concept with a 0.6 damping ratio. It is interesting to note the linear nature of the change, particularly for the pitch response. Also, in certain regions, the amount of roll motion produced by firing to the side, is unaffected by small changes in the CG height.

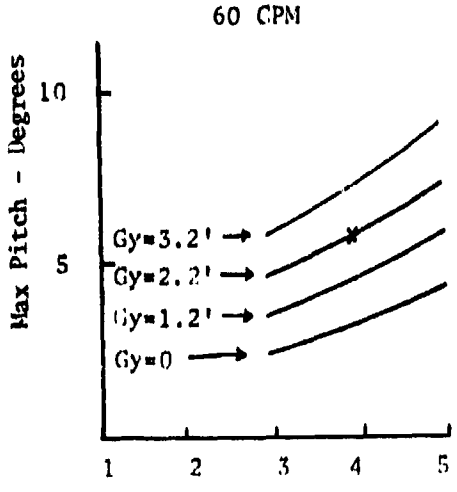



Figure 8
C.G. Height Effects

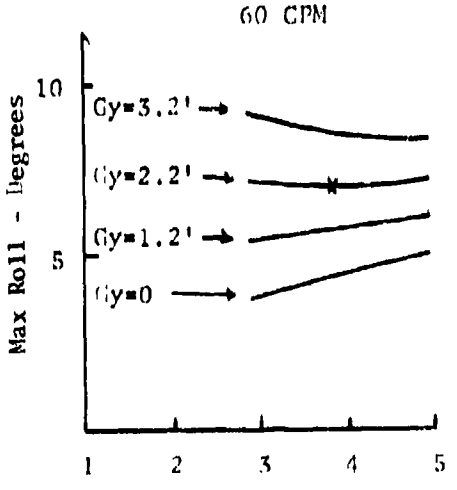



Figure 9
Trunnion & C.G.
Height Effects

The data points marked with an "x" on Figures 8 and 9 denote the baseline static CG height and trunnion position. These were the values used on the runs depicted in Figures 6 and 7.

## DESIGN CASE STUDY

The M551 was selected as a trial case for application of the various parameter "sensitivities". This vehicle was chosen because it has a relatively poor cross country ride performance when compared to other tracked vehicles in its weight class. The objective was to improve the overall vehicle performance (as described by ride comfort and platform stability), but to stay within certain physical constraints that could not easily be changed on an existing vehicle.

The rules governing the effort were as follows. The mass distribution and geometric characteristics were considered fixed. In other words, vehicle parameters such as CG position, roadwheel spacings, and gun trunnion position were not allowed to be changed. Also the amount of jounce travel allowed for each roadwheel was considered fixed. Therefore the only design modifications that could be considered, were variations in the spring and shock absorber characteristics.

It is clear from the previous results that the ride quality could be easily improved by simply softening the suspension system of the M551. It is equally clear, however, that application of this tactic alone would result in a severe degradation of the platform stability. Since it is difficult to assess the comparative tradeoff values of the two performance measures, it was desired to significantly improve the ride quality while having little or no impact on the platform stability characteristics. The ride quality, rather than the platform stability, was selected for improvement since the extreme stiffness of the original suspension system provided a reasonably stable platform considering the magnitude of the reaction impulse of the 152mm cannon.

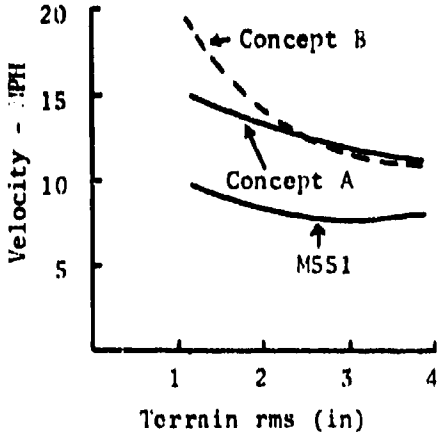



Figure 10
Case Study

Two distinct directions were taken in an attempt to improve on the ride performance. The first attempt concentrated on the damping characteristics only. The original springs were retained for this effort. It was conjectured that the stiffness of the suspension system would be suffi-

cient to restrain the vehicle's response to firing while modifications to the damping could possibly improve the ride quality.

The shock absorbers for each modification had lower rates in rebound than in jounce (the rebound rate was about 60% of the jounce rate) and blowoff was provided at the same velocity in each direction. The damping rates and blowoff velocities were varied and eventually an additional shock absorber was added at the number 2 roadwheels. Several design iterations were evaluated before arriving at the configuration referred to as concept A in Figure 10 and in Table 4.

The second approach was to soften the springs on the front three roadwheels (on each side) while maintaining the original springs on roadwheels 4 and 5. The softer springs were calculated to give a vertical natural frequency of 80 cpm (if they were used on all wheels).

This resulted in an effective natural frequency of 103 cpm for each of the design iterations leading to concept B. Shock absorbers were again used on the first, second, and last roadwheels.

| Vehicle | Natural Freq. cpm | Damping Ratio | Shock Locations | Maximum Pitch degrees | Maximum Roll degrees |
|---|---|---|---|---|---|
| A | 120 | 0.39 | 1,2,5 | 4.14 | 5.43 |
| B | 103 | 0.54 | 1,2,5 | 4.30 | 5.84 |
| M551 | 120 | 0.62 | 1,5 | 4.13 | 5.66 |

Table 4 M551 and Derivatives Comparison

The results portrayed in Figure 10 and Table 4 indicate that it is possible to improve the overall vehicle performance. Concept A has an improved ride performance over the entire spectrum of terrain roughnesses and has, at the same time, not increased the maximum hull rotations experienced due to firing the gun. The results obtained with concept B show an even more dramatic improvement in ride quality over the standard M551, though the platform stability properties are slightly degraded in this case.

The real thrust of these results is, of course, not the "improvement" of a long established vehicle, the M551. It is rather the demonstration of the utility and power of the methodology presented. The models used in obtaining the results discussed in this paper are

extremely efficient, easy to use analysis tools. With these models the designer can evaluate various design tradeoffs between ride comfort and platform firing stability, or simply compare the performance of his concept against that of some known combat vehicle.

## REFERENCES

1. Foss, Christopher F., "Armoured Fighting Vehicles of the World", Charles Scribner's Sons, 1977.

2. Lins, William F., "Human Vibration Response Measurement", TACOM Report No. 11551, June 1972.

3. Lee, R. A., Pradko F., "Analytical Analysis of Human Vibration", SAE Paper No. 680091, Jan. 1968.

4. Hoogterp, Francis B., "Interactive Vehicle Dynamics and Ride Evaluation Package", TARADCOM Report No. 12413, Nov. 1978.

5. Haug, E. J., Wehage, R., and Barman, N. C., "Dynamic Analysis and Design of Constrained Mechanical Systems", University of Iowa Technical Report No. 50, Nov. 1978.

6. Hoogterp, Francis B., "Platform Firing Stability Model", TARADCOM Report No. 12459, Sept. 1979.

7. Parent, R., "Integrated Engineering Service Test of XM551", TECOM Report No. DPS-1897, March 1966.

# CODING AND PROCESSING FOR RELIABLE DATA TRANSMISSION(U)

MR. WILLIAM A. HUBER
US ARMY COMMUNICATIONS RESEARCH
AND DEVELOPMENT COMMAND
FORT MONMOUTH, NEW JERSEY 07703

The coding and processing techniques discussed in this paper are designed to improve the reliability of digital message reception in the presence of noise. High noise conditions occur quite frequently during radio reception especially when these signals originate at forward area military positions. The usual cause for this poor radio reception is related to the constraints imposed by low transmitter power, limits on air time, and unfavorable terrain conditions. The importance of improving the reliability of forward area message reception is stressed because the information so obtained provides essential descriptors that are pertinent to command and control functions.

In an effort to determine the cause of message delivery failure in the presence of increasing noise, laboratory tests were performed under controlled noise conditions. The results of these tests indicated that as the signal-to-noise power ratio decreased errors first appeared in the data character framing information. It should be noted that the message synchronizing procedure used to decode received digital data is hierarchical in nature involving bit synchronization, data character framing, and message block identification. If any one of these functions is not correctly implemented the transmitted message cannot be successfully decoded. Ideally, as increased noise degrades the data communications channel the synchronizing system should not fail before the message information becomes unintelligible; a condition that is presently not satisfied because data character framing requires error-free code reception.

The reason for the premature failure of the data character framing function in the presence of increasing transmission channel



PRECEDING PAGE BLANK-NOT FILMED

noise lies in the inherent weakness of the method used to establish data character framing. Data character framing is accomplished by template matching techniques where a given bit sequence is transmitted during the message preamble that must be perfectly matched at the receiver terminal. With this method a single bit error will prevent the correct identification of data character framing thereby causing loss of the complete message. The vulnerability of data character framing to noise cannot be reduced by employing error correction coding because subsequent error correction decoding requires a priori knowledge of data character framing.

The identification of data character framing as the most noise vulnerable function in the data communications synchronizing procedure led to the development of a data character framing technique that does not use noise sensitive template matching. This technique involves the introduction of a new code and associated data processing system. Before describing the new system, techniques used to obtain bit synchronization will be discussed.

## II BIT SYNCHRONIZATION

The following brief discussion of bit synchronization is provided to illustrate that this function can be maintained under noise conditions that are far more severe than can be tolerated by the present data character extraction process. By increasing the performance of the data character framing process in the presence of noise, the weak link in the data reception reliability chain can be strengthened.

Figure 1 illustrates a typical data message structure. Bit synchronization data consists of a series of 1's and 0's. This cyclic signal structure is convenient for obtaining the timing information necessary to synchronize the receiver clock with the message signal. Receiver synchronization is usually accomplished by means of a Phase-Locked Loop (PLL) such as illustrated in Figure 2. In its basic form, the PLL (1) consists of a phase detector, a loop filter/amplifier and a receiver clock which consists of a voltage controlled oscillator (VCO). The phase detector is used to compare the phase of the incoming signal with that produced by the VCO, and then generate a resultant error voltage proportional to the difference. The loop filter provides short term memory and usually contains a linear amplifier. Amplification is provided so that oscillator control voltage can be derived from minimal error voltages thus maintaining close tracking between the incoming signal and the VCO. The VCO is designed to have a nominal frequency stability within a few cycles of that of the incoming signal frequency. In operation this frequency differential

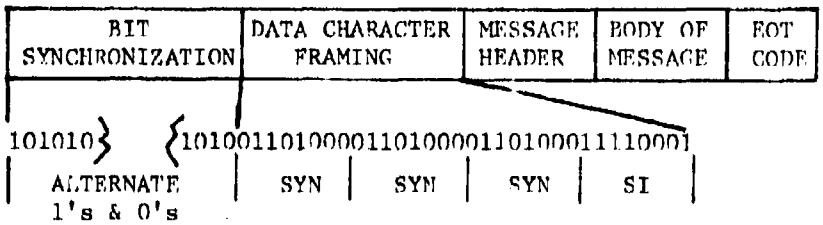



Figure 1. Typical data message structure.

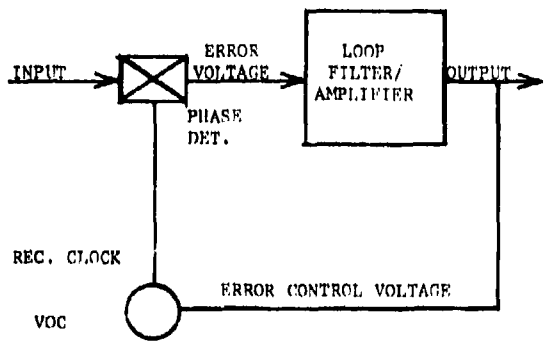



Figure 2. Phase-locked loop.

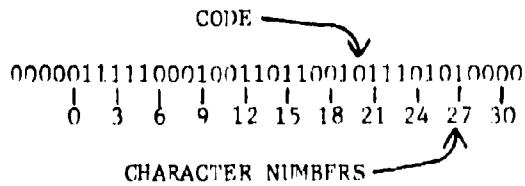



Figure 3. Proposed data character framing code.

is used to derive an error control voltage that drives the frequency of the VCO into lock with the incoming signal.

The PLL's superior performance during high noise reception conditions is related to the following: The PLL acts as an extremely narrowband filter which rejects sideband noise. It does this by supplying a local coherent signal that enhances its ability to extract narrowband signals from noise. The memory provided by the PLL supplies additional flywheel stability to the VCO during the presence of random noise.

## III DATA CHARACTER FRAMING CODE

A typical 32-bit data character framing code illustrated in Figure 1, consists of three American Standard Code for Information Interchange (ASCII) "SYN" characters, and one ASCII "SI" character. It should be noted that the weakness in the present method for extracting data character framing information is not related to the ASCII code characters utilized, but is related to the inability of the technique used to cope with bit errors. This weakness is illustrated by the following: In operation the receiver contains a replica of the 32-bit data character framing code which it attempts to match with the continuous flow of a 32-bit data sequence encoded in the incoming signal stream. When a 32-bit match is obtained it is used to establish the timing period for the start of the first bit of the first data character of the message. The start of successive data characters is then identified by means of a local counter driven by the bit synzhronized clock. It is readily seen from the above procedure that a single bit error in the received data character framing code will cause the loss of the transmitted message.

It is emphasized that the present technique for processing the data character framing code is a serial bit-by-bit procedure with no tangible results until the entire code has been processed. Not until this point is reached can a GO or NO-GO decision be made with respect to the starting period for message data character framing. It would therefore be advantageous to have a technique that would provide the required data character framing by sampling less than the complete 32-bit code. This feature would reduce the chance of error in direct proportion to the reduction of the bit sample size. It would also be advantageous to make the samples sufficiently small so that a multiplicity of samples could be processed during one 32-bit data character framing period. Such a technique would be especially effective during noisy reception conditions as it would provide several opportunities to obtain the required error-free bit sample necessary to establish

message data character framing.

A data character framing code that possesses the noise immunity features discussed above has been developed, based on viewing data framing as a countdown rather than a pattern. This code is illustrated in serial form in Figure 3. The pertinent noise immunity features that make this code ideally suited to the data character framing application are: it contains a high information density, it provides for simultaneous parallel and serial scanning, it identifies error-free bit sequences, and it establishes data character framing from only one small error-free bit sequence within the frame.

The high information density of the new code can be illustrated by first examining its structure. This is done with reference to Figures 4 and 5. The code is derived from the 32, 5-bit character code shown tabulated in paralled form in Figure 4. For identification purposes numerals 0 - 31 have been arbitrarily assigned to the codes as indicated in Figure 4. It will be noted from Figure 4 that successive characters of the code are formed by shifting each of the bits in columns 1,2,3 and 4, one bit position to the left and judiciously selecting bits for the vacated column one positions. If the 5-bit code shown in Figure 4 were to be used for randomly transmitting numerals 0 through 31, its efficiency with respect to information density would be no better than if the standard 5-bit binary code had been used. However, the data character framing operation is not a random function; instead, it is a precisely defined countdown procedure. Present data character framing techniques cannot take advantage of this countdown feature because no method is available for identifying the progression of the countdown process.

The structure of the data character framing code shown in Figure 3 contains the necessary data to meet this countdown criteria. That is, as the countdown progresses each new bit must contain the additional data necessary to identify the position of that bit within the countdown sequence. To demonstrate this, it should be noted that all the information contents of the 32, 5-bit characters of Figure 4 reside in the 36 bits comprising row 0 and column 1. All the remaining 124 bits in columns 2,3,4 and 5 are redundant. This can be seen by reference to Figure 4, where for any 5 by 5-bit array, the bits contained in the bottom row and left column of the array are identical. This structural characteristic was used to derive the data character framing code illustrated in Figure 3, which consists of the four 0 bits of row 1, plus the 32 bits of column 1. If this sequence of bits is scanned with a 5-bit parallel aperture, then at any point in the scanning process the bits within the aperture identify the

| Character Number | Bit Number | Character Number | Bit Number |
|---|---|---|---|
| | 54321 | | 54321 |
| 0 | 00000 | 16 | 11011 |
| 1 | 00001 | 17 | 10110 |
| 2 | 00011 | 18 | 01100 |
| 3 | 00111 | 19 | 11001 |
| 4 | 01111 | 20 | 10010 |
| 5 | 11111 | 21 | 00101 |
| 6 | 11110 | 22 | 01011 |
| 7 | 11100 | 23 | 10111 |
| 8 | 11000 | 24 | 01110 |
| 9 | 10001 | 25 | 11101 |
| 10 | 00010 | 26 | 11010 |
| 11 | 00100 | 27 | 10101 |
| 12 | 01001 | 28 | 01010 |
| 13 | 10011 | 29 | 10100 |
| 14 | 00110 | 30 | 01000 |
| 15 | 01101 | 31 | 10000 |

Figure 4. Proposed code data structure.

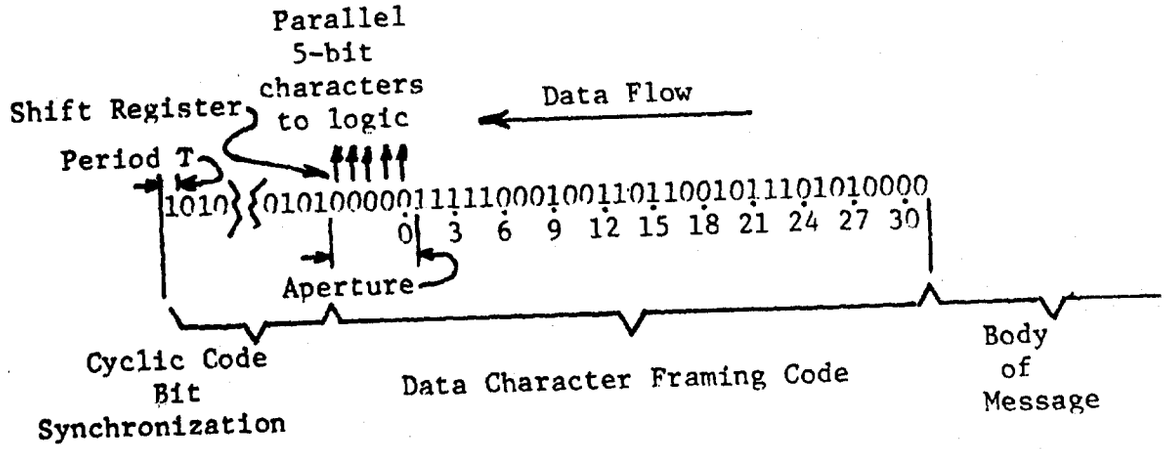



Figure 5. Processing the data character framing code.

position of the aperture within the code. To establish data character framing code countdown it is only necessary to pass the input message data stream through a 5-bit shift register and extract the 5-bit countdown codes from the parallel output of this shift register. A simplified schematic of this procedure is illustrated in Figure 5.

In the above discussion of data character framing code countdown it was tacitly assumed that the received code contained no errors. If the received code contains errors then the associated extracted countdown code will also be in error. An example of the countdown procedure where the signal stream contains errors is illustrated in Figure 6. When referring to Figure 6, it should be noted that for simplicity of illustration it is assumed that the data signal stream is stationary and that the 5-bit aperture scans from left to right. The bit identification numbering system used in Figures 3 and 5 is retained in Figure 6. This numbering system is also consistant with that of Figure 4 when it is considered in conjunction with the corresponding character in the scanning 5-bit aperture.

No generality will be lost if we start the discussion on code processing by assuming that the 5-bit scanning aperture is reading code (00000) corresponding to bit number "0" on the 0-31 scale. For the present we will assume that the first six bits of the code are error-free. As the 5-bit aperture scans one bit position to the right code (00001) corresponding to bit number "1" will be read. Up to this point the two codes read have been in consecutive order, namely: "0" and "1". If it is now assumed that bit number "2" is in error, as is indicated in Figure 6 by underlining the bit in error, then this code will be read as character number "10". Because of the symetrical nature of the code count, this abrupt jump in count order can only be attributed to an error in either or both codes forming the discontinuous transition. At this point it is not necessary to locate this error; it is sufficient to know that an error exists thus making the data unreliable for processing.

As the 5-bit aperture continues to scan the right, it will be noted from Figure 6 that the effect of the error is prevalent for the five bit numbers "10","21","22","23", and "24". From this example it can be seen that the minimum number of 5-bit aperture samples required to establish an error-free countdown sample set is six. That is, if six successive 5-bit aperture scans result in the generation of six consecutive bit numbers, then this sample set is error-free and the last bit of this set can be considered correct and used as a reference to specify the number of bit periods remaining before the start of the first character of the body of the message. That this is true can be seen by noting that the set of six, 5-bit aperture samples

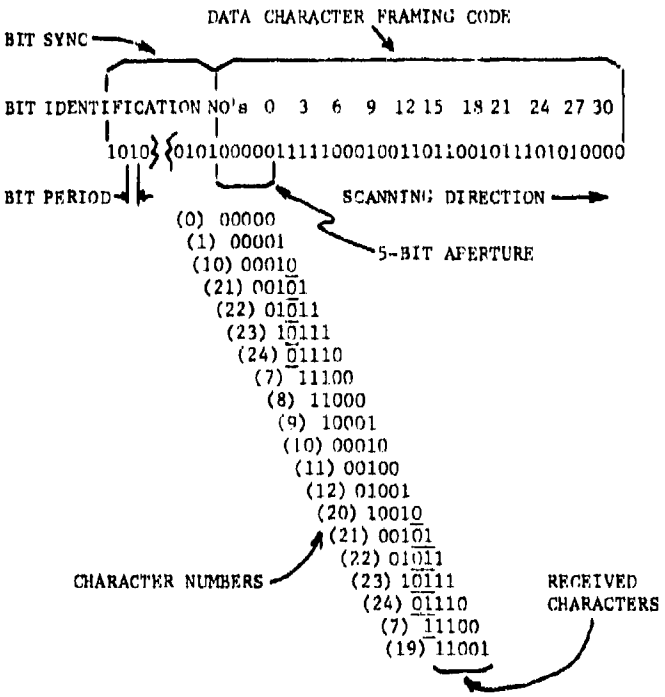



Figure 6. Countdown procedure during error conditions.

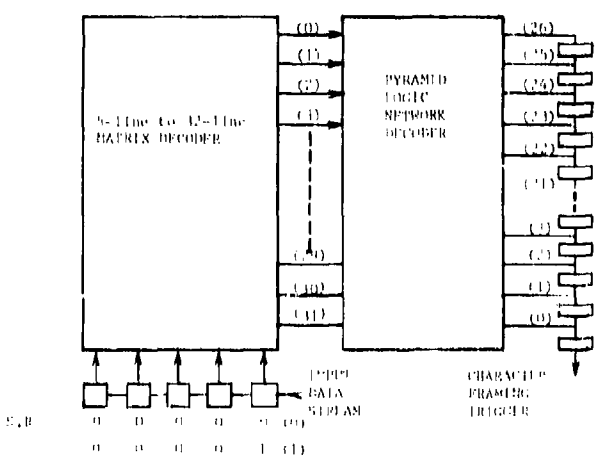



Figure 8. Functional diagram for processing the data character framing code.

indicated by bit numbers "7","8","9","10","11" and "12" contain no discontinuity of count. Bit number "12" of this sample set can therefore be used to initiate a countdown of the remaining 19-bit periods of the data character framing code. The completion of the 19-bit countdown marks the beginning of the first character of the body of the message. This information can then be used to synchronize an internal counter for use in maintaining data character framing, message block identification, and error detection and correction.

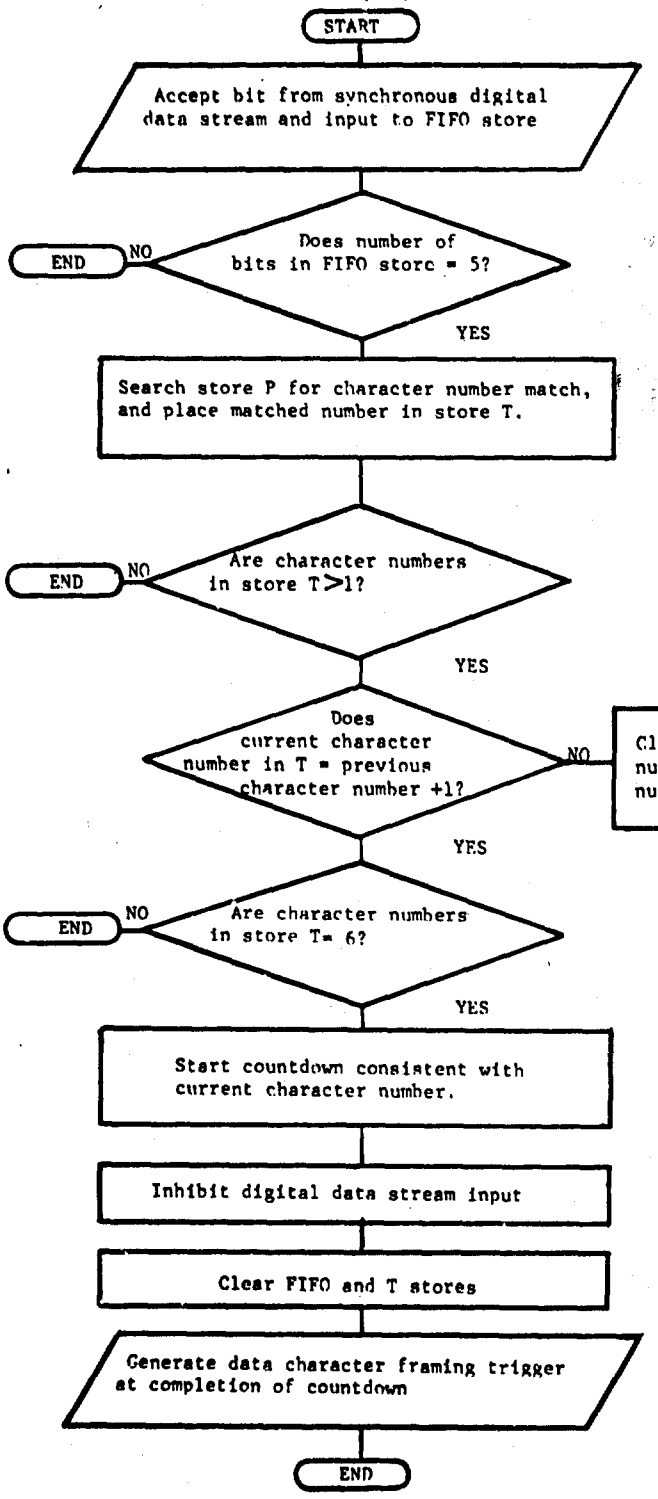
## IV DATA CHARACTER FRAMING LOGIC

A flowchart illustrating the requirements for processing the data character framing code is contained in Figure 7. The input processing requirement is for a 5-bit store of the first-in-first-out (FIFO) type. This store must operate at the synchronous transmission rate, and be capable of accepting a binary digital data stream, while providing a synchronous 5-bit parallel output.

A functional diagram of a system for firmware implementation of the data character framing code is presented in Figure 8. The FIFO store of Figure 7 is implemented in Figure 8 by a 5-bit shift register. The data signal input is shifted through the register at a synchronous rate by the internal clock. The 5-bit parallel output of the shift register is fed to a 5-line to 32-line matrix decoder. A more detailed look at a portion of the 5-line to 32-line matrix decoder is shown in Figure 9. Input lines of the matrix decoder are activated by a logical-1; output lines of the matrix decoder respond by producing a logical-1.

A simplified partial schematic of the pyramid logic network decoder is shown in Figure 10. The function of this logic is to filter out all random inputs; producing a single output only in response to six consecutive inputs. The input latches (2) of Figure 10 provide the necessary data storage so that during each bit period a constant logical-1 or logical-0 is applied to the associate AND gates. In Figure 10 it is assumed that consecutive character number codes (0),(1),(2),(3),(4) and (5) have been received. This places a logical-1 on the pyramiding AND gates resulting in a logical-1 appearing at the output of AND gate "$E_1$", where it is applied to the input of a 26 stage countdown shift register. The output of the shift register initiates data character framing.

An additional feature not shown in Figure 10 that must be provided for is the ability to automatically reset to logical-0 any latch that has been set to logical-1, but does not belong to a consecutive set of six character numbers. Reset logic to accomplish




Maximum processing time must be less than the bit interval.

P is permanent store that contains the 32,5-bit code characters.

T is temporary store.

NOTE: After END's system remains quiescent until activated by next synchronous trigger.

Figure 7. Flowchart for processing the data character framing code.

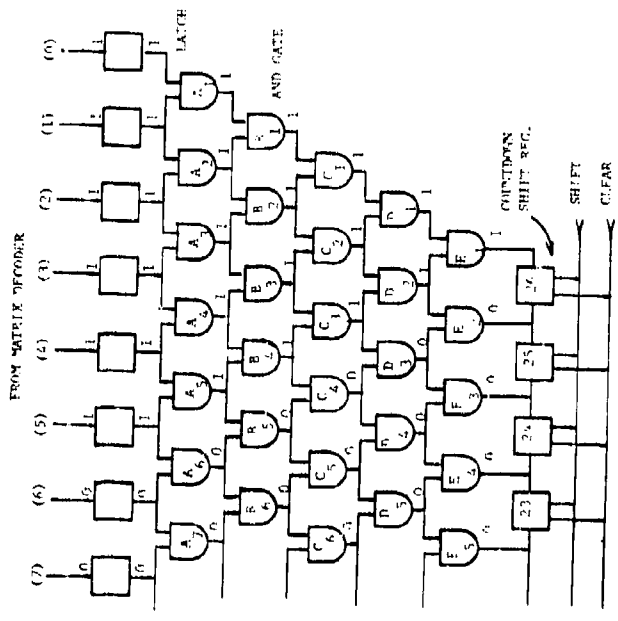



Figure 10. Simplified pyramid logic network decoder.

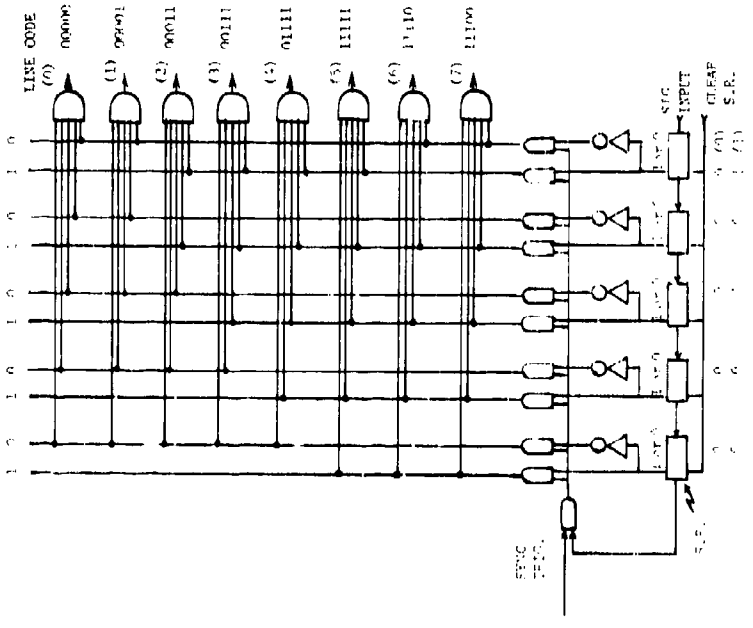



Figure 9. 5-line to 32-line matrix decoder.

this is shown in Figure 11, and the associate time logic of operational events is tabulated in Table I. Only three input lines from the matrix decoder are illustrated in Figure 11 as the operation of the remaining 29 lines is identical to those presented. Referring to Table I, the horizontal lines represent circuit activating triggers with time delays increasing when going from the top to bottom of the columns. The column numbers identify circuit nodes in Figure 11.

## V CONCLUSIONS

A new binary digital code and associated data processing technique has been described that is capable of extracting data character framing information from a noise contaminated digital data stream. The need for such an approach has been emphasized because of the inherent weakness of the present method for extracting data character framing, which is based on a noise sensitive template matching technique. The ineffectiveness of the present method for obtaining data character framing was substantiated during simulated operational tests where, as the reception conditions deterioriated because of increased noise, the data character framing function was first to fail. Further tests indicated that if the ability to extract data character framing could be extended down in the noise to a point where the message contents become unintelligible, message delivery rates could be significantly increased. The structured code of the proposed data character framing technique satisfies the above condition by virtue of its ability to operate in noisy environments. It accomplishes this desirable result without resorting to increased transmitter powers or receiver sensitivities.

The effectiveness of the proposed data character framing method is indicated by noting that the system's noise immunity features are such that it is only necessary to receive ten consecutive error-free bits (six consecutive character numbers) to establish correct data character framing. In the limit, all remaining bits in the data character framing code could be in error. These ten consecutive error-free bits may occur anywhere within the data character framing message preamble period.

The magnitude of improved reliability in message delivery that can be realized by using the proposed data character framing technique is related to the reduction in the size of the data sample. If a S/N power ratio of 7.5 dB is assumed, then the probability of receiving an error as read from the curve of Figure 12, (3) is $10^{-2}$, or on the average of one error per hundred bits. Present data character framing requires a string of 32 error-free bits, so there is a

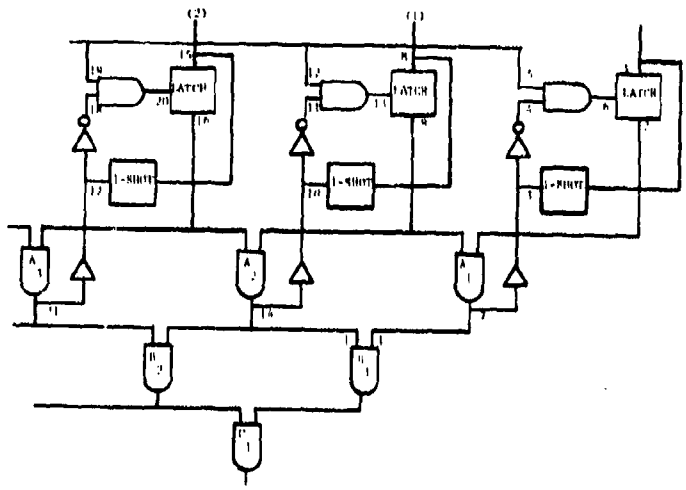


Figure 11. Parallel reset logic.

TABLE I

Time logic table

| | | (1) DATA | | 1-SHOT | | RESET | | | (2) DATA | | 1-SHOT | | RESET | | | (3) DATA | | 1-SHOT | | RESET | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| RESET | $t_1$ | 0 | 0 | 0/0 | 1 | 1 | 1 | 0 | 0 | 0 | 0/0 | 1 | 1 | 1 | 0 | 0 | 0 | 0/0 | 1 | 1 | 1 | 0 |
| DATA | $t_2$ | 1 | 1 | 1/0 | 0 | 0 | 0 | 0 | 0 | 0 | 0/0 | 1 | 0 | 0 | 0 | 0 | 0 | 0/0 | 1 | 0 | 0 | 0 |
| 1-SHOT | $t_3$ | 0 | 1 | 1/0 | 0 | 0 | 0 | 0 | 0 | 0 | 0/0 | 1 | 0 | 0 | 0 | 0 | 0 | 0/0 | 1 | 0 | 0 | 0 |
| RESET | $t_4$ | 0 | 1 | 1/0 | 0 | 1 | 0 | 0 | 0 | 0 | 0/0 | 1 | 1 | 1 | 0 | 0 | 0 | 0/0 | 1 | 1 | 1 | 0 |
| DATA | $t_5$ | 0 | 1 | 1/1 | 0 | 0 | 0 | 1 | 1 | 1 | 1/0 | 0 | 0 | 0 | 0 | 0 | 0 | 0/0 | 1 | 0 | 0 | 0 |
| 1-SHOT | $t_6$ | 0 | 1 | 0/1 | 0 | 0 | 0 | 1 | 0 | 1 | 1/0 | 0 | 0 | 0 | 0 | 0 | 0 | 0/0 | 1 | 0 | 0 | 0 |
| RESET | $t_7$ | 0 | 1 | 0/1 | 0 | 1 | 0 | 1 | 0 | 1 | 1/0 | 0 | 1 | 0 | 0 | 0 | 0 | 0/0 | 1 | 1 | 1 | 0 |
| DATA | $t_8$ | 0 | 1 | 0/1 | 0 | 0 | 0 | 1 | 0 | 1 | 1/1 | 0 | 0 | 0 | 1 | 1 | 1 | 1/0 | 0 | 0 | 0 | 0 |
| 1-SHOT | $t_9$ | 0 | 1 | 0/1 | 0 | 0 | 0 | 1 | 0 | 1 | 0/1 | 0 | 0 | 0 | 1 | 0 | 1 | 1/0 | 0 | 0 | 0 | 0 |
| RESET | $t_{10}$ | 0 | 1 | 0/1 | 0 | 1 | 0 | 1 | 0 | 1 | 0/1 | 0 | 1 | 0 | 1 | 0 | 1 | 1/0 | 0 | 1 | 0 | 0 |
| DATA | $t_{11}$ | 0 | 1 | 0/1 | 0 | 0 | 0 | 1 | 0 | 1 | 0/1 | 0 | 0 | 0 | 1 | 0 | 1 | 1/1 | 0 | 0 | 0 | 1 |
| 1-SHOT | $t_{12}$ | 0 | 1 | 0/1 | 0 | 0 | 0 | 1 | 0 | 1 | 0/1 | 0 | 0 | 0 | 1 | 0 | 1 | 0/1 | 0 | 0 | 0 | 1 |
| RESET | $t_{13}$ | 0 | 1 | 0/1 | 0 | 1 | 0 | 1 | 0 | 1 | 0/1 | 0 | 1 | 0 | 1 | 0 | 1 | 0/1 | 0 | 1 | 0 | 1 |

probability of success of 0.67. The proposed data character framing technique requires a string of only 10 error-free bits, so there is a probability of success of 0.89, which results in 33% improvement in message delivery.

To adapt the proposed data character framing technique to present military data communications systems all that is required is to substitute, in the message preamble, the new data character framing code for the presently used data character framing code. The new code is then processed by methods such as have been described. These methods could be implemented by either software or firmware techniques. Firmware implementation could be in the form of an integrated circuit chip. Both the effectiveness of the data character framing method and its simplicity of implementation suggest its universal use as a standard method for obtaining data character framing.

## ACKNOWLEDGEMENTS

The author wishes to acknowledge his indebtedness to Messrs. David K. Ruppe and Thomas J. Wheeler, USA CORADCOM, for reviewing the manuscript and providing suggestions for improving its contents. The author would also like to thank Mr. Richard E. Loiselle, USA CORADCOM for his efforts during the early development of the data character framing code concept.

## REFERENCES

1. Gardner, Floyd M., "Phaselock Techniques", 2nd Ed., John Wiley & Sons, Inc., N. Y. 1979.

2. Peatman, John B., "The Design of Digital Systems", McGraw-Hill Book Co., N. Y., 1972.

3. Stein, S. and Jones, J. J., "Modern Communications Principles", McGraw-Hill Book Co., N. Y., 1967.

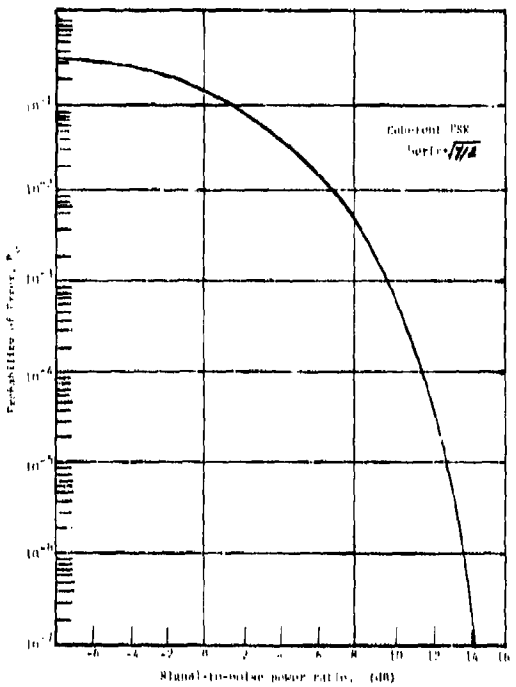


Figure 12. Error rates for coherent FSK system.

## APPENDIX

$P_e = 1/2 \operatorname{erfc}\sqrt{\gamma/2}$ Probability of error

The complementary error function is defined by

$$\operatorname{erfc} x = 1 - \operatorname{erf}$$

$$= \frac{2}{\pi}\int_{x}^{\infty} e^{-t^2} dt$$

where

$$\gamma = \frac{u^2}{2N}$$ Power ratio

N = average noise power level at filter output at instant of sampling

u = signal level at filter output momentarily containing the data bit at the same instant of sampling N.

The above analysis (3) applies to a Frequency Shift Keying (FSK) signal using synchronous (coherent) detection. This type of detection requires an exact replica of each of the frequencies representing the binary states to be available at the receiver. To satisfy this requirement PLL's are used. As the signal and noise add in the filters, it is expected that the filter output with the signal will be algebraically larger than the filter without the signal. If this is not true and error is indicated.

# FUNDAMENTAL LIMITATIONS OF ELECTRON BEAM LITHOGRAPHY FOR FUTURE MILITARY IC DEVICE FABRICATION (U)

G.J. IAFRATE, PhD., J.N. HELBERT, PhD., A.T. BALLATO, PhD.,
C.F. COOK, Jr., Mr. and W.S. McAFEE, PhD,
USA ELECTRONICS TECHNOLOGY AND DEVICES LABORATORY (ERADCOM)
FORT MONMOUTH, NEW JERSEY 07703

## INTRODUCTION

Future military requirements for real-time information acquisition and tactical electronic warfare (EW) information processing have established a need for very large scale (VLSI) and very high spped (VHSI) integrated circuit microelectronics. Both VLSI and VHSI technologies require a high areal density of integrated circuit (IC) elements thereby dictating a reduction of device size into the submicron and ultra-submicron regions.

The requisite miniaturization of military IC devices into the submicron and ultra-submicron region raises serious questions regarding the use of electron beam lithography (EBL) as an appropriate fabrication technique for this end. In EBL an intermediate energy (5-20 keV) electron beam exposes a radiation sensitive target (such as the polymer PMMA) overlaid on an IC device substrate; a subsequent chemical development treatment removes the exposed region leaving an outline of the pattern sketched by the electron beam (e-beam) when employing a positive-acting resist. The e-beam technique is attractive because it leads to the production of micron to submicron resolved channels and also lends itself to computer-controlled production-line automation. However, we find that the use of an e-beam introduces certain difficulties in that electrons penetrating the target scatter away from the incident electron beam thereby causing unintended exposure to other regions of the target; also, the penetrating electrons are generally energetic enough to reach and backscatter from the IC substrate thus tending to expose unintended regions of the resist as well. This scattering and backscattering leads to a widening and distortion of the exposed region and results in a degradation of the resolution of two



PRECEDING PAGE BLANK-NOT FILMED

parallel e-beam lines. It is clear from this study that the ultimate resolution of the e-beam process is determined by electron scattering effects in the resist material and by electron backscattering from the device substrate, and not by the resolution of the electron-optical system. In this paper we assess the merit of the e-beam fabrication technique for use in very high speed integrated circuit (VHSIC) technology. A study is presented which describes the electron scattering and backscattering processes in electronic materials. A theoretical study describing primary electron backscattering from single- and double-layered substrates is presented; also attention is focused on the question of the spatial region exposed by a scattered electron beam in a lithographic resist material. Experimental electron backscattering, e-beam and scanning electron microscopy studies are presented to corroborate theoretical predictions.

## ELECTRON BACKSCATTERING

Many phenomena, such as polymer bond breaking, x-ray production, cathodoluminescence, electron-hole pair generation in semiconductors and insulators, etc., result from the inelastic interaction of a high energy (1-50 keV) electron beam with matter. Numerous primary electron backscattering studies, both experimental and theoretical, have been carried out in an effort to understand the dominant mechanisms for electron scattering and energy loss arising from the electron beam-target interaction. Most theoretical treatments (1-3) of electron backscattering are limited in usefulness, suffering either from severe physical approximations or mathematical complexity. However, the recent theoretical extension (4) of Everhart's theory (1) to include the calculation of energy distribution spectra has stimulated a renewed interest in this simple analytic approach.

In this paper we make use of Everhart's theory (1,4) to describe electron backscattering from solids, double layers, and supported thin films. [Due to limitations in space, only salient features of the theory will be discussed throughout the paper. Detailed theoretical analyses and discussions can be found in the references cited.] Everhart's theory assumes that primary electrons, upon entering a solid target, suffer energy loss in accordance with the Thomson-Whiddington (T-W) law and undergo changes in direction via single large-angle (greater than $90^{\circ}$) Rutherford scatters. Moreover, electrons that are scattered through angles less than $90^{\circ}$ are treated as if they are not scattered at all.

The T-W energy-loss law asserts that the energy loss associated with a typical particle at a given depth into the target is inversely proportional to the energy of the particle at that depth. Thus, in making use of the assumptions of Everhart's theory, the relative number of electrons backscattered from a solid target of atomic number Z with

energies in the range $(\varepsilon E_0, E_0)$, where $E_0$ is the primary electron energy and $\varepsilon$ is some fraction of $E_0$, can be calculated as

$$\eta_s(\varepsilon,Z) = \frac{(a+1)y_0 - 1 + (1-y_0)^{a+1}}{(a+1)y_0} \tag{1}$$

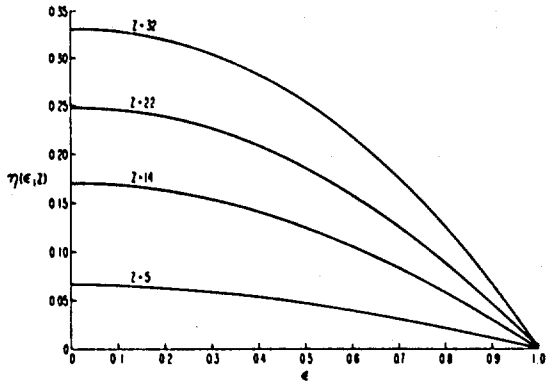
where

$$y_0 = \tfrac{1}{2}(1-\varepsilon^2) \tag{2}$$

and

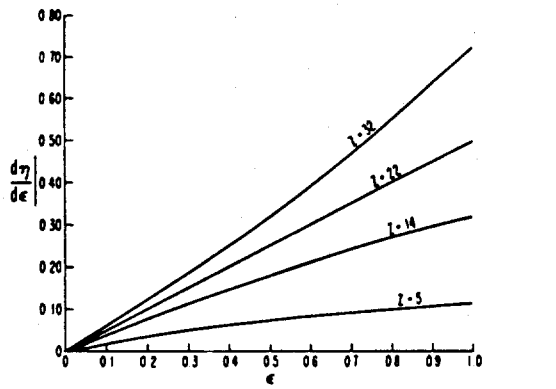
$$a = (0.045)Z. \tag{3}$$

The energy distribution curve (EDC) can be otained, to within a factor $(1/E_0)$, directly from the absolute value of the first derivative of $\eta_s(\varepsilon,Z)$ with respect to $\varepsilon$. Figures 1 and 2 show the dependence of $\eta_s(\varepsilon,Z)$ and the EDC on $\varepsilon$ for various values of Z; since $\varepsilon$ is the reduced energy with which the least energetic electrons in the backscattering current, $\eta_s(\varepsilon,Z)$, escape from the target, it is then also thought of as the reduced potential energy required to retard backscattered electrons with reduced energies less than $\varepsilon$ from being collected by an electron analyzer. Figure 2 indicates two salient features of the EDC as predicted by Everhart's theory, namely, the general "triangular" shape of the curve and the change in curvature of the EDC's from concave downward to concave upward with increased Z.




ELECTRON BACKSCATTER COEFFICIENTS FOR B, Si, Ti & Ge FOR VARIOUS RETARDING VOLTAGES

Fig. 1.




ENERGY DISTRIBUTION CURVES FOR ELECTRONS BACKSCATTERED FROM B, Si, Ti & Ge

Fig. 2.

These two features have been observed in the experiments of Kulenkampff and Spyra (5) as shown in Fig. 3. The poor agreement between theory and experiment for Z = 78 (Pt) in Fig. 3 is expected since diffusion as

given by Archard's theory (2), rather than single scattering in accordance with Everhart's theory (1), is believed to be dominant for $Z \gtrsim 40$. In addition, the lack of agreement between theory and experiment for Cu and Al when $\epsilon \lesssim 1$ is believed to be due to the failure of the T-W continuous energy loss law (indeed, any continuous energy loss law) near the target surface. (6,7)

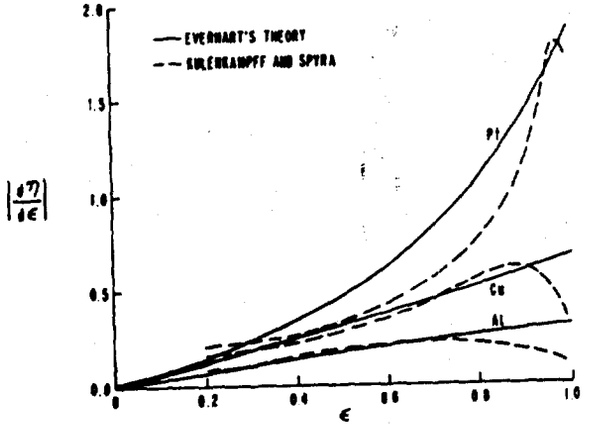



Fig. 3. Comparison of EDC's of Everhart's theory with those experiments obtained in ref. (6) for Al, Cu, and Pt.

We point out in Fig. 1 that the relative number of electrons backscattered from reasonably-high-Z solids such as silicon (Z = 14) and gallium arsenide (Z = 33), can be quite high, ranging from 17% to about 35% respectively. For solids with higher nuclear charge, such as gold (Z = 79), the relative number of backscattered electrons can be as high as 65%.

Finally, we note that the backscattering coefficient derived in this analysis does not depend explicitly on the incident electron energy when the retarding potential energy, $\epsilon$ is zero. This is a result which seems to be consistent with the experimental data reported by Sternglass (8) and others for primary electron energies above 3-5 keV.

Everhart's theory can easily be extended (4) to describe electron backscattering from double layers and supported thin films. In the analysis the target is considered to be a double layer consisting of an infinitely thick substrate covered by an overlying thin film of thickness D. The thin film and substrate are assumed to have mass density and average atomic number ($\rho_1$, $Z_1$) and ($\rho_2$, $Z_2$), respectively. When the maximum depth of electron penetration is less than or equal to the thin film thickness, D, the electrons energetically can only backscatter from the thin film; the appropriate backscattering co-

efficient is then given by $\eta_s(\varepsilon,Z)$ in Eq. (1) with Z replaced by $Z_1$. On the other hand, when the maximum depth of electron penetration is greater than D, electrons then have sufficient energy to backscatter from both thin film and substrate. In this case the backscattering coefficient is given by

$$\eta(\varepsilon,Z_1,Z_2) = \eta_{TF} + \eta_{TFS} \tag{4}$$

Here, $\eta_{TF}$ represents the reflection coefficient for electrons backscattered from the thin film of thickness D and $\eta_{TFS}$ is the reflection coefficient for electrons transmitted through the thin film and then backscattered from the substrate with sufficient energy to escape from the target.

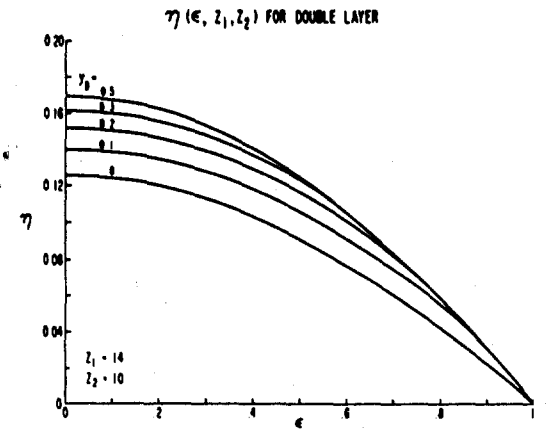



Fig. 4. Backscattering coefficient for double layer targets. $Z_1$ and $Z_2$ refer to the average nuclear charge of the thin film and substrates, respectively; $y_D$ is the reduced thin film thickness.

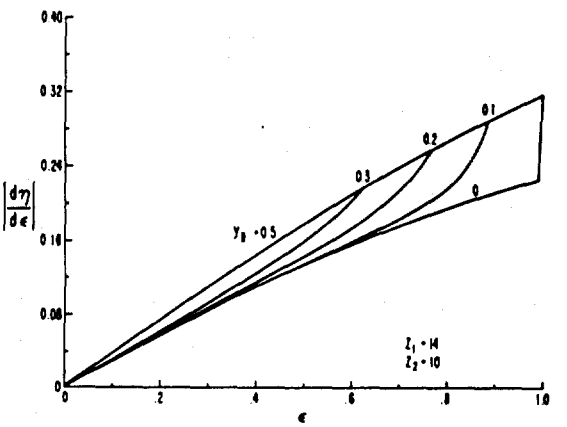
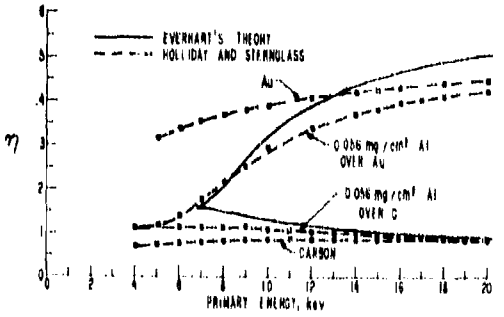



Fig. 5. EDC's for double-layer targets. $Z_1$ and $Z_2$ refer to the average nuclear charge of the thin film and substrates, respectively; $y_D$ is the reduced thin film thickness.

The backscattering coefficient and the EDC derived therefrom are plotted as functions of ε in Figures 4 and 5 for $Z_1 = 14$ and $Z_2 = 10$, and for various values of $y_D$ [$y_D$ is the reduced thin film thickness $D/R_1$, where $R_1$ is the range of electrons in the thin film]. For $\varepsilon \lesssim \varepsilon_c = (1-2y_D)^{\frac{1}{2}}$, the backscattering coefficient veers away from its "bulk" thin film value to accommodate the composite thin film-substrate scattering. This variation in the backscattering coefficient

is greatly enhanced in the EDC's of Fig. 5 and is due to the assumed discontinuous nature of the interface; a sigmoidal variation in Z and ρ across the interface should lead to a smooth variation in the EDC's.

Experimentally measured EDC's for double-layer supported thin film targets appear not to be available in the literature so that comparison with theory could not be made. There are, however, many experiments, for double layers and supported thin films which relate the backscattering coefficient to the thin film mass thickness ρD and primary electron energy $E_0$. Holliday and Sternglass (9) have reported backscattering coefficients as a function of primary electron energy for various double-layer targets. In Fig. 6 we compare the theoretical backscattering coefficient with their results to show good agreement.




COMPARISON OF η FOR MULTILAYER FILMS OF ALUMINUM OVER GOLD AND ALUMINUM OVER CARBON

Fig. 6.

In addition, Cosslett and Thomas (10) have reported electron backscattering coefficients as a function of mass density and primary electron energy for thin films supported on rings. In Fig. 7 we compare our backscattering coefficient with experimental results to show reasonable agreement.

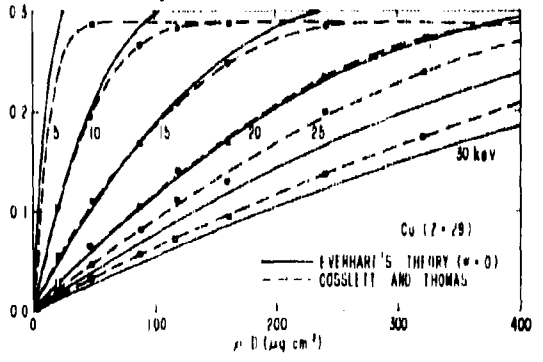



Fig. 7. Comparison of $\eta_{TF}$ with experiment for Cu.

Comparison of theoretically obtained backscattering coefficients and EDC's with results obtained by experiment yielded good agreement, which is remarkable in view of the simplicity of Everhart's assumptions and the adopted model for the double-layer interface. Moreover, we have pointed out that, within this model for which we have established reasonable confidence, the percentage of incident electrons reflected from a solid or double-layered substrate will increase monotonically with the atomic number of the reflecting substrate. It will be shown in a later section that this behavior affects the resolution of closely spaced e-beam lines rather dramatically.

## ELECTRON SCATTERING

In this section we address the question of the spatial extent of electrons scattered in the target. We do this by making use of the multiple scattering theory of Bethe et al. (11,12) to calculate the average cosine between the actual direction of motion and the direction of the primary beam in terms of the average atomic number of the target and the depth of electron penetration into the target. Moreover, we make use of the order-of-magnitude expression for the depth at which two e-beam lines overlap in terms of the average atomic number of the target.

Bethe et al. (11,12) have established the average cosine between the actual direction of motion and the direction of the primary beam for fast electrons to be

$$\langle\cos\theta\rangle = \exp\left[(-2)\int_{E_d}^{E_0} \frac{dE}{\lambda\left|\frac{dE}{dx}\right|}\right] \tag{5}$$

where $\lambda$ is the "transport mean free path", $|dE/dx|$ is the energy loss law for primary electrons moving through the target, $E_0$ is the incident electron energy, and $E_d$ is the energy of electrons at depth d. It follows from reference (2) that the $\langle\cos\theta\rangle$ can be expressed as

$$\langle\cos\theta\rangle = (1-y_d)^{7Z/40} \tag{6}$$

where the normalized depth $y_d$ is the depth "d" penetrated divided by the range, R, of the electrons in a solid of atomic number Z. In Fig. 8 we plot the $\langle\cos\theta\rangle$ of Eq. (6) as a function of normalized depth, $y_d$, while treating the atomic number, Z, as a parameter. It is clear from this figure that, as a function of normalized depth, the $\langle\cos\theta\rangle$ tends to zero with increasing rapidity as the target nuclear charge increases. This means that, for a fixed normalized depth of penetration, the electrons will scatter through an "average" angle which increases

monotonically with target nuclear charge.

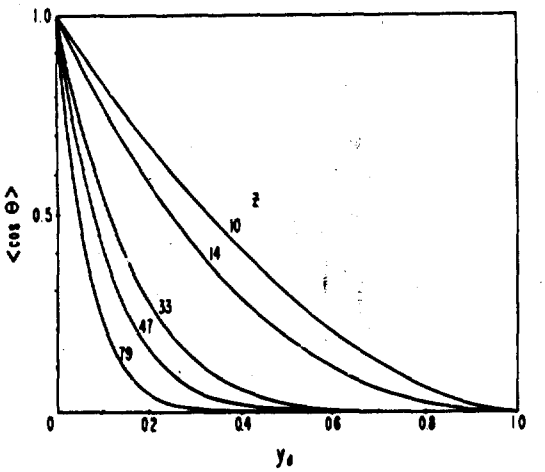



Fig. 8. Plot of $\langle\cos\theta\rangle$ as a function of normalized depth, $y_D$, with average atomic number, Z, as a parameter.

This finding has significant importance in the light of recent work reported by IBM (13) on high atomic-numbered (Z) acrylic polymer resist compositions. IBM has studied Thallium (Z = 81) and cesium (Z = 55) substituted methacrylic acid copolymers which have average atomic numbers (Z) higher than Z-values of well-known resist polymers such as poly(isobutyleno)(Z = 2.6), poly(methyl methacrylate)(PMMA) (Z = 3.6), and poly(batene-1-sulfone)(Z = 4.1); the value of Z for 100% cesium-substiuted acrylic polymer resist is 11. In a later section electron scattering in the resist will be shown to be a factor governing e-beam pattern resolution.

One can make use of the $\langle\cos\theta\rangle$ calculation to establish the dependence of the undercutting on target nuclear charge of two equally exposed parallel electron-beam lines. We define L, somewhat arbitrarily, as the depth at which electrons scattered from each beam mutually expose the region of the target midway between the two electron-beam lines. The depth L is then determined by setting the $\langle\cos\theta\rangle$ equal to the geometrical cosine, $\cos\theta_g = 1/[1 + 1/4(D/L)^2]^{1/2}$, depicted in Fig. 9. We set the $\langle\cos\theta\rangle$ in Eq. (6) [with d = L] equal to $\cos\theta_g$ to obtain an analytical expression for the normalized depth, $\mathcal{L}$, as

$$(1-\mathcal{L})^{7Z/40} = 1/[1 + 1/4(\mathcal{L}/\mathcal{L})^2]^{1/2} \qquad (8a)$$

with

$$\mathcal{D} = D/R, \quad \mathcal{L} = L/R, \qquad (8b)$$

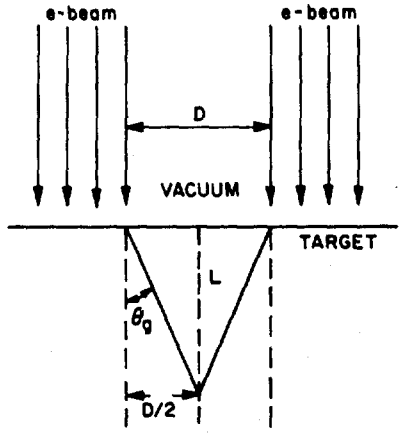



Fig. 9. Diagram depicts two e-beams, separated by a distance D, striking the target; the geometrical angle, $\theta_g$, and the undercutting length, L, are visually defined.

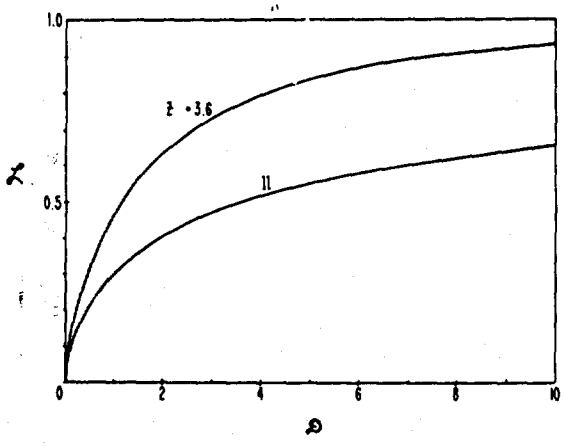



Fig. 10. Plot of reduced depth, $\mathcal{L}$, as a function of reduced separation, $\mathcal{D}$, with average atomic number, Z, as a parameter.

Here, as before, R is the range of the electron in the target material. In Fig. 10 we plot the reduced depth $\mathcal{L}$ as a function of reduced separation between electron beams with average atomic number as a parameter. The results predict a marked decrease in the depth $\mathcal{L}$ with increased target nuclear charge for a fixed separation between electron beams. We also note from Fig. 10 that increasing the average atomic number of the resist from 3.0 to 11 radically changes the dependence of undercutting on the spacing of the electron beams.

## ELECTRON-BEAM LITHOGRAPHIC STUDIES

In the previous sections of this paper, we have outlined two theoretical models which adequately describe electron backscattering and scattering phenomena in solids. Moreover, we made several observations based on the theoretical results, namely, that a substantial fraction of incident electrons can be reflected from a semiconductor surface or interface and that the spatial extent of penetrating electrons scattered from an incident electron beam is strongly dependent on the average nuclear charge of the target. We now show that both of these phenomena dramatically affect the resolution of closely spaced e-beam lines.

Experimentally, we have studied electron scattering and backscattering effects in "infinitely thick" resist materials and in thin

film resist materials mounted on various IC substrates. [The terms "infinitely thick" and "thin" film refers to the thickness of the film relative to the range of the electrons.] In this study we have irradiated a variety of substrates by EBL techniques; the resulting e-beam exposures were then photographed by use of scanning electron microscopy (SEM). In our EBL process a 20 keV electron beam exposes a radiation-sensitive target [PMMA was used in this study] overlaid on an IC substrate; a subsequent chemical development treatment removes the exposed region leaving an outline of the pattern created by the scattered electron beam. Figure 11

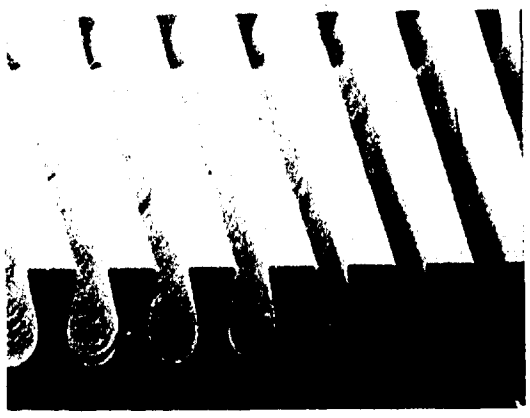


Fig. 11. SEM photograph of a typical EBL exposure of thick PMMA resist material.

represents an SEM photograph of a typical EBL exposure of thick PMMA resist material. For this particular pattern, the e-beam machine was programed to "write" ten equally spaced one micron thick lines of two micron spacings. The teardrop-shaped hollows represent developed regions of e-beam exposure. The geometrical shape of a given hollow is directly related to the spatial extent of the scattered e-beam in the resist material. It is clear from this figure that e-beam scattering in the resist will limit the e-beam line separation achieveable in EBL; as the e-beam line separation is reduced, the teardrops below the surface overlap to undercut the resist material between the lines thereby destroying the e-beam pattern.

In order to study the depth of undercutting as a function of e-beam separation in PMMA without destroying the e-beam pattern, we have developed a unique technique which makes use of the misalignment of the e-beam deflection fields. The deflection field is the maximum area of e-beam projection achieveable without translating the e-beam specimen stage. Since the e-beam machine in our laboratory writes with a deflection field of 0.5 x 0.5 squared millimeters, to write larger patterns covering areas greater than the deflection field

requires precise matching (butting) of the adjacent deflection fields. If adjacent deflection fields are not well matched, then an e-beam pattern which is written across the deflection field boundry will give rise to a "butting" error. The e-beam pattern resulting from a lateral (transverse) adjacent deflection field mismatch is seen in Fig. 13c. In our study we intentionally misalign the adjacent deflection fields in the longitudinal direction with respect to ten equally spaced e-beam lines. In this case the e-beam lines in the array are doubly exposed at the deflection field boundry. Thus, when the pattern is developed, the doubly exposed regions at the field boundries produce much deeper crevasses whose internal shapes resemble the e-beam patterns receiving twice the dose. For our purposes, we write the entire pattern with an underexposing e-beam dose; since the field boundry receives double exposure, the PMMA samples fracture very easily at the field boundry thus making visable the e-beam radiation pattern without destroying the pattern. An SEM photograph of a typical sample is seen in Fig. 12. In this figure it should be noted that a triangular region is defined between the underexposed e-beam lines as theoretically visualized in Fig. 9. We have written a series of arrays, varying the spacings between one micron thick lines, while keeping the e-beam exposure rate and misaligned deflection field constant, to obtain the depth of undercutting as a function of line spacing. The results indicate excellent agreement with the shape of the curve predicted in Fig. 10 for PMMA (Z = 3,6).

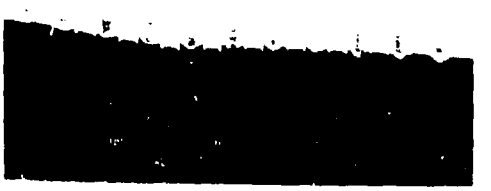


Fig. 12. SEM photograph of e-beam radiation pattern.

Finally, we have experimentally studied e-beam proximity effects [overexposure effects arising from e-beam writing of closely-spaced lines and from electron backscattering from substrates] on thin films of PMMA due to substrate electron backscattering. We have already established in a previous section that the percentage of incident electrons backscattering from a substrate will increase with the atomic

number of that substrate. We have e-beam irradiated three different samples to show dramatic proximity effects.

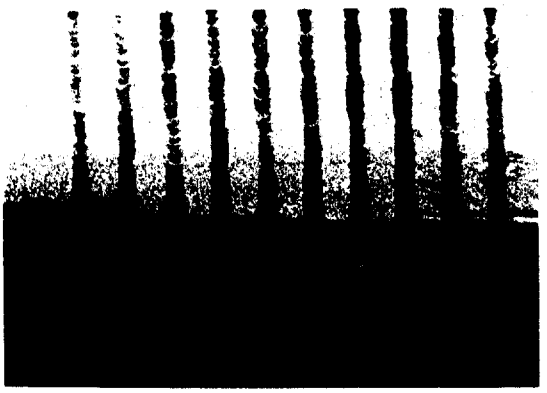


Fig. 13a. SEM photograph of e-beam exposure on PMMA.

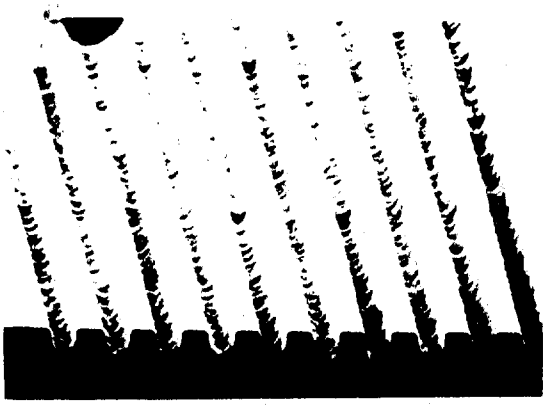


Fig. 13b. SEM photograph of e-beam exposure on PMMA/silicon substrate.

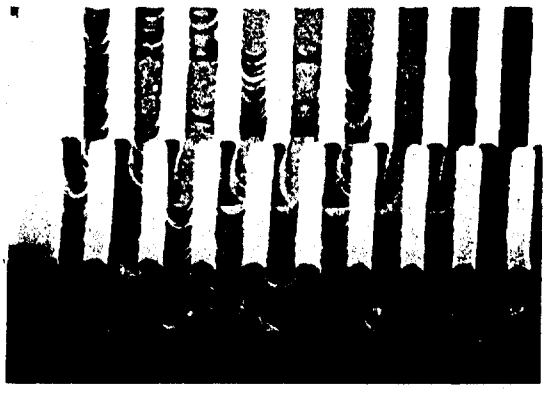


Fig. 13c. SEM photograph of e-beam exposure on PMMA/gold substrate.

In Figure 13a we show an e-beam exposure on "infinitely thick" PMMA. In Figures 13b,c, we show typical e-beam exposures of three microns thick PMMA on silicon (Z = 14) and gold (Z = 79), respectively. All three ples were exposed to equal e-beam doses, yet the samples show . . increased e-beam over-exposure with increased substrate nuclear charge. Thus, we have clearly shown that e-beam scattering effects play a major role in limiting the resolution of closely spaced e-beam lines.

## SUMMARY

In this paper two theoretical models have been outlined which adequately describe electron backscattering and scattering phenomena in solids. The theoretical results indicated that a substantial fraction of incident electrons can be backscattered from a semiconductor substrate and that the spatial extent of penetrating electrons from an incident electron beam is strongly dependent on the average nuclear charge of the target. Experimentally, we have shown that both of these phenomena dramatically affect the resolution of closely spaced e-beam lines. As part of our experimental study, we have developed a technique for studying closely-spaced e-beam patterns without destroying the pattern.

In future studies we plan to expand our theoretical efforts to include the simulation of e-beam patterns by using a further extension of Everhart's theory as well as Monte Carlo methods. Experimentally, we plan to extend our present studies to include a wide variety of resist materials and substrates by using the excellent EBL and SEM facilities available in ET&DL.

## ACKNOWLEDGMENTS

The authors wish to thank J. Kwiatkowski and D. Eckart for technical assistance; we also thank C. Mauser for typing this manuscript.

## REFERENCES

1. T. Everhart, J. Appl. Phys. 31, 1483(1960).

2. G. Archard, J. Appl. Phys. 32, 1505(1961).

3. R. Dashen, Phys. Rev. 134, A1025(1964).

4. G.J. Iafrate, W.S. McAfee, and A. Ballato, J. Vac. Sci. Technol., 13, 843(1976).

5. H. Kulenkampff and W. Spyra, Z. Phys. 137, 416(1954).

6. D. Brown, D. Wittry, and D. Kyser, J. Appl. Phys. 40, 1627(1969).

7. L. Spencer and V. Fano, Phys. Rev. 95, 1172(1954).

8. E.J. Sternglass, Phys. Rev. 95, 345(1954); P.C.R. Palluel, C.R. Seances Acad. Sci. Roum. 224, 1492(1947).

9. J. Holliday and E. Sternglass, J. Appl. Phys. 28, 1189(1957).

10. V. Cosslett and R. Thomas, Br. J. Appl. Phys, 16, 779(1965).

11. H. Bethe, Rose, and Smith, Proc. American Philosophical Society, 78, 573(1938).

12. H. Bethe, Ann. Physik, 5, 325(1930).

13. R. Feder, T. Haller, M. Hatzakis, L.T. Roman kiw, and E. Spiller, IBM Technical Disclosure Bulletin, 18, 2346(1975).

# THE RESPONSE OF TURBINE ENGINE ROTORS TO INTERFERENCE RUBS

*ALBERT F. KASCAK, MR.
PROPULSION LABORATORY, ARMY AVRADCOM
LEWIS RESEARCH CENTER, NASA
CLEVELAND, OHIO 44135

## INTRODUCTION

In a typical aircraft gas turbine there are many instances in which rotor rubs occur. Two of the most common are blade tip and seal rubs, which are caused by thermal mismatch, rotor imbalance, high "g" maneuver loads, aerodynamic forces, etc. Current interest in fuel efficiency is a consideration which drives the engine design toward closer operating clearances. Thus increasing the probablity of rotor rubs. The interaction of a rotor with its case, (rotor rubs), has been studied in ref 1 and 2. Ref 1 studied a steady state interaction between a rotor with a rigid case neglecting friction at the interface and Ref 2 studied a steady state interaction between a linear flexible rotor and case including friction at the interface. Ref 1 and 2 did not consider the critical transient situation in which the rotor bounces off the case.

It is known that rotor rubs can have an important effect on the rotor dynamics. When a rotor rubs on the case, a frictional force is generated which can drive a rotor to whirl in a direction opposite to the direction of rotation, (backward whirl). This frictional force is relatively constant up to the backward whirl speed at which the rotor rolls around the case. Since this rolling contact speed is proportional to the rotational speed of the rotor times the ratio of the diameter to the rotor clearance, the whirl speed can be hundreds of times the rotational speed of the rotor; and thus be potentially very dangerous.

There are two basic methods for studying transient rotor dynamics. One is the modal method (ref 3 and 4) which expands the

solution in terms of a few of the lower frequency mode shapes. If the transient under study is localized (like a blade loss or a rotor rub), the high frequency components are, at least initially, dominant. Thus the modal method is not applicable to this type of transient. The other method involves the direct integration of the equations of motion, which can be done in either of two ways, explicit or implicit integration. For example, ref 5 used explicit integration of the equation of motion, but this solution is plagued with numerical stability problems. Further, ref 6 showed that explicit integration of the equation of motion was unstable when the product of the critical frequency (for any mode numerically possible) and the time step was large. Therefore, the explicit integration can only be done for simple rotors.

In contrast, the implicit integration tends to be stable (ref 7 and 8); but it requires the solution of a large number of nonlinear simultaneous equations at each time step. Ref 9 used a technique similar to ref 7 except that it was applied directly to the second order equation of motion. Ref 9 also noted that the generalized forces on a rotor were functions of the generalized position and velocity of the point where the forces were applied and its nearest axial neighbors. This allowed the variables to be arranged so that the Jacobian of the set of nonlinear equations was block tridiagonal. Therefore, computing time became proportional to the number of elements in the rotor dynamics model rather than to the cube of the number of elements. The objective of this study is to refine the method used in ref 9 to include an automatic time step routine; and then apply the technique to study blade loss induced rotor rubs. The automatic time step routine is necessary so that the time step can be varied as the rotor impacts the case. Also, the numerical stability of the method used in ref 9 will be investigated.

## SYMBOLS

| | |
|---|---|
| a | reference amplitude |
| c | radial clearance |
| E | absolute error estimate |
| F | force |
| O | order of error in Taylor series |
| q | order of Taylor series |
| r | radial displacement |
| S | stability matrix |
| t | time |
| $\Delta t$ | time step |
| u | defined in eq.(4) |

| | |
|---|---|
| $z$ | independent variable |
| $\alpha$ | given set of constants |
| $\zeta$ | damping ratio |
| $\lambda$ | eigenvalue of stability matrix |
| $\mu$ | coefficient of friction |
| $\omega$ | frequency |

## ANALYSIS

Numerical integration:

Given an arbitrary vector function $\vec{z}_k(t)$ whose derivatives exist, $\vec{z}_k^{(j)}(t)$, a Taylor series expansion can be written:

$$\vec{z}_k(t + \Delta t) = \sum_{j=0}^{q-k} \frac{(\Delta t)^j}{j!} \vec{z}_k^{(j)}(t) + \vec{o}_{q-k} \tag{1}$$

with remainder of order $\vec{o}_{q-k}$. If the arbitrary function is chosen as:

$$\vec{z}_k = \frac{(\Delta t)^k}{ak!} \vec{r}(k) \tag{2}$$

the Taylor series for this function becomes:

$$\vec{z}_k(t + \Delta t) = \sum_{j=0}^{q} \binom{j}{k} \vec{z}_j(t) + \vec{o}_q \tag{3A}$$

where the binomial coefficients are defined as:

$$\binom{j}{k} = \begin{cases} \dfrac{j!}{k!(j-k)!} & \text{for } j \geq k \\ 0 & \text{for } j < k \end{cases} \tag{3B}$$

If the form of the remainder is chosen as:

$$\vec{o}_q = a_k \vec{u} \tag{4}$$

the Taylor series becomes:

$$\vec{z}_k(t + \Delta t) = \sum_{j=0}^{q} \binom{j}{k} \vec{z}_j(t) + a_k \vec{u} \tag{5}$$

where the alphas are given in ref 7 and $\vec{u}$ can be determined from the equations of motion at the advanced time. The form of the set of the equations of motion at the advanced time is:

$$\sum F(\vec{r},\ \dot{\vec{r}},\ \ddot{\vec{r}},\ t + \Delta t) = 0 \tag{6}$$

From the definition of $\vec{z}$ , the various derivatives become:

$$\vec{r}(k) = \frac{ak!}{(\Delta t)^k}\vec{z}_k \tag{7}$$

Substituting for the various derivatives into the equations of motion; and knowing the values at the previous time, result in the equations of motion being a function of:

$$\sum F(\vec{u},\ t + \Delta t) = 0 \tag{8}$$

This set of equations can be solved for $\vec{u}$ and, from this value of u, the remainder can be used as an error estimate to control the time step. From the definition, $\vec{z}_1$ represents a nondimensional form of $\vec{r}_1$. Therefore an estimate of the maximum absolute error is:

$$E = \alpha_1 ||\vec{u}|| \tag{9}$$

where $||\vec{u}||$, the vector norm is the maximum component of $\vec{u}$. The computer code used in ref 9 was modified to include the following automatic time step algorithm. If E>.01%, re-do the calculation with the time step reduced by a factor of 10. If .01%>E>.001%, accept the calculation but decrease the time step by a factor of 2. If .001%>E>.0001%, accept the calculation and maintain the same time step. If .0001%>E, accept the calculation but increase the time step by a factor of 2.

Numerical stability:

The analysis of the stability of the numerical integration technique assumes a model of a rotor bearing system that is linearized at some instant of time. The homogeneous equation of motion for any mode is:

$$\ddot{r} + 2\omega\zeta\dot{r} + \omega^2 r = 0 \tag{10}$$

where omega is the natural frequency and zeta is the damping ratio for the mode. For every mode that is numerically possible, with

nonnegative damping ratio, the amplitude must either remain constant or decay in time. The numerical integration is defined as unstable if the amplitude grows in time.

Substituting the Taylor series for the various derivatives into the modal equation of motion at the advanced time results in:

$$u = -\sum_{j=0}^{q} \left[\frac{j(j-1) + 2j\omega\,\Delta t\,\zeta + (\omega\,\Delta t)^2}{2\alpha_2 + 2\alpha_1\omega\,\Delta t\,\zeta + \alpha_0(\omega\,\Delta t)^2}\right] z_j(t) \qquad (11)$$

For this value of $u$, the Taylor series expresses the solution at the advanced time in terms of the solution at the present time as:

$$z_k(t+\Delta t) = \sum_{j=0}^{q} \left\{\binom{j}{k} - \frac{\alpha_k\left[j(j-1) + 2j\omega\,\Delta t\,\zeta + (\omega\,\Delta t)^2\right]}{\left[2\alpha_2 + 2\alpha_1\omega\,\Delta t\,\zeta + \alpha_0(\omega\,\Delta t)^2\right]}\right\} z_j(t) \qquad (12)$$

Defining the matrix $S$ to be:

$$s_{kj} = \binom{j}{k} - \frac{\alpha_k\left[j(j-1) + 2j\omega\,\Delta t\,\zeta + (\omega\,\Delta t)^2\right]}{\left[2\alpha_2 + 2\alpha_1\omega\,\Delta t\,\zeta + \alpha_0(\omega\,\Delta t)^2\right]} \qquad (13)$$

and the vector $\vec{z}$ whose $k$th element is $z_k$, results in the finite difference equation:

$$\vec{z}(t+\Delta t) = S\vec{z}(t) \qquad (14)$$

This equation has a solution of the form:

$$\vec{z}(t+\Delta t) = \lambda\vec{z}(t) \qquad (15)$$

where lambda is an eigenvalue of:

$$S\vec{z} = \lambda\vec{z} \qquad (16)$$

If the $|\lambda| > 1$, the amplitude grows and the method is numerically unstable.

Rub model:

The interaction of a rotor with its case is a complicated phenomenon. It can involve non-linear deformation of both the rotor and the case. Rotor-case rubs were experimentally studied in ref 10. Analytically only simple rotor-case rub models are available; therefore, the case was assumed to be linear with dry friction interaction with the rotor. The radial and tangential forces on the rotor are then:

$$F_r = 0, \quad F_\theta = 0 \qquad |\vec{r}| < C \tag{17A}$$

$$F_r = -k(|\vec{r}| - C), \quad F_\theta = \mu F_r \qquad |\vec{r}| > C \tag{17B}$$

## RESULTS AND DISCUSSION

The numerical method of ref 9 employed a second order integrator with a constant time step. However, to study blade loss induced rotor rubs, it is necessary to modify the method of ref 9 to include higher order integrators with an automatic time step routine. The automatic time step routine is necessary so that the time step can be varied as the rotor impacts the case. In order to calculate high frequency components accurately, the time step must be less than the period of the high frequency component. When only low frequency components are important the time step can be increased to decrease computing time. The algorithm discribed in the analysis section keeps the maximum error in the displacement at less than .01%. It tries to maintain the error between .001% to .0001% by either decreasing or increasing the time step.

Another way to decrease computing time is to use a higher order integrator. Ref 7 studied the numerical stability of up to a sixth order integrator applied to a first order differential equation. The numerical stability of these integrators applied to a second order differential equation was given in the analysis section. The numerical stability of an integrator is based on modal rotor dynamics analysis. If the integrator is applied to a mode which is not driven and has damping, the amplitude must decay in time. Figure 1 shows a stability map for the integrators used in ref 7 applied to a second order differential equation. The abscissa is the damping ratio and the ordinate is the product of the time step and natural frequency for the mode. The stability map has contours on it for which the amplitude does not change from one time to the next. On one side of the contour

the amplitude grows; (unstable region), and on the other side it decays, (stable region).

Figure 1 shows the stable regions for a fourth through sixth order integrator. The second and third order integrators were stable everywhere. For the regions where the integrators were unstable, the amplitude grew by a few percent per time step. It would take on the order of a hundred time steps for the amplitude to double, and it would take on the order of a thousand time steps for the amplitude to increase by a factor of a thousand. Due to round off errors, every mode that is numerically possible in the rotor dynamics model, has a finite amplitude. These amplitudes may be small; but if they are in an unstable region, in a few thousand time steps they can become very large. For this reason, only the second and third order integrators were used. This is still a vast improvement over other types of integrators such as the one used by NASTRAN. NASTRAN uses an implicit form of the Newmark-Beta integrator, ref 8. This integrator is second order and does not have an error estimate.

The rotor-bearing system described in ref 11, (which dynamically simulates a typical small gas turbine), was used as the example problem. This rotor bearing system consisted of a shaft with three disks mounted on two axially preloaded ball bearings (fig 2). In this rotor-bearing system the bearings were mounted in squeeze-film damper journals, and the journals had centering springs.

The first three critical speeds for the rotor bearing system without oil in the dampers are shown in figure 3. Note that all the modes are bent- shaft modes. The "classical" hierarchy only applies to stiff shafts; therefore, the classical mode shapes do not characterize the actual mode shapes. The first mode, about 7600 rpm, classically would be the cylindrical mode. But in this case, it has a large amount of bending outward near the shaft center. The second mode, about 9200 rpm, classically would be the conical mode. In this case, it has a slight amount of bending outward near the shaft ends. The third mode, about 11200 rpm, classically would be the bending mode. In this case, it has a large amount of bending throughout the shaft.

The rotor-bearing system was modeled by using 23 elements. Prior to the blade loss simulation the rotor was assumed to be balanced and operating at 9500 rpm. The blade loss was simulated by an instantaneous application of 5 mils of mass excentricity in the far disk. The equations of motion for this system were directly integrated by the method used in ref 9 with a variable time step. The output was interpolated to equal time steps; (100 time steps per shaft

rotation), and displayed on a CRT, figure 4. The display showed an oblique view of the rotor bearing system, with the bearing center line as the oblique axis. The transverse vibration is indicated by the position of the rotor centerline. The scale of the transverse vibration exaggerates the amplitude of the vibration. The display on the CRT was photographed at each time step. These photographs were then shown as a motion picture.

Figure 5 shows the superposition of the first ten frames of the blade loss simulation without a rub. Initially the rotor, the bearing, and the mass center line coincided. After the blade loss, a traveling wave starts at the blade loss disk and travels down the rotor. During the time high frequency components are dominant, (because the rotor as a whole is not moving). A model analysis which only uses the lower modes cannot discribe the motion during this time period.

Figure 6 shows the position of the rotor for the first six rotations of the rotor after blade loss without a rub taking place. During the first rotation, the blade loss disk spirals out. During the second rotation, the disk on the other end of the shaft spirals out. During the third rotation. the center disk spirals out. After this the envelope of the rotor positions, seems to oscillate in a conical fashion, with a frequency of about 1/4 operating speed. This beating seems to be at a frequency difference between the operating speed and the 1st critical speed. (Ref 12 experimentally showed a similiar beat frequency between the operating speed and the critical speed.) During this time the rotor shape resembles the third critical, except that the bearing center line is not in the plane of the rotor. The maximum amplitude occurred on the blade loss disk on the sixth rotation and on the opposite disk on the fourth rotation. The conclusion drawn from this figure is that if there is clearance space down the rotor and a rub occurs, it does not necessarily occur at the blade loss disk first.

The rotor-case rub was simulated by surrounding each disk with a shroud that had a 2 mil radial clearance and a stiffness of 100,000#/in. The rub was induced by a repeat of the blade loss simulation with the clearance restrain. Two rub simulations were run, one with a coefficient of friction of .1 and the other with a coefficient of friction of .2.

Figure 7 shows the first 6 rotations of the shaft after blade loss for a coefficient of friction of .1. During the first shaft rotation the blade loss disk spirals outward and bounces off the case four times. Each collision of the rotor with the case sends out

traveling waves down the rotor. These waves interact with each other causing the envelope of the rotor motion to be very complicated. On the second shaft rotation both outboard disks bounce off the case four times. As the rotor continues to turn the orbit becomes more circular. That is, the rotor-case interaction becomes less of a bouncing nature and more of a continuous contact. The envelope of the rotor motion seems to be oscillating in a conical nature; but both outboard disks seem to remain in contact with the case. The rotor continues to whirl about the bearing centerline in the rotational direction (forward whirl). The frictional drag forces are not large enough to drive the rotor into backward whirl.

Figure 8 shows the first 4 rotations of the shaft after blade loss for a coefficient of friction of .2. The motion of the rotor on the first rotation is similar to the .1 coefficient of friction case. On the second rotation, the blade loss disk has a very hard collision with the wall, causing the rotor to bend considerably. On the third rotation the rotor whirl direction changes from forward to backward whirl and the rotor whirl begins to accelerate in the backward direction. On the fourth rotation, the rotor motion becomes very large and it continues to grow on succeeding rotations.

This example problem has shown that small changes in the coefficient of friction, (from .1 to .2) can change a rotor response to a blade loss condition from a relatively safe response to a catastrophic response. For seal rubs the coefficient of friction is probably between .1 to .2. For blade tip rubs, this rub model is not accurate. This type of rub involves material removal, phase changes, and or non-elastic deformations. If this model were to be used in a general manner, then the coefficient of friction would probably be greater then .2.

In conclusion, this computer code allows us to look at blade loss induced rotor rubs and displays the rotor motion in a motion-picture format. A 10-minute, 16-millimeter, color, sound motion-picture supplement is available, on loan, that shows the computer made motion picture for the blade loss induced rotor rubs.

## SUMMARY OF RESULTS AND CONCLUSIONS

A method for direct integration of a rotor dynamics system experiencing a blade loss induced rotor rub was developed. The following conclusions were drawn:

1. The method was numerically stable for any time step up to a third order integrator.

2. The time step was controlled so that the maximum error was less then .01% and the probable error was between .001% to .0001%.
3. For the rotor typical of small gas turbines a small change in the coefficient of friction, (from .1 to .2), caused the rotor to change from forward to backward whirl and to destroy itself in a few rotations.

This method provides an analytical capability to study the susceptibility of rotors to rub induced backward whirl problems.

## REFERENCES

1. Johnson,D.C.: Synchronous Whirl Of A Vertical Shaft Having Clearance In One Bearing. J. Mech. Eng. Sci. vol.4, No.1, 1962, pp 85-93.
2. Black,H.F.: Interaction Of A Whirling Rotor With A Vibrating Stator Accross A Clearance. J. Mech. Eng. vol.10, No.1, 1968, pp 1-12.
3. Childs,D.W.: A Rotor-Fixed Model Simulation Model Of Flexible Rotating Equipment. J. Eng. ind., Vol.96, No.2, May 1974, pp 659-669.
4. Gunter,E.J.;et al.: Transient And Stability Analysis Using The Modal Method. UVA/528144/ME77/102, University of Virginia, 1977.
5. Shen,F.A.: Flexible Rotor Dynamics Analysis. (R-9252, Rocketdyne; NASA Contract NAS3-14422,) NASA CR-121276, 1973.
6. Kascak,A.F.: Stability Of Numerical Integration Techniques For Transient Rotor Dynamics. NASA TP-1092, 1977.
7. Gear,C.W.: The Automatic Integration Of Stiff Ordinary Differential Equations. Proc. IFIP Congress Information Processing 68 (Edinburgh 1968), A.J.H. Morrell ed., Vol.1, Mathematics Software, North Holland Publ. Co., Amsterdam 1969, pp 187-193.
8. Newmark,N.M.: A Method Of Computation For Structural Dynamics. J. of the Eng. Mech. Div., Proc. of the American Soc. of Civil Eng. Paper No.2094, Vol.85, No. EM3, July 1950, pp 67-94
9. Kascak,A.F.: Direct Integration Of Transient Rotor Dynamics. NASA TP 1597, AVRADCOM TR 79-42, Jan 1980.
10. Bill,R.C. and Wisander,D.W.: Friction And Wear Of Several Compressor Gas-Path Seal Materials. NASA TP 1128, 1978.
11. Cunningham,R.E.; Fleming,D.P.; and Gunter,E.J.: Design Of A Squeeze-Film Damper For A Multi-Mass Flexible Rotor. ASME Trans. J. Eng. Ind., Vol.97, No.4, Nov. 1975, pp 1383-1389.

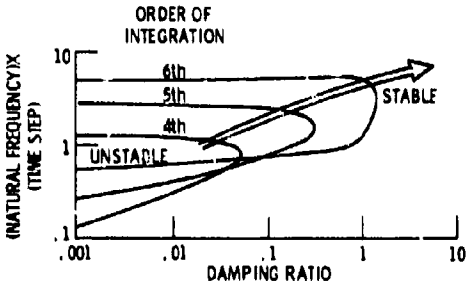



Figure 1. - Numerical stability of Gear's integration method applied to a second order differential equation for a 2nd thru 6th order of integration. The 2nd and 3rd order methods are always stable.

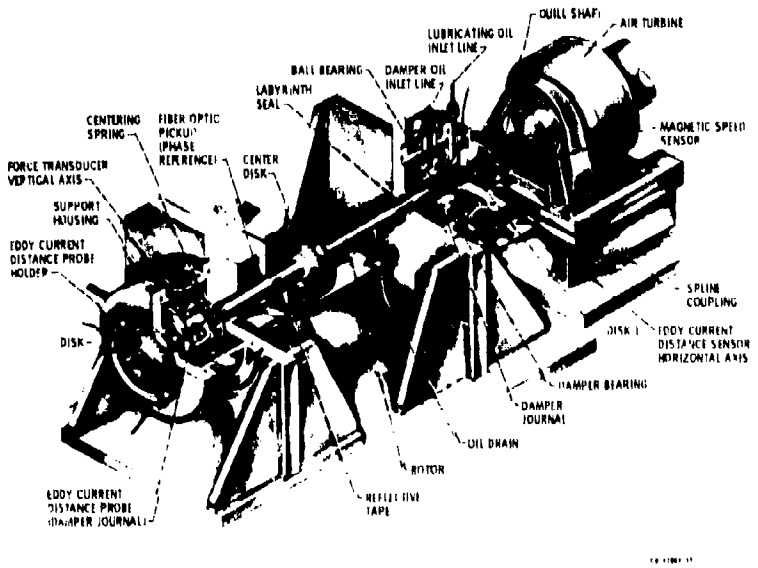



Figure 2. - Schematic of test apparatus used in experiments on steady-state unbalance response.

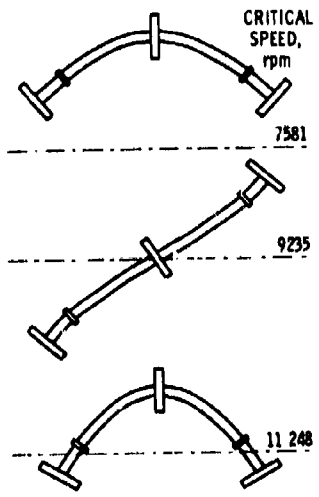



Figure 3. - Critical speeds and mode shapes.

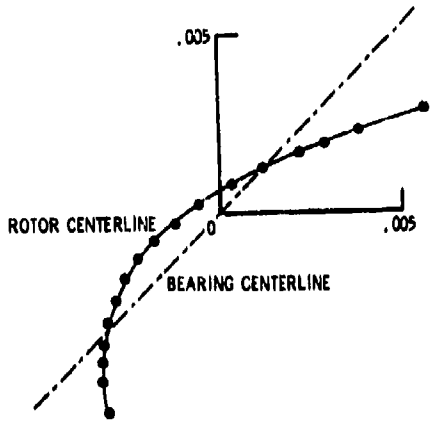



Figure 4. - Oblique view of rotor centerline whirling about the bearing centerline.

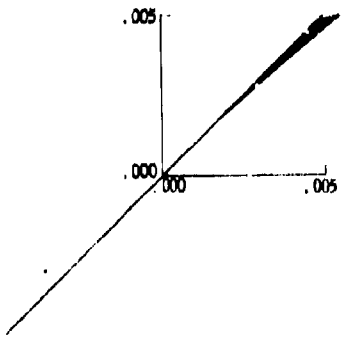



Figure 5. - Initial movement of rotor after blade loss.

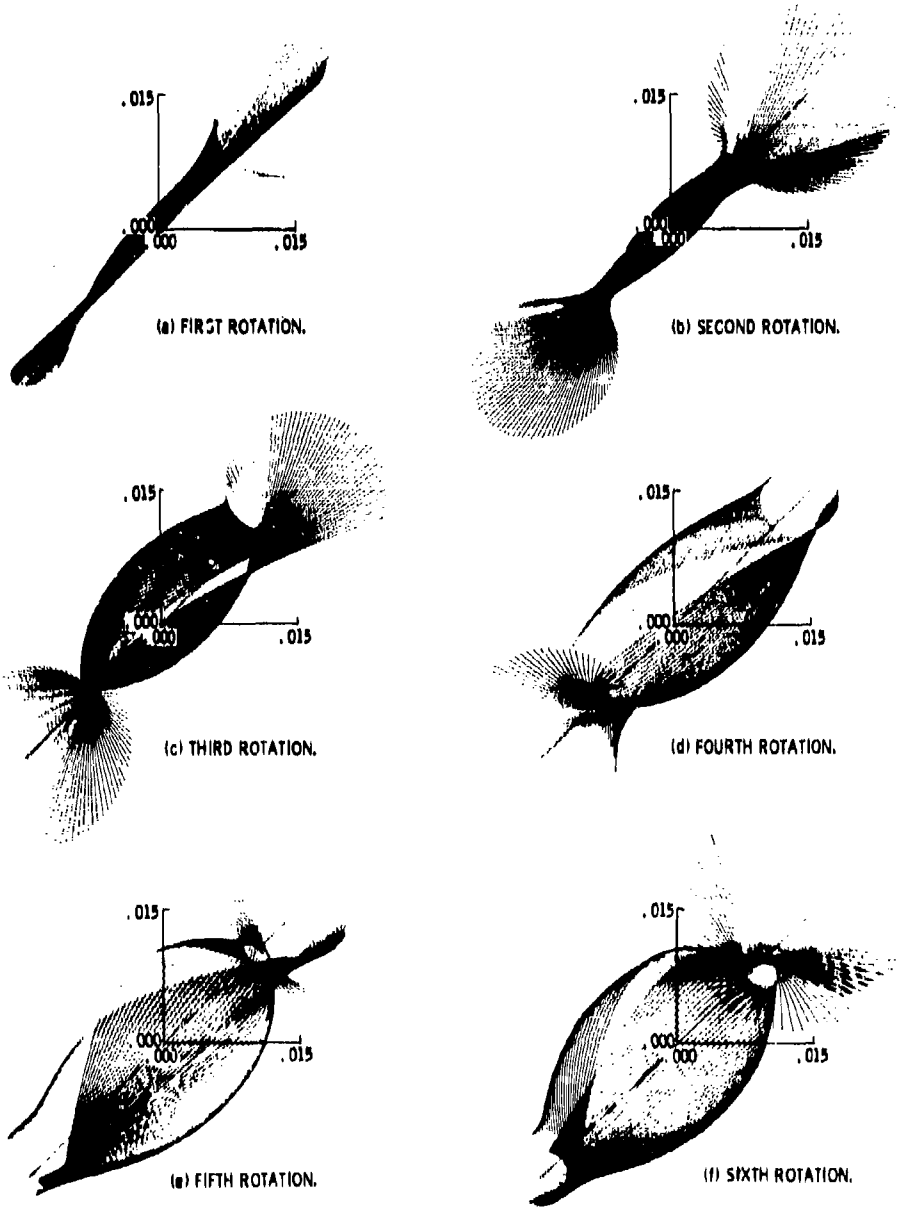



Figure 6. - Envelope of rotor motion for first six rotations of rotor after blade loss (without a rub taking place).

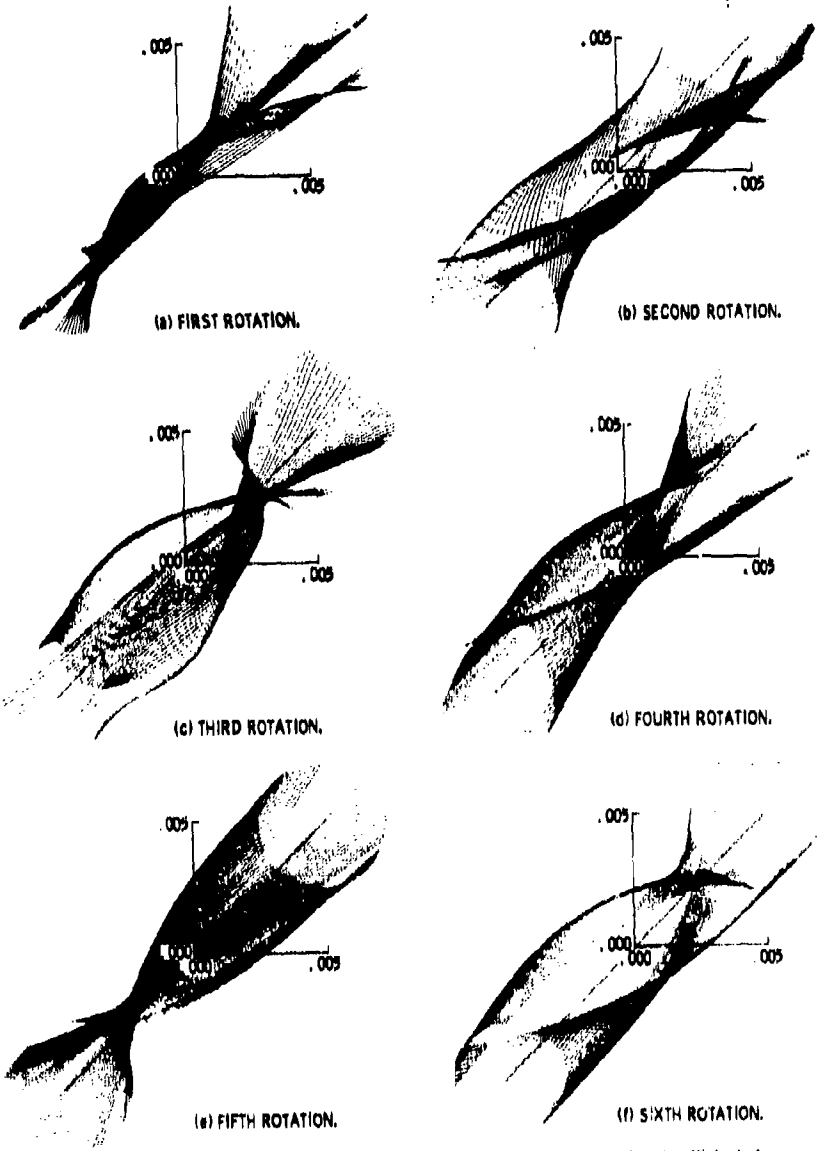



Figure 7. - Envelope of rotor motion for first six rotations of rotor after blade loss (coefficient of friction equals 0.1).

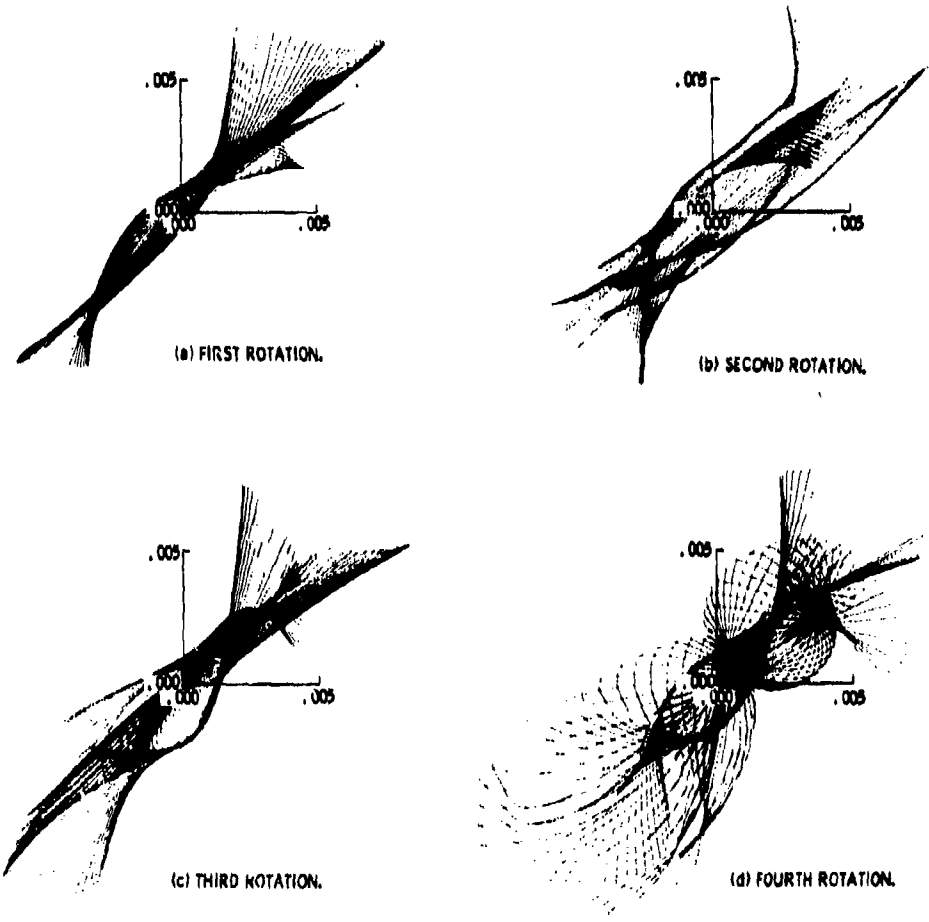



Figure 8. - Envelope of rotor motion for first four rotations of rotor after blade loss (coefficient of friction equals 0.2).

# NON-DESTRUCTIVE TESTING FOR FIELD WELDS: REAL-TIME WELD QUALITY MONITOR

*FRANK KEARNEY
CONSTRUCTION ENGINEERING RESEARCH LABORATORY
CHAMPAIGN, IL 61820

## 1. INTRODUCTION

### Background

During the welding process, changes in arc voltage, travel speed, and heat input can occur without the operator's knowledge. These changes can cause defects such as porosity, slag inclusions, incomplete fusion, and undercut in the deposited weld metal. The cost of locating and repairing these defects can be a major portion of construction costs; welding inspection can constitute 25 to 40% of the weld fabrication cost. In addition, weld defects decrease service life of welded joints.

Consequently, it is necessary to monitor the welding parameters to detect, identify, and locate possible defects. A weld monitor with real-time output would aid the inspector in designating suspect areas for non-destructing testing after welding. Further, a real-time weld quality monitor could be used to interrupt welding when defects are occurring thus precluding costly rework. To address this need, the US Army Construction Engineering Research Laboratory (CERL) is conducting research to develop a field-portable real-time weld quality monitor (WQM).

In the initial phase of the study, the following requirements were established for the device:



PRECEDING PAGE BLANK

1. Monitor the three primary signals from the weld system: arc voltage, current, and travel speed; compare them to preset limits; and alert the operator if the limits are exceeded.

2. Calculate the heat input, nugget area, and cooling rate from the three primary signals; compare these values with preset limits; and alert the operator if these limits are exceeded.

3. Be field portable.

4. Interface easily with in situ welding equipment.

Essentially, the WQM is intended to provide a mechanism to merge the welding engineer's design intent with the actual field welding process.

Following delineation of these requirements, a prototype WQM was designed, fabricated, and tested using input from a fully automatic gas metal-arc (GMA) welding machine. The automated GMA process was chosen to obtain close control and reproducibility of the welding variables for initial testing.

## Objective

The objectives of the first phase of the study were (1) to configure a portable, real-time WQM system, and (2) conduct laboratory and field tests to determine the adequacy and field applicability of the design.

## Approach

The design of the breadboard WQM was modified to incorporate improvements indicated by actual welding situations in the laboratory. Hardware was assembled and packaged for field use.

In the transitional period from laboratory prototype to field prototype, personnel in Government and the private sector were consulted and their suggestions were used to further improve the unit. The unit was then installed in a welding situation that would thoroughly test all modes of operation.

## 2. FACTORS AFFECTING WELD MECHANCIAL PROPERTIES

In the development, testing, and evaluation of the various generations of the weld quality monitor, certain basic welding parameters and relationships were used as guidelines.

Defects

Changes in the welding parameters of arc voltage, travel speed, and heat input can cause several types of defects of deposited weld metal.

*Porosity* is a void or gas pocket trapped in solidifying weld metal. The reduced solubility of the gas in the metal caused by the decreasing temperature forces the gases out of solution. The gases are originally introduced either by poor shielding, which contains air, or by chemical reactions in the molten weld metal. With stick electrodes, too long an arc resulting from excessive arc voltage can reduce the shielding effectiveness, thus introducing gas.

*Slag inclusion* is the entrapment of an oxide or other nonmetallic material under the weld beam. The major source of slag is the coatings on stick electrodes. This defect is related to heat input.

*Incomplete fusion* is the failure of adjacent layers of the weld metal or weld base plate to fuse. Incomplete fusion may result when the adjacent metal is not heated to the melting point because of insufficient heat input.

*An undercut* is a groove melted into the base plate at the toe of the weld and is caused primarily by excessive travel speed in relation to the welding current.

In addition to the defects caused by improper control, the heat generated by the welding process can cause the following changes in the base metal:

1. Grain Coursening
2. Softening ("Annealing effects")
3. Hardening (Phase precipitation or transformation)
4. Segregation of constituents
5. Grain boundary melting
6. Loss of ductility
7. Loss of toughness

8. Residual stresses causing distortion or cracking.

The type of change which occurs depends on the chemical composition of the base metals and electros and heat history of the base plate.

In the two commonly used field welding processes -- shielded metal-arc (stick electrodes) and gas metal-arc (bare wire) -- the source of heat for melting the material is an electric arc. Control of the arc parameters will control the amount of heat generated, the length of time at an elevated temperature, and the cooling rate of the weld zone.

Base Metal Microstructure

The cooling cycle after a weld pass determines the microstructure of the weld metal and the heat-affected zone. With fast cooling rates, some steels become very hard because of a martensitic transformation. If the cooling is sufficiently slow, the metal may be more ductile and the structure ferritic and pearlitic. The type of steel generally determines which of these structures is desired. For low-carbon and low-alloy steels, the pearlitic structure is desirable, while for high-strength quenched and tempered steel, the martensitic structure is desirable.

Martensite is undesirable in low-carbon and low-alloy steels designed for yield strengths less than eighty ksi (552 $MN/m^2$) because of its hardness and low solubility for hydrogen at ambient temperatures. The combination of characteristics increases the likelihood of hydrogen cracking in the joints. Use of low hydrogen (stick electrodes) the gas metal-arc welding system reduces this tendency towards hydrogen-induced cracking.

Cooling Rate Control

Control of the cooling rate is essential in preventing undesirable microstructure in the weld and heat-affected base plates. A mathematical combination of arc voltage, current, and travel speed known as heat input (HI) has been used as a means of controlling cooling rates for many years. The equation for calculating heat input is:

$$HI(J/in.) = \frac{\text{VOLTAGE} \times \text{AMPERAGE} \times 60}{\text{TRAVEL SPEED (in./min.)}} \qquad \text{[Eq 1]}$$

The normal maximum has been 55 thousand to 60 thousand Joules/inch. (21,654 to 23,622 J/cm) for the field processes mentioned above. Another means of controlling cooling rate has been preheat treatments. Dorschu [1] has shown that the relationship between heat input, preheat temperature, and cooling rate is:

$$CR = \frac{m\ (T - T_o)^2}{HI} \qquad \text{[Eq 2]}$$

Where CR = cooling rate

T = test temperature

$T_o$ = preheat temperature

CR = cooling rate

c = constants

HI = heat input

Equation 2 indicates that the higher the preheat temperature and heat input, the slower the cooling rate.

Shultz and Jackson [2] have shown that the cross-sectional area of the weld bead is a useful indicator of weld metal mechanical properties and that a relationship exists between the area and cooling rates. They also found that arc voltage has little or no effect on the nugget area and cooling rate. The relationship that Schultz and Jackson have developed for nugget area, arc current, and speed is:

$$na = 122 \times 10^{-7}\ \frac{i^{1.55}}{S^{.0903}} \qquad \text{[Eq 3]}$$

where na = nugget area (sq. in.)

i = arc amperage

S = arc travel speed (in./min.)

## 3. CIRCUIT DESCRIPTION

Figure 1 is a block diagram of the weld quality monitor showing the input signals from the welding arc. These signals are conditioned to standard values and sent to the comparator module, which compares the input signals with a set of limit signals. If the input signals are too high or too low, the appropriate alarm is triggered. Input signals are also transmitted to the analog computer module for calculation of the heat input, cooling rate, and nugget area. The calculated values are then compared to reference signals and the appropriate alarm is triggered if needed.

## 4. LABORATORY TESTS

### Procedure

Each channel of the laboratory prototype monitor was individually tested with a variable signal similar in current and voltage level to the signal from a welding machine. The limits for each channel were set, and the test of voltages were varied to simulated changes in the primary signal.

After each channel was tested successfully, the three simulated primary signals were fed into the monitor simultaneously. The limits were again set and the input voltages varied. All circuits including the analog computer section were checked for accuracy and reproducibility.

The monitor was then connected to the CERL welding machine to test the circuitry with actual signals after the limits were set; the welding arc was established on a test plate.

### Results

Results of the laboratory testing showed that all channels performed satisfactorily both independently and in conjunction with each other. The warning lights were triggered when the input signal exceeded the limits set by the reference signal, and no difficulties were encountered when the limit spans were changed.

While investigating the signals of the three parameters (voltage, current, and speed), it was found that the voltage and amperage signals contained spurious noise signals. These signals were removed by (1) incorporating filters in the data channels to

eliminate the peaks and smooth out the signals, thus reducing the chance of damage to components, and (2) replacing the shunt as the amperage signal source with a Hall effect solid state transductor. (The advantage of using the transductor is that it is not directly connected to the welding machine as the shunt is; instead, it fits around the cable and measures the magnetic field generated by the current passing through the cable.) The transductor minimized amperage transient signal problems; filters were installed in all channels in field contingencies. The modifications indicated by the laboratory testing program were incorporated into the monitor before field testing.

## 5. FIELD TEST

### Site Selection

The general types of welding operations considered for field testing were shop fabrication which uses automated welding equipment and field fabrication/repair which involves manual or semi-automatic welding and is more dependent on the operator's subjective judgment.

In addition, it was decided that field tests would be more conclusive if the weld quality monitor were used in conjunction with some other form of non-destructive testing. Two sites were available that offered these combinations: Flint Steel Corp., Tulsa, OK, and a hydro-electric turbine shaft repair job at Ozark Hydro-Electric Plant, Ozark, AR. The field repair job at Ozark power plant was chosen since it would entail situations that could not be simulated during the CERL laboratory evaluations. It was felt that the time and space constraints of the field repair situation would assess the unit's adaptability most rigorously. In addition to the hardware evaluation, the field test would provide an opportunity for welding personnel from industry to appraise the WPM.

### Test Operations

The WQM and auxiliary equipment were transported from CERL to the Ozark plant in a conventional automobile with no special handling. The equipment was set up by maintenance personnel and was ready for operation in less than one day.

The WQM was set up approximately 50 feet (15 m) from the repair location. Installation of this device involves simple

disconnection and reconnection of one of the leads from the welder power unit; no hard wiring is required.

Since no speed measurement system was available for this test, a precision voltage source was used to provide an equivalent signal to compute heat input and nugget area. For this mode of operation, a voltage corresponding to a particular welding speed is input to the analog computer module to compute heat input in nugget area (equations 1 and 3). For example, if the analog module were scaled for one volt equal one inch per minute, then a six-volt signal from the precision voltage source would be input for a welding speed of six inches per minute.

The signals taken at the output of the signal conditioners before filtering for inputting the comparators are not distorted, and the response of the sensors to the voltage and current variations incurring in the arc is preserved in the transduction and the conditioning process. Thus, several data utilization options are possible -- from simple alarms to adaptive control systems.

## Results

The central unit and associated sensors were interconnected and energized without disrupting the welding contractor or requiring welding equipment modification. This verified the adaptability and flexibility of the design objective. Installation of the system was accomplished by a ceramic engineer; an electrical engineer was not required.

The WQM was operated by non-electronic personnel (a welding engineer) with minimal instruction. The data display and printout were understandable to both laboratory and contractor personnel.

During the start up of the WQM, erratic operation was indicated by the visual display; the modular packaging method enabled the problem to be diagnosed and repaired rapidly by interchanging modules. Again, this was accomplished by non-electronic personnel using predefined trouble-shooting procedures.

With the feasibility of the design intent of the WQM demonstrated, a concerted program was initiated to develop: (1) suitable speed-measuring systems for manual welding situations, (2) specific radiometric measurement techniques involving acoustic emission and thermal spectral analysis, and (3) digital processing features using microprocessor to facilitate programming the WQM.

## 6. OPTOELECTRONIC WELD EVALUATION

Direct quantitative measurements of certain parameters of welds in process have not been possible for several reasons. In particular:

a. The high weld temperatures consume and destroy sensors proximate to the weld area.

b. Contacting sensors introduce a discontinuity of the weld process causing data of uncertainty.

c. In the case of manual welding, the subjectiveness peculiar to the welder is indeterminant and variable.

Presently, some indirect measurements are utilized such as thermocouples, etc., but these techniques exhibit time lags, averaging effects, and other factors that mitigate the validity and reproducibility of the information obtained.

Because of this inability to measure directly and instantaneously the quantities relevantly to a satisfactory weld, a research program was implemented to produce non-contacting instrumentation techniques that will be field applicable to directly monitor pertinent weld measurements such as cooling rate, weld speed, and heat input, to serve as input data to the CERL weld quality monitor.

Optoelectronic technology is used to detect the amplitude and wave length of radiation emitted by the welding arc. A photodector, or an array of photodectors is the primary sensor with appropriate circuitry to provide the required output information.

### Physics of Welding Arc

For all practical purposes, the welding arc can be thought of as a gaseous conductor which changes electrical energy into heat. The welding arc can be defined as a particular group of electrical discharges that are formed and sustained by the development of a gaseous conductive media. The current carriers for the gaseous medium are produced by thermal means and field emission.

The arc current is carried by the plasma, the ionized state of a gas composed of nearly equal numbers of electrons and ions.

Mixed with the plasma are other states of matter, including molten metals, slags, vapors, neutral and excited gaseous atoms and molecules.

Measured values of welding arc temperatures normally range between 5,000 and 30,000 degrees K, depending on the nature of the plasma and the current conducted by it.

The amount and character of spectral radiation emitted by arcs depend fundamentally upon the atomic mass and chemical structure of the gas, temperature, and the pressure. Spectral analysis of arc radiation will show bands, lines, and continua. The analysis of radiation from organic-type covered electrodes shows molecular bands due to the existence of vibrational and rotational states as well as line and continuum emission from excited and ionized states. The inert gas arcs radiate predominately by atomic excitation and ionization. As the energy input to arcs increase, higher states of ionization occur, giving radiation from higher energy levels.

The fundamental method utilized in the development of non-contacting sensors in this study is to separate and quantify segments of weld spectra correlatable to specific weld parameters. The visible spectrum and a portion of the infrared spectrum emanating from the argon-shielded gas tungsten arc are shown in Figure 2.

Optical Electronic Transduction Methods

Two methods of segmenting or partitioning weld spectrum are: Selection of photosensors having a spectral response only in sections of the spectrum to be measured and, use of optical filters to limit the wave length of radiation impinging on the photodector. For this work, the latter method was used; the radiation physics and adaptation of the optical electronics to the problem is illustrated in Figure 3. Extreme flexibility was provided by various combinations of commercial photographic filters which made it possible to segment the arc spectra into approximately five bands which provided adequate resolution to quantify weld flaws. Two examples will briefly illustrate the procedure.

An analysis of the metallurgical phase diagrams associated with weld nugget area suggested that the normal (acceptable) weld spectrum and a deviant spectrum characterizing a flaw would have wave lengths greater than 700 nanometers. To implement the "front end" of the sensor system, a Wratten type 89b filter was

selected coupled with a type TIL-63 phototransistor. This provided a sensor system with a photometric "window" of approximately 700-1,050 nanometers; thus, the desired spectra were detected while extraneous spectra were attenuated. A fiber optic light pipe was the transmission device between the arc and the phototransistor. Weld arc instability or "sputtering" is one of the most common flaw-inducing conditions encountered in practice. Laboratory testing using radiometers indicated that spectral lines omitted by an unstable arc were very dense in the visible range; to quantify this, a ratton 57 cylinder was used in the front end with notable results. Another flaw-inducing condition that was detected by this "poor man's photometer" was magnetic arc blow. Present work is concentrated on developing rugged temperature high optical systems to provide durability for field use. To date, the results are most encouraging; the fiber optic bundles are 1/16 to 1/8 inch diameter and fortuitously have a pass band in the range required, specifically .4 nanometers to 1.9 nanometers.

7. Large Scale Integration (LSI) and the WQM

The primary factor that makes the WQM a practical and ubiquitous tool is the confluence of welding engineering and large-scale integration electronics technology. Measurements and recordings of voltage, current and, more recently, acoustic emission data is becoming quite standard. The CERL WQM is innovative and unique in that it utilizes this data for <u>in-situ</u>, <u>real-time</u> analysis for <u>continuous</u> and <u>instantaneous</u> quality assurance.

8. Epilogue

Concurrent with the submission of this paper, the first successful tests of a prototype optoelectronic system configured for field use were conducted at CERL.

In this unit the spectrum is segmented and quantified by a grooved spectrograph and linear photodiode array. A high temperature fiber optic bundle is routed along the flexible cable/hose assembly to the welding gun and does not interfere with normal welding operations. Because of this versatility, the system is applicable to manual welding which is the principal method used by the Army.

Flaws caused by slag inclusions, loss of flux, loss of cover gas and magnetic arc blow have been distinctly characterized.

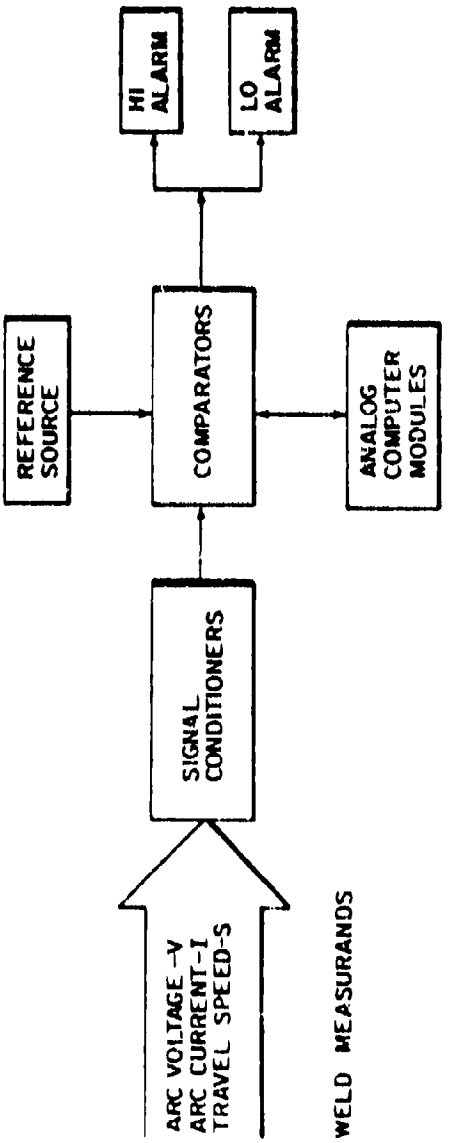



FIG. 1 BLOCK DIAGRAM OF WELD QUALITY MONITOR

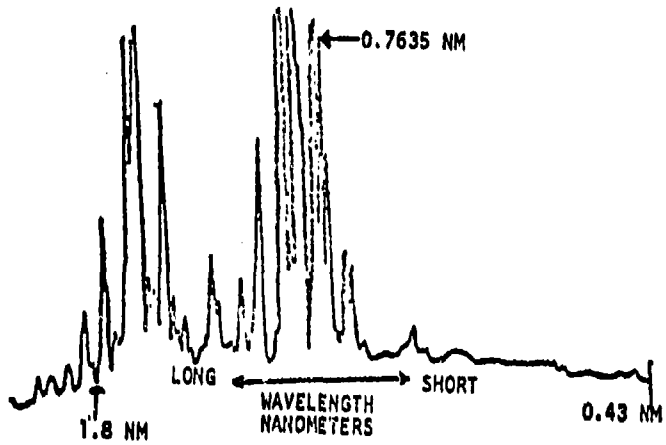



FIG. 2 SPECTRUM OF THE ARGON SHIELDED GAS TUNGSTEN ARC
ARGON ARC 3mm (1/8 in.) DIAMETER PURE
TUNGSTEN ELECTRODE 200 AMP (ELECTRODE POSITIVE)

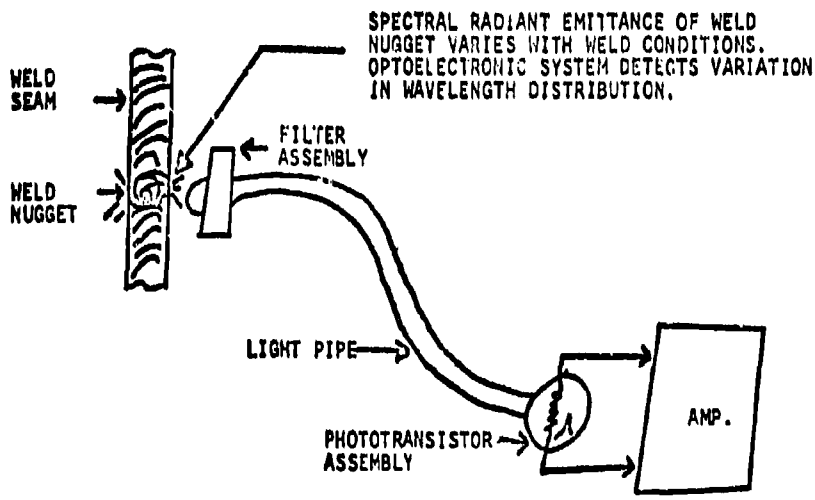



FIG. 3 OPTOELECTRONIC SYSTEM AND APPLICATION TO WELDS

## REFERENCES

[1] Dorschu, K. E., "Control of Cooling Rates in Steel Weld Metal," Welding Research Supplement (February 1968).

[2] Schultz, B. L. and C. E. Jackson, "Influence of Weld Bead Area on Weld Metal Mechanical Properties," Welding Research Supplement (January 1973).

[3] Weber, R. A. and C. E. Jackson, Review of Weldability of Construction Materials, Interim Report M-168/ADA027383 (CERL, 1976).

[4] Weber, R., F. Kearney, and S. Joshi, Development of a Weld Quality Monitor, Interim Report M-183/ADA027644 (CERL, July 1976).

[5] American Welding Society, Welding Handbook, AWS, 1976.

[6] Hackforth, H. L., Infrared Radiation, McGraw-Hill Book Co. 1960.

[7] Sharpe, R. W., Research Techniques in Nondestructive Testing, Academic Press, 1970.

[8] Simon, R. E., Electro-Optics Handbook, RCA, 1974.

*KEHS, BRANDT, BROMBORSKY and
LASCHE

# THE GENERATION OF GIGAWATT POWER LEVELS OF MICROWAVE RADIATION

R. ALAN KEHS, MR.
HOWARD E. BRANDT, DR.
ALAN BROMBORSKY, MR.
GEORGE LASCHE, CAPT, USA
U.S. ARMY ELECTRONICS RESEARCH AND DEVELOPMENT COMMAND
HARRY DIAMOND LABORATORIES
ADELPHI, MD 20783

## Introduction

Recent dramatic increases in the ability to produce ultrahigh power bursts of microwave radiation[1] both in the United States and in the Soviet Union are causing a complete reevaluation of our use of electronic systems on the battlefield and the susceptibility of current devices to new forms of electronic warfare. At the Harry Diamond Laboratories, a reflex triode has been used to produce peak powers as high as 3 GW in the X-band. Although this device was originally developed as an ion source,[2,3] the oscillating dipole motion of the electrons makes it an obvious candidate for a source of electromagnetic radiation. A fully relativistic, time-dependent, one-dimensional simulation code was written to investigate these collective oscillations and to predict the microwave energy spectral density.

## The Experiment

The basic geometry of the triode is shown in figure 1. When connected to the FX-45 Flash X-Ray machine at the Harry Diamond Laboratories,[4] the carbon cathode delivers 20 kA average current during a 25 ns wide pulsed accelerating potential of 1 MV peak. The anode is a 6.4 m thick aluminized Mylar film located 1 cm from the emission cathode. To keep the random energy and self-fields of the electrons from blowing up the beam (radially), a slow pulsed magnetic field of variable (0 to 4 kG) peak amplitude was applied along the axis of the triode. The final section of the vacuum coaxial line and

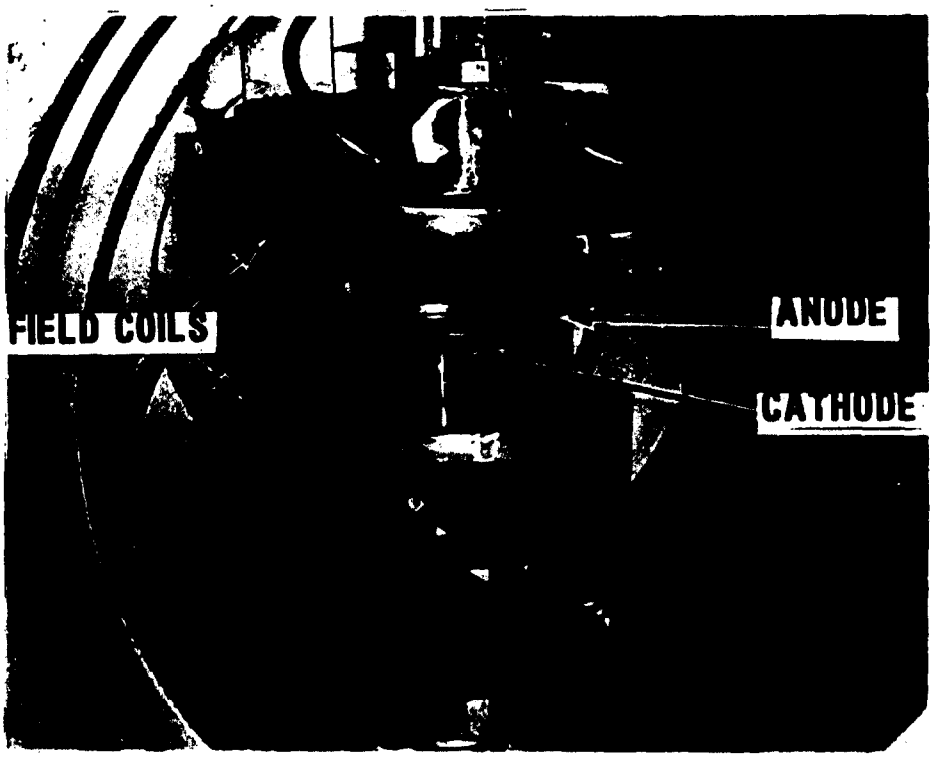



Figure 1. Reflex Triode Attached to FX-45 Pulse Time.

the cathode support structure were fabricated from copper clad G10 circuit board to maximize the intensity of the applied magnetic field and minimize the eddy current losses. The cathode diameter was 5 cm, the cathode-ground plane gap was 10.5 cm, and the anode-cathode gap was 1 cm.

The placement of the microwave diagnostics is shown in figure 2. X-band and Ku-band waveguides were used to carry signals to the detection apparatus. For radiation perpendicular to the triode axis, both horizontal and vertical polarizations were recorded by inserting and removing twists in the waveguide. Radiation parallel to the triode axis also was measured. Crystal detectors were used to determine power magnitudes, and long dispersive waveguides were used to examine the frequency content of the radiation.

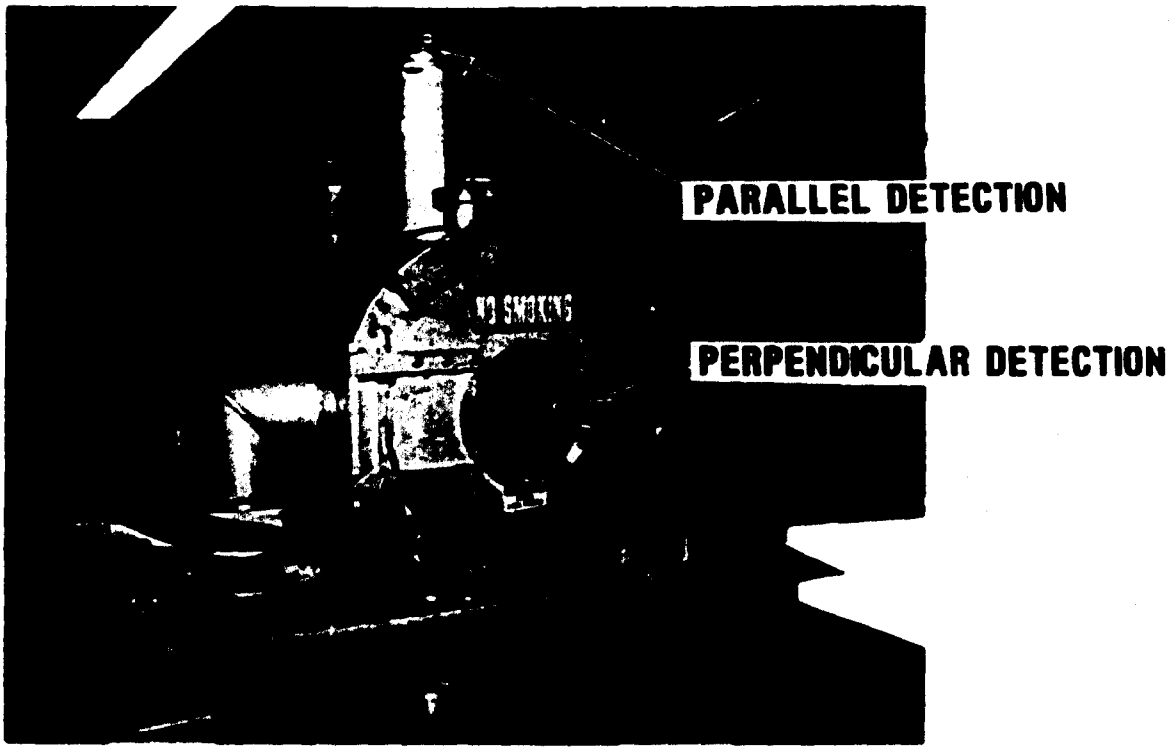



Figure 2. The Reflex Triode Vacuum Chamber with Microwave Diagnostics in Place.

Theory of Operation

The actual operation of the reflex triode is demonstrated by the particle trajectory shown in figure 3. As current from the cathode flows through the anode, space charge effects lead to the formation of charge bunches. These charge bunches reflect the electrons, and give rise to multiple modes of oscillation about the anode. The resultant space charge oscillations are similar to the Barkhausen oscillations that were observed in the early radio tubes of the 1920's.[5] These waves have a rich frequency content in which the dominant modes are determined by gap spacing, gap voltage, and applied magnetic field.

To gain insight into the parameters that affect microwave radiation from a reflex triode, a computer simulation of the electron motion was undertaken.[6] Based on a relativistic extension of the techniques used by Birdsall and Bridges,[7] the code used superparticle

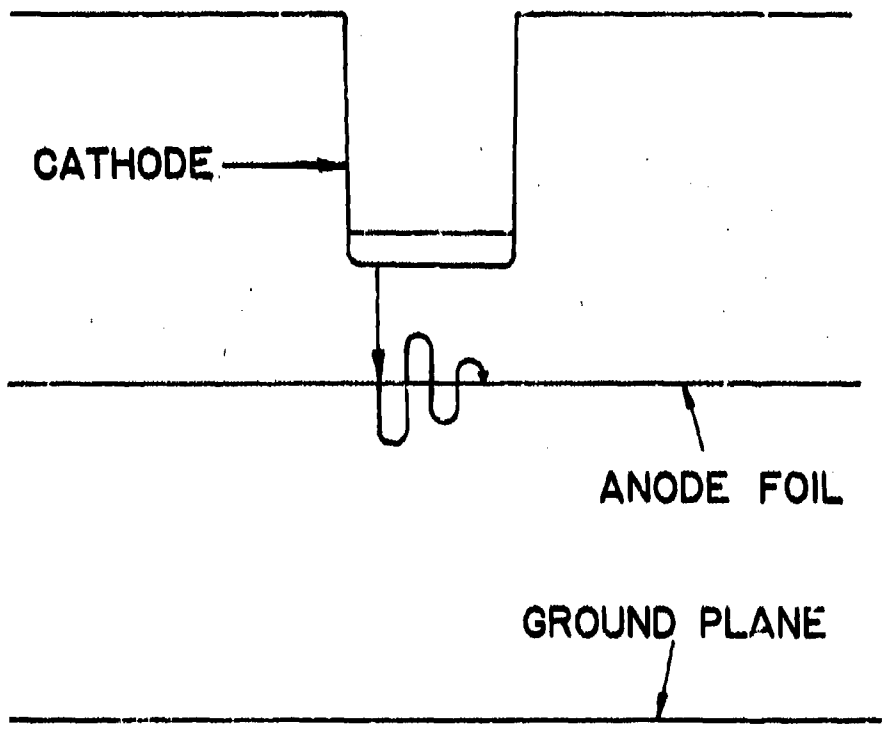



Figure 3. A Particle Trajectory in the Reflex Triode.

space charge sheets to represent the flow of charge in the triode. It was found that seven parameters completely define the electron dynamics of a reflex triode simulation run. The three geometric parameters are the distances from the anode to the near and far cathodes (shown as a ground plane in figure 3) and the area of the emitting cathode. The machine parameter is the waveform of the voltage applied to the anode-cathode gap as a function of time. The material parameter is the stopping power of the anode foil as a function of electron energy. The two internal code simulation parameters are the time step between electric field calculations and particle pushes and the available current for injection into the anode-cathode region. Time steps are chosen to be a small fraction of an anode-cathode light transit time. Typically, a time step of 0.26 ps is used, giving a Nyquist cutoff frequency of 2000 GHz.

The simulation results show that, at the onset of the voltage pulse across the anode-cathode gap, electrons are explosively pulled out of the cathode plasma and fill the triode region. The total space charge, however, does not monotonically increase but oscillates in time while rising to a saturated time averaged total charge. Electrons entering the gap with favorable phase relative to the given frequency component of the resultant space charge fields give up energy to the fields and remain in the triode region, while those with unfavorable phase will gain energy from the fields and are ejected from the triode region. Favorably phased electrons tend to be grouped together spatially since they enter at nearly the same time. Also, space charge periodically limits the subsequent entrance of electrons at the emission cathode. These mechanisms give rise to electron space charge bunching. While the electron bunches oscillate back and forth about the anode, they also interact with one another ejecting and capturing electrons from one another and thereby depleting and growing in size. This behavior is similar to strong Langmuir turbulence in a relativistic electron plasma.

Once the triode has been saturated, the total charge oscillates about the mean value. A spatially fluctuating virtual cathode forms on the far side of the anode at a time averaged distance approximately equal to the anode-cathode gap. Transient bunching mechanisms lead to high frequency oscillations about the anode. Spectral analysis of these space charge oscillations reveals a broad spectrum of peaks in the gigahertz regime separated by 0.2 to 0.4 GHz and peaking at 10 GHz.

The actual spectral density of the emitted radiation was computed in the far field dipole approximation by using the calculated current density of the space charge cloud. An important

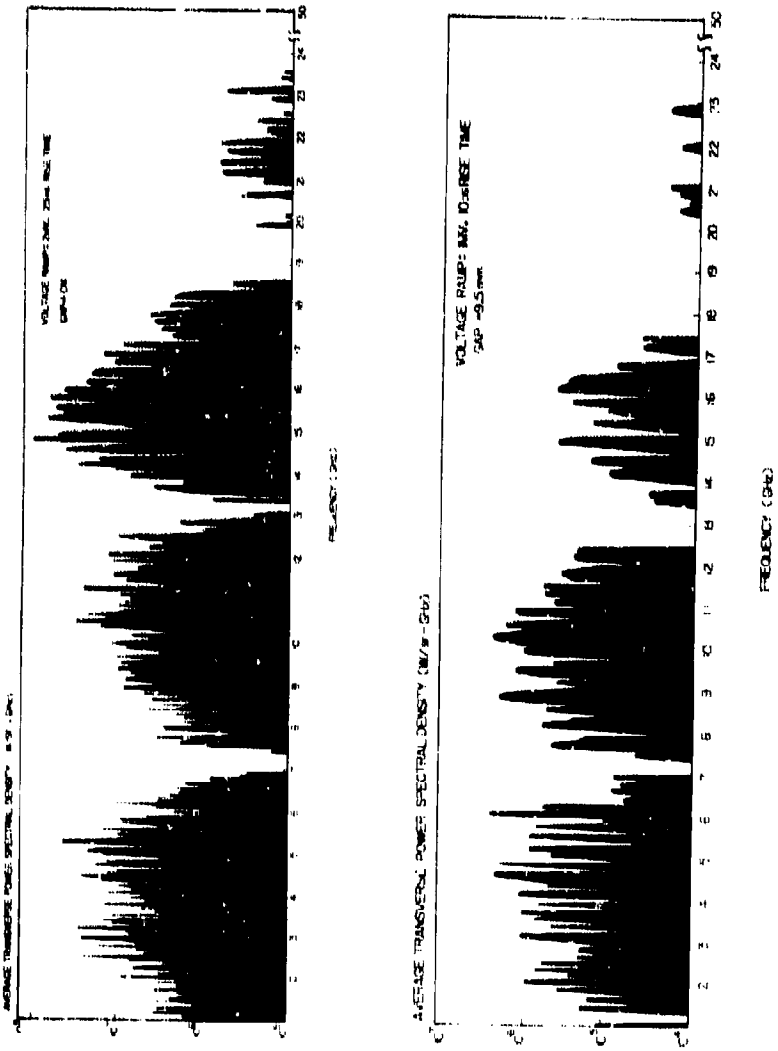


Figure 4. Calculated Microwave Spectra from a One-Dimensional Reflex Triode Computer Simulation

consequence of this treatment is an interference form factor for the oscillating space charge cloud geometry with zeroes in the region of interest (7.2 and 13.2 GHz).

Discussion

Figure 4 shows two calculated average power spectra for typical experimental parameters. On the upper plot, the predicted average power in the X-band is approximately 50 MW. This prediction compares favorably with experimental shots during which a high (several kilogauss) magnetic field was applied along the axis of the triode and confirms the validity of the model in that parameter regime. The high magnetic field confines particle movement to the one-dimensional motion seen in the computer simulations. The lower plot shows the computed radiation spectrum that would be expected if a larger (higher peak and longer rise time) pulse were applied across the anode-cathode gap. The power output is seen to be sensitive to changes in the size of the applied voltage pulse, which (in our machine) typically varies by 5 to 10%. Such variations can at most account for power outputs of 100 to 150 MW, while peaks as high as 3 GW have been recorded experimentally.

For all shots on which more than 200 MW of peak power was measured, the applied field was less than 500G. At these low magnetic field levels, the electron motion is no longer constrained to one dimension and the system may allow multi-dimensional motion and bunching effects which would influence the production of gigawatt peak power microwave pulses. To better understand and exploit these effects, a fully relativistic, time dependent, $2\frac{1}{2}$ - dimensional superparticle simulation code is being prepared and should be operational by mid 1980. It is hoped that a more sophisticated model of triode operation will be constructed that will predict the production of higher peak and average output powers with a substantial increase in efficiency (presently 5%). Already, 30% efficiency at gigawatt power levels in S-band has been achieved with a triode at the Tomsk Polytechnical Institute in the Soviet Union.[8]

Near term experiments will attempt to operate without a physical anode foil, to match the impedance of the triode to the electron beam generator, and to carefully map the dependence of the microwave radiation on the applied magnetic field. Long term goals will concentrate on using the results of our simulation effort to improve our understanding of the reflex triode and to increase its radiation output.

Table I lists the values of various parameters associated with the reflex triode and allows comparison with values from two other experimental devices that have produced gigawatt power levels of microwave radiation. The magnetron is an attractive device, but the small size required for microwave operation limits the voltage and the current that can be applied and therefore the power that can be extracted. The pulsed Gyrotron is a large, inefficient experiment that requires (in its present form) a huge electron beam generator which would prohibit practical use on the battlefield. However, the reflex triode requires only a modest electron beam source, and optimization should be able to increase both its output and its efficiency.

Table I
Parameters for Ultrahigh Power Microwave Generating Experiments

| Parameter | Magnetron[9,10] | Gyrotron[11] | Reflex Triode[6] |
|---|---|---|---|
| Accelerating potential | 360 kV | 3.3 MV | 1 MV |
| Beam current | 12 kA | 80 kA | 20 kA |
| Pulse length | 30 nsec | 70 nsec | 20 ns |
| Axial field | 8 kG | 10 kG | 75 G |
| Power out | 1 GW | 1 GW | 1 GW |
| Frequency | S-band | X-band | X-band |
| Efficiency | 35% | 0.4% | 5% |

To place a device like the reflex triode experiment in a proper practical perspective, it is necessary to consider its actual use on the battlefield. Historically, one of the weakest links preventing exploitation of these ultrahigh power microwave sources has been the electron beam generators. However, capabilities for producing intense relativistic electron beams have also increased dramatically over the last several years.[12,13] Although most of the industry's attention has been devoted to the superhigh power generators like the (10 MV, 1 MA) AURORA and the (1.5 MV, 4.5 MA) Proto II facilities, a great deal of work has been done on improving reliability, reproducibility, and the repetition rate of the smaller (0.25 MV, 5 to 30 kA) machines.

At the Sandia Laboratories, an electron beam machine has been designed[14] to deliver 350 kV, 30 kA, 30 nsec pulses at a continuous rate of 100 per second. The machine has actually operated in this

mode for several minutes without suffering any type of breakdown. Improved versions of this machine have delivered voltages as high as 1 MV to diodes on a sustained repetition rate basis. Any of the Sandia repetition rate machines would easily fit on a flatbed trailer, and with some minor component redesign, they could probably be squeezed onto a pickup truck. In short, these machines are on the verge of becoming exceptionally transportable. With such electron beam generators already being developed, practical use of the ultrahigh power microwave generation schemes is already becoming possible.

There are four major areas where microwaves find use on the battlefield: communications, radars, electronic warfare, and directed energy weapons.

In communications applications, high power means longer ranges, better signal to noise ratios, and immunity to jamming and interference. Practical monopulse radars could be developed to cover problem-situations in which the transmitter must remain hidden or the target must be acquired rapidly and there is no time to process multiple pulses. These sources are useful not only as jammers, but also as simulators of enemy jamming equipment. Finally, the outputs of these devices are reaching levels where they can be considered as directed energy weapons.

## Conclusion

Although several laboratory experiments have demonstrated the capability of generating gigawatt power levels of microwave radiation, the Army's work on the reflex triode has established it as the clear front-runner in the race to put practical ultrahigh power microwave sources on the battlefield in the 1980's. We should be ready to incorporate these new sources into our defense technology, and we should be prepared for the problems that these devices can cause when they are used against us.

## References

1. D. A. Hammer et al, Microwave Production with Intense Relativistic Electron Beams, Annals of the New York Academy of Sciences, Vol. 251, 1975.

2. S. Humphries, Jr., J. J. Lee, & R. N. Sudan, Advances in the Efficient Generation of Intense Pulsed Proton Beams, Laboratory of Plasma Studies, Cornell University, LPS154, Aug 1974.

3. S. Humphries, Jr., R. N. Sudan, & W. C. Condit, Jr., The Production of Intense Megavolt Ion Beams with a Vacuum Reflex Discharge, Laboratory of Plasma Studies, Cornell University, LPS 161, Jan 1975.

4. H. E. Brandt, A. Bromborsky, H. B. Bruns, & R. A. Kehs, Microwave Generation in the Reflex Triode, Proc. of the 2nd Int. Topical Conf. on High Power Electron & Ion Beam Research & Technology, Oct 1977.

5. H. Barkhausen & K. Kurz, Shortest Waves Obtainable with Valve Generators, Phys. Zeit., Vol. 21, Jan 1920.

6. H. E. Brandt, A. Bromborsky, H. B. Bruns, R. A. Kehs, & G. P. Lasche, Gigawatt Microwave Emission from a Relativistic Reflex Triode, Harry Diamond Laboratories, HDL-TR-1917, 1980.

7. C. K. Birdsall & W. B. Bridges, Electron Dynamics of Diode Regions, Academic Press, New York, 1966.

8. A. N. Didenko, Y. Y. Krasik, S. F. Perelygin, G. P. Fomenko, Pis'ma Sh Teknichesky Fiziki, Vol. 5, No. 6, 1979.

9. G. Bekefi & T. J. Orzechowski, Giant Microwave Bursts Emitted from a Field-Emission, Relativistic-Electron-Beam Magnetron, Phys. Rev. Letters, Vol. 37, No. 6, 9 Aug 1976.

10. A. Palevsky & G. Bekefi, Microwave Emission from Pulsed Relativistic E-Beam Diodes. II. The Multiresonator Magnetron, Phys. Fluids, Vol. 22 (S), May 1979.

11. V. L. Granatstein, M. Herndon, P. Sprangle, Y. Carmel, & J. A. Nation, Gigawatt Microwave Emission from an Intense Relativistic Electron Beam, Plasma Physics, Vol. 17, 1975.

12. H. H. Fleischmann, High-Current Electron Beams, Physics Today, May, 1975.

13. J. A. Nation, High Power Electron & Ion Beam Generation, Particle Accelerators, Vol. 10, No. 1, 1979.

14. G. J. Rohwein, M. T. Buttram & K. R. Prestwich, Design and Development of a 350 kV, 100 pps Electron Beam Accelerator, Proc. of the 2nd Int. Topical Conf. on High Power Electron & Ion Beam Research & Technology, Oct 1977.

*KELSO & MARLEY

# NETWORK TRAFFIC ANALYSIS MODEL (U)

*DAVID H. KELSO, MR.
US ARMY ELECTRONIC PROVING GROUND
FT HUACHUCA, ARIZONA 85613

JOHN J. MARLEY, MR.
HQ US ARMY TEST AND EVALUATION COMMAND
ABERDEEN PROVING GROUND, MD 21005

## I. INTRODUCTION

The Network Traffic Analysis Model is a fully documented time and event oriented dynamic model which simulates the actions of individual equipment at the macro decision level. All multichannel related equipment in a communications system and their interactions are simulated so as to arrive at an analysis tailored to each individual problem.

The basic program is titled the "Simulation Model for Interference Analysis of Nodal Systems" (SIMIANS). SIMIANS consists of many programs which simulate a "black box" or group of "black boxes" for each general type of equipment, (for example, a message switch).(1) The black boxes are defined with specific equipment characteristics, such as delays, capacities, and other parameters. It is in this manner that specific equipment are simulated, (such as an AN/TYC-39 using step 2 software[1]).

[1]SIMIANS does not emulate software when modeling a communications system. The software and hardware of a system is assumed to work as intended by the designer unless empirical data is avilable.



PRECEDING PAGE BLANK-NOT FILMED

Simulated tactical equipment is modeled in a simulated tactical scenario which includes both friendly and enemy communication-electronic emitters. The deployment currently used is a Scenario-Oriented Recurring Evaluation System (SCORES).(2) With a SCORES simulated deployment, SIMIANS can simulate communications systems up to and including the size normally used for the support of a US Army Corps actively engaged with the enemy. This simulation includes the effects of electromagnetic compatibility and vulnerability.

Traffic Loading of the Network Traffic Analysis Model is accomplished by modifying data from the Communications Support Requirements Studies (COMSR) supplied to the US Army Electronic Proving Ground (USAEPG) by the US Army Signal School.(3) COMSR data (which is based upon input from all US Army Service Schools) contains the average times per day and lengths of conversations and messages between any two or more individuals in a deployment, as well as other relevant data such as precedence, and information content. The times and lengths are then modified by statistical processes to provide unique traffic (calls & messages) between individual subscribers in the simulation.

## II. SIMULATION METHODOLOGY

### a. General

SIMIANS is a dynamic discrete time event oriented model written in an extension of FORTRAN called "the Simulation Language for the Analysis of Communications Systems," (SLACS).(4,5,6) SLACS, (which has been designed to provide efficient, simplified coding for dynamic simulation modeling an extensive electromagnetic environment), simulates dynamic interaction using a general methodology similar to GPSS and SIMSCRIPT II.5 in that it utilizes external and internal event files as well as a driver program.(7,8) (See Figure.) The SIMIANS program, written in SLACS, diverges from other tactical communication simulations in the level of detail, the size of the modeling effort, and the consideration of the effects of the electromagnetic environment.(9,10,11,12,13)

### b. Creation of External Events

External events are stored in a computer file before the simulation begins. The computer file is then used to drive the simulation by providing distinct calls and messages with their respective phone numbers and routing indicators as input to the

simulation. These telephone calls and messages are derived from real-time tactical communications requirements which in turn have been derived from COMSR needlines[2] to be simulated by the model. In addition to communication requirements, the file may contain event data on attritions, failures, and repairs which have been derived from DA statistics. An individual external event identifies the time, origin, destination of call or message, precedence, information content, and length of the call or message.

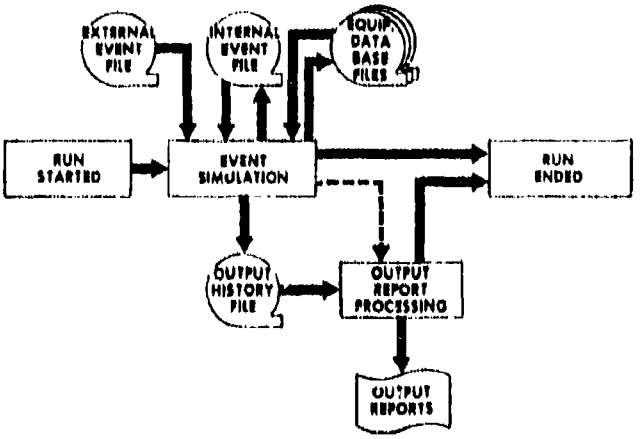



Figure. SLACS Simulation Methodology.

COMSR data contains the message or call originator and destination, the average number of communications per 24-hour period, and the average message length as well as the other information pertaining to messages and calls discussed in the previous paragraph. SIMIANS assumes that the COMSR needline traffic can be represented by a non-homogeneous stochastic process. Needline data described by COMSR data is in terms of 24-hour intervals. SIMIANS considers each needline in terms of a nonuniform arrival rate throughout a given 24-hour period (described in more detail, next para). Each period is considered a unique increment statistically independent of other increments, in

[2]A needline is defined as the requirement of two individuals to transfer information for a specific purpose by a specific mode. In COMSR, each interunit COMSR record and intraunit COMSR record is a needline.

that the calls or messages generated during one 24-hour period are not influenced in any manner by traffic generated in a previous period. Arrival rates as defined by COMSR remain constant, and will remain so in the generation of external events with the exception of a change of combat posture to an other than normal condition. The arrival rate is defined as the frequency with which a message or call is attempted.

It has been assumed in the writing of SIMIANS, that the average number of calls or messages per 24-hour period reflected in COMSR data has been derived from traffic which follows a Poisson Distribution. The traffic is then assumed to be uniformly placed throughout a 24-hour period at the average rate supplied by COMSR. A random number is generated between zero and one, and converted to a time scale between 0.00 and 86,400.00 seconds (86,400 sec = 1 day). Thus, the unique time of each call and message attempted is derived.

Next the length of each call and message is defined. Each message and telephone call simulated in SIMIANS assumes that a telephone call or message can be broken into two parts, an identification, or preamble section of fixed length and the text which is of variable length. The variable length is assumed to be exponential defined with a mean (u) minus the length of the fixed portion 1. (The value u+1 is equal to the average length found in COMSR.) A random number from 0.00 to 1.00 is used to vary the length of the text and thus, the entire message length. The message length is determined in SIMIANS by the following equation:

$$L = 1 - (u-1)\ln R$$

Where L = message length, 1 = fixed portion of length, u = the mean of the text length, R = the random number generated from 0.00 to 1.00.

The calculation of unique messages and calls is performed for each unit in the simulated tactical deployment. Then, using Army doctrine, telephone numbers and message routing indicators are assigned to each telephone call or message.

Although not part of the current effort, equipment-oriented events can be calculated in a manner similar to traffic events. Instead of using COMSR data as the basis for calculating equipment events, DA statistics, studies, test results, and equipment specifications are used. Thus, mean time between failure data is used to calculate the time of equipment failures and the mean downtime is used to calculate unique lengths of equipment downtime. In addition,

attrition due to the simulated combat can be injected either on a one to one basis or on a statistical basis dependent upon a unit's combat posture.(14) Equipment events are calculated for each communication device in the deployment. Then each equipment event is associated with the pertinent routing indicator(s) or telephone numbers they would affect.

Once the equipment oriented events have been produced, they are combined with the traffic events and then sorted by time. This time ordered sequence of events is then loaded into the external event file.

c. Details of Simulation

All known and future multichannel communications can be essentially broken down into three main categories, common user voice traffic, dedicated circuit traffic, and message traffic. SIMIANS handles each of these three systems as independent systems sharing the same and/or different radio-equipment cable link resources. Should a system be designed where one or more of these categories are no longer independent, SIMIANS will be modified.

(1) Common User Voice Traffic Simulation. SIMIANS breaks the common user telephone call down into a series of events at which delays and decisions are calculated. The simulation process for each call starts when the telephone is lifted up and placed in an off-hook condition (at a time determined by an external event). The status of equipment is then checked, and depending upon circuit wiring, a delay is incurred in the dialing of the phone or ringing of the operator. The equipment added (switch or switchboard) has its status checked. If the status is operational, the call is connected to the first switch or switchboard. Here at the switchboard, delays are incurred for the routing and alt-routing of the call, (the decision as to the routing and alt-routing of the call is based solely upon information available at that piece of equipment), decision as to preemption, and the addition of cryptographic equipment if the call is secure. This process continues until the telephone call either reaches its destination and/or fails due to excess dB loss, equipment nonavailability, preemption, or failure of the called party to answer. (Note: The telephone call may fail for any of these reasons after it reaches its destination before it is completed.) If the call is interrupted or never connected, a decision is made whether or not the call will be reinitiated based upon a probability dependent upon the failed call's precedence times a random number from 0.00 to 1.00, drawn from a

uniform distribution. The time of reinitiation is generated statistically, based upon the call precedence, by generating an exponentially distributed random number with a mean value equal to the average delay assumed for the precedence of that call. The length of the reinitiated call will remain the same as the original if never connected or will be calculated as follows if it was interrupted before completion. First, the percentage of call lengths remaining is calculated. Then the time to reinitiate the conversation is the time to identify oneself and re-start the discussion, added to the unfinished length. This reinitiation time is based upon 10% of the original telephone conversation. For example, a telephone conversation is interrupted when it is 30% complete; 70% remains to be completed. Thus, the new call length is (0.7 + 0.1) L, where L is the length of the original call. In the manual switching system (and possibly some automatic switches) it is possible that the initiator of a high precedence telephone call may desire to wait at a particular switching node for a circuit to become available. SIMIANS evaluates this possibility by assuming that the higher the call precedence and the more nodes already connected, the more willing the caller will be to hold rather than to place the call at a later time. However, the longer the caller has been on hold, the less likely he is to be willing to remain on hold. The formula used by SIMIANS in calculating this probability is:

$$P(\text{hold}) = \frac{((15P) + (5N))}{100} \text{ where } 0 \leq P(\text{hold}) \leq 1.$$

P = call precedence
1 = Routine
2 = Priority
3 = Immediate
4 = Flash
5 = Flash Override
N = Numer of nodes already connected.

If the caller had been on hold previously, the probability of remaining on hold is decreased by 0.1. This computed probability (P(hold)) is compared to a uniformly distributed random number between the interval of 0.00 and 1.00 to decide whether the caller will hang up or hold. (The decision to hold is made if P(hold) is less than the generated number.) If the caller hangs up, the probability of reinitiation is calculated as described previously.

In the manual system a process called ring-off occurs. Ring-off is the process which allows the caller to notify his local operator that he has completed his conversation. This notifies the operator that the caller's circuit is free. Should this not be done, depending upon operator availability, false indications as to trunk and circuit availability are used in the calculation of routings, preemptions, and alt-routings. Failure to comply with this procedure could thus noticeably affect the performance of a node or switchboard. SIMIANS calculates the probability that ring-off will occur at the end of all telephone calls and compares it with a random number from 0 to 1. If the number generated is less than the calculated probability, a ring-off will occur. Should a ring-off not occur, the availability of a circuit would be discovered in SIMIANS the same as it would be in reality.

Although much of the discussion above is addressed toward the manual switching process, the same basic decisions occur in the automatic switch. However, unlike the manual system each automatic switch has its decisions based upon its hardware and software rather than just hardware capacity. Delays and probabilities will have to be reevaluated for each type of automatic software/hardware combination, which may result in additional specific algorithms for a certain type of equipment. Thus, an average delay (T) for an AN/TTC-39 circuit switch will step 0 software is expected to be different when compared with an AN/TTC-39 circuit switch with step 2 software.

(2) Dedicated Circuit Traffic Simulation. Dedicated circuits unlike common user circuits are not connected through an active switching process. In a manual system, through circuits are connected only through a patching panel and are routed via appropriate communication links. In an automatic system the dedicated circuit call may be routed only through a patch panel, but more than likely will also be routed over a preprogrammed route by a circuit switch. This type of call is called an essential user bypass call. In both the automatic and manual systems, preemption, routing, busy circuits, and other related problems of the common user system do not impact upon the dedicated circuit. However, the automated system has direct access service (hotline) which does not exist in the manual system. SIMIANS treats this type of call as a pre-dialed common user call because normal precedence and preemption procedures are followed.

The dedicated circuit call simulation by SIMIANS starts when the phone is picked up. (This time like the common user is determined by an external event.) A check is made to determine if the circuit

is operational. If it is, a check then follows to determine if the circuit is "busy." (Note: Busy here refers to the probability of another person in the same office as the person originating the call already using the phone. For example, the S-3 is talking to a G-3 when the assistant S-3 desires to talk to the assistant G-3.) Should this "busy" condition exist, then it is assumed by SIMIANS that the call in question will immediately follow the call in progress by some small delay. The caller then signals the called party and, after a delay, the called party answers and the call continues until completion or until the circuit fails. If the dedicated circuit fails, the conversation will be reiniated over the common user system in the same manner as a reinitiated common user call.

(3) Message Circuit Traffic Simulation. The essential differences between the telephone system and message system are, for the purposes at hand, that the former may require a number of links all in use simultneously, and that call preemption and alt-routing can occur frequently; the latter may involve a number of one-link circuits used in sequence and queing is used to resolve most circuit contention conflicts.

The major difference between message and voice traffic is that many messages contain multiple addresses, thus, one external event may initiate many different addressed messages of different precedences over different circuits.

An external event starts the simulation of a message. (SIMIANS assumes that the message is ready for transmission at this time, i.e., already typed, etc.) The operational status of the circuit to the next node is checked. If the circuit is not available, then the message is placed in a queue. The message is then sent to the next node with an appropriate transmission delay added. If the next node is an automatic switch, the format is checked for correctness. If it is in error, then the message is re-initiated. (The probability of error is determined differently for each item of equipment.) This process continues until the message is received by the address node when an acknowledgement is sent to the sender (if required). When messages are awaiting transmission, a que will build up in SIMIANS, based upon order of precedence and first in/first out. Although the probability of human error is determined on an individual basis (if determined at all), the probability of bit errors being injected by the transmission media is very real. (This injection of bit error rates is addressed by SIMIANS as described in paragraph IIc(4).)

The above discussion is pertinent to both automatic and manual message switching, however, manual switching delays can be readily defined where the automatic delay is dependent upon the software/hardware combination. Thus, a particular hardware/software combination may have more or less macro decisions resulting in different delays than the manual system. SIMIANS will address this difference by the addition and modification of algorithms as necessary.

(4) Simulation of Electromagnetic Effects. The simulation of electromagnetic effects within SIMIANS is based upon Environmental Interference Effects Model, which itself is a combination of programs used to predict the electromagnetic compatibility and vulnerability of communications-electronics systems operating in a tactical environment.(15)

At the run initialization, SIMIANS evaluates all radio-links of interest, their present status and status change times are calculated. This process is repeated every 0.1 second (real time). Link status is recalculated for all lines when jamming events occur. A determination of jamming effectiveness is made as a function of jammer azimuth and frequency when it is turned on.

SIMIANS performs this analysis as follows: Based upon the assigned frequency or frequency band of each link transmitter, a sort is made of all potential interference generators to determine which of these can possibly cause interference on each link based upon frequency compatibility. Those interference generators with compatible frequencies are further scanned to determine which of them are close enough to the link to interfere with the signal of interest. For each potential interferer a detailed test is made to determine the signal to interference ratio (S/I) when the potential interferer is emitting. (Signal strengths are calculated based upon the Longley-Rice propagation model as modified by the Electromagnetic Environmental Test Facility.(16,17)) The list of links interfered with for a given jammer are compiled in a computer file for that jammer. Thus, when that jammer is turned on, appropriate internal events are generated, (internal events are simulated events generated as a result of other internal and external events). The turn-on cycle or duty cycle of a jammer is supplied as input to the model which is determined from available threat data. Once the S/I ratio has been calculated, the bit error rate and other rates are calculated based upon test data which has been inputted into the model. This data is then used to determine the effects of electromagnetic emitters on the system.

d. Output from the Simulation

Details of each event process will be saved in an output history file. The output history file contains a complete record of each call and message processed. However, for most purposes, such bulk data is useless, therefore, several reports have been devised.(18) (Others are possible and may be added as part of specific projects.) The first report available is a call status report which will provide a detailed listing of each unique communication message that was attempted by the model, for each completed call the average delay is computed. The second report is a call failure report which describes failed attempts, consisting of a detailed listing and a statistical summary by time period. A third report provides a statistical summary by time received and a distribution of call attempts throughout a given block of time. A fourth, and the last currently available report, is the delay summary. It is the accumulation of delays as determined for each of the $\Delta t$'s accumulated. These delays reflect the real-world delay patterns of network operations.

In addition to the reports produced after the simulation, reports can be generated by external events during the simulation to provide such data as the current use of all internodal channels in the system, or the current contents of all store and forward queues.

## III. VALIDATION AND VERIFICATION

The SIMIANS program is continually being assessed to determine how well it faithfully represents actual tactical communications equipment working as a system.(19,20) This is being done four different ways; using sensitivity analysis, comparison of model output with historical records, comparison of model output with data collected explicitly for the validation, and finally, from the comparison of model output with the experiences of experienced tactical Army communicators.

The sensitivity analyses are conducted to determine how much the model output varied with small changes in delay parameters. When a parameter can be changed considerably without any noticeable change in output, its probability based calculation is replaced by a constant. Where the model has proved to be sensitive to a parameter, the most accurate data is used, and the method of delay calculation reexamined to insure that the calculation was correct.

The history of past field exercises by tactical Army units when available is compared to the output of SIMIANS when possible. Where large differences occur, the model is examined to determine if a basic process is correctly modeled.

Since the accuracy of data from past exercises is not known, it was decided to conduct a simple field exercise using three manual switchboards to compare the model's design logic with that of the real world. This was done by comparing the number of telephone calls attempted, completed, lost, interrupted, and preempted. No attempt was made at this time to obtain information or call delays, due to the time and cost involved.

In addition, all of the model's outputs are reviewed by experienced Army tactical communicators to determine if correct procedures and expected results are obtained. When the model deviates from experience, the cause of this difference is extensively investigated. When the fault is with the model, the logic is corrected.

## IV. CONCLUSIONS

The Network Traffic Analysis program, SIMIANS, can be used to compare magnitudes of delays for different equipments and networks to determine a doctrine of optimal efficiency and stability through a series of trade-off sensitivity studies. This type of analysis can be used to determine how to best deploy a type of equipment in a given theater by using a simulated tactical deployment and various levels of expected traffic loading. Additionally, the results of trade-off sensitivity studies may be used as input to cost effectiveness studies.

Although SIMIANS is extremely valuable during the various phases of equipment developmental testing, the use of SIMIANS could be of as great or greater value in determining the efficiency of proposed equipment specifications at the beginning of the developmental life cycle.

## REFERENCES

1. Computer Program Specification Document for the Simulation Model for Interference Analysis of Nodal Systems (SIMIANS) Network Traffic Analysis, US Army Electronic Proving Ground, Ft. Huachuca, AZ, Publication No. USAEPG-SR-1062, December 1979. (UNCLASSIFIED)

2. Conceptual Approach for the Analysis of Distributed Jamming Systems, Appendix B, US Army Electronic Proving Ground, Ft. Huachuca, AZ, January 1980. (UNCLASSIFIED)

3. Revised User Instruction Booklet, Communications Support Requirements (COMSR) Data Base, COMSR Technical Operations Element (CTOE), US Army Signal School, Fort Gordon, GA, 1 April 1977. (UNCLASSIFIED)

4. Computer Specification Program Document for the Simulation Language for the Analysis of Communications System (SLACS) Precompiler, Network Traffic Analysis Model, US Army Electronic Proving Ground, Ft. Huachuca, AZ, Publication No. USAEPG-SR-1062(1), April 1979. (UNCLASSIFIED)

5. Simulation Language for the Analysis of Communications Systems (SLACS) User's Guide, Network Traffic Analysis Model, US Army Electronic Proving Ground, Ft. Huachuca, AZ, Publication No. USAEPG-SR-1062(2), April 1979. (UNCLASSIFIED)

6. Simulation Language for the Analysis of Communications Systems (SLACS) Implementation Manual, Network Traffic Analysis Model, US Army Electronic Proving Ground, Ft. Huachuca, AZ, Publication No. USAEPG-SR-1062(3), August 1979. (UNCLASSIFIED)

7. Schriber, Thomas J. "Simulation Using GPSS," John Wiley & Sons, New York, NY, 1974.

8. Kiviat, Villanueva, and Markowitz, SIMSCRIPT II.5 Programming Language, Consolidated Analysis Centers, Inc., Arlington, VA, October 1975.

9. COMMEL II Users Manual, Volume I - Model Overview, Evaluation Technologies, Inc., Arlington, VA, Report No. CAA-D-76-6-VOL I, October 1976. (UNCLASSIFIED)

10. Conceptual Design for the Army in the Field Alternative Force Evaluation CONAF Evaluation Model IV Part I - Model Description, General Research Corporation, McLean, VA, Report No. OAD-CR-60, September 1974. (UNCLASSIFIED)

11. Conceptual Design for the Army in the Field Alternative Force Evaluation CONAF Evaluation Model IV Part III - Users Handbook, General Research Corporation, McLean, VA, Report No. OAD-CR-60, December 1974. (UNCLASSIFIED)

12. Theater Force Evaluation by Combat Simulation, Vol I - Process Descriptions, Vector Research, Inc., Ann Arbor, MI, Report No. VRI-CAA-2-FR-77-1 Volume I, 30 June 1977. (UNCLASSIFIED)

13. Theater Force Evaluation by Combat Simulation, Volume II - Methodology, Vector Research, Inc., Ann Arbor, MI, Report No. VRI-CAA-2-FR-77-1 Volume II, 30 June 1977. (UNCLASSIFIED)

14. Methodology Investigation Final Report: Investigation of Traffic Analysis Models, US Army Electronic Proving Ground, Ft. Huachuca, AZ, Publication No. USAEPG-FR-1057, July 1978. (UNCLASSIFIED)

15. The Environmental Interference Effects Model of the Electromagnetic Environmental Test Facility, Volume II - Theory, US Army Electronic Proving Ground, Ft. Huachuca, AZ, Publication No. USAEPG-FR-1092, October 1978. (UNCLASSIFIED)

16. Longley, A. G. and P. L. Rice, "Prediction of Tropospheric Radio Transmission Loss Over Irregular Terrain--a Computer Method--1968," ESSA Technical Report ERL 79-ITS 67, July 1968 (Revised September 1972). (UNCLASSIFIED)

17. Detailed Capabilities of the Electromagnetic Environmental Test Facility, US Army Electronic Proving Ground, Ft. Huachuca, AZ, January 1980. (UNCLASSIFIED)

18. Specification Document Network Traffic Analysis History Report, US Army Electronic Proving Ground, Ft. Huachuca, AZ, Publication No. USAEPG-SR-1062(7), March 1980 (DRAFT). (UNCLASSIFIED)

19. Validation Concept for the Interference Analysis of Nodal Systems (SIMIANS), Network Traffic Analysis (NTA) Model, US Army Electronic Proving Ground, Ft. Huachuca, AZ, March 1980 (DRAFT). (UNCLASSIFIED)

*KELSO & MARLEY

20. Law, Averall M., Validation of Simulation Models I: An Overview and Survey of Real Practice, University of Wisconsin-Madison, August 1979. (UNCLASSIFIED)

## ACKNOWLEDGMENT

Credit for the technical work in this program which extends over a three year period, belongs to a significant number of Bell Technical Operations, TEXTRON, and Government professionals working at the US Army Electronic Proving Ground's Electromagnetic Environmental Test Facility. In addition to the authors (Government Employees), significant contribution to the NTA effort has been provided by: Mr. D. Chandler, Mr. M. McNamee, Mr. H. Stevenson, Dr. T. Tooman, Mr. K. Weise, Mr. E. Weiss, and Mr. T. Wooley. Our deep appreciation goes to them and others not mentioned, without whose effort, the Network Traffic Analysis Model would not exist.

# ELECTRON MICROSCOPY USED FOR FRACTURE MODE IDENTIFICATION (U)

*CHARLES R. KLARICH, R. L. HUDDLESTON, AND M. J. DRABO
Materiel Testing Directorate
Aberdeen Proving Ground, Maryland 21005

## INTRODUCTION

Metallurgical failure analysis represents a technology which combines the knowledge of many scientific and engineering disciplines. During these times of increasing product liability litigation, interest in failure analysis has intensified. The failure analyst is asked to identify the cause of catastrophic failures in many different materials. During past years at the proving ground, the list of materiel evaluated includes such diverse items as a cargo truck firewall and 175 mm gun tubes. The disciplines required include, metallography, chemical analysis, light and electron microscopy, stress analysis, and the measurement of mechanical properties.

Effective failure analysis depends on a combination of factors; these include the assessment of prior manufacturing and service history, a knowledge of materials and available test methodology and finally, how these factors interrelate with the applied loads and design intent (1). Conventional failure analysis begins with a visual examination of the item to evaluate the material response to fracture. The analyst uses all the disciplines outlined above to focus the investigation on the cause of failure. A powerful tool in the analysis is the Scanning Electron Microscope (SEM) and the discipline called fractography.

Fractography then, may be described as the study of a fracture surface to determine the micromechanisms of fracture. This information can help to determine the cause of the failure. Visual examination of metal fractures goes back to the sixteenth century.

An optical microscope was first used in the examination of metallic fractures by Reaumer in 1722 (2), but the first detailed use of optical fractography was done years later by Zappfe and Clogg in 1945 (3). They studied numerous fractures made in steels and characterized their appearance. The introduction of commercial Transmission Electron Microscopes (TEM) in the 1960's accelerated and development of fractography. The actual fracture surface cannot be viewed with a TEM because the electron beam must pass through the sample, but replication techniques were devised. Sample preparation was difficult and artifacts on the fracture surface were a problem. The SEM is an instrument that essentially bridges the gap between the optical microscope and the TEM. Today fracture surfacescan be viewed directly after the proper cleaning techniques are used by the microscopist. A second advantage is the depth of focus obtained with the SEM compared to optical techniques. A rough fracture surface will remain in focus at magnifications of 5000X.

This paper describes the analysis of two barrel support jacket failures from the M2 machine gun analyzed at the Materiel Testing Directorate, Aberdeen Proving Ground (APG), utilizing the SEM. Electron microscopy is a diagnostic tool primarily used to view fracture topography and identify fracture mode. The fracture modes are classified primarily as ductile or brittle plus fatigue, and intergranular variations. The following examples illustrate ductile, brittle, intergranular and fatigue failure mechanisms.

Ductile failure is characterized by the formation of voids on the fracture surface that plastically coalesce to form dimples. This micromechanism is termed microvoid coalescence. The fractograph of a tensile sample machined from a plain carbon AISI 1018 steel shows this type of failure. The size of the dimples is determined by number and location of the microvoids that are nucleated. The nucleation sites are inclusions, second phase particles, or grain boundaries.



Figure 1. A Fractograph of a 1018 Plain Carbon Steel Tensile Bar Which Shows Ductile-Dimpled Rupture. Magn: 2400X

An example of a brittle failure is the fracture surface observed on a sintered tungsten material (4). The slug was fabricated from a tungsten powder that was liquid sintered with a nickel-iron binder. The material was then swagged 25%. A slug of this material was fractured with a stress wave and then viewed with the SEM.





Figures 2 and 3. A Fractograph of the Liquid Sintered Tungsten Material Which Shows Cleaved Tungsten Particles and Round or Oblong Flat Facets in a Ductile-Iron Matrix. Magn: 1400X and 3500X

Electron fractography showed the fracture mode was brittle. Three different micromechanisms were observed in the material. The tungsten particles failed by cleavage and intergranular fracture. Cleavage steps were observed on numerous cleavage planes. The flat oblong or round areas were regions of intergranular rupture. These facets can result from air being trapped in the liquid sintered material. The third type of micromechanism was the ductile failure of the nickel-iron matrix.

A brittle fracture with an intergranular mechanism was noted on a spider universal joint from the Infantry Fighting Vehicle (IFV). The U-joint was case carburized and was in service for a total of 474 miles when failure occurred (Fig 4).

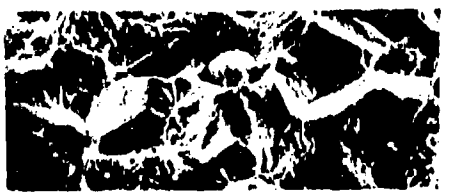


Figure 4. Intergranular Fracture, U-Joint. Magn: 700X

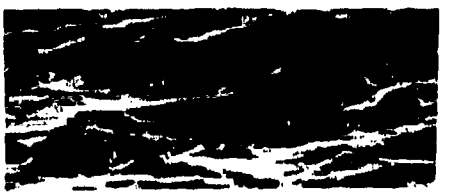


Figure 5. Fatigue Striations Observed on Axle Shaft. Magn: 350X

Visual and low power optical examination did not reveal the intergranular fracture mechanism. The extent of the brittle cracking was a function of the case depth.

The next type of fracture to be illustrated is the fatigue failure. These failures are characterized by a striated fracture surface. The number of cyclic load applications are approximated from the striation spacing at different locations on the fracture surface. A recent example of a fatigue failure occurred during the initial production test (IPT) of a rough terrain forklift (5). Several rear axle shafts failed catastrophically during reliability testing. The axle shafts broke in the splined end from reverse torsional fatigue (see Figure 5).

The variety of material failures encountered at the proving ground range from alloy steels to sintered tungsten. Electron fractography is used to determine the micromechanisms on the fracture surface. This information is correlated with mechanical property information and the microstructure of the failed item to determine the cause of failure. Corrective action is then recommended. A typical example of the failure technique used at APG is shown in the analysis of the barrel support jacket failures. The analysis uses an integrated approach that evaluates background data, service history, and visual examination before employing other analytical methods to focus on the cause of the failure. The scanning electron microscope is the central core of the analysis.

## BACKGROUND - BARREL SUPPORT JACKET FAILURES

A Blank Firing Attachment (BFA) M19 was developed for the M2 Heavy Barrel Machine Gun (HBMG) Cal .50 weapon.



Figure 6. View of M2 Machine Gun with Blank Firing Attachment.

The BFA is used in tactical field training exercises as a Weapon Effects Signature Simulator (WESS) to support the real train engagement simulation system (6-8). The blank firing attachment provides an effects signature from the position of the firing weapon. The blank firing attachment caps the barrel and generates a recoil with the blanks. This activates the bolt mechanism to eject the spent blank and feeds an additional blank into the chamber. The M2

is designed to fire 450 rounds/minute with conventional ammunition. It operates at 700 rounds/minute with the blanks. More force is generated firing the blank ammunition and the barrel also has more vertical whip. The barrel support jackets each had 5-6000 rounds fired with the blank firing attachment in place. The total number of rounds on the barrel supports was not determinable. The machine guns are periodically rebuilt so the failed components were probably on several different weapons since manufacture. Two barrel support jackets failed during testing of blank firing attachment. The cracked barrel supports caused misalignment of the front barrel bushing in the M2 .50 cal HBMG. This misalignment caused the barrel to bind in the bushing thereby causing weapon stoppages.

## TEST TECHNIQUES

A. Electron Microscopy.

The barrel support jackets were visually examined and sectioned before sputter etching. The fracture surfaces were coated with black residue which was the result of a gas wash from the firing of the M2 machine gun. Repeated sputter etchings were required. The sputtering device provides etching-cleaning of samples by immersing the work in a charged plasma field. Organic material is removed as is some of the base material itself.

The fractographic examination of all samples was conducted with an Advanced Metals Research (AMR) Model 1200 Scanning Electron Microscope (SEM) with an accelerating voltage of 25 KV. Resolution of the instrument is 100 angstroms. Qualitative chemical analysis and X-ray mapping was done with an energy dispersive X-ray spectrometer attached to the SEM.

B. Microscopy.

A microstructural analysis was conducted on samples of each broken barrel support jacket with a metallograph. A sample used for microscopic examination was also examined with the SEM to chemically identify a substance on the fracture surface.

C. Hardness Measurements.

Brinell hardness, 500 Kgm load, was determined on each barrel support jacket near the failure.

D. Quantitative Chemical Analysis.

An X-ray spectrometer was used to determine the chemical analysis of the barrel support jackets. The carbon level was determined with a gasometric carbon analyzer.

## RESULTS

A. Visual Examination/Electron Microscopy.

BSJ labeled Number 1 had a through crack in a web section near the center of the part.

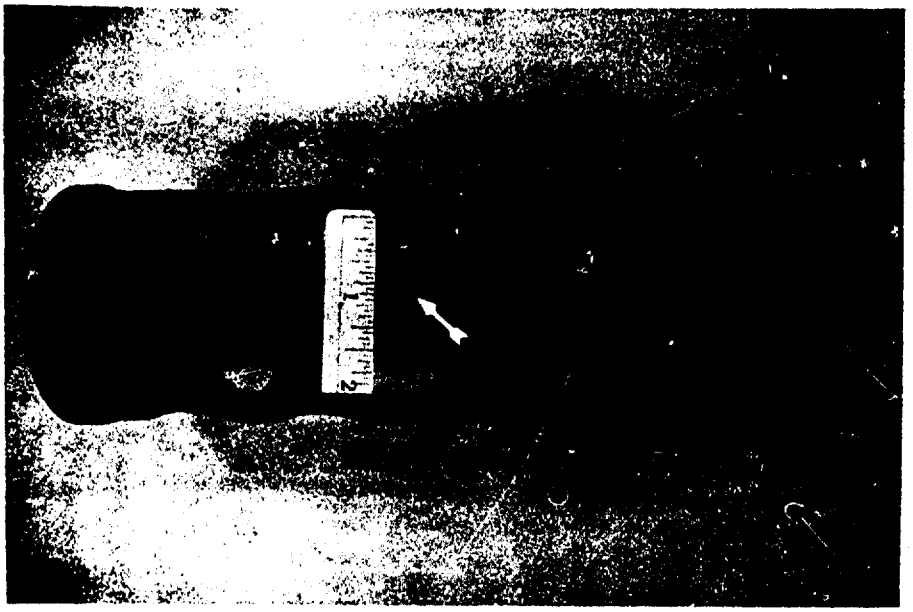


Figure 7. A View of BSJ Number 1 Which Shows the Location of the Through Crack.

The second BSJ had three sequential cracks in the webs at the forward end of the part. Each of these cracks was a through crack.

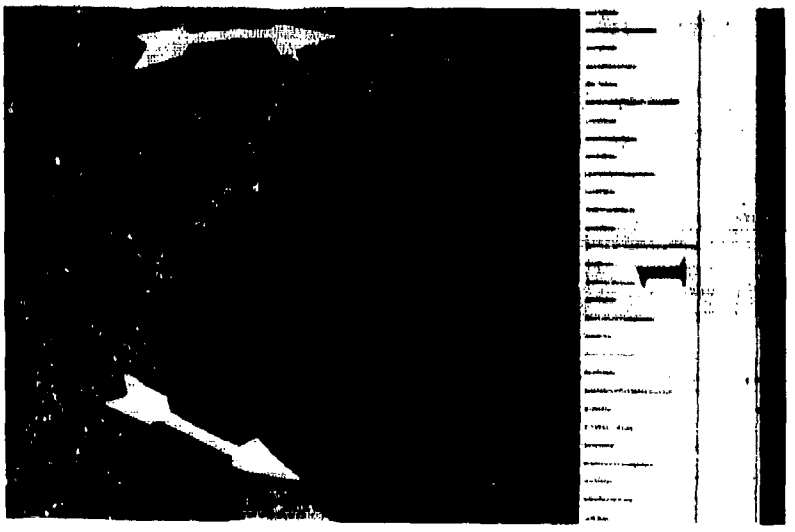


Figure 8. A Photograph Which Shows Two of the Cracked Webs in BSJ Number 2.

Visual examination of the fracture surfaces did not lead to a positive identification of the fracture mode. SEM analysis of the fracture surfaces was hampered by a black residue which resulted from the gas wash of the M2 machine gun. The fracture surface on BSJ Number 1 contained three distinct zones. The origin of failure on BSJ Number 1 was a flat zone on the fracture surface. Evidence of fatigue striations was obliterated by burnishing of the mating fracture surface. The remainder of the fracture surface visually appeared as an intergranular fracture. The periphery of the fracture surface on the side away from the flat zone contained a minute shear lip. Electron fractography showed the mating fracture surfaces had rubbed together and obliterated the fracture topography.



Figure 9. A Fractograph of BSJ Number 1 Which Shows the Fracture Surface Damage. The Fracture was not Intergranular. Magn: 800X

The fracture surfaces from BSJ Number 2 showed a ductile-dimpled fracture surface formed by microvoid coalescence. Manganese sulfide inclusions were visible in several of the dimples. This fracture topography is indicative of a tensile failure.

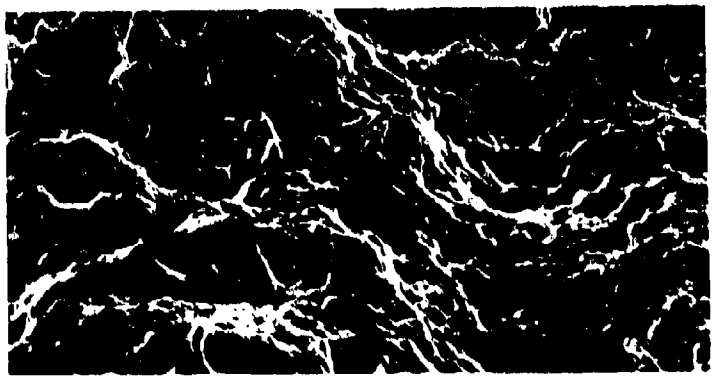


Figure 10. The Fracture Surface of BSJ Number 2 Showed a Ductile-Dimpled Fracture Surface Formed by Microvoid Coalescence. Magn: 1320X

B. Microstructural Analysis.

The microstructure of BSJ Number 1 at the fracture surface showed a non-metallic inclusion at the fracture surface and manganese sulfide inclusions throughout the matrix. The etched matrix material showed the BSJ Number 1 was made from a pearlitic malleable iron. The pearlitic matrix was partially spheroidized by a tempering treatment in the manufacturing sequence. The irregular nodules of temper carbon graphite are typical of a malleable iron. The graphite conforms to ASTM Type III temper carbon.

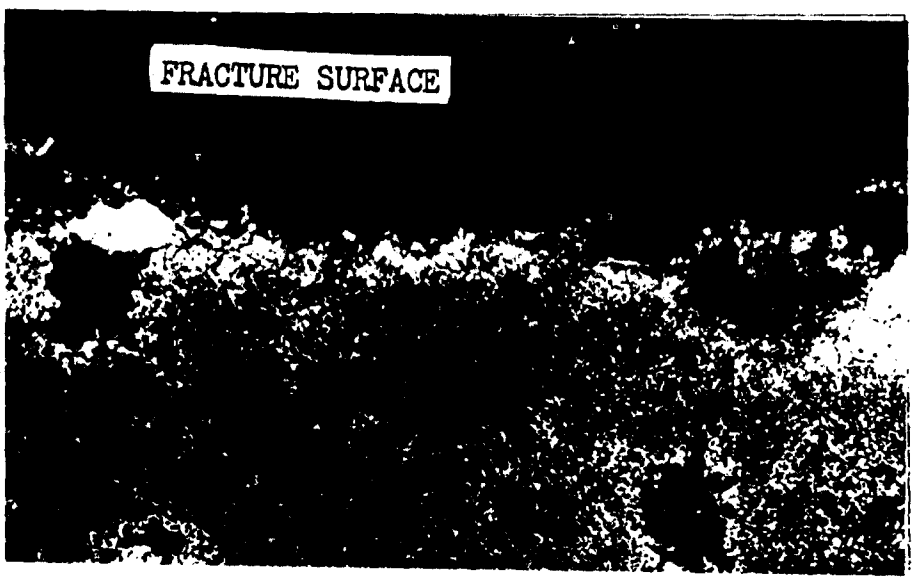



Figure 11. A Micrograph of BSJ Number 1 Which Shows a Pearlitic Malleable Iron Matrix. Note the Non-Metallic Inclusion on the Fracture Surface. Magn: 200X Etchant: 2% Nital

Note the area of ferrite adjacent to the inclusion on the fracture surface. This indicates the defect was present during the annealing of the material. The inclusion or the ferritic layer extended 87% across the length of the fracture surface. This caused a zone of weakness across the web that failed in BSJ Number 1. This area was the flat region viewed on the fracture surface. The white constituent around the temper carbon nodules is ferrite. The formation of a complete bull's-eye structure was minimized by a short annealing time, so the envelopes of free ferrite around the graphite nodules are not completely formed.

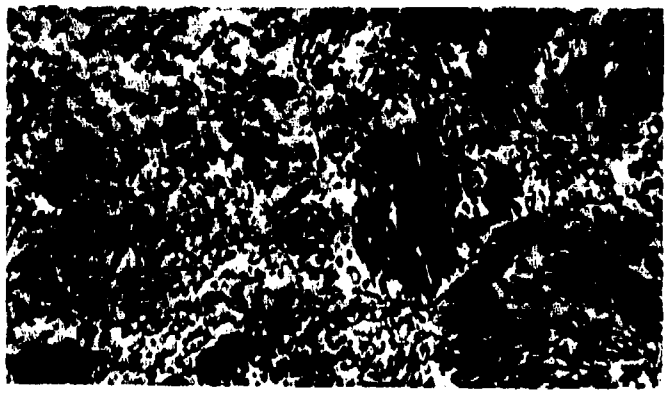


Figure 12. A Micrograph of the Matrix Material of BSJ Number 1 Shows a Pearlitic Structure that was Partially Speriodized by a Tempering Treatment. The Gray Particles are MNS Inclusions. Magn: 1400X Etchant: 2% Nital

Microstructural examination of BSJ Number 2 also showed it was pearlitic malleable iron. The pearlitic matrix was partially spheroidized by the tempering treatment. No large non-metallic inclusions were seen on the fracture surface of BSJ Number 2.

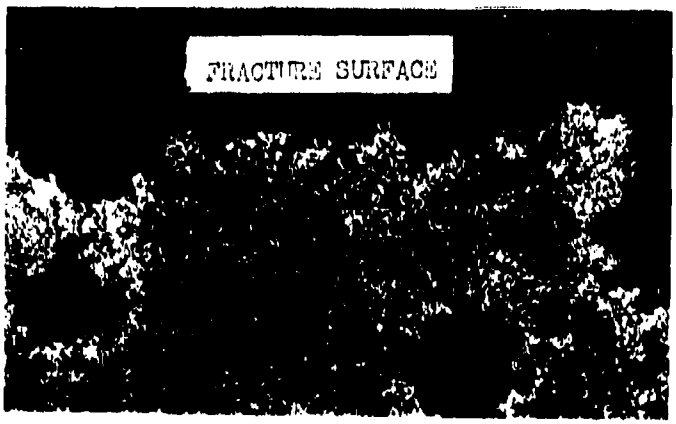



Figure 13. A Micrograph of the Fracture Surface on BSJ Number 2 Which Shows the Irregular Fracture Path Typical of a Ductile-Dimple Rupture. Magn: 200X Etchant: 2% Nital

The fracture path followed the low strength ferrite grains and temper carbon nodules across the cross-section of the web in BSJ Number 2.

C. Chemical Analysis.

The X-ray spectrochemical analysis of the barrel support jackets showed they were both a malleable iron material. The carbon and manganese content were typical for a pearlitic malleable iron. The silicon content was slightly higher than normal.

Table 1. Chemical Analysis of Barrel Support Jackets No. 1 and 2

| BSJ Part No. | % Element, Wt C | MN | P | S | Si | Ni | Cr | Mo | V | Cu |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2.46 | 0.41 | 0.07 | 0.14 | 1.78 | 0.14 | 0.04 | 0.01 | 0.03 | 0.08 |
| 2 | 2.45 | 0.43 | 0.07 | 0.15 | 1.82 | 0.14 | 0.04 | 0.01 | 0.03 | 0.08 |
| Typical Pearlitic Malleable | 2.00-2.70 | 0.25-1.25 | 0.05 max | 0.03-0.18 | 1.00-1.75 | - | - | - | - | - |

D. Elemental Mapping.

A sample that was used for microstructural analysis was also for elemental mapping. The area of interest was the inclusion seen on the fracture surface on BSJ Number 1.

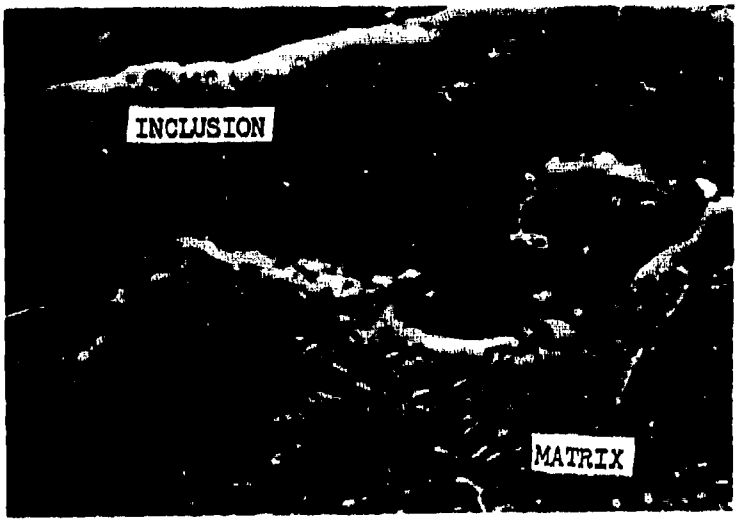



**Figure 14. A SEM View of the Inclusion at the Fracture Surface on BSJ Number 1. Magn: 1800X**

The platelets seen in the matrix are pearlite lamellae. These are alternate layers of ferrite and iron carbide. The globular particles in the matrix are manganese sulfide inclusions. Acquiring an X-ray energy dispersive spectrum from the fracture surface inclusion showed the elements Fe, Mn and Si were present. Elements on the periodic chart below atomic Number 13 (aluminum) are not detected. Elemental mapping was conducted for the elements Fe, Mn, and Si.

The iron concentration was distributed uniformly throughout the malleable iron and the inclusion. The iron concentration was the highest of the three elements mapped. Manganese was noted over the entire area that was mapped with the highest concentrations at the inclusions in the malleable iron. The silicon was concentrated near the fracture surface in the region examined.

E. Hardness.

The Brinell hardness (500 Kgm load) was taken on various sections of BSJ's 1 and 2 near the through cracks. The hardness of BSJ Number 2 was lower than the typical pearlitic malleable iron.

Table 2. Brinell hardness (500 Kgm Load) of BSJ's Labeled 1 and 2

| BSJ Part No. | Section I. D. | BHN (500 Kgm Load) |
|---|---|---|
| 1 | A | 158, 158, 150 |
| 1 | B | 158, 158, 158 |
| 2 | A | 130, 130, 140 |
| 2 | B | 124, 143, 136 |
| 2 | C | 143, 136 |
| 2 | D | 130, 143, 136 |
| 2 | F | 136, 150, 136 |
| 2 | E | 136, 143, 150 |

Note: The typical hardness range for a pearlitic malleable-iron grade 40010 ranges from BHN 149-197. Other grades of pearlitic malleable iron have higher hardness ranges.

## DISCUSSION

The fatigue process normally proceeds in two stages - initiation of the crack and crack propagation. Barrel support jacket Number 1 failed by fatigue in a web section thicker than the failed webs in BSJ Number 2. A casting defect weakened the cross-section and allowed a fatigue crack to initiate in the web. The layer of ferrite adjacent to the defect indicates that it was present during the annealing process. The inclusion was identified by the elemental mapping technique as an iron compound with Mn and Si in solid solution. The initiation of cracking at the inclusion most probably occurred during the initial use of the BSJ before the addition of the blank firing attachment. The rate of fatigue crack growth is controlled primarily by the intensity of loading which is described by the stress intensity factor (K). The crack growth rate was low during the initial use of the BSJ on the M2 HBMG, but addition of the blank firing attachment increased the stress intensity factor at the defect and allowed the crack to propagate to failure in less than 6000 cycles.

The stresses on the BSJ are increased when the M2 machine gun is operated with the blank firing attachment. The gun is operated at 700 rounds/minute compared to the design rate of 450 rounds/minute and has the extra weight of the attachment. The increased stress in combination with the thin web cross-section and lower than average hardness resulted in a tensile failure of BSJ Number 2. SEM

examination confirmed the tensile micromechanism of failure. The BSJ was also bent by the forces generated during testing the blank firing attachment.

The entire material failure analysis process is an integration of the bits of information gained from electron microscopy, microstructural analysis, service conditions, hardness testing, visual examination, and X-ray elemental mapping. Analysis of field failures is not as straight forward as the analysis of laboratory produced fractures. The discovery of a casting defect at the fracture surface of BSJ Number 1 by microstructural structure analysis pinpointed the origin of failure. This information plus the knowledge of the service conditions in combination with the other data lead to the determination the BSJ Number 1 failed in fatigue. The electron microscopy data was conclusive in the analysis of BSJ Number 2. These three webs failed in a tensile mode and no burnishing of the fracture surface was in evidence. The other information, expecially the hardness data, corroborates the SEM analysis. Electron microscopy is essential in the analysis of materiel field failures to reduce the uncertainty and also make positive identifications of the fracture mode.

## CONCLUSIONS

Barrel support jacket Number 1 failed in fatigue caused by an increased loading due to the blank firing attachment and a casting defect that weakened the web.

Barrel support jacket Number 2 failed in a ductile rupture mode caused by a combination of a higher than normal loading, a thin cross-sectional area and lower than normal hardness for a pearlitic malleable iron part. The micromechanism of the failure was microvoid coalescence.

## CORRECTIVE ACTION RECOMMENDED

An improved casting technique be used in the manufacture of the barrel support jackets.

Only thick webbed barrel support jackets be used on the M2 machine gun when the blank firing attachment is used.

The depot rebuild program for the M2 machine gun use only the thick web barrel support jackets.

Cracks not be allowed in the barrel support jackets used in the depot rebuild operation.

SUMMARY

Electron microscopy is essential in the analysis of materiel service failures to reduce the uncertainty and also to make positive identifications of the fracture modes. The examples given illustrate the way the SEM is used to characterize fracture topographies. The SEM is at the core of the analytical procedure which focuses diverse bits of information into determining the cause of failure. The analysis includes prior service history, manufacturing methods, mechanical properties, metallurgy, and the SEM. Each failure requires a specific combination of tests to solve the problem. The tests are complimentary and result in an intense examination of the critical areas of interest. The payoff is correct identification of the cause of failure and corrective action which results in improved Army materiel.

REFERENCES

(1) Drabo, M. J., "Analytical Processes for Failure Analysis", Minutes 1979 TECOM Instrumentation Conference, p. 271.

(2) Reaumer, "L'Art De Converter Le Fer En Acier", Paris 1722.

(3) Zappfe, C. A., and Clogg, M., "Fractography - A New Tool for Metallurgical Research", Trans ASM Vol 34 (1945) p. 71.

(4) Klarich, C. R., "Ballistic Technology - Fragmenting Warheads", Physical Test Branch Report 76-M-75, MTD, APG, Maryland.

(5) Faller, J. G., "IPT of 4000 Pound Rough Terrain Forklift", Physical Test Branch Report 79-M-55, MTD, APG, Maryland.

(6) Letter Requirement for Blank Firing Adaptor for Cal .50 Machine Gun M2 Between Commander, U.S. Army Training and Doctrine Command and Commander, U.S. Army Materiel Development and Readiness Command, 1976.

(7) Maule, M., "Development Test II (PQT-G) of Blank Firing Adaptor for M2 Machine Gun and Product Improvement Test of Cal .50 Blank Ammunition", Test Report APG-MT-5346, Aberdeen Proving Ground, Maryland.

(8) Klarich, C. R., "Metallurgical Analysis of Two Failed Barrel Support Jackets", Physical Test Branch Report 79-M-47, MTD, APG, Maryland.

*KUZANEK

# IMPROVED METHODS FOR COMPUTING DRAC CORRECTED MISSILE IMPACT PREDICTIONS IN REAL TIME (U)

*JERRY F. KUZANEK, Ph.D.
NATIONAL RANGE OPERATIONS DIRECTORATE
WHITE SANDS MISSILE RANGE, NM 88002

1. Introduction. During missile flight tests at White Sands Missile Range (WSMR), position data on the current status of the missile is transmitted 20 times per second from radar sites to a Univac 1108 computer. Consecutive pairs of such data are averaged 10 times per second and computations for plotting displays such as current position, range verses altitude, or impact prediction are based upon this averaged data.

In the event that a missile veers from its planned trajectory, it will be necessary to terminate thrust to prevent the missile from impacting in a populated area. For this reason, the Range Safety Officer (RSO) requires that for each computational cycle (10 per second) an instantaneous impact prediction (IIP) of the missile be computed. This point is the intersection of the missile trajectory, should thrust be terminated, with the Clarke Spheroid (of 1866) model of the Earth at an altitude of 4000 feet.

If the effects of atmospheric drag are neglected then a vacuum IIP can be rapidly computed using Kepler's central force equations, since the missile trajectory is an ellipse. This method is applicable for a variety of low drag vehicles whose exit from and reentry to the Earth's atmosphere occur at high angles. Since approximately one millisecond is required to compute a vacuum IIP, it is always computed at a 10 per second rate. Details of this computation are given in [1].

For low altitude missiles, whose trajectory is significantly affected by atmospheric drag, a drag-corrected IIP calculation



PRECEDING PAGE BLANK-NOT FILMED

is also required by the RSO. Since the ordinary differential equations describing the missile trajectory in the atmosphere are not solvable in closed form, as they were in the case of an elliptical trajectory for a missile in a vacuum, these equations must be numerically integrated until the resulting trajectory pierces the Earth's surface. The numerical method currently being used is a fourth-order Runge-Kutta (RK) method with approximately ten steps per trajectory length or a second-order RK method with approximately twenty steps per trajectory length. The fourth-order RK method is used whenever the ballistic coefficient $\beta$ of the missile is greater than 200 lbs/$ft^2$, whereas the second-order RK method is used whenever $\beta$ is less than or equal to 200 lbs/$ft^2$. In either case, approximately forty evaluations of the equations of motion are required per drag IIP computation. Since the Runge-Kutta method is a single step method, the step size at each integration step is independent of the step size at all previous steps. The step size used is adjusted at each step so that it decreases as drag increases, in order to obtain a more accurate drag IIP in the same number of integration steps. A derivation of the equations of motion and the method of solution is described in [2], whereas the step size adjustment is described in [3].

2. Statement of Problem. Some missions at WSMR involve several simultaneous missiles. A drag IIP computation is required for each such missile. Experimental runs have indicated that only one drag IIP could be computed at a ten per second rate and at most four drag IIPs at a five per second rate [4]. Until recently drag IIPs were computed at a five per second rate, with a linear extrapolation of the last two computed drag IIPs being used to approximate the drag IIP at the next intermediate time. Additional mission requirements have necessitated the implementation of a variable rate for computing drag IIPs, ranging from ten per second to two per second [4]. At a two per second rate, four drag IIP approximations via linear extrapolation are required for each computed drag IIP, in order to output IIPs at a ten per second rate. Consequently, the output drag IIPs at a two per second rate are not as smooth and accurate as they are at a five or ten per second rate.

The purpose of this paper is to present an alternative method of computing drag IIPs and alternative methods of obtaining approximations to the drag IIPs at intermediate times, in order to decrease computation time and/or increase accuracy. In order to put this paper in its proper perspective, we briefly mention some previous investigations toward achieving these goals.

3. Previous Investigations. In 1965 a method for obtaining a drag-corrected Kepler IIP for Athena missiles was investigated [5]. Briefly this method consisted of using position and velocity data from a previous Athena mission to calculate, in non-real time, a table of differences between Kepler IIPs and drag-corrected IIPs as a function of velocity. During the next Athena mission a linear extrapolation of this data was used to obtain a drag-corrected Kepler IIP in real time.

In 1975 various numerical integration methods for computing drag IIPs using digital and/or analog computers were investigated [6], which included Adams-Moulton, Milne-Hamming, Euler, a variable order Adams method called DIFSUB developed by C. W. Gear [7, 8, 9], and 2nd, 3rd, and 4th order Runge-Kutta methods. The conclusions reached were that the RK methods were better than the other methods, except possibly the method of Gear. Large errors were associated with the analog solutions.

Subsequently, alternative methods for expressing the equations of motion, using Encke's method [10, pp. 29-35] and two different versions of the method of variation of parameters [10, pp. 116-120 and 11], were investigated, as well as alternative numerical integration methods, such as an improved variable order, variable step Adams method [12], a rational extrapolation method [13, 14, 15, 16], and a Gauss-Jackson ($\Sigma^2$) method [10]. The conclusions drawn were that the method of variation of parameters took about twice as long as Cowell's method for the computation of the same drag IIPs, whereas Encke's method took about four times as long [17, p. 131 and 18, p. 15]. Furthermore, the alternative numerical integration methods investigated offered little if any improvement over the variable step RK method currently being used [17, p. 134].

Recently, an "f and g series" impact predictor algorithm, which is based upon a Taylor-series-in-time representation of a missile's position and velocity, was developed for which a "two to ten fold reduction in computer execution time for satellite orbits and a seven to ten fold reduction in execution time for ICBM trajectories" could be achieved over conventional numerical integration [19, p. 59]. A second report uses the f and g series technique to determine the "geographical distribution of debris impact coordinates that would result if the missile were destroyed," [20, p. 287]. This technique is an extension of the classical f and g series used in the solution of Kepler's equations of motion [21, pp. 107-111].

Such a Taylor series type of solution to an initial value problem is "generally impractical from a computational point of view," [22, p. 365]. In fact "... the necessity of calculating the higher derivatives makes Taylor's algorithm completely unsuitable on high-speed computers for general integration purposes," [23, p. 330]. However, in "comparison with fourth-order predictor-corrector and Runge-Kutta methods, the Taylor series method can achieve an appreciable saving in computer time, often by a factor of 100," [24, p. 389]. In view of these statements, the equations necessary to implement the Taylor series method are derived in this paper in order to determine whether or not the claims made for satellite orbits and ICBM trajectories are equally valid for short and medium range trajectories, such as those experienced by the missiles tested at WSMR.

4. Equation of Motion. The vector equation of motion of a missile is

$$\ddot{\bar{r}} = \ddot{\bar{r}}_u + \ddot{\bar{r}}_g + \ddot{\bar{r}}_d , \tag{4.1}$$

where $\ddot{\bar{r}}$ is the total acceleration of the missile, $\ddot{\bar{r}}_u$ is the unperturbed (Keplerian) acceleration, $\ddot{\bar{r}}_g$ is the perturbative acceleration due to higher order harmonics of the Earth's gravitational field, and $\ddot{\bar{r}}_d$ is the perturbative acceleration due to the Earth's atmospheric drag.

Missiles which are launched from one end of WSMR and which impact at the other end do not travel more than 143 miles, whereas missiles which are launched from Green River, Utah and impact on WSMR do not travel more than 500 miles. For such short and medium range missiles the perturbative acceleration $\ddot{\bar{r}}_g$ is insignificant compared with $\ddot{\bar{r}}_d$. Therefore, $\ddot{\bar{r}}_g = \bar{0}$ in this paper.

The expression for $\ddot{\bar{r}}_u$ is

$$\ddot{\bar{r}}_u = (-\mu/r^3)\bar{r} , \tag{4.2}$$

where $\bar{r} = (x, y, z)$ is the position vector of the missile, $r = (\bar{r}\cdot\bar{r})^{1/2}$

is the distance of the missile from the origin of an inertial rectangular coordinate system, and $\mu$ is a gravitational constant. If r is measured in feet and time in seconds, then $\mu = 1.406559714 \times 10^6$ $ft^3/sec^2$.

The coordinate system used is right-handed, with origin at the Earth's center (geocentric), x-y plane in the equatorial plane, z-axis positive through the South Pole, and the x-axis aligned at 106°20' west longitude at the instant missile position and velocity data is obtained. Since this data is obtained in a non-inertial (relative) coordinate system which rotates with the Earth, the velocity components $\dot{x}_r$, $\dot{y}_r$, and $\dot{z}_r$ of the relative velocity vector $\bar{v}_r$ must be converted to corresponding components $\dot{x}$, $\dot{y}$, and $\dot{z}$ of the inertial velocity vector $\dot{\bar{r}}$ by

$$\dot{x} = \dot{x}_r + \omega y, \quad \dot{y} = \dot{y}_r - \omega x, \quad \dot{z} = \dot{z}_r \ , \tag{4.3}$$

where $\omega = 7.29211583 \times 10^{-5}$ rad/sec is the Earth's angular rate of rotation.

The vector acceleration for atmospheric drag is given by

$$\dot{\bar{r}}_d = -(\rho(h) v_r / 2\beta)\bar{v}_r \ , \tag{4.4}$$

where $\bar{v}_r$ is the velocity of the missile relative to the Earth's atmosphere in ft/sec, $v_r = (\bar{v}_r \cdot \bar{v}_r)^{1/2}$, $\beta$ is the ballistic coefficient of the missile in $lbs/ft^2$, and $\rho(h)$ is the density of the atmosphere in $slugs/ft^3$ at the current missile altitude h. The Earth's atmosphere is assumed to rotate with the Earth, the effects due to the wind are neglected, and $\beta$ is assumed to be a constant for each missile.

If $\dot{\bar{r}}$ denotes the velocity of the missile in an inertial coordinate system and $\bar{\omega} = (0, 0, -\omega)$, then the relative velocity $\bar{v}_r$ is given by

$$\bar{v}_r = \dot{\bar{r}} - \bar{\omega} \times \bar{r} \ . \tag{4.5}$$

The atmospheric density is approximated by

$$\rho(h) = A_0 \exp (A_1 h + A_2 h^2) , \tag{4.6}$$

where $A_0 = 8.1283549 \times 10^{-2}$, $A_1 = -3.0319838 \times 10^{-5}$, and $A_2 = -1.6214665 \times 10^{-10}$ for $h < 70{,}000$ ft; $A_0 = 1.7237156 \times 10^{-1}$, $A_1 = -5.4682878 \times 10^{-5}$, and $A_2 = 3.7544187 \times 10^{-11}$ for $70{,}000$ ft $< h <$ $150{,}000$ ft; and $A_0 = 0$ for $h > 150{,}000$ ft. The altitude h is given by

$$h = r - R + 4000 , \tag{4.7}$$

where $R = 20929831.0 - 71303.68411(z/r)^2$ (cf. [1, p. 41]) is the Earth radius of the Clarke Spheroid of 1866 at the same latitude as the missile but at the 4000 ft altitude of WSMR.

5. Taylor Series Method. The Taylor series expressions the missile's position $\bar{r} = \bar{r}(t)$ and velocity $\dot{\bar{r}} = \dot{\bar{r}}(t)$ at time t are

$$\bar{r} = \Sigma_{i=0}^{n} (\tau^i/i!)\bar{r}_0^{(i)} + \bar{E}_1(t) , \tag{5.1}$$

$$\dot{\bar{r}} = \Sigma_{i=0}^{n-1} (\tau^i/i!)\bar{r}_0^{(i+1)} + \bar{E}_2(t) , \tag{5.2}$$

where $\tau = t - t_0$ is the time interval or step size for re-initialization of the series and $\bar{r}_0^{(i)} = \bar{r}^{(i)}(t_0)$ is the $i^{th}$ time derivative of the position evaluated at epoch $t_0$. The error vectors $\bar{E}_1(t)$ and $\bar{E}_2(t)$, due to the truncation of these Taylor series at the $\bar{r}_0^{(n)}$ term, are given by

$$\bar{E}_1(t) = (\tau^{n+1}/(n+1)!)\, \bar{r}^{(n+1)}(\xi_1), \quad t_0 \leq \xi_1 \leq t, \tag{5.3}$$

$$\bar{E}_2(t) = (\tau^n/n!)\, \bar{r}^{(n+1)}(\xi_2) , \quad t_0 \leq \xi_2 \leq t. \tag{5.4}$$

Following a rule of thumb, stated by Moore [25], of choosing n to be approximately equal to the number of significant decimal digits that can be carried by the computer, we choose n to have a maximum value

of 6. However, for small values of the ballistic coefficient $\beta$, smaller values of n with smaller step sizes $\tau$ generally are preferred.

The derivatives $\bar{r}_0^{(i)}$ in (5.1) and (5.2) are obtained by differentiating $\ddot{\bar{r}}_u$ in (4.2) and $\dot{\bar{r}}_d$ in (4.4) with respect to time and adding. Hence from (4.1), with $\dot{\bar{r}}_g = \bar{0}$,

$$d^m \ddot{\bar{r}}/dt^m = d^m \ddot{\bar{r}}_u/dt^m + d^m \dot{\bar{r}}_d/dt^m. \tag{5.5}$$

In the development of the f and g series algorithm in [19], the expressions for the second and higher derivatives of $\ddot{\bar{r}}_u$ failed to include the acceleration $\dot{\bar{r}}_d$ and its derivatives [19, p. 23, (6)], resulting in the omission of a term in the expression for $d^2\ddot{\bar{r}}/dt^2$ [20, p. 301, (15)], and the omission of terms in all higher derivatives of $\ddot{\bar{r}}$. Therefore, it was decided to determine these higher derivatives of $\ddot{\bar{r}}_u$ and $\dot{\bar{r}}_d$ independently from [19] and [20], in order to correct this omission.

Applying Leibnitz' Theorem for finding the $n^{th}$ derivative of the product of two functions to (4.2), we obtain

$$\frac{d^m \ddot{\bar{r}}_u}{dt^m} = -\mu \sum_{k=0}^{m} \binom{m}{k} \frac{d^k}{dt^k}\left(\frac{1}{r^3}\right) \frac{d^{m-k}\bar{r}}{dt^{m-k}} , \tag{5.6}$$

in which

$$\begin{cases} d(1/r^3)/dt = 3\dot{r}/r^4 , \\ d^2(1/r^3)/dt^2 = 3\ddot{r}/r^4 - 4\dot{r}^2/r^5 , \\ d^3(1/r^3)/dt^3 = 3\dddot{r}/r^4 - 36\,\dot{r}\ddot{r}/r^5 + 60\,\dot{r}^3/r^6 , \\ d^4(1/r^3)/dt^4 = 3r^{iv}/r^4 - 48\,\dot{r}\dddot{r}/r^5 - 36\,\ddot{r}^2/r^5 \\ \qquad\qquad + 360\,\dot{r}^2\ddot{r}/r^6 - 360\,\dot{r}^3/r^7 . \end{cases} \tag{5.7}$$

Since $\dot{r} = (\dot{\bar{r}} \cdot \dot{\bar{r}})^{1/2}$, we easily obtain

$$(5.8)\quad \begin{cases} \ddot{r} = (\dot{\bar{r}} \cdot \ddot{\bar{r}}) / \dot{r} , \\ \dddot{r} = (-\ddot{r}^2 + \ddot{\bar{r}} \cdot \ddot{\bar{r}} + \dot{\bar{r}} \cdot \dddot{\bar{r}}) / \dot{r} , \\ r^{iv} = (-3\ddot{r}\dddot{r} + 3\ddot{\bar{r}} \cdot \dddot{\bar{r}} + \dot{\bar{r}} \cdot \bar{r}^{iv}) / \dot{r} . \end{cases}$$

The corresponding higher derivatives of $\dot{\hat{\bar{r}}}_d$ are determined by applying Leibnitz' Theorem to (4.4), to obtain

$$(5.9)\quad \frac{d^m \dot{\hat{\bar{r}}}_d}{dt^m} = -\frac{1}{2\beta} \sum_{k=0}^{m} \binom{m}{k} \frac{d^k(\rho v_r)}{dt^k} \frac{d^{m-k} v_r}{dt^{m-k}} ,$$

where

$$(5.10)\quad \frac{d^k(\rho v_r)}{dt^k} = \sum_{i=0}^{k} \binom{k}{i} \frac{d^k \rho}{dt^k} \frac{d^{k-i} v_r}{dt^{k-i}} .$$

From (4.5), we have

$$(5.11)\quad d^m \bar{v}_r / dt^m = d^m \dot{\bar{r}} / dt^m - \bar{\omega} \times d^m \bar{r} / dt^m .$$

Since $v_r = (\bar{v}_r \cdot \bar{v}_r)^{1/2}$, we easily obtain

$$(5.12)\quad \begin{cases} \dot{v}_r = (\bar{v}_r \cdot \dot{\bar{v}}_r) / v_r , \\ \ddot{v}_r = (-\dot{v}_r^2 + \dot{\bar{v}}_r \cdot \dot{\bar{v}}_r + \bar{v}_r \cdot \ddot{\bar{v}}_r) / v_r , \\ \dddot{v}_r = (-3\dot{v}_r \ddot{v}_r + 3\dot{\bar{v}}_r \cdot \ddot{\bar{v}}_r + \bar{v}_r \cdot \dddot{\bar{v}}_r) / v_r , \\ v_r^{iv} = (-4\dot{v}_r \dddot{v}_r - 3\ddot{v}_r^2 + 3\ddot{\bar{v}}_r \cdot \ddot{\bar{v}}_r + 4\dot{\bar{v}}_r \cdot \dddot{\bar{v}}_r + \bar{v}_r \cdot \bar{v}_r^{iv}) / v_r . \end{cases}$$

The higher time derivatives of $\rho$ in (5.10) are obtained by repeated applications of the chain rule of differential calculus to $\rho(h(t))$ in (4.6). We first note from (4.7) that the time derivatives of the Earth's radius R are small compared with the time derivatives of $r = \bar{r}$. Therefore, we have

$$(5.13) \qquad d^m h/dt^m = d^m r/dt^m + \varepsilon_m ,$$

where $|\varepsilon_m| << |d^m r/dt^m|$.

Applying the chain rule to $\rho(h(t))$ and using (5.13), we obtain

$$(5.14) \quad \begin{cases} d\rho/dt = \dot{r}\partial\rho/\partial h , \\ d^2\rho/dt^2 = \dot{r}^2\partial^2\rho/\partial h^2 + \ddot{r}\,\partial\rho/\partial h , \\ d^3\rho/dt^3 = \dot{r}^3\partial^3\rho/\partial h^3 + 3\,\dot{r}\,\ddot{r}\,\partial^2\rho/\partial h^2 + \dddot{r}\,\partial\rho/\partial h , \\ d^4\rho/dt^4 = \dot{r}^4\partial^4\rho/\partial h^4 + 6\,\ddot{r}\dot{r}^2\ddot{r}\,\partial^3\rho/\partial h^3 \\ \qquad + (3\,\ddot{r}^2 + 4\,\dot{r}\,\dddot{r})\partial^2\rho/\partial h^2 + r^{iv}\,\partial\rho/\partial h . \end{cases}$$

Differentiating (4.6), we have

$$(5.15) \quad \begin{cases} \partial\rho/\partial h = (A_1 + 2A_2 h)\rho , \\ \partial^2\rho/\partial h^2 = 2A_2\rho + (A_1 + 2A_2 h)\partial\rho/\partial h , \\ \partial^3\rho/\rho h^3 = 4A_2\partial\rho/\partial h + (A_1 + 2A_2 h)\,\partial^2\rho/\partial h^2 , \\ \partial^4\rho/\partial h^4 = 6A_2\partial^2\rho/\partial h^2 + (A_1 + 2A_2 h)\partial^3\rho/\partial h^3 . \end{cases}$$

All quantities required for the computation of the derivatives $\bar{r}_0^{(i)}$, $i = 1, \ldots, 6$, in the Taylor series (5.1) and (5.2) can now be determined from the preceding equations. To implement the Taylor series method, a value of n is choosen between 2 and 6 with lower values corresponding to smaller values of $\beta$. The step size $\tau$ is determined such that the maximum number of steps permitted per trajectory integration is not exceeded, and such that the local truncation errors $\bar{E}_1$ in (5.3) and $\bar{E}_2$ in (5.4) are not exceeded. Therefore, given the position $\bar{r}_0$ and velocity $\dot{\bar{r}}_0$ of a missile at epoch $t_0$, we obtain the position $\bar{r}$ and velocity $\dot{\bar{r}}$ of a missile at time t from (5.1) and (5.2). This process is repeated until impact time T, for which $r(T) = R(T)$.

Since the preceding computations were made in an inertial coordinate system, the true impact coordinates $x_I$, $y_I$, and $z_I$ can

be obtained from the coordinates $x_T$, $y_T$, and $z_T$ of the position vector $\bar{r}(T)$ at impact by a rotation through $\omega T$ radians as follows

$$(5.16)\qquad \begin{cases} x_I = x_T \cos \omega T - y_T \sin \omega T , \\ y_I = x_T \sin \omega T + y_T \cos \omega T , \\ z_I = z_T . \end{cases}$$

6. Methods for Approximating Drag IIPs between Computational Cycles. Ten times a second, or once every 100 milliseconds, all data on a missile, such as its position, velocity, and drag IIP, is updated. However, the computation of drag IIPs for several missiles cannot be completed in fewer than 100 milliseconds even using the Taylor series method, because much of the computer's time is spent making many other computations during each 100 millisecond time period. For example, if drag IIPs can be computed only once every 500 milliseconds, then approximations to the drag IIPs are required four times per 500 millisecond computational cycle. The current method for obtaining these approximations is by a linear extrapolation of previously computed drag IIPs, considered as functions of range time. Unfortunately, this method does not account for new values of the missile's initial position $\bar{r}_0$ and velocity $\dot{\bar{r}}_0$ at these intermediate times and is much less accurate than the following improved methods.

The first improved method for approximating drag IIPs between computational cycles is by a quadratic extrapolation of the components of previously computed drag IIPs, considered as functions of their corresponding components of the vacuum IIPs. Since vacuum IIPs are always computed every 100 milliseconds anyway, no additional computer time is required for their use in this method. In fact, it requires about the same amount of computational time as the current linear extrapolation method (one millisecond), yet is more accurate by an average factor of 13. Furthermore, it does account for new intermediate time values of the missile's initial position $\bar{r}_0$ and velocity $\dot{\bar{r}}_0$ since the vacuum IIPs at these times are functions of $\bar{r}_0$ and $\dot{\bar{r}}_0$. However, if consecutive pairs of components of vacuum IIPs are "close" together, then this quadratic extrapolation method may yield erroneous drag IIP approximations. Therefore, the following method which avoids this problem is recommended.

The second improved method for approximating drag IIPs between computational cycles is by a quadratic extrapolation of the differences between previously computed drag IIPs and their corresponding vacuum IIPs, considered as functions of range time. It too requires about the same amount of computational time as the current linear extrapolation method, yet is more accurate by an average factor of 6.

In general, higher order extrapolations may yield increasingly worse approximations to the drag IIPs as the order is increased. This would be especially true if the radar-determined position and velocity of the missile were not following a smooth trajectory, in which case an accurate drag IIP would be needed most. In fact, "If polynomial extrapolation must be done with poorly behaved functions, then very low degree extrapolation is usually the safest, but even this should be carried out only for values of x very close to the tabulated region," [26, p. 58].

7. Conclusions. The use of the Taylor series method resulted in equivalent drag IIPs being computed in two to ten times less time than by the currently used Runge-Kutta method. The use of the method of quadratic extrapolation of the differences between previously computed drag IIPs and their corresponding vacuum IIPs resulted in approximations of drag IIPs between computational cycles being computed over six times more accurately than by the currently used linear extrapolation method, with about the same amount of computational time, and without the possibility of erroneous approximations.

## REFERENCES

[1] Berry, F.N. and W.W. Page, "Athena Impact Prediction for Range Flight Safety," Army Missile Test and Evaluation Directorate, Electro-Mechanical Laboratories Division, Technical Note #103, June 1964.

[2] Baker, C.M., "IMPCTR - An Automated Impact Prediction Procedure," Analysis and Computation Directorate, Systems Programming Division, Internal Memorandum NR. 79, April 1968.

[3] Falke, J.F., "RTMIN Drag Corrected IIP," NR-AM-R, April 1974 (Rev. June 1975).

[4] Falke, J.F., "RTMIN Drag IIP Control Logic," NR-AM-R, August 1977.

[5] Berry, F.N., "A Brief Survey of the Differences Between Kepler and Integrated Impact Predictions," Computer Directorate, Computer Sys Div, Internal Memo NR. 9, Nov 1965.

[6] Falke, J.F., "Development of Missile Instantaneous Impact Prediction Schemes for Digital and Hybird Computers," Project Report for CICE 699, NR-AM-R, April 1975.

[7] Gear, C.W., "Algorithon 407 - DIFSUB for Solution of Ordinary Differential Equations [D2]," Communication of the ACM, Vol. 14, No. 3, March 1971, pp. 185-190.

[8] Gear, C.W., "The Automatic Integration of Ordinary Differential Equations," Communication of the ACM, Vol. 14, No. 3, March 1971, pp. 176-179.

[9] Gear, C.W., Numerical Initial Value Problems in Ordinary Differential Equations, Prentice-Hall, Englewood Cliffs, N.J. 1971.

[10] Herrick, S., Astrodynamics: Orbit Correction, Perturbation Theory, Integration, Vol. 2, Van Nostrand Reinhold Co., London, 1972.

[11] Pines, S., "Variation of Parameters for Elliptic and Near Circular Orbits," The Astronomical Journal, Vol. 66, No. 1, February 1961, pp. 5-7.

[12] Shampine, L.F., and M.K. Gordon, Computer Solution of Ordinary Differential Equations: The Initial Value Problem, W.H. Freeman, San Francisco, 1975.

[13] Bulirsch, R. and J. Stoer, "Numerical Treatment of Ordinary Differential Equations by Extrapolation," Numer. Math., Vol. 8, 1966, pp. 1-13.

[14] Fox, P.A., "DESUB: Integration of a First-order System of Ordinary Differential Equations," Mathematical Software, J. Rice, ed., Academic Press, New York, 1971, Chapter 9.

[15] Clark, N.W., ANL D250-DIFSUB (3600 Fortran Program), Argonne National Laboratories, 1966.

[16] Clark, N.W., ANL D251-DIFSUB (3600 Fortran Program for use with ANL D250), Argonne National Laboratories, 1966.

[17] Kuzanek, J.F., "An Investigation of Various Methods to Improve Drag Corrected Instantaneous Impact Predittions for Range Flight Safety," National Range Operations Directorate, Analysis and Computation Division, Internal Memorandum No. 160, January 1979.

[18] Kuzanek, J.F., "A Comparison of Various Methods for Computing Instantaneous Impact Predictions of Missiles for Range Flight Safety", ARO Report 79-3, Proceedings of the 1979 Army Numerical Analysis and Computers Conference, September 1979, pp. 43-61.

[19] Baker, R.M.L., Jr. and N.H. Jacoby, Jr., "A Time Series (F&G Series) Approach to Missile Trajectory Prediction," J. of the Astronautical Sciences, Vol. XXIII, No. 1, January-March 1975, pp. 19-59.

[20] Baker, R.M.L., Jr., T.J. Mucha, D.R. Darby, N.H. Jacoby, Jr., A.W. Johnson, and R.E. Ryan, "Range-Safety Debris-Pattern Analysis," J. of the Astronautical Sciences, Vol. XXIII, No. 4, October-December 1975, pp. 287-323.

[21] Escobal, P.R., Methods of Orbit Determination, John Wiley & Sons, Inc., New York, 1965.

[22] Dorn, W.S. and D.D. McCracken, Numerical Methods with Fortran IV Case Studies, John Wiley & Sons, Inc., New York, 1972.

[23] Conte, S.D. and C. de Boor, Elementary Numerical Analysis: An Algorithmic Approach, 2nd ed., McGraw-Hill, New York, 1972.

[24] Barton, D., I.M. Willers, and R.V.M. Zahar, "Taylor Series Methods for Ordinary Differential Equations-An Evaluation," Mathematical Software, J. Rice, ed., Academic Press, New York, 1971, pp. 369-390.

[25] Moore, R.A., Interval Analysis, Prentice-Hall, Englewood Cliffs, New Jersey, 1966.

# APPLICATION OF PROCESS MODELING TO SHELL DRAWING OPERATIONS

GOVERDHAN LAHOTI, PhD
BATTELLE COLUMBUS LABORATORIES, COLUMBUS, OH 43085
*FEE M. LEE, Mr.
U. S. ARMY/ARRADCOM, DOVER, NJ 07801

The manufacture of artillery shells involves a number of metalforming, machining, and heat treatment operations. Cold or hot drawing of shells, where the wall thickness of the shell is reduced while keeping the inner diameter essentially unchanged, represents an important step in the forming operations. From the point of view of deformation mechanics, the analysis of metal flow and stress in shell drawing is very complex. The effect of friction at the tool-material interface, the heat generation and heat transfer during deformation, and the effect of strain, strain rate and temperature on the flow properties of material are difficult to analyze and predict, since a process model must be realistic and should not neglect the effect of significant process variables.

The mathematical models and the computer programs capable of optimizing the shell drawing process for actual artillery shells and cartridge cases were developed earlier.[1] These mathematical models were based on the analysis of plastic deformations and included the effects of the various process variables (such as punch speed, billet temperature, and lubrication), the properties of shell, die, and punch materials, and the die configuration (both conical and streamline). These models were computerized so that they can be used to analyze the mechanics of the process and predict potential material failure (such as punch-through), and optimize the die configuration and process variables. These mathematical models were expanded to simulate shell drawing through multiple dies in tandem with a tapered punch, as shown in Figure 1. Graphical display capabilities were also included during computerization of the expanded math models.



PRECEDING PAGE BLANK-NOT FILMED

## Stress Analysis

The present model simulates the actual drawing process by dividing the punch movement into a finite number of discrete steps. The Sachs' method[2] of analysis was used to calculate stresses and drawing loads at each step. This analysis is valid for both conical and streamlined dies, since a complex die profile can be approximated by a series of straight lines.

Figure 2 shows a segment of the shell between the die and the punch, where the die is stationary and the punch is moving to the right. The operation consists of drawing a tube of outside radius $R_o$ and inside radius $R_i$ by a punch through a die to a tube of outside radius $r_o$ and inside radius $r_i$. The details of the analysis of the stresses and the loads in this operation are given in Reference (3).

In calculating the stresses, the flow stress of each element is considered as a function of strain, strain rate, and temperature. The strain in an element is the cumulative strain, and strain rate is calculated from a velocity field developed during earlier studies.[1] The temperature of an element will depend upon the heat generated from plastic deformation, and friction at the tool-material interfaces, and the heat conduction to the colder dies and punch. Expressions for strain, strain rate and temperature of an element inside a die are also given in Reference (3).

## Computer Model

Based on this stress analysis, a system of computer programs, named DRAWNG , was developed to simulate the drawing operation for artillery shell through multiple dies in tandem that employs a tapered punch. These computer programs are coded in FORTRAN IV and are applicable to both the cold and hot drawing of shells. A functional flow chart of DRAWNG is given in Figure 3. In its present form, DRAWNG is operational on a CDC system in interactive mode using a Tektronix graphics terminal. The die, billet and punch geometry are inputed to the program. This input can be read either through a data file prestored in the computer as a cataloged file or through the keyboard.

DRAWNG simulates the tandem drawing operation on a real-time basis, and the step-by-step results are displayed on the computer's graphic display terminal, as shown in Figure 4. On the top one-third of the screen, first the title is printed, and then the dies are drawn showing specified spacings between them, and the billet and punch are positioned for the beginning of simulation. Once the simulation begins, the step-by-step movement of the punch and the billet is shown on the top one-third of the screen. At the same time, the total ram load versus punch displacement and wall stress versus punch displacement are

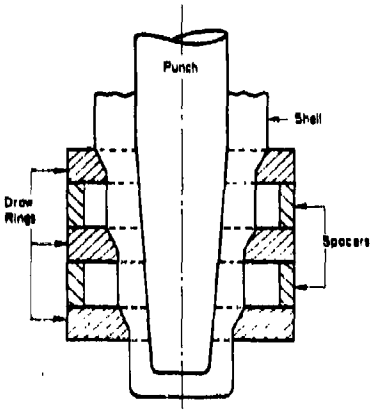



Figure 1. Tandem drawing with a tapered punch.

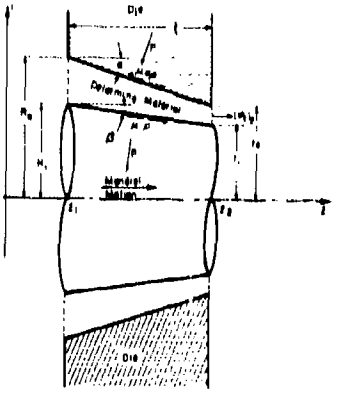


Figure 2. Drawing through conical die with a tapered punch.

In order to enhance the value of these mathematical models, confirmation tests under production or near-production conditions were conducted to substantiate them. The confirmation tests of the shell drawing operation at room temperature were conducted with 106.68mm (4.2-inch) M335 shell at Chamberlain Manufacturing Corporation's Waterloo, Iowa division under production and near-production conditions. The confirmation tests at hot forging temperatures were conducted with 155mm M107 shell at Chamberlain Manufacturing Corporation's Scranton, Pennsylvania division under production conditions. In both cases, tests were conducted using conventional conical and computer-designed streamlined dies. Finally, the evaluation of the mathematical models with the test results was conducted at Battelle.

## MATHEMATICAL MODELING OF TANDEM DRAWING WITH A TAPERED PUNCH

The mathematical models developed earlier[1] were valid only for drawing through a single conical or streamlined die with a straight punch. Actual shell drawing operations, however, very often use more than one drawing die in tandem. Further, in these operations, the first portion of the punch is invariably tapered, as shown in Figure 1. Therefore, these existing mathematical models were modified to include drawing through tandem dies with a tapered punch.

shown on the left half and right half of the lower two-thirds of the screen, respectively. During the simulation, the computer programs calculate the correct flow stress in the deformation zone corresponding to local strain, strain rate, and temperature, and utilize appropriate equations for stresses, depending upon whether the element is free, or within a die, or in between two dies. In addition, the tensile strength of the product is plotted on the wall stress-versus displacement diagram to show whether punch-through is predicted at any stage of the drawing operation. At the end of simulation, the computer program provides messages which enable the user to enlarge any of the three diagrams on the screen.

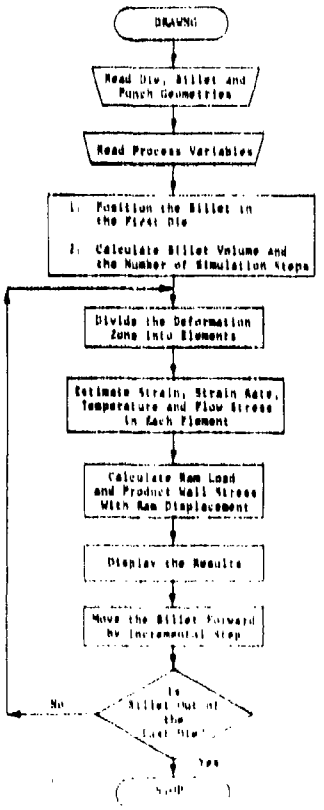



Figure 3. Functional flow chart of the computer program DRAWNG.

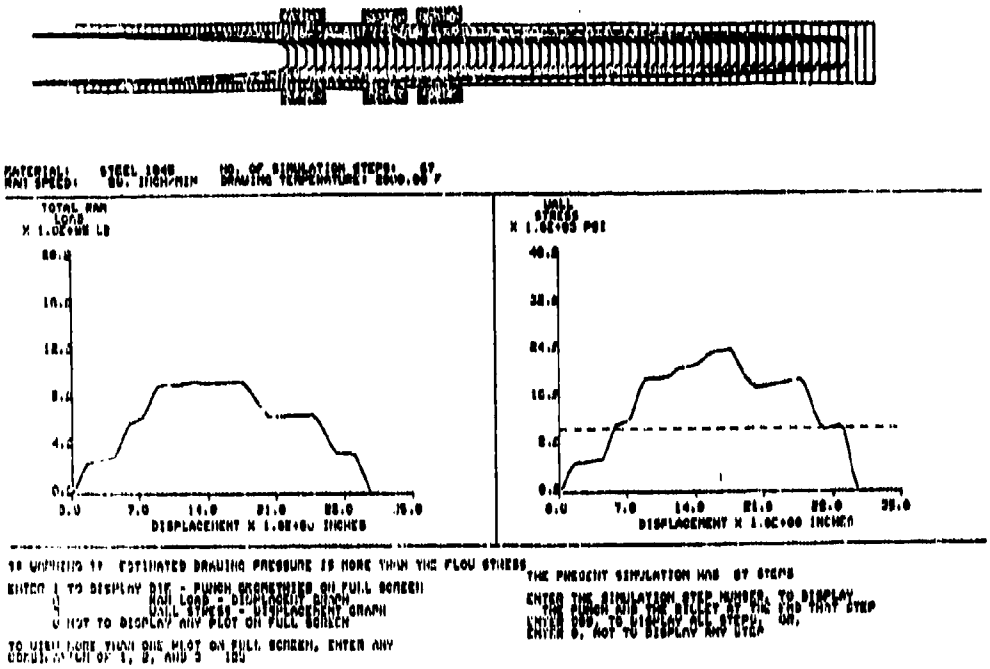



Figure 4. Simulation of tandem drawing as displayed on a CRT screen.

## CONFIRMATION TESTING OF THE SHELL DRAWING OPERATION

Confirmation tests of the shell drawing operation were conducted under production or near-production conditions to evaluate the accuracy of the predictions. These tests were conducted at two different shell manufacturing plants. The cold drawing tests (at room temperature) were performed at Chamberlain Manufacturing Corporation's Waterloo, Iowa division under either production or near-production conditions. Hot-drawing tests (at hot-working temperatures) were conducted at Chamberlain Manufacturing Corporation's Scranton, Pennsylvania division under actual production conditions. In both cases, the tests were conducted with standard army shells which are under current production, and except for the drawing dies, the existing tooling (punch, die holder, etc.) were used. The tests were first conducted using conventional conical dies, and later they were repeated using the optimally designed streamlined dies with double-curvature profiles (designed based on prior mathematical modeling work on drawing of shell.[1])

## Cold Drawing Tests

Confirmation tests of the cold drawing operation with conventional dies were conducted under an actual production environment at Chamberlain Corporation's Waterloo, Iowa division. For this purpose, the final double draw operation of 4.2 inch M335 shell (Figure 5) on a Bliss production hydraulic press was selected. The press is rated at 1.78 MN (200 tons), and has a 1020 mm (40-inch) stroke. The nominal ram speed is 68 mm/s (160 inch/min), and the ram diameter is 368.3 mm (14.5 inch).

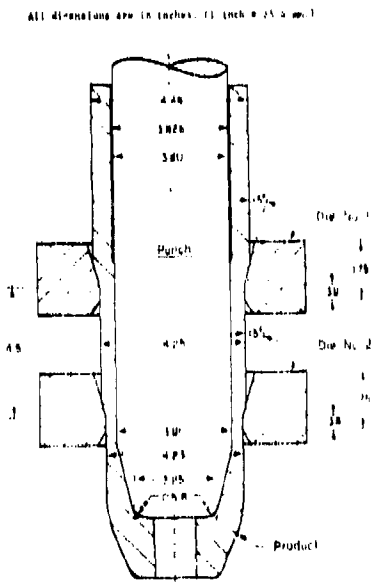
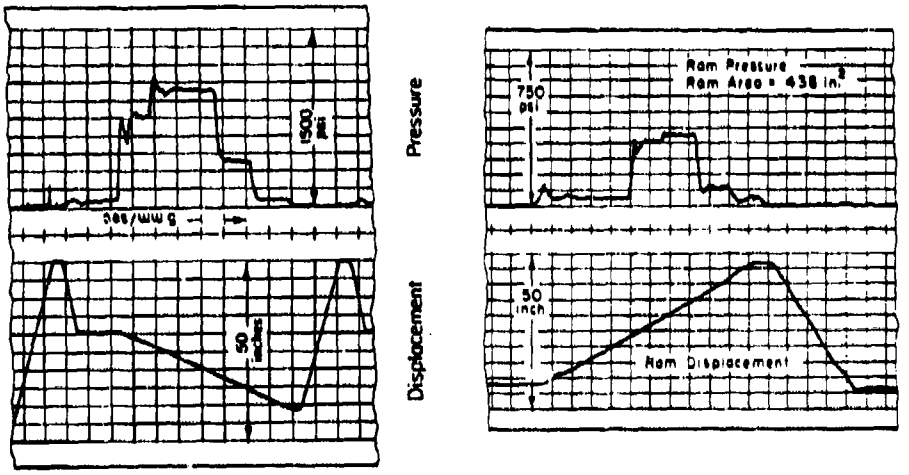



Figure 5. Schematic representation of final double draw of M335 shell.

Confirmation tests of the cold drawing operation with streamlined dies were conducted under near-production conditions in Chamberlain Corporation's Research and Development division at Waterloo, Iowa. These tests were performed on a Verson hydraulic press rated at 5.34 MN (600 tons) with a maximum ram displacement of 1.676 m (66 inch). This press has a nominal ram speed of 91.4 mm/s (216 inch/min) and ram diameter of 0.60 m (23-5/8 inch) ram area = 0.283 sq m (438 sq in.) . These presses were instrumented with pressure and displacement trans-

ducers, and the ram pressure and the ram displacement were recorded simultaneously on two separate channels of a two-channel brush recorder. At a selected interval, the ram pressure and ram displacement were recorded. The typical ram displacement plots for the conical and streamlined dies are given in Figure 6.




(a) Conical dies (b) Streamlined dies

Figure 6. Typical ram pressure and ram displacement recordings during cold drawing of M335 shell. (1 in. = 25.4 mm; 1 psi = 6.895 KN/m$^2$)

A few preforms and a few drawn shells were picked for detailed investigation. All the important dimensions, eccentricity and tensile properties were measured. The eccentricity of the drawn shells was almost always better than that of the preform. For the conical dies, the maximum, the minimum and the average peak loads were 0.804 MN (90.41 tons), 0.738 MN (82.98 tons), and 0.771 MN (86.69 tons), respectively. The measured loads for the deformation through streamlined dies were consistently lower. The maximum, the minimum and the average peak loads in this case were, 0.731 MN (82.19 tons), 0.673 MN (75.61 tons), and 0.687 MN (77.26 tons) respectively. Further, as seen in Figure 6(b), the breakthrough peak was typically absent here due to a more uniform deformation. The billet temperature after deformation was approximately 93.3°C (200°F) as compared to 104.4°C (220°F) in conventional drawing, also indicating a less severe deformation through streamlined dies.

## Hot Drawing Tests

Confirmation tests of hot drawing operations with both conventional conical and streamlined dies were conducted for 155 mm M107 shell under actual production conditions at Chamberlain Corporation's Scranton, Pennsylvania division. These tests were conducted on the No. 1 Bliss press line which consists of a vertical hydraulic press. It has a load capacity of 3.56 MN (400 tons), a stroke of 3035 mm (120 inch), a nominal press speed of 0.466 m/s (1100 inch/minute), and a ram diameter of 511 mm (20.125 inch).

To measure ram load during the drawing operation, the hydraulic line of the press was tapped with a pressure transducer, and its conditioned output was connected to one of the two channels of a strip recorder. The second channel of the recorder was connected to the output from a displacement transducer to measure the ram displacement. The temperatures of the incoming billet (after cabbaging and piercing) and of the product were measured with an infrared digital pyrometer. The temperatures of the dies were measured periodically with a contact pyrometer.

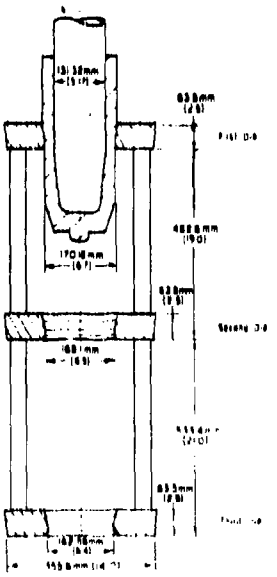


Figure 7. Schematic representation of hot drawing of 155 mm M107 shell.

A schematic representation of the hot drawing operation is shown in Figure 7. As in the case of cold drawing of shells, the general practice in industry for hot drawing of shells is to use dies with a conical entrance. The general configuration of the three dies used for hot drawing of 155 mm M107 shell under actual production has two entrance angles. Incoming material first touches the dies on a 10-degree segment at the die entrance. The 12-degree segment on the die entrance is primarily for easy centering of the workpiece in the top die. Prior to the beginning of the first shift on the day of the testing, the dies were replaced with a new set and were flame heated as usual and the press was instrumented to record ram pressure and ram displacement on a two-channel strip recorder. The load, displacement and temperature measurements were made every 15 minutes. A typical ram pressure and ram displacement recording is shown in Figure 8.

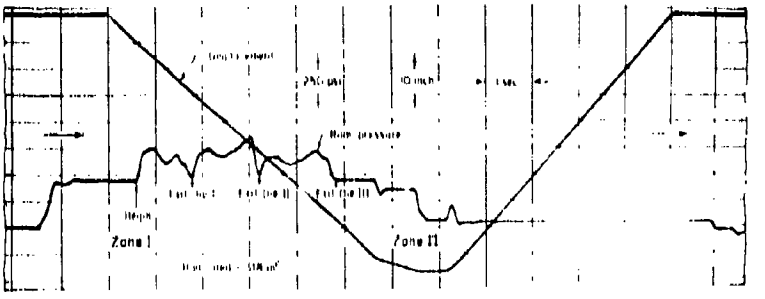



Figure 8. Typical ram pressure and ram displacement recording during hot drawing of M107 shell through conventional conical dies. (1 in.= 25.4 mm; 1 psi = 6.895 KN/m$^2$)

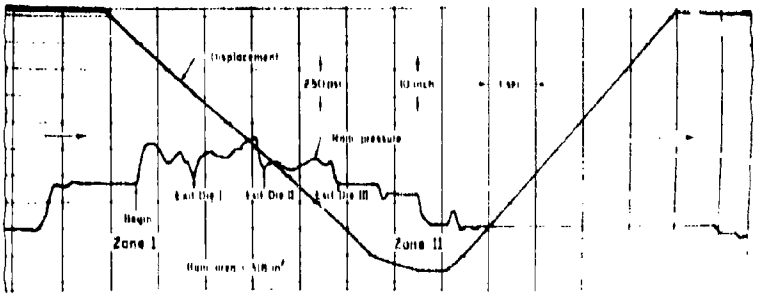



Figure 9. Typical ram pressure and ram displacement recording during hot drawing of M107 shell through streamlined dies. (1 in. = 25.4 mm; 1 psi = 6.895 KN/m$^2$)

In the present mathematical modeling studies, the interface friction shear stress is defined by

$$\tau = \frac{m}{\sqrt{3}} \bar{\sigma} = f\bar{\sigma} \qquad (2)$$

where $f = m/\sqrt{3}$ is the friction factor and its value could be between 0 and 0.577.

The interface friction shear factor in cold drawing was characterized using the well known ring test.[4] For phosphate and soap lubrication used in cold drawing, the friction factor m=0.06 (f=0.35) was used.

The load-displacement curves, predicted by the computer program DRAWNG for both conical and streamlined die arrangements were evaluated. For this purpose three cases were selected corresponding to the maximum, minimum, and average peak ram load recorded during each of the tests. For each of these selected cases, the theoretical load-displacement curve was generated using actual dimensions of the preform and the product. Theoretically predicted and experimentally measured load-displacement curves (corresponding to maximum recorded peak load), for drawing through two conical dies in tandem, are compared in Figure 10. Overall agreement between predicted and measured curves is very good. The predicted peak loads are somewhat higher because the upper bound approach was used in the analysis. The slight mismatch on the displacement axis is primarily due to initial positioning of the preform in the second die, since its forward end is contoured (see Figure 5) and does not match the die contour. Similar agreement were found for curves corresponding to the minimum and average peak loads.

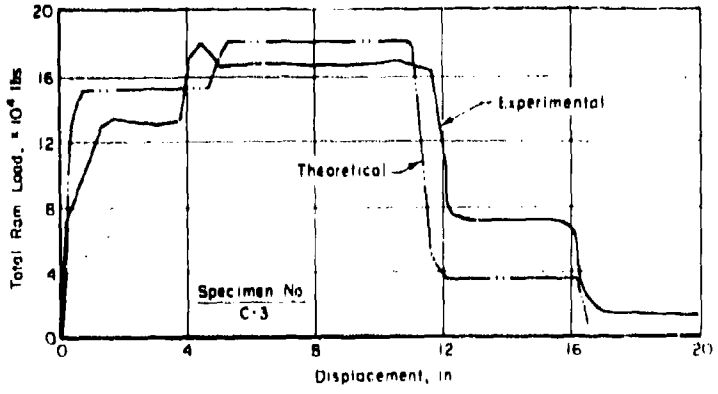



Figure 10. Theoretical and experimental load-displacement curves for cold drawing of M335 shell through two conical dies in tandem. (1 in. = 25.4 mm; 1 lb. = 4.448 N)

Following the tests with conventional conical dies, hot drawing tests with streamlined dies were conducted during a production run on the next day. Again, the dies were replaced during the night shift prior to the day of tests. As in the case of tests with conventional conical dies, the press was instrumented to record ram pressure and displacement, and temperature. Ram pressure and ram displacement were recorded every 15 minutes. The production with the streamlined dies was monitored for about 6 hours, and the dies consistently produced good parts. A typical ram pressure and ram displacement recording during these tests is shown in Figure 9. Although the ram pressure curve in Figure 9 does not look much different than that in Figure 8, on the overall, the operation with streamlined dies appeared quieter and smoother compared to the operation with the conical dies. This set of dies produced a total of 15,836 parts before they needed replacement. Engineers at Chamberlain stated that this die life was short, but it may not be totally attributable to the die design.

## EVALUATION OF MATHEMATICAL MODELS

The predictions from the computer program DRAWING were evaluated with respect to the results from cold and hot drawing confirmation tests. For this purpose, it became necessary to characterize the flow stress and the interface friction factor under both cold and hot working conditions. The flow stress data were obtained from tension tests, and the friction factor was determined by conducting ring tests.[4]

### Cold Drawing Operations

Interface friction and flow stress are two basic inputs required in the program DRAWNG. Since deformation in shell drawing occurs primarily due to tensile stress, the material flow stress should be determined under tensile loading. Therefore, a randomly selected preform was cut and tension specimens of 6.35-mm (0.25-inch) gage diameter were machined. These specimens were tested in an Instron testing machine at a crosshead speed of 0.04 mm/s (0.100 inch/min). Load versus displacement was recorded at a constant chart speed, and later it was reduced into true stress ($\bar{\sigma}$) versus true strain ($\bar{\epsilon}$) curve. In order to use this information in the computer program DRAWNG, a least square mean fit to the experimental points was developed as given below:

$$\bar{\sigma} = 772\ (\bar{\epsilon})^{0.1822}\ (MN/m^2). \qquad (1)$$

Experimental load-displacement curves, corresponding to the maximum, peak load for cold drawing of M335 shell through two streamlined dies (designed by the computer program CDVEL[1], are compared with the theoretically predicted load-displacement curves in Figure 11. The mathematical model in these cases predicts higher peak loads compared to measured peak loads. However, overall agreement between predictions and measurements is good for all engineering purposes. In addition, these dies proved to produce good parts, required approximately 13 percent less force and energy, and the breakthrough peak was typically absent compared to the conical dies in a similar tooling arrangement.

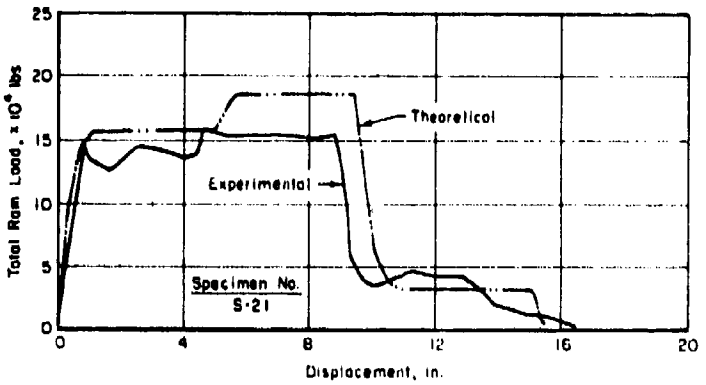



Figure 11. Theoretical and experimental load-displacement curves for cold drawing of M335 shell through two streamlined dies in tandem. (1 in. = 25.4 mm; 1 lb. = 4.448 N)

## Hot Drawing Operations

The flow stress for AISI 1045 steel at hot working temperatures and strain rates was taken from Reference 4. An average value of the friction factor f = 0.35 was selected based on an independent ring compression study for hot forging of steels.

To evaluate the load-displacement curves predicted by the computer program DRAWNC for hot drawing of M107 shell through conical and streamlined dies, three different cases for each type of dies were selected. Hot dimension of the preform was calculated from cold dimensions of the preform measured during trials. However, the effective length of the preform varied considerably, from 342.9 mm (13.5 inch) to 368.3 mm (14.5 inch). Punch dimensions were taken from the shape of the base of the finished product. Tool assembly, as shown in Figure 7,

was considered.

Length of the preform has considerable effect on overall shape of the load-displacement diagram. If the preform length is on the shorter side, the load drops nearly to zero as the product exits the first die and the peak near the exit from the second die is small. This behavior is shown in the theoretical curve in Figure 12, where the preform length was taken as 342.9 mm (13.5 inch). On the other hand, if the length of the preform is on the higher side, the product is picked up by the second die as soon as it exits from the first die, and the peak near the exit from the second die is relatively high, since the shell is in the second and the third die at the same time. This behavior is shown in the theoretical curve in Figure 13, where the effective preform length is taken as 368.3 mm (14.5 inch). Therefore, for evaluation purposes, an average preform length (effective) of 355.6 mm (14.00 inch) was taken. With this length of the preform, the trend observed in experiments were also reproduced by computer predictions, as shown in Figure 14. Overall agreement is good; peaks and valleys in the theoretical curves are somewhat sharper than measured. This is basically due to the simplified heat generation and heat transfer analysis used in the present model. However, the peak loads correlate well with the experimental measurements, except under the third die. This is, again, believed to be due to simple heat transfer analysis to a certain extent and due to the complex preform shape used in production compared to one assumed in the present analysis. In general, agreement between the theory and experiment appears good.

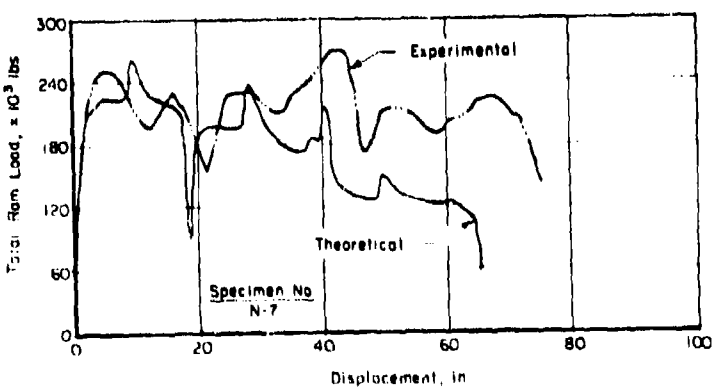



Figure 12. Theoretical and experimental load-displacement curves for cold hot drawing of M107 shell through conical dies for preform length = 342.9 mm (13.5 inch).

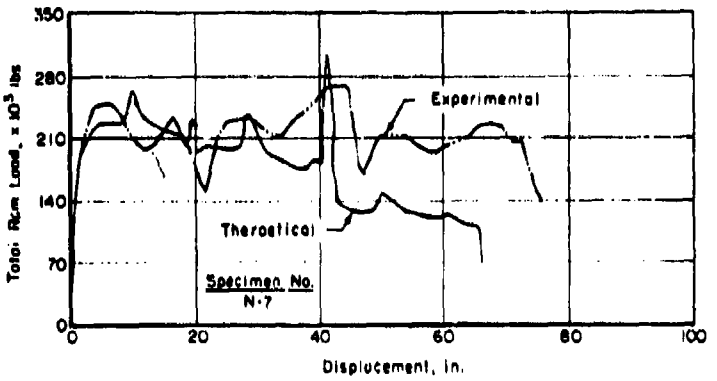



**Figure 13.** Theoretical and experimental load-displacement curves for hot drawing of M107 shell through conical dies for preform length = 368.3 mm (14.5 inch).

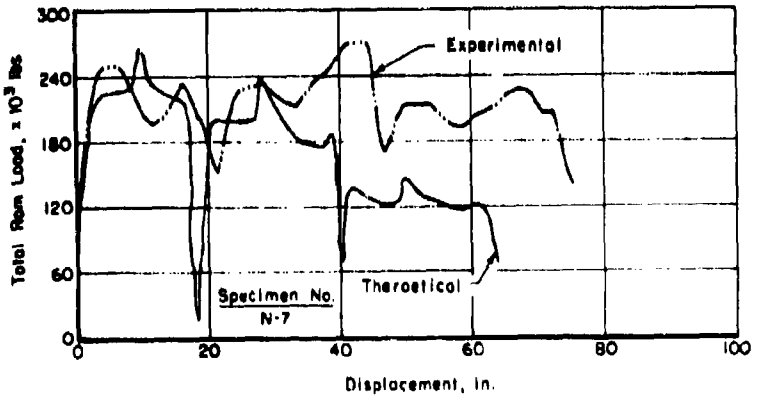



**Figure 14.** Theoretical and experimental load-displacement curves for hot drawing of M107 shell through conical dies for preform length = 355.6 mm (14.0 inch).

Similar comparisons of load displacement curves for hot drawing through streamlined dies were made. Considering the fact that preform dimensions and friction at the interfaces (due to different amount of scale) in hot drawing vary from piece to piece, correlation between the theoretically predicted and experimentally measured load displacement curves is good. Again, predicted loads through the third die were lower than measured values. This discrepancy is partially due to the difference between theoretical and actual preform shape, as discussed earlier. However, this error is not considered of any significance since process design and equipment selection in reality is primarily based on the peak loads in the process.

## CONCLUSIONS

Mathematical models for optimization of the shell drawing process were expanded to consider drawing through multiple dies in tandem and tapered punch and coded in computer program DRAWNG. DRAWNG is capable of simulating the shell drawing process, both hot and cold, and generate the ram load and the product wall stress versus ram displacement diagrams during simulation on a Cathode Ray tube (CRT). Predictions from these models were evaluated with respect to hot and cold drawing confirmation testings conducted under actual or near production conditions. In addition, these tests included an evaluation of the streamlined dies, designed by the computer program CDVEL developed earlier.[1]

Conclusions of the present study are:

(1) The system of computer programs DRAWNG is capable of simulating both the cold and hot shell drawing using single or multiple dies (both conical and streamlined) in tandem with a tapered punch.

(2) Comparisons between theoretically predicted and experimentally measured ram load versus ram displacement curves indicate that computer programs DRAWNG are capable of predicting load-displacement curves, both under hot and cold drawing conditions.

(3) Success of all experiments was predicted by the computer program since good parts were produced in the cold and hot drawing operations and no material failure had occurred.

(4) Streamlined dies designed by using the computer program CDVEL produced good parts, both dimensionally and property-wise, under actual production environment in the manufacturing of M335 and M107 shells.

## REFERENCES

1. G. D. Lahoti, T. L. Subramanian, and T. Altan, "Development of a Mathematical Model and Computer Programs Capable of Optimizing the Drawing Process for Actual Artillery Shells and Cartridge Cases", Final Report, Contract No. DAAA25-74-C0557, Frankford Arsenal, May 1975.

2. E. G. Thomsen, et al., Mechanics of Plastic Deformation in Metal Processing, McMillan, N.Y., 1965.

3. C. H. Lee, and T. Altan, "Influence of Flow Stress and Friction Upon Metal Flow in Upset Forging of Rings and Cylinders", Trans. ASME, J. Engrg. Industry, Vol. 94, August, 1972, p. 775.

4. T. Altan, and F. W. Boulger, "Flow Stress of Metals and Its Application in Metal Forming Analysis", Trans. ASME, J. Engrg. Industry, Vol. 95, November, 1973, p. 1009.

# TERRAIN THERMAL MODELING FOR CAMOUFLAGE AND TARGET ACQUISITION (U)

LEWIS E. LINK, JR., PH.D
U. S. ARMY ENGINEER WATERWAYS EXPERIMENT STATION
VICKSBURG, MISSISSIPPI 39180

## INTRODUCTION

Camouflage and target acquisition have opposing functions, one to hide and the other to seek. They have a common denominator, however, in that the features that surround the target to be camouflaged or identified (sometimes called the background) are critical in both the hide and seek role. An equally intimate knowledge is needed of the characteristics of both the target and the background. In essence, making something match the background and discriminating something from the background are inverse problems that require the same technology.

In the past decade thermal infrared (IR) technology has come of age providing sensors with new capabilities for target acquisition and presenting a new threat for camouflage. Optimizing IR sensors for target acquisition or optimizing camouflage measures to defeat such sensors requires a quantitative understanding of the thermal IR signatures of both targets and backgrounds.

The Army-Wide Ground Target Signature Program (AWGTSP) is addressing the need for a target-background design data base for sensor design and evaluation through a three-part program: development of a battlefield IR signature model that will allow extrapolations of target and background signatures to varying environmental, climatic, and seasonal conditions throughout the world; updating a tactical signature library to fill critical gaps in the existing empirical signature data base; and susceptibility analyses designed to ensure that vulnerability of Army tactical materiel is known so that effective camouflage can be brought to bear. An equally

important problem is the camouflage of key elements at fixed installations for which similar background information is needed.

Work on the AWGTSP has resulted in considerable progress in computer codes for prediction of the performance of surveillance, target acquisition, and terminal homing devices and prediction of target signatures. The target models have ranged from simple to complex, the more sophisticated approaches using combinatorial geometry. To date, targets have received considerably more attention than background; a compatible and equally capable background modeling capability is needed.

## OBJECTIVES AND APPROACH

The study described herein was designed to generate a capability to realistically predict the temperature histories of natural and cultural features that commonly comprise the backgrounds to targets. With such a capability, it would be possible to examine the temperature contrasts that occur between targets and background features both with time and changing weather conditions. This in turn provides basic information needed to examine the performance of existing or proposed target acquisition devices and the effectiveness of alternative camouflage measures.

In the following paragraphs, two temperature prediction models are presented, one for terrain surface features and one for vegetation canopies. The terrain surface model was developed at the U. S. Army Engineer Waterways Experiment Station (WES), while the vegetation canopy model was developed at the Colorado State University under contract to WES. A brief description is given of each model followed by a discussion of model sensitivities and sample applications.

## TERRAIN SURFACE TEMPERATURE MODEL

### Philosophy

The Terrain Surface Temperature Model (TSTM) was developed to estimate the temperatures of actual or hypothetical material systems and for actual or hypothetical weather conditions. A premium was placed on simplicity and flexibility with respect to operational constraints. In short, a model was needed that considered the dominant physical phenomena that influence material temperatures and yet be reasonable to use. The model handles sensible heat transfer, latent heat transfer, the impact of cloud type and cover, and seasonal/geothermal heat fluxes. A brief description of the model framework is given in the next section; a complete description is available in Reference 1.

Mathematical Framework

The model predicts surface temperatures for a multilayered (1-6 layers) system by determining energy transfer in, out, and through the system. The model assumes that the major energy fluxes are vertical (i.e., perpendicular to the layers) and that the layers are horizontally uniform. Temperature estimates result from solving the one-dimensional heat transfer equation:

$$\alpha^2 \frac{\partial T(z, t)}{\partial z^2} = \alpha(z) \frac{\partial T(z, t)}{\partial t}$$

subject to the boundary conditions

$$\sum_{i=1}^{n} b_{it} = o \quad \text{at} \quad z = 0 ; \qquad \sum_{i=1}^{n} B_{it} = o \quad \text{at} \quad z = B$$

where the observable surface is $z=0$; the lower surface is $z=B$; $\alpha(z)$ is the diffusivity; and both $b_{it}$ and $B_{it}$, $i=1, 2, \ldots n$, denote heat fluxes at time $t$. An example geometry is shown in Figure 1. Within a layer, a simple explicit finite-difference technique is used, while at boundaries and interfaces a Newton-Raphson iteration scheme is applied.

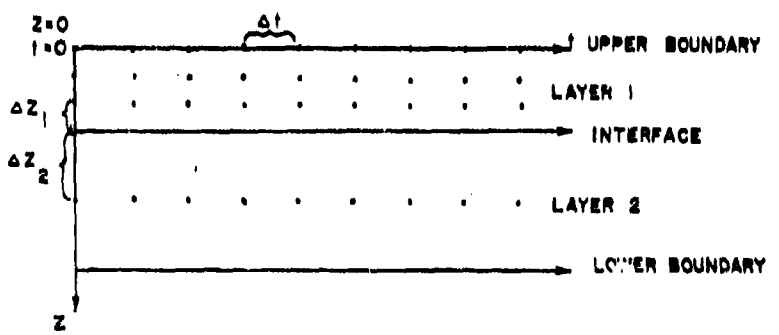



Figure 1. Sample layer geometry for TSTM

The surface boundary condition is estimated with the following heat balance equation:

$$S + I\downarrow - H - E - I\uparrow + G = 0$$

where

S = solar insolation
I↓ = radiant energy from the atmosphere and clouds
H = sensible heat
E = latent heat
I↑ = radiant energy emitted by the top surface
G = heat conduction into the material

Solar insolation can be input as measured values or estimated using a procedure that was adapted by Balick (1) from the work of Small (2) and Sellers (3).

The Brunt equation (3) is used to estimate radiant energy from the atmosphere. Cloud contributions are treated with an empirical factor adapted from Geiger (4), and both cloud type and the amount of cloud cover are considered. Ground radiation energy loss is treated using conventional grey body emitter theory. Sensible heat loss is estimated by an equation following Lamb (5) that provides the operational advantage of not requiring roughness characteristics for the surfaces being modeled. Latent heat loss is modeled after Lamb (5) with the addition of a saturation factor which allows dry to saturated moisture conditions on the surface.

The bottom boundary can be described with three options: a constant temperature; a constant heat flux or a constant heat flux with an airspace below the bottom boundary; and an additional constant radiating surface below the airspace.

Model Inputs

Inputs to the model include atmospheric constants, atmospheric hourly data, surface-sun orientation, initial temperature profile, and material properties. Atmospheric constants required are atmospheric pressure (mb), dust content (particles $cm^{-3}$), precipitable water (mm), wind speed (m $sec^{-1}$), cloud type index and meteorological instrument shelter height above the surface (cm). The shelter height value represents the height above the ground that air temperature and wind speed are measured.

Atmospheric hourly data required for the 24-hour diurnal cycle forecast include air temperature ($^{o}C$), relative humidity (%), cloud cover (tenths, 0.0 - 1.0), wind speed (m $sec^{-1}$), and total insolation (cal $cm^{-2}$ $min^{-1}$). Solar insolation can also be computed as previously mentioned.

Material properties are needed for the surface and each of the layers. Surface properties required are thermal emissivity, optical absorptivity, and percent saturation of the surface. Each layer is defined by its thickness (cm), thermal diffusivity ($cm^2$ $min^{-1}$), and thermal conductivity (cal $min^{-1}$ $cm^{-1}$ $^{o}K^{-1}$).

## Model Output

The principal model output is the temperature of the surface with time. Any time step could conceivably be used; however, 15 to 60 minutes appears to be the most useable range. Input or computed values of solar insolation, energy absorbed, atmospheric IR emission, surface convection, and evaporative heat loss are printed out to assist in evaluating the predicted temperature data.

## Parameter Sensitivity and Example Output

A sensitivity analysis was accomplished to examine model output behavior with systematic changes in model inputs. A study of the results showed that the model output was by far most sensitive to air temperature. Other parameters that when changed created significant changes in the output included the conductivity and emissivity of the surface layer, cloud cover, and surface absorptivity. Changes in conductivity and emissivity equally affected daily minimum and maximum temperatures as did changing cloud cover type, while changes in absorptivity affected the daily maximum temperatures much more severely than the daily minimums.

The TSTM has been validated using weather data from various locations in the United States and the Federal Republic of Germany (FRG). Figure 2 shows a comparison of predicted and measured temperatures for a 15-cm-thick concrete pad. The concrete pad was modeled as a two-layer system with the pad underlain by a 700-cm layer of soil. The measured data were obtained with a thermistor

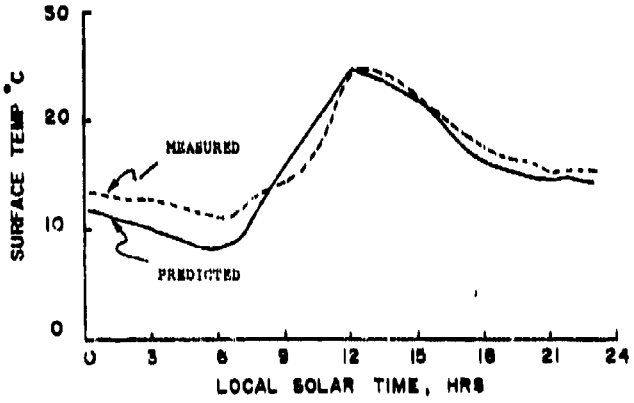



Figure 2. Predicted and measured surface temperatures for a 15-cm-thick concrete pad, 4 Oct 1979, in Federal Republic of Germany

attached directly to the surface of the concrete pad. The agreement between the predicted and measured curves in the figure is quite good, especially between about 0800 and 2400 hours. The difference in the curves from approximately 0-time to 0800 hours could have resulted from an inaccurate estimate of the concrete conductivity (no measured values were available) or increased cloud cover at night (cloud cover data were not available for the nighttime).

## VEGETATION CANOPY THERMAL MODEL

### Philosophy

The Vegetation Canopy Thermal Model (VCTM) was developed to approximate the thermal behavior of a layered vegetation canopy by a mathematical abstraction of the material and geometry characteristics of the canopy and the energy transfer mechanisms that occur there. The model is physically based and considers geometric arrangement of canopy elements, scattering of direct and emitted energy within the canopy, increased absorption of elements due to the thermal emissions of neighboring elements, and the directional variation of energy radiated from the canopy.

The vegetation canopy is abstracted as three statistically independent infinite horizontal layers as illustrated in Figure 3. Within each layer, the leaves, branches, and other canopy elements are described as a statistical ensemble giving their orientations and number densities. An energy budget equation is formulated for each layer that accounts for inflow and outflow of energy. The roots of the resulting system of equations are the average surface temperatures in the layers.

The VCTM assumes steady-state conditions. Time-dependent events are modeled by incremental changes in steady-state energy flow. Spectral structure in the thermal wavelengths is not considered and scattering of thermal energy within the canopy is neglected. In addition, individual canopy elements are considered to radiate thermal energy in an isotropic manner. A detailed discussion of the VCTM is given in Reference 6.

### Mathematical Framework

The mathematical framework for the VCTM is designed to handle individually the effects of canopy geometry, thermal radiation transfers, solar radiation absorption, thermal existance, transpiration, and convection. The values computed from these operations allow calculation of the total energy budget for each canopy layer and the thermal existance from the canopy.

The most important aspect of canopy geometry for describing radiation transfer is the frequency of gaps in the canopy and the extinction of radiation within the canopy. To compute gap frequency

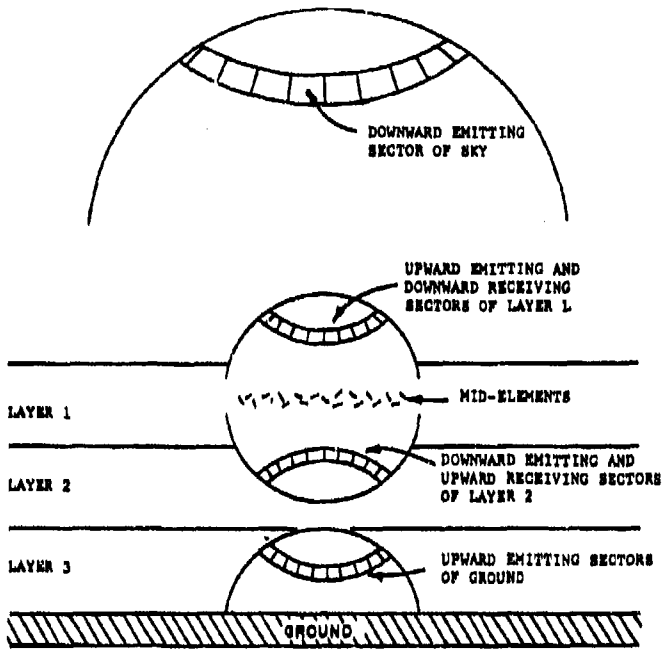



Figure 3. Abstraction of VCTM showing canopy layers, sky, ground, and hemispherical sector concept for computing energy transfer

and extinction, the hemispheres above and below a canopy layer are discretized into 9 hemispherical inclination bands. Each of these bands are further discretized into 18 azimuthal sections (see Figure 3). The radiation transfers between the three canopy layers, the ground, and the sky are calculated within each sector. An in-depth discussion of the theory involved is given in Reference 7.

Thermal radiation transfer is handled by allowing each layer to emit and receive thermal radiation in the hemisphere occurring above and below it. The computations are first made for a component in the middle of a layer, termed the mid-element. The equation that calculates for a particular sector, the flux density absorbed by a mid-element at a particular inclination angle from any given source layer $\frac{\phi_{ijkim}}{m^2}$ is

LINK

$$\frac{\phi_{ijkim}}{m^2} = \sigma \cdot \varepsilon_i \cdot T_i^4 \cdot CONT_{jim} \cdot SECTOR_j \cdot ABSORB_m \cdot COS_{ijk}$$

where:

$\sigma$ = Stefan-Boltzmann constant

$\varepsilon_i$ = average emissivity of elements in layer $i$

$T_i$ = true average surface temperature ($^{o}K$) of the mid-elements in layer $i$ (unknown)

$CONT_{jim}$ = contributing coefficient for mid-elements in layer $m$ absorbing flux from elements in layer $i$ for all sectors within hemispherical band $j$

$ABSORB_m$ = average thermal absorption coefficient for elements in layer $m$

$SECTOR_j$ = quantity $(\sin \Theta_2 - \sin \Theta_1)/9$ defining the inclination limits of sector $i$ in hemispherical band $j$

Within source sector $i$ in hemispherical band $j$, $\frac{\phi_{ijkim}}{m^2}$ is the thermal flux density absorbed by a mid-element in layer $m$ inclined at inclination angle $k$ from source elements in layer $i$ represents the sky and ground in addition to the three canopy layers.

The total flux density emitted by elements in layer $i$ and absorbed by a particular mid-element in layer $m$ at an inclination $k$ is computed by summing the product $CONT_{jim} \cdot SECTOR_j$ over $j$ (from 1 to 9) and $COS_{ijk}$ over $i$ (from 1 to 18). The total flux density absorbed by a mid-element in layer $m$ at inclination $k$ is computed by summing all sources

$$\frac{\phi_{km}}{m^2} = \sum_{i=1}^{5} \frac{\phi_{kim}}{m^2}$$

where $i$ = 1, 2, 3, 4, 5 represents the sky, layer 1, layer 2, layer 3, and the ground, respectively. Nine equations for each layer are constructed. For each layer the appropriate equation is weighted by the frequency of occurrence of elements within the corresponding inclination class. The nine equations are summed to compute the average absorbed thermal flux density within the three canopy layers.

Solar radiation absorption is handled using a stochastic model, Solar Radiation Vegetation Canopy Model (SRVC), developed at the

Colorado State University (7). The SRVC predicts the diurnal apparent directional reflectance of a vegetation canopy and allows realistic consideration of the complex scattering and absorption of light as a function of canopy geometry. The thermal exitance of all canopy mid-elements and the ground is calculated by the Stefan-Boltzmann Law.

Transpiration is handled using a relation developed by Gates (8). The driving force is the difference between water vapor density within the leaf and in the free atmosphere beyond the boundary layer. Physically based formulas are available to compute the necessary parameters with the exception of leaf resistance to water vapor diffusion.

The convection equation developed by Tibbals et al. (9) is used to describe forced convection. Sky thermal exitance is calculated using an empirical equation dependent only on air temperature near the ground surface.

The pieces for computing the total energy budget for each canopy layer have not been described. These relations result in a system of three nonlinear equations and three unknowns. A least quadratic convergent numerical routine (10) is used to solve for the roots of the equations, the average temperature of the canopy layers. The model also predicts the effective radiant temperature (ERT) and equivalent thermal exitance in the nine viewing inclination bands at ten-degree intervals above the canopy. The contributions of each canopy layer and the ground are considered in the calculation. The ERT for a sensor looking horizontally into the canopy is also predicted using the Stefan-Boltzmann Law with appropriate emissivity and average layer temperature values.

Model Inputs

Inputs to the model include environmental factors, canopy geometry descriptors, and thermal and optical properties of canopy components. The environmental factors are entered on an hourly basis; all other parameters are considered static.

Environmental factors required are air temperature (within the canopy), ground temperature, relative humidity (within the canopy), wind velocity (within the canopy), and total incoming solar irradiance above the canopy.

Canopy geometry parameters include Leaf Area Index (LAI), Leaf Angle Distribution (LAD), Branch Area Index (BAI), and Branch Angle Distribution (BAD). The LAI is the total one-sided leaf area that occurs over a unit of ground area. Values can be measured or derived from the literature. The BAI is derived by measuring the length and width of tree limbs at various points in the canopy and using conical and cylindrical approximations. The values for LAD and BAD are obtained from photography using optical Fourier transform techniques. In practice, values for LAI, BAI, LAD, and BAD can be estimated from

the literature although values for noncrop vegetation types are limited.

Inputs describing thermal properties of canopy and ground components are thermal absorption and emissivity and leaf resistance to water vapor diffusion. Optical properties needed are spectral absorption and transmission of leaves (or needles) and the spectral reflectance of the ground surface.

Model Outputs

The principal output of the model is the average temperature for each canopy layer and the ERT of the canopy as a function of view angle above the canopy for each time increment. In addition, values of ground thermal exitance, sky thermal exitance, absorbed solar flux density of each layer, thermal exitance of each layer, absorbed thermal flux of each layer, convectional exchange, and transpirational exchange are displayed in the output.

Parameter Sensitivity and Example Output

Sensitivity analysis were run to examine the impact of systematic changes in parameter values for both daytime and nighttime conditions. The data used were obtained at Leadville, Colorado, on 15 and 16 July 1977.

Results of the sensitivity analysis showed the following:

a. Within a reasonable range (0.96 - 1.00), changes in emissivity did not significantly change the average layer temperature (less than 1-deg change).

b. Within a range of 0.3 - 1.2 min $cm^{-1}$, a change in leaf resistance to water vapor diffusion had only a small impact on average layer temperature. At lower values, such as 0.15 min $cm^{-1}$, which may be appropriate for a full sun condition on a summer day for conifers, the model output becomes much more sensitive to this parameter.

c. A significant change in canopy geometry for lodgepole pine did not significantly impact the thermal radiation transfer within the canopy. However, canopy geometry did clearly affect the contribution of thermal radiation from each layer to the ERT above the canopy as a function of view angle.

d. Air temperature is the single most important parameter for predicting the average temperature of components in each layer.

An example of the model output is shown in Figure 4. The example shows simulated average layer temperatures for a lodgepole pine using environmental data obtained in Leadville, Colorado, on 15 and 16 July 1977. The simulated values are compared to radiometric temperature measurements obtained with a hand-held radiometer. Additional outputs are available for Douglas fir and oak-hickory canopies.

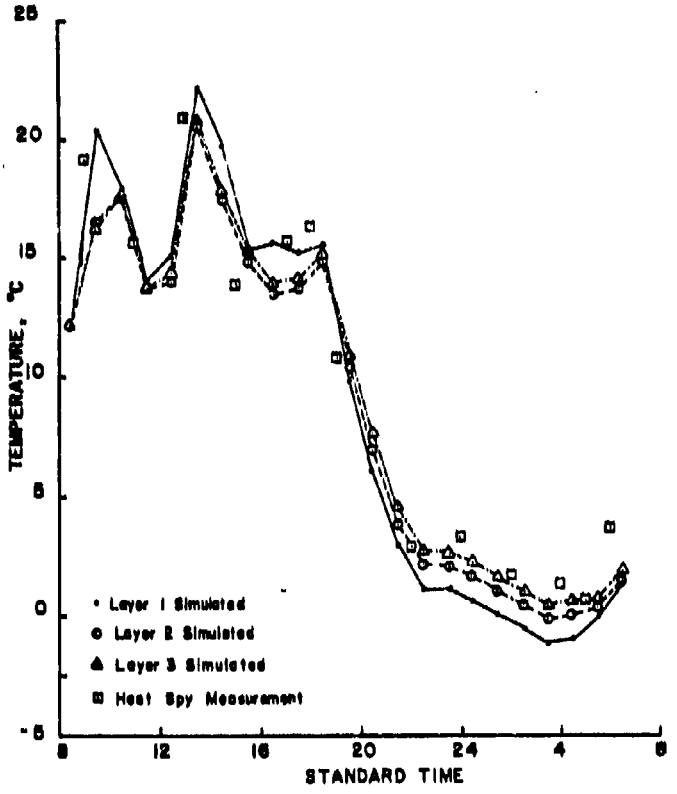



Figure 4. Simulated and measured (radiometric) temperatures for lodgepole pine for 15-16 Jul 1977 in Leadville, Colorado (horizontal perspective of canopy)

## MODEL APPLICATIONS

Models such as the TSTM and VCTM open new avenues to study target acquisition and camouflage problems. An initial concept and prototype products for a comprehensive data base-modeling capability for target acquisition sensor development and evaluation is presented by the author in reference 11. The concept includes a terrain and weather data base from which the basic inputs to the TSTM and VCTM can be formulated. The weather data and associated terrain input are used to compute the range of temperature expected for specific terrain features in a given season. Figure 5 shows an example generated for a grassy area in Fulda, FRG, for the summer season. Curves, such as those shown in Figure 5, can be compared with similar forecasts or measured data on targets to determine under what conditions and time-of-day the target will contrast most with the background.

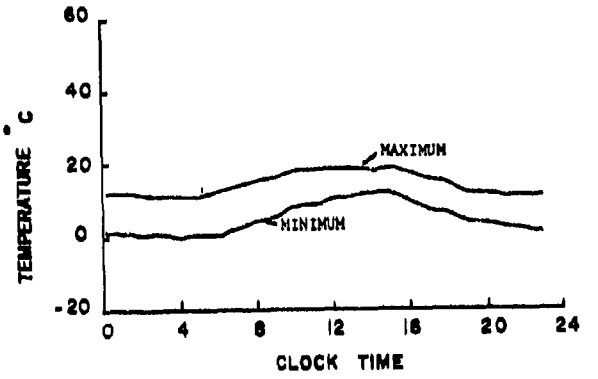



Figure 5. Example of output of VCTM for grass in the Fulda, FRG, area. The curves define the range of temperatures that the grass is likely to experience during summer conditions

Thermal models have a valuable place in developing fixed-installation camouflage design criteria. Figure 6 illustrates how the TSTM has been used to study the effectiveness of alternative camouflage measures in reducing temperature of key targets on fixed facilities. In the example, the impact of painting a roof surface with solar-reflecting paint is demonstrated.

More complex applications of the TSTM and VCTM are also ongoing. The WES is directing a NATO thermal camouflage field trial in the FRG.

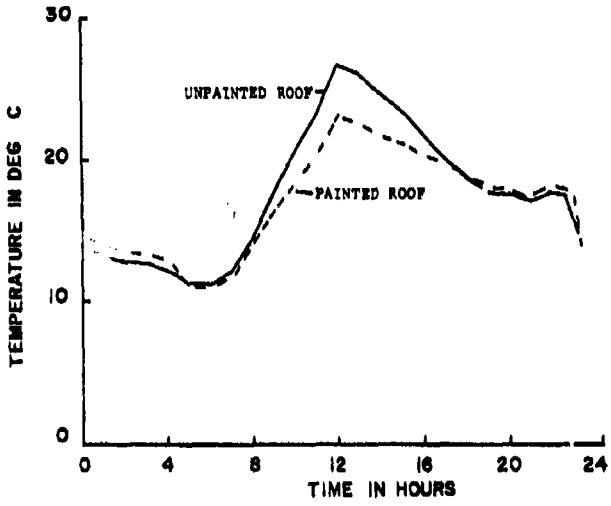



Figure 6. Use of TSTM to examine effectiveness of solar-reflecting paint for camouflage of fixed installations

In this experiment, the above models are being used in conjunction with image transformation techniques to study the effectiveness of alternative camouflage measures. Thermal images, obtained by ground-based and airborne sensors and showing key facilities at a military base as they appear uncamouflaged, are transformed to illustrate how they would appear with alternative camouflage measures and under different time and weather conditions.

## CONCLUSIONS

The models presented herein represent a new dimension in the quantitative consideration of background features for target acquisition and camouflage applications. The model inputs include environmental factors that are for the most part available at recording meteorological stations and geometry and material properties that can be measured or derived from the literature. A gap exists in specific geometric descriptors for vegetation canopies; however, emphasis is being placed on generating the needed data.

Although the TSTM and VCTM are one-dimensional models, they can be used to realistically examine background features in a variety of situations. Finite-element methods are being examined to handle vertical walls, and a three-dimensional vegetation model has been

initiated under a U. S. Army Research Office grant to handle sophisticated open canopy vegetation conditions.

Current applications of the models emphasize development of criteria for thermal camouflage of fixed installations. The thermal models provide a major advantage in this effort because the potential effectiveness of alternative camouflage measures can be judged before expensive field applications.

## References

1. Balick, L., "Thermal Modeling of Terrain Surface Elements," USAE Waterways Experiment Station, Technical Report in Preparation, Vicksburg, MS.

2. Small, J. T., "A Theoretical Analysis of Changes in Thermal Signatures Caused by Physical and Climatological Factors," Masters Thesis, Air Force Institute of Technology, Air University, 1977.

3. Sellers, W. D., "Physical Climatology," University of Chicago Press, Chicago, IL, 1965.

4. Geiger, R., "The Climate Near the Ground," Translated by Scripts Technical, Inc., Harvard University Press, Cambridge, MA, 1966.

5. Lamb, R. C., "The Radiation and Energy Balance on a Burned vs. an Unburned Natural Surface," Conference on Fire and Forest Meteorology, American Meteorological Society and Society of American Foresters, Lake Tahoe, CA, 1974.

6. Kimes, D. S., Smith, J. A., and Ranson, K. J., "Terrain Feature Canopy Modeling," Final Report to U. S. Army Research Office, Grant No. DAAG 29-78-G-0045, Colorado State University, Fort Collins, CO, 1979.

7. Oliver, R. E., and Smith, J. A., "A Stochastic Canopy Model of Diurnal Reflectance," Final Report prepared by Colorado State University for the U. S. Army Research Office, Durham, NC, 1974.

8. Gates, D. M., "Energy Exchange in the Biosphere," Harper and Row, Inc., NY, 1968.

9. Tibbals, E. C., Carr, E. K., Gates, D. M., and Kreith, F., "Radiation and Convection in Conifers," American Journal of Botany, Vol. 15, No. 5, 1964.

10. Brown, K. M., "A Quadratically Convergent Newton-Like Method Based Upon Gaussian Elimination," SIAA Journal on Numerical Analysis, Vol. 6, No. 4, 1969.

11. Link, L. E., Jr., "Thermal Model of Battlefield Scence Components," Miscellaneous Paper EL-79-5, USAE Waterways Experiment Station, Vicksburg, MS, 1979.

# FINDING EDGES IN NOISY SCENES

*RAUL MACHUCA, PhD.
ALTON L. GILBERT, PhD.
INSTRUMENTATION DIRECTORATE
WHITE SANDS MISSILE RANGE, NEW MEXICO 88002

Research into methods of identifying edges in a noisy scene has been an active field of investigation for many years. Treatment of the subject may be found in many books written over the past decade ([1], [2], and [3]) and many different approaches are proposed. Recently a survey and comparative analysis of the methods was made [7].

The body of this paper is segmented into 4 parts. In the first we derive and define a "Moment Operator" which we show to work well for step and ramp edges. Then, we define and characterize second order edges using the concept of the rotation of a point in a vector field and develop the detector analytically. In Section 3 we develop the algorithms for implementing the previously defined operators. Finally, in Section 4, these algorithms are evaluated using ROC curves and compared with previously known techniques.

The detection of edges to isolate objects in a scene is motivated by many distinct problems. One such problem arises in a tracking system where the input video image is analyzed and the object to be tracked identified. Subsequent input and feedback to the drive controls causes the sensor to re-orient to a new position in an attempt to maintain the same x-y coordinate position for the object in the field of view. While this problem motivated the research that led to this paper, the results herein discussed are much broader in scope and application. The constraints imposed by this problem led to a method that is useful in high data throughput systems.



PRECEDING PAGE BLANK-NOT FILMED

## SECTION 1. EDGES FROM MOMENTS

First order edge detection methods work in the following way: A picture function f(x,y) is transformed to another picture function F(x,y) = Tf(x,y) in such a way that the edges of objects in the scene will be in the set $\{(x,y):F(x,y)\geq W\}$ for some W. The usual method is to transform the picture using T equal to the gradient operator. Different edge detection methods correspond to different numerical approximations to the gradient.

The method used in our edge detection program is not based on derivatives. To reduce the effect of noise, this edge detection method uses integrals.

The reasoning for the use of moments to find edges is as follows. A digitized picture can be thought as a lamina whose density at each point is f(x,y), so points of high intensity correspond to points of high density. A point (a,b) on an edge in the original function (see Figure 1) would correspond to a point in this lamina (digitized picture) with high densities on one side and lower densities on the other side. Thus if we look at a small lamina centered at point (a,b) and compute the center of mass of this small lamina, we can expect the center of mass to lie within an area of high densities.

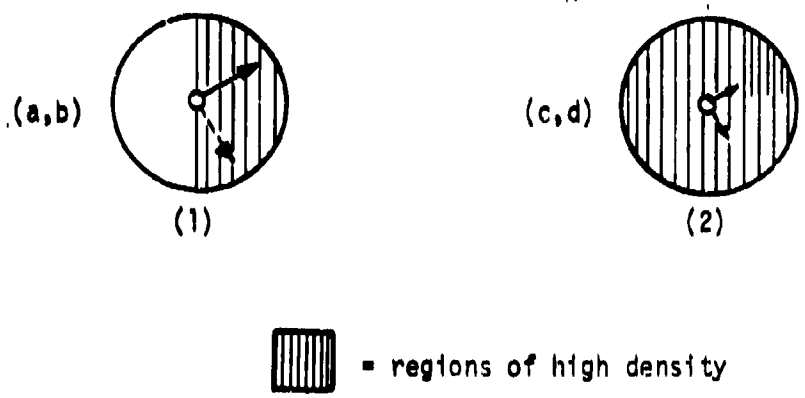



Figure 1. Example Center of Mass Vectors for (1) and Edge and (2) a Region of Uniform Intensity.

Suppose we now look at a point (c,d) such that the densities around it are fairly constant. Then the center of mass of a small lamina about it would be close to (c,d). In this case, a vector from (c,d) to the center of mass would be very small compared to a vector from (a,b) to the center of mass in the previous case.



(a) IMAGE OF ROCKET AND PLUME. THE PLUME IS THE LARGE REGION OF HIGHEST INTENSITY.



(b) RAMP AND STEP EDGES FOUND BY USING THE MOMENT OPERATOR.



(c) THE VECTOR FIELD GENERATED BY THE MOMENT OPERATOR.



(d) SECOND ORDER EDGES DETECTED BY USING THE VECTOR FIELD.

Fig. 2

We conclude that one way to transform f(x,y) to F(x,y) such that edges of the original picture lie in the set $F(x,y) \geq W$ is to replace every f(x,y) by the length of the vector from (x,y) to the center of mass of a small lamina centered about (x,y). That is, F(x,y) is the magnitude of the vector from (x,y) to the center of gravity of a square lamina centered at (x,y) whose density is given by the picture function f(x,y).

Figure 2(b) is an example of how this method works on a scene (Figure 2(a)) typical of those we study at WSMR.

Once the coordinated $(\bar{X},\bar{Y})$ of the center of mass of a lamina about (x,y) are calculated, the direction of the edge (if any) can easily be found. Since $(\bar{X},\bar{Y})$ points to where the intensity of the picture is the highest, the direction of the edge is perpendicular to the direction of the vector from (x,y) to $\bar{X},\bar{Y}$. If we take (x,y) = (0,0), then the direction of the edge is $\theta = \text{Arctan}\ (\bar{Y}/\bar{X}) + \pi/2$.

Thus this model gives for each point in the scene a quantity that measures the probability that a point is an edge point and a direction which is the direction of a possible edge through that point.

The model introduced in Section 1 will not work for roof edges. This is because at the very peak of the roof, exactly where the edge is situated, both $\bar{X}$ and $\bar{Y}$ are equal to zero. In order to detect roof edges we need to take advantage of the direction information, and as Figures 6(a), (b) and (c) show we need to detect the shearing cause by the change in direction of the vector field at the edge points. One way of doing this is by using a tool from the theory of vector fields, namely the rotation of a vector field about a point.

## SECTION 2. SECOND ORDER EDGES

After a scene is processed by the moment edge detector, each point is assigned a direction and a magnitude. In effect this specifies a vector at each point of the plan in question; i.e., these vectors define a vector field over the scene. An important tool in the study of vector fields is the rotation of a vector field (see [4] and [5]). To define the rotation of a vector field, suppose a vector of the vector field $\Phi$ at the point (x,y) is given by

$$\Phi(x,y) = \{\phi(x,y),\ \psi(x,y)\}$$

$$\phi(x,y) = \bar{X}(x,y)$$

$$\psi(x,y) = \bar{Y}(x,y)$$

If a curve $\Gamma$ on the plane (scene) is given in the form

$$\Gamma: \quad x = x(t),\ y = y(t) \qquad a \leq t \leq b$$

then $\Phi(t) = \{\phi[x(t), y(t)], \psi[x(t), y(t)]$ is defined on the interval $[a,b]$ (see Figure 3).

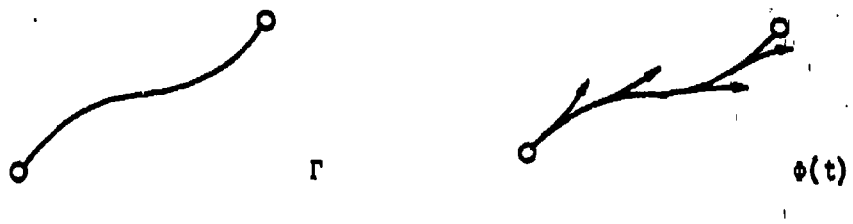



Figure 3. A Curve $\Gamma$ and It's Corresponding Vector Field $\Phi(t)$

For each $t \in [a,b]$ there is determined an angle, the angle in radians between $\Phi(t)$ and $\Phi(a)$ measured from $\Phi(a)$ to $\Phi(t)$. This angle is a many valued function of t. The continuous branch of this function (vanishing for $t = a$) is designated by $\Theta(t)$ and called an angular function of the field $\Phi$ on a curve $\Gamma$. The rotation of the field $\Phi$ on the curve $\Gamma$ is defined to be

$$\gamma(\Phi,\Gamma) = \frac{1}{2\pi}\ [\Theta(b) - \Theta(a)]$$

If $\Gamma$ is a closed Jordan curve, then the rotation is found by subdividing $\Gamma$ into two curves (not closed), computing the rotation of each, and adding. In the following, $\Gamma$ is taken to be a small circle about a point.

We can write the rotation as

$$\gamma = \frac{1}{2\pi}\ [\Theta(b) - \Theta(a)] = \frac{1}{2\pi} \int \frac{d\Theta(t)}{dt}\ dt.$$

With $\Theta(t) = \text{Arctan } \overline{Y}/\overline{X} + \pi/2$, we make the following observations:

(1) If $\Theta(t) =$ constant, then $\frac{d\Theta(t)}{dt} = 0$ and $\gamma = 0$. So $\gamma = 0$ when $x =$ a point on the edge of an object in a scene (see Figure 4).

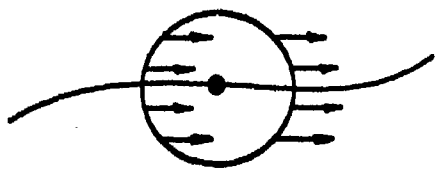


Figure 4. Vector Field at a Step or Ramp Edge Point

(2) If $\Theta$ is symmetric about x and $\Gamma$ is a small circle about x = edge point on a roof edge, see Figure 5, then with $\Gamma = \Gamma_1 + \Gamma_2$ (where $\Gamma_1$ = one half of the circle and $\Gamma_2$ = the other half)

$$\int_{\Gamma} \frac{d\Theta(t)}{dt}\, dt = \int_{\Gamma_1} d\Theta(t) + \int_{\Gamma_2} d\Theta(t) = \pi + \pi = 2\pi$$

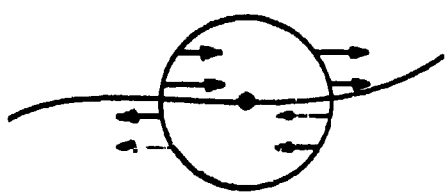


Figure 5. Vector Field at a Roof Edge Point.

An example of how these observations can be used to detect second order edges appears as Figure 6(c) and (d).

(a) ORIGINAL IMAGE OF DISK WITH A ROOF EDGE.

(b) VECTOR FIELD GENERATED BY APPLYING MOMENT OPERATOR TO 2(a).

(c) EDGE POINTS OF 2(a) FOUND BY IDENTIFYING THOSE POINTS FOR WHICH $\int_C d\theta = 2\pi$

Fig. 6

## SECTION 3. ALGORITHMS FOR IMPLEMENTATION

### A. Calculation of Moment

Since we are interested in real time applications of these methods we simplify the calculation of $\bar{X}$ and $\bar{Y}$ by setting

$$M = \int_{-h}^{h} \int_{-k}^{k} f\ (x + T,\ y + u)\ dt\ du = 1$$

This can be justified by observing that M/4hk is the average of the intensities over a small neighborhood of (x,y) and so this value can be approximated by the average value of intensities over the entire picture. This would then be just a scale factor and so could be left out.

To calculate the integrals involved we use an integral formula [6] of order $O(h^6)$. The formula for integration is

$$\int\int F(x,y) = \sum_{i=1}^{9} Wi*Di \text{ with } W_1 = 25/324,\ W_{21} = 10/81$$

and if we apply this to the integrals for $\bar{X}$ and $\bar{Y}$ and factor out all scale factors we get

$$\bar{Y} = 5 * (D1 - D5) + 4 * (D8 + D2 - D6 - D4)$$

$$\bar{X} = 5 * (D7 - D3) + 4 * (D8 + D6 - D2 - D4)$$

and use abs $(X)^2$ - abs $(Y)^2$ for the associated magnitude. If we sweep a 3 x 3 window across digitized scene D7 can be taken as the upper left hand corner while D3 is the lower right hand corner. In this case the direction of a possible edge is equal to

$$\Theta = \text{Arctan}\left(\frac{\bar{Y} + \bar{X}}{\bar{Y} - \bar{X}}\right) + \pi/2$$

### B. Calculation of the Rotation

The vector field of a roof edge will look like the vector field of Figure 5. So to find roof boundary points we have to find points for which in a neighborhood of such a point $\int_c d\Theta = 2\pi$

The smallest region, in the discrete case over which we can take an integral is a 2 x 2 window. Thus our algorithm sweeps a 2 x 2 window

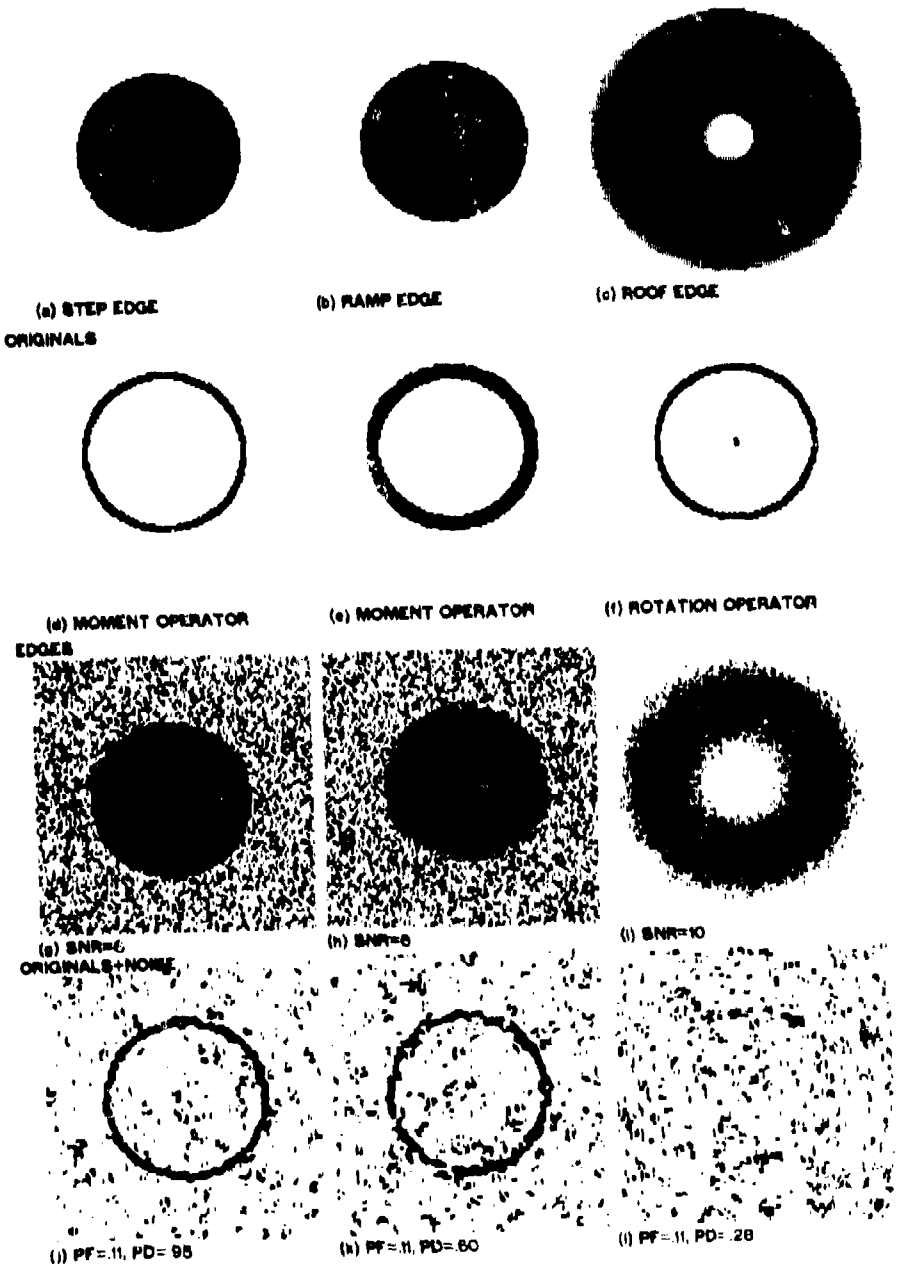
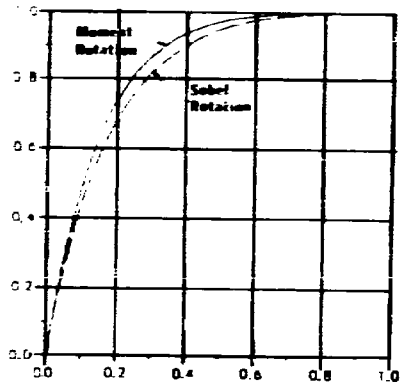
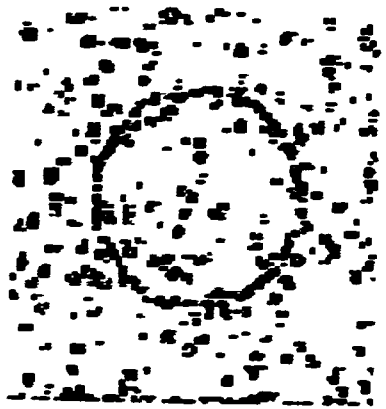

(a) STEP EDGE
(b) RAMP EDGE
(c) ROOF EDGE
ORIGINALS
(d) MOMENT OPERATOR
(e) MOMENT OPERATOR
(f) ROTATION OPERATOR
EDGES
(g) SNR=6
(h) SNR=6
(i) SNR=10
ORIGINALS+NOISE
(j) PF=.11, PD= 98
(k) PF=.11, PD=.60
(l) PF=.11, PD=.28

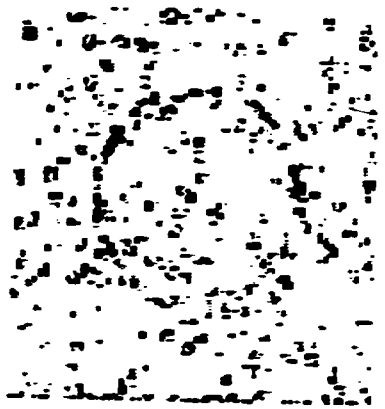
Fig. 7




(a) ROC CURVE FOR ROOF EDGE AT SN=13. MOMENT EDGE OPERATOR AND SOBEL EDGE OPERATOR APPLIED AFTER A 3x3 AVERAGING.



(b) THIS ROOF EDGE WAS FOUND USING MOMENT OPERATOR. SNR=13, PF=.10, PD=.58.



(c) ROOF EDGE WHICH WAS FOUND USING SOBEL OPERATOR. SNR=13, PF=.10, PD=.51.

Fig. 8

across a scene and computes the integral $\int_c d\Theta$ for each of these windows. If it turns out that this integral is equal to $2\pi$ then those 4 points which make up the window are classified as boundary points. To calculate the integral of a 2 x 2 window we use an approximation

$$\int d\Theta = \sum_{\ell=1}^{4} \Delta_\ell \Theta$$

computed by a computer program given in Reference [10].

For the purposes of this experiment the procedure used to generate a file which is the file of detected second order edges in the following:

1. From the original file (scene) two files are generated; one (ACI) contains SQRT $(\overline{X})^2 + (\overline{Y})^2$; and the other (ANG) the angle of $(\Theta,\ 0 \leq \Theta \leq 2555)$ a possible edge.

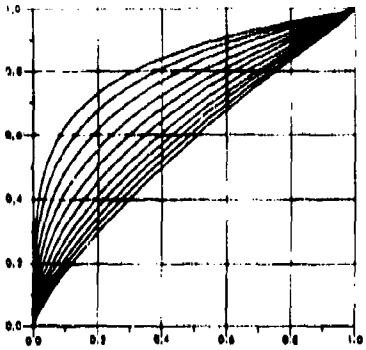
2. From the ANG and ACI files one new file AAA is created. AAA is created by sweeping a 2 x 2 window across the ANG file. The rotation is calculated, and if a point is classified as boundary then to the corresponding point of AAA (initialized at zero) is added the average of those elements of ACI that have the same subscripts as those of the 2 x 2 window being swept across ANG.

Examples of how this method works are Figures 6(c) and 2(d).

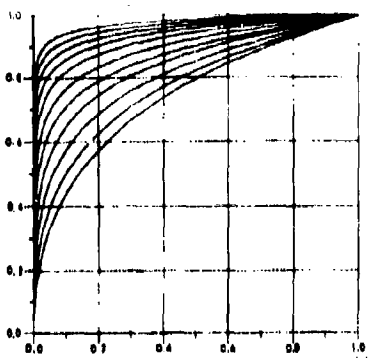
## SECTION 4. EVALUATION

The methods described above were tested on disks whose edges were step, ramp and roof edges. The step and ramp edges had edge height equal to 16 while the roof edge was constructed by beginning at the center with gray value equal to 100 incrementing by one to gray value equal 132 and then decrementing by one to gray value 100. All files were 128 x 128 x 8.

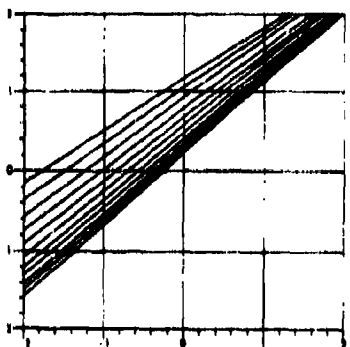
To test the effectiveness of the different operations considered here we added Gaussian noise of different standard deviation to achieve a given signal to noise ratio and then tested the algorithms (Figure 7). The SNR ratio was measured in db; that is, we used SNR = $10 \log_{10}(\frac{16}{\sigma_n})^2$ where $\sigma_m$ = standard deviation of the noise. For the ramp and step edges we used SNR = 4, 5, 6, ...., 14 while for the roof edge the



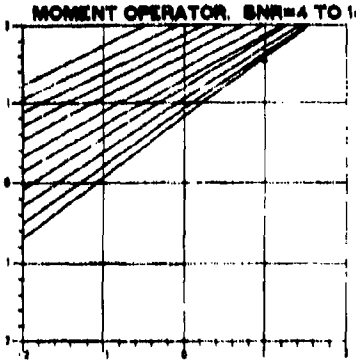
(a) ROC CURVES FOR SOBEL OPERATOR ON RAMP EDGE DISK. SNR=4 TO 14.



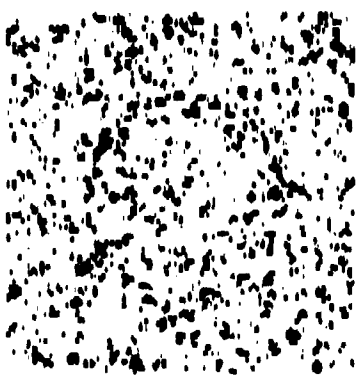
(b) ROC CURVES FOR RAMP EDGE DISK FIRST PROCESSED BY TAKING AVERAGE (3x3) AND THEN BY MOMENT OPERATOR. SNR=4 TO 14.



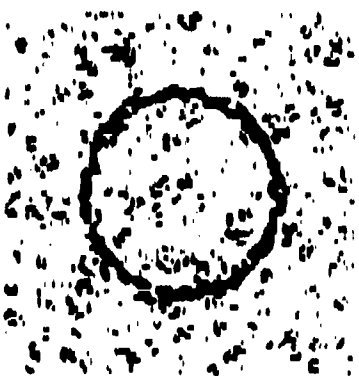
(c) ROC CURVES OF FIGURE 11(a) AFTER BEING NORMALIZED.



(d) NORMALIZED ROC CURVES FOR AVERAGE-MOMENT OPERATOR FIGURE 11(b).



(e) EDGE DETECTED USING SOBEL OPERATOR ALONE. SNR=6, PF=.11, PD=.23.



(f) EDGE DETECTED BY FIRST DOING 3x3 AVERAGING AND THEN USING MOMENT OPERATOR. SNR=6, PF=.11, PD=.60.

Fig. 9

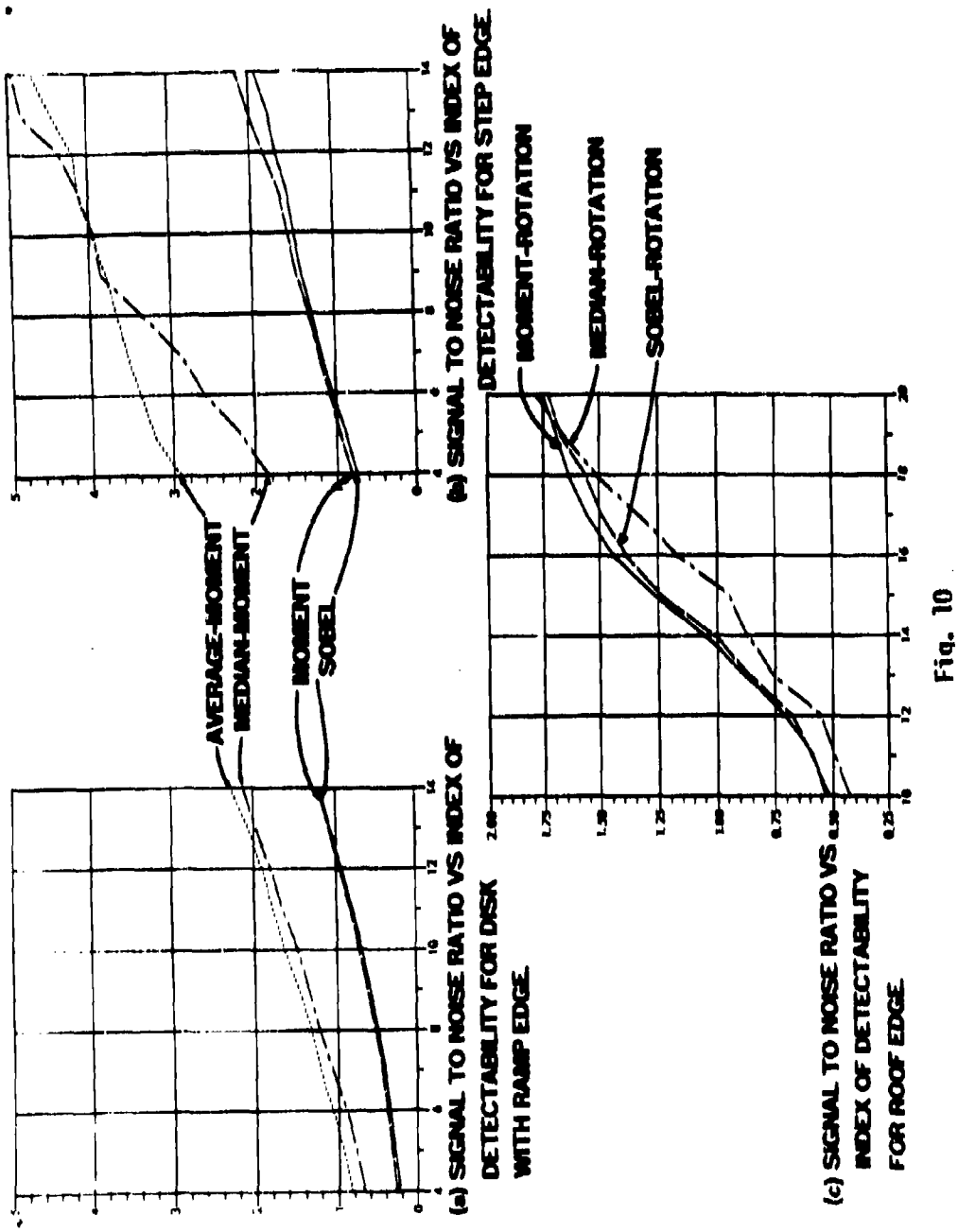

AVERAGE-MOMENT
MEDIAN-MOMENT
MOMENT
SOBEL
(a) SIGNAL TO NOISE RATIO VS INDEX OF DETECTABILITY FOR DISK WITH RAMP EDGE
(b) SIGNAL TO NOISE RATIO VS INDEX OF DETECTABILITY FOR STEP EDGE
MOMENT-ROTATION
MEDIAN-ROTATION
SOBEL-ROTATION
(c) SIGNAL TO NOISE RATIO VS INDEX OF DETECTABILITY FOR ROOF EDGE

Fig. 10

signal to noise ratios used were 10, 11, 12, ..., 20. To measure the effectiveness of the different algorithms we graphed PF = the probability of false alarms vs PD = the probability of detection (Figure 9(a) and 9(b), for details see [7]). Figures 8(b) and 8(c) contain examples of processed roof edge disks with SNR = 13. The graphs of PF vs PD (ROC Curves) for the corresponding operators appears in Figure 8(a).

The results for different operators and step, ramp and roof edges appear respectively in Figures 10(a), 10(b), and 10(c). These graphs show that the performance of the moment operator is in all cases better than that of the Sobel operator. A significant improvement is obtained by first applying the average and then the moment operator. When the signal to noise ratio is high the median gives better results that the average, but there is a crossover point at which the average filter gives better results than the median.

## REFERENCES

1. A. Rosenfeld and A. Kak, "Digital Picture Processing", Academic Press, New York, NY, 1976

2. B. Lipkin and A. Rosenfeld, "Picture Processing and Psychopictorics", Academic Press, New York, NY, 1970

3. W. Pratt, "Digital Image Processing", John Wiley and Sons, New York, NY, 1978

4. M. Krasnoselsky, A. Perov, P. Zabreiko, "Plane Vector Fields", Academic Press, New York, NY, 1961

5. J. Milnor, "Topology from the Differentiable Viewpoint", University Press of Virginia, Charlottesville, Virginia, 1965

6. A. H. Stroud, "Approximate Calculation of Multiple Integrals", Prentice Hall, Englewood Hills, NJ, 1971

7. I. Abdou, "Quantitative Methods of Edge Detection", Image Processing Institute, University of Southern California, Los Angeles, CA, 1978

8. R. Angus and T. Daniel, "Applying Theory of Signal Detection in Marketing: Product Development and Evaluation", American Journal of Agricultural Economics, Vol 56, No 3, August 1974, (pp 573 - 577)

9. M. Giles, "Grating Detectability: A Method to Evaluate Abberration Effects in Visual Instruments", Optical Engineering, Vol 18, No 1, Jan 79, (pp 33 - 38)

10. R. Machuca and A. Gilbert, "Finding Edges in Noisy Scenes", IEEE Transactions on PAMI, to appear

# HIGH PERFORMANCE VEHICLES

MARTIN, LYNN A., MR.
TANK-AUTOMOTIVE RESEARCH AND DEVELOPMENT COMMAND
WARREN, MICHIGAN 48090

Survivability on the modern battlefield has become increasingly difficult. The development of more lethal armament has pushed armor to the limit. Historically the increased armament threat has been challenged by improving the armor of the fighting vehicles, especially tanks, but the current situation requires a different approach. The proliferation of effective weapons on the battlefield has pushed tank weights to 60 tons, which is very near the practical weight limit.

There are several ways to increase survivability on the battlefield including reducing crew sizes and re-packaging the crew and equipment. Another possibility that is being explored is to determine the effect that mobility and agility have on vehicle survivability. The determination of the influence of mobility/agility on battlefield survivability is the subject of this paper.

The Armored Combat Vehicle Technology (ACVT) program was started several years ago and includes both the development of vehicle concepts with varying weights and performance and the testing of hardware to validate vehicle performance characteristics. The two test beds that have been built and are being tested are the High Mobility Agility Test Vehicle (HIMAG) and the High Survivability Test Vehicle (Light), (HSTV(L)). The HIMAG vehicle can accommodate many variations in engineering properties (vehicle weight, wheel travel, wheel base, etc.) and has been tested at vehicle weights from 32 to 45 tons. The HSTV(L) had a design weight of 19 tons and uses a gas turbine engine to obtain high performance at the



PRECEDING PAGE BLANK-NOT FILMED

light weight. Both of these vehicles were designed to have high mobility performance with complementing automotive and suspension systems. Each of these vehicles are equipped with a 75mm automatic cannon and the HIMAG vehicle will be used to evaluate several different fire control systems. The power/weight ratios of these test beds are at least equal to that of the XM-1 and have the same wheel travel, 15-16 inches from static to bump stop. Although these programs are interacting, they will be used to supplement the concept study, which is being pursued concurrently.

The concept study has many phases. Using the concepts developed by TARADCOM, individual vehicle performance will be predicted over a variety of terrains. These performance predictions will be made using the U.S. Army Mobility Model (1, 2) over selected terrains in West Germany and Jordan. These vehicle speeds are used as inputs to several other analytical models. The Army Systems Analysis Activity (AMSAA) uses the information in their DUEL model, which evaluates the probability of hits/kills when the vehicle and weapon performances are integrated. The BDM Corporation has developed a model (MOBAG) to evaluate the effectiveness of evasive maneuvers on survivability. A more complete description of this model is a mobility/agility armor anti-armor survivability model. The output from these various analytical models will provide input data for TRASANA, U.S. Army TRADOC's System Analysis Group, when they conduct a force-on-force war game. These war game results will assist in evaluating the effectiveness of mobility/agility on battlefield survivability.

Preliminary to all of these evaluations, TARADCOM developed a total of 25 vehicle concepts which met various requirements and would present a sufficient variation in performance for the mobility/agility evaluation procedure. The guidelines for the design of the 25 concepts were specific weight, automotive performance, and weapon selection. The range in vehicle weights were 16 tons, 23 tons, 40 tons, and 60 tons. The 16 ton vehicles are an attempt to accommodate the requirement of being helicopter transportable. The upper weight limit of 60 tons was selected to be compatable with the XM-1. The original goal was to have automotive performance that equaled the XM-1 and went up to the HIMAG level. The combination of power/weight ratio and wheel travel was designed to achieve these mobility levels, at a minimum. This was interpreted to mean horsepower/ton ratios of 25/1(XM-1) to 37.5/1(HIMAG) and 15-16 inch wheel travels. To meet these power levels, the engines and transmissions which would be available within the expected time frame were examined. The engines and their power outputs used in this evaluation were XM-1 (1500 HP),

XM-2 (500 HP), a single rotor rotary (375 HP), a two rotor rotary (600 HP), and a turbo-compounded XM-2 (at 800 HP and 1000 HP). Using these engines, the desired goals of HP/ton ratios of 25/1 minimum were not met as some of the concepts increased in weight beyond the pre-selected values. The wheel travel of the concepts vary from 13 inches to 16 inches.

The weapon selection included both a high velocity 75mm and 90mm, along with several missile systems. The cannons were configured in both externally mounted and turreted. Several crew sizes were used, with 2, 3 and 4 man crews depending upon the missions and the weapon selection. Vehicle concepts were designed to meet several specific mission requirements and these requirements dictated the weapon selection. The concepts were designed to meet at least one of six missions:

1. USMC Mobile Protected Weapon System
2. Main Battle Tank Replacement in Calvary Units
3. Calvary Vehicle with Anti-Armor Capability
4. Follow On to the Improved TOW Vehicle (ITV)
5. Light Tank in Light Division
6. Infantry Vehicle with Anti-Armor Capability

The protection level of each concept is different and is dictated by the gross weight limitation. The protection level was limited to the available "extra" weight after the essential elements have been included. The essential elements are defined as hull, suspension, power train, crew, weapon and ammunition (gun or missile), fuel, OVE, etc. Obviously, the lighter weight vehicles have less protection. Within the same weight range, crew size, weapons, and packaging may make the protection different.

One of the key elements for improving survivability through mobility/agility is the capability of achieving a high rate of speed and maintaining this rate of speed while performing a mission. The vehicle suspension system is very important to achieving the goal of high cross country speed, and adequate wheel travel is essential to high speed operation. The suspension system must be of a balanced design with appropriate combinations of wheel travel, spring rates, and damping rates. The suspension systems were designed to the latest guidelines from TARADCOM's Suspension Group. It is important to establish a ride limiting speed for each concept to ensure that the crew can still function while the vehicle is moving. For each concept, this ride limiting speed as a function of crew comfort was determined using a two-dimensional ride dynamic simulation model (3). The various vehicle parameters were estimated

and the concepts were simulated while traveling over a variety of terrain roughnesses at various speeds. To evaluate the differences in vehicle/crew performance, the criteria of absorbed power in the vertical direction was selected based upon previous laboratory and field tests. In the early 1960's, the U.S. Army TARADCOM conducted many tests on people to determine the maximum vibration that could be tolerated comfortably (4, 5). There were over 1400 hours of testing with 31 volunteers being subjected to both random and sinusoidal inputs (6). When all of this data was reduced, it was concluded that an average absorbed power level of 6 watts in the vertical direction was the maximum comfortable limit for a sustained period of time. The absorbed power criteria is a method where the frequency of the induced acceleration is as important as the magnitude of the acceleration. This criteria of 6 watts average absorbed power at the driver's position has been used in field tests and simulations to evaluate ride dynamic performance of vehicles.

During these many field tests, several other observations have been made. Although the ride limiting speed was determined by using the 6 watts absorbed power criteria, it was noted that vehicle operators were willing to absorb higher levels of energy for a short period of time - up to 30 minutes. Using this data as a guideline, it was determined that an absorbed power level of 12 watts would be used in the situation where the driver was highly motivated, probably for self-survival. Another operational limit was noted during product improvement testing of the M60A1. At these tests, the M60A1 gunner did not attempt to fire the main weapon when he was subjected to more than 2 watts absorbed power. The exact reasons for this gunner response are not known and have not been documented but it provides a starting point for evaluating vehicles for fire-on-the-move capability. To summarize the ride dynamic evaluation, crew functioning vehicle speeds were determined for three different absorbed power conditions:

1. 6 watts at the driver's station.
2. 12 watts at the driver's station.
3. 2 watts at the gunner's station.

All of these absorbed power levels are taken in the vertical direction only. The extremes in vehicle performance, based upon ride quality, are given in Figures 1 through 5. Figures 1 and 2 show performances based upon the driver limit of 6 and 12 watts absorbed power.

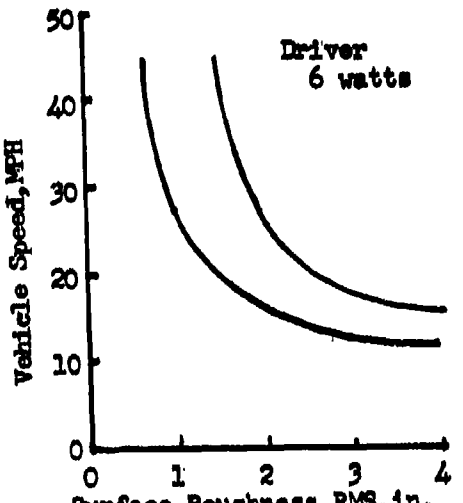



Fig. 1

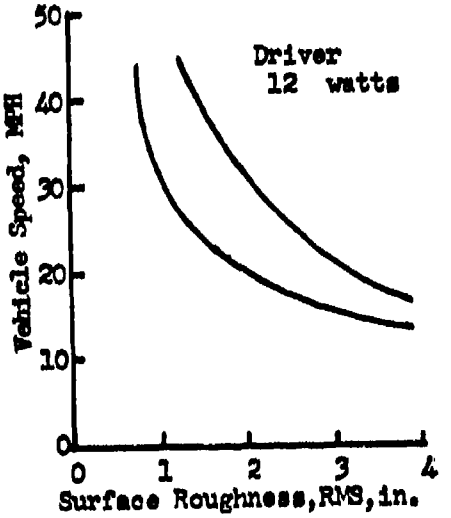



Fig. 2

The 25 vehicle concepts have wheel travels that range from 13 inches to 16 inches, measured from static to bump stop. It would seem that this difference in wheel travel would not make much difference in the ride quality. Also, the vehicle with the largest wheel travel should have the best ride dynamic response. These expected results are not always true as shown in Figures 1 and 2. All 25 concepts fall within the performance bands outlined but there is a significant difference between the best and the worst. In some cases, the speed difference is 15-20 miles per hour. There are several critical parameters that are more important than a wheel travel difference of three inches. The two most critical paramaters are the location of the sprung mass center of gravity with respect to the center of the suspension system and the driver location. The driver position does not vary significantly in relation to the hull, but does vary up to 28 inches relative to the center of gravity. The vehicle center

of gravity is located at a maximum variation of ±10 inches from the center of suspension. The ideal location is for the center of gravity and center of suspension to have the same relative position in the longitudinal direction (fore-aft). This results in equal wheel loadings and minimum disturbances to the hull when travelling cross-country.

The largest percentage of terrains that have been mapped fall within the surface roughness range of 1.5-2.5 inches root-mean square (RMS) elevation. Over the very rough terrains, those with RMS values greater than 3.0, all vehicles travel at approximately the same speed. This is because the roadwheels hit the bump stops regardless of the available wheel travel because the terrains have low frequency wave lengths. The same general trend of vehicle performance follows for driver limited speeds at 12 watts absorbed power.

Figure 3 shows the spread between the concepts at the fire-on-the-move (FOM) limit of 2 watts at the gunner's seat. The vehicle speeds are much lower for this FOM capability and the differences between vehicles is much greater due to the variety in gunner location. In some of the vehicles, the gunner is located very near the sprung mass center of gravity ( and receives the best ride), and other, two-man vehicles have the gunner sitting next to the driver, which gives him the maximum vehicle motions. The driver and gunner in a side-by-side configuration is the least desirable location for maximum FOM vehicle speed.

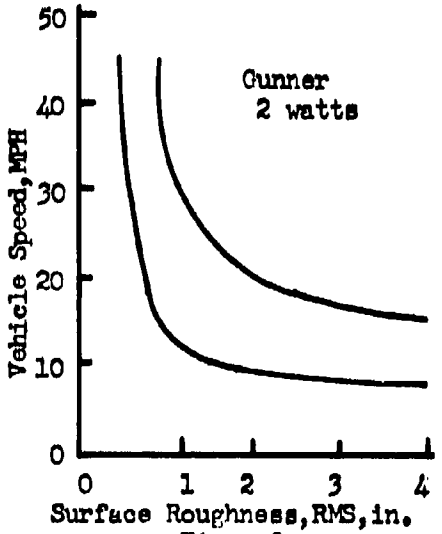



Fig. 3

The ride dynamic performance of all 25 concepts was included in the performance bands shown in Figures 1, 2 and 3. To illustrate the performance of individual concepts, Figures 4 and 5 represent the speed variations for given concepts. With some of the vehicles, the ride limiting speeds are nearly the same for the driver at 6 watts and the gunner at 2 watts absorbed power. Conversely, many concepts will have to slow down significantly when it is desired to fire-on-the-move accurately.

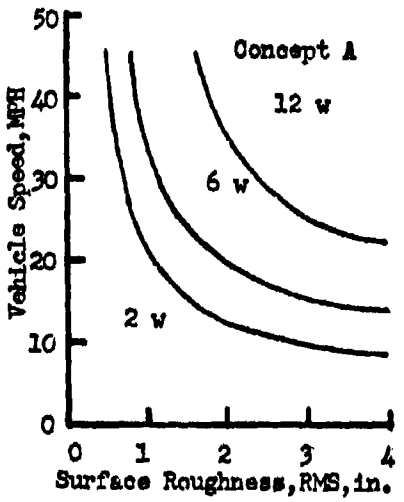


Fig. 4

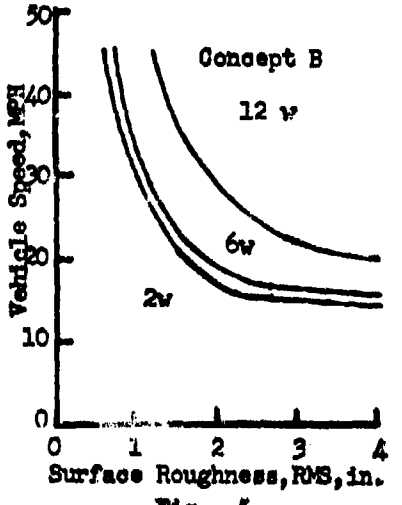


Fig. 5

The hypothesized advantage of the lightweight, high mobility vehicles is in the capability to reduce exposure time between defilade positions and perform evasive maneuvers to avoid being hit when fired upon. The reduced exposure time is a function of vehicle acceleration and elapsed time required to dash between cover positions.

For the concepts being evaluated, the performance envelopes for these two characteristics are shown in Figures 6 and 7. The difference in accelerations is significant and a function of both the vehicle weight and available power (7). The exposure windows used in the war games are based on field visibility data, but any reduction in time to cross these "windows" will increase survivability. It should be emphasized that the performance shown in

Figures 6 and 7 are for hard surface, good traction soil conditions. Any deterioration in these soil conditions will degrade vehicle performance.

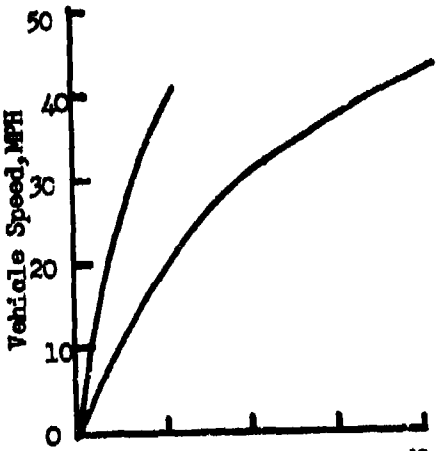


Fig. 6

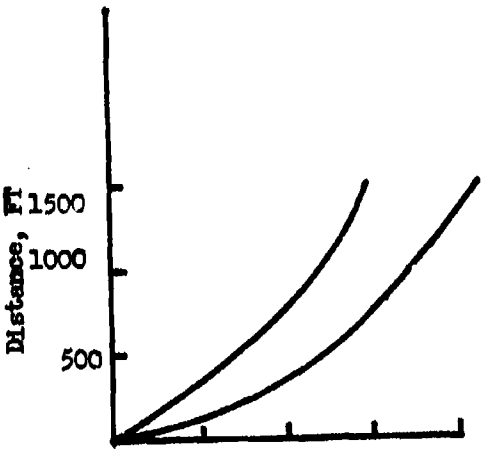


Fig. 7

The automotive performance of the concepts has been addressed, but how this performance will be used for evasive maneuvers has not. To evaluate the potential of performing evasive maneuvers while being fired upon, the BDM mobility/agility (MOBAG) model will be used. This model integrates the total systems, including detection, being acquired by a threat, firing on by the same threat, and the probability of being hit/killed. This model uses line of sight and visibility data taken during field exercises to determine the exposure windows during an attack upon a fixed defender position. The maneuvers employed while crossing these exposure windows will be varied to evaluate increased performance. As a pre-processor to this MOBAG evaluation, the vehicle concepts were simulated doing several sine wave maneuvers over the terrain that might be encountered. The particular sine wave maneuvers vary in wave length from 50 to 250 meters and in amplitude from 3.5 to 15 meters. For each of these maneuvers, the vehicle longitudinal and acceleration, lateral acceleration, and vehicle attitude with respect to direction of travel are recorded and used by MOBAG. The MOBAG model uses the basic assumption that the most significant parameter to avoiding being hit is the vehicle lateral acceleration. Therefore, for each exposure window, the vehicle path and defender

position are used to determine the angle for which the vehicle velocity and accelerations must be corrected. All potential evasive maneuvers are evaluated for each concept/exposure window combination to determine which single maneuver maximizes survivability.

These data will provide the basis for evaluating the effectiveness of mobility/agility on vehicle survivability. In conclusion, the capability to design vehicles with wide ranging weight, power train performances, and armor to meet specific U.S. Army requirements has been demonstrated. The potential of these concepts to perform the task for which they were designed is currently being evaluated. There are many ways to approach the problem of increasing battlefield survivability, and you now have been exposed to a system for evaluating survivability as influenced by high performance vehicles.

## REFERENCES

1. Staffs of US Army Engineer Waterways Experiment Station and US Army TACOM, "AMC '71 Mobility Model", TACOM Report No. 11789, July 1973.

2. P. Jurkat, C. Nuttall, P. Haley, "AMC '74 Mobility Model", TARADCOM Report No. 11921, May 1975.

3. F. Hoogterp, "Interactive Vehicle Dynamics and Ride Evaluation Package", TARADCOM Report No. 12413. November 1978.

4. F. Pradko, R. Lee, V. Kaluza, "Theory of Human Vibration Response", ASME Paper 66-WA/BHF-15, December 1966.

5. R. Lee, F. Pradko, "Analytical Analysis of Human Vibration", SAE Transactions, Vol 77, Paper 680091, 1968.

6. W. Lins, "Human Vibration Response Measurement", TACOM Report No. 11551, June 1972.

7. J. Wollam, "Interactive Drive Train Model", TARADCOM, unpublished.

*McCAULEY & CORBIN

# TRANSPARENT, POLYCRYSTALLINE CUBIC ALUMINUM OXIDE

*James W. McCauley, PhD
Normand D. Corbin, Mr.
Army Materials and Mechanics Research Center
Watertown, MA 02172

## I. INTRODUCTION

The means used to observe or sense the enemy have progressed from actual eye-to-eye observation to extensive use of radar and sonar, and now include using infrared (IR) signals. At the same time, various forms of armor, from face shields to sophisticated electromagnetic (EM) windows and domes (radomes, IR domes), have been developed to transmit signals and also to protect the sensing mechanisms - either the human eye or intricate electronic devices. Countermeasures such as smoke and radar-jamming systems have concurrently evolved to defeat the various sensing devices. In order to minimize the effectiveness of dedicated (single-mode) or even broadband countermeasure tactics, sensing devices of the future, therefore, must be able to simultaneously function over a large region of the EM spectrum, including visible light, IR, microwave and millimeter wave radars. It is imperative, then, that new materials must be developed to transmit a wide range of the EM spectrum, while at the same time protecting the fragile sensing equipment in wide-ranging types of severe battlefield environments. Consideration of the above criteria results in the conclusion that only a select few materials, if any, can provide all the requirements. Further, the material should be isotropic to minimize any distortion of the transmitted signal. In lieu of isotropic materials, recourse has been made to the use of single crystals of anisotropic materials, oriented to transmit the signals in isotropic or near-isotropic direction. Aluminum oxide is a candidate material for these applications, but has been neglected from serious consideration because of its property anisotropy, relatively high coefficient of thermal expansion, and high cost for the production of



PRECEDING PAGE BLANK-NOT FILMED

single-crystal shapes (1). True optical transparency of polycrystalline $Al_2O_3$ is impossible, unless all the grains are identically oriented. Further, significant strain can result at grain boundaries due to thermal expansion mismatch of misoriented grains (2). An alternate approach to this inherent problem is the stabilization of a cubic $Al_2O_3$ structure. A defect cubic spinel, $\gamma$-$Al_2O_3$, (3) can be prepared in powder form, but fully dense ceramics have not been reported due to the ease of conversion to the more stable alpha form at moderate temperatures ($\sim$1000 C) (4,5). However, it has been known for some time (6) that nitrogen additions to $Al_2O_3$ in the form of AlN can produce spinel-like structures. Since that time various efforts have been made to understand the phase equilibria in this system (7-9). The phase diagrams do not, however, indicate the temperature limits of stability for the $\gamma$-$Al_2O_3$ type oxynitride material, nor has single-phase material been successfully sintered. This report describes the results of a program concerned with refining the temperature-composition stability limits of cubic aluminum oxynitride spinel (ALON - nitrogen-stabilized cubic aluminum oxide) so that fully dense, single-phase ceramics could be sintered.

## II. EXPERIMENTAL

Sintering and phase equilibria studies were carried out in an inductively heated graphite furnace using flowing $N_2$ (1/2 liter per minute at about 1 atm). The starting powders of $\gamma$-$Al_2O_3$* (1.1 μm at 50%) and AlN* (14 μm at 50%) were ball-milled for 24 hours using an ethanol fluid media, isostatically pressed at 25,000 psi, and pre-reacted at 1200 C for 24 hours in gas tight flowing $N_2$, prior to final sintering studies. Only small amounts of impurities were picked up in the ball-milling procedure. However, more sophisticated sintering studies would require an improved mixing procedure.

Final reaction and sintering runs at elevated temperatures were conducted for one hour so that direct comparison between runs could be made. Weight loss was determined for each specimen and each was characterized by X-ray diffraction and reflected light microscopy. Density measurements and transmitted light microscopy were also carried out on selected products. Our fundamental premise is that we are attempting to deduce the high temperature equilibrium relations and sintering mechanisms from the resultant products. Without the use of sophisticated apparatus, volatility is extremely difficult to suppress, so quantitative measurements of weight loss will yield useful information on vapor phase formation. The reaction samples were contained in a covered BN crucible with a sight hole for pyrometric

*Cerac/Pure Inc., Butler, Wisconsin

temperature measurement. Neutron activation analysis of the AlN indicated 1.7 wt% (≈1.5 mole % $Al_2O_3$) oxygen. The AlN powder also contained about 1 to 2 wt % of unnitrided Al metal powder.

## III. RESULTS AND DISCUSSION

Figure 1 illustrates the temperature-composition stability limits for cubic aluminum oxynitride spinel (ALON) in the $Al_2O_3$-AlN system for 1 atm of flowing $N_2$ gas. The phase relationships were deduced from careful analyses of both microstructural and X-ray diffraction data. The $Al_2O_3$-AlN system is a pseudobinary composition join in the Al-N-O system. Hence, the phase rule allows for phase fields with up to three coexisting phases. The designations 12H, 21R, and 27R refer to the so-called "AlN" polytypes (10). In this system there seems to be an intimate relationship between liquid formation and the appearance of the various polytypes. Further, the morphology of the polytypes are variable and seem to reflect the difficulty in attainment of equilibrium. It is our conclusion from this work in the vicinity of the ALON stability field and also other parts of this system that some of the polytypes are metastable products of quenched or poorly quenched liquids. We have not yet determined how to differentiate between AlN polytypes which are metastable from those which are not. Hence, in Figure 1 we have "dashed in" all the phase boundaries dealing with polytypes and in some cases have not differentiated between a liquid and the polytypes.

There is a relatively wide range of compositional stability, roughly centered at 35.7 mole % AlN, and a maximum in thermal stability at about 2050 C. At this point ALON seems to melt incongruently into one alumina-rich, stable liquid and one nitride-rich, unstable (volatile) liquid. At about 2000 C on the AlN-rich side of the single-phase field, vaporization increases dramatically which kinetically seems to influence reactions in the single-phase region.

As previously indicated by McCauley (11) a constant anion spinel model seems to predict an ALON composition at 35.7 mole % AlN. Using the chemical formula $Al_{(64 + x)/3} \square_{(8 - x)/3} O_{32 - x} N_x$ obtained from this model, the following composition for N = 5 can be calculated:

$$Al_{23}O_{27}N_5 \equiv 5\ AlN \cdot 9\ Al_2O_3.$$

Crystalline solution stability limits were determined by detailed reflected light microscopy and refined lattice parameters.

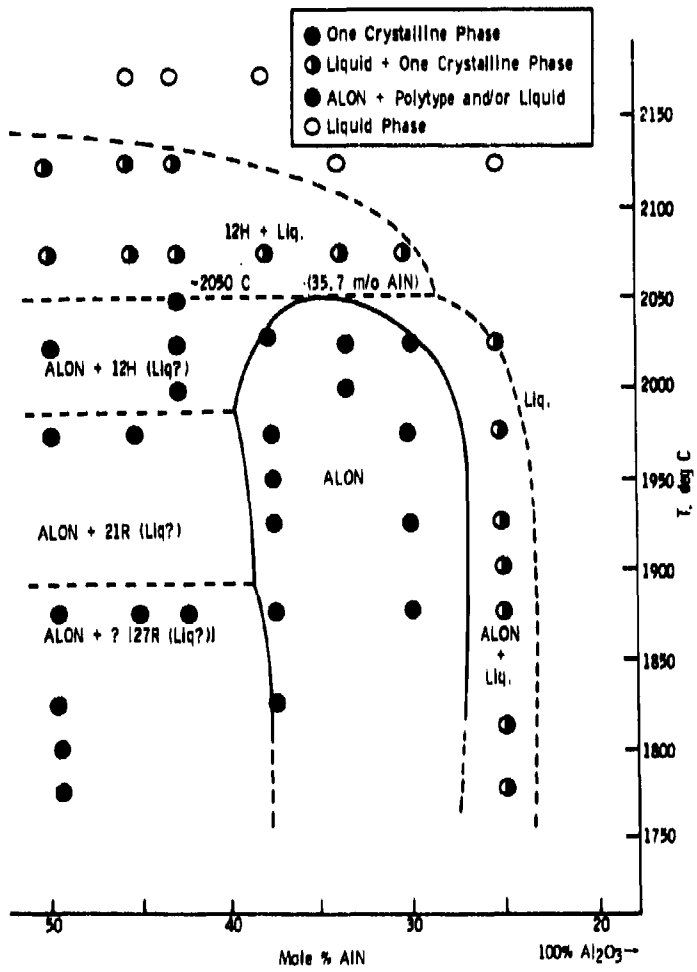



**Figure 1. Proposed isobaric (1 atm of flowing $N_2$) high-temperature phase relationships in the region of ALON stability in the pseudobinary $Al_2O_3$-AlN composition join.**

Figure 2 illustrates the phase assemblages on the 37.5 mole % AlN composition line on either side of an apparent phase boundary line between the ALON single-phase field and the liquid plus ALON and 12H polytype (and/or liquid) region. Note the dramatic disappearance of porosity in Figure 2a and the concurrent appearance of liquid. The liquid and 12H polytype appear as the lighter colored intergranular phases in Figure 2a; in Figure 2b the darker circular areas are remnant porosity. This liquid occasionally quenches to a noncrystalline phase, but also crystallizes into various types of AlN polytypes,



a. 2025 C - in liquid + ALON(SS) field; 10.04 wt% loss

b. 1975 C - in ALON(SS) field; 2.88 wt% loss

Figure 2. Microstructures of phase assemblages in the $Al_2O_3$-AlN system; 37.5 mole % AlN.
19-066-500/AMC-77

in this case the 12H polytype. Figure 3 illustrates the microstructure of the ALON plus liquid region on the $Al_2O_3$-rich side of the single-phase field.

In order to confirm the microstructurally determined range of ALON crystalline solution, refined lattice parameters were determined by a least-squares technique on X-ray powder diffracting data. These data are illustrated in Figure 4 along with those obtained by Lejus (7) on material fabricated at 1700 C. There is about a five mole % AlN difference between a least-squares line through the 1975 C data and the Lejus line. The difference can possibly be ascribed to a higher oxygen content of the latter material or the sluggish reaction ratos at 1700 C. The equation for this line is indicated on the figure.

ALON ceramics of various nitrogen compositions can be sintered using the experimental conditions previously indicated to 99% of theoretical density. The theoretical densities were calculated using the constant anion model and the refined lattice parameters. Figure 5 shows a typical set of characterization data on a series of runs at 37.5 mole % AlN. By relating these data to the microstructures illustrated in Figure 2 it can be seen that the appearance of liquid results in a dramatic increase in volatility, represented by weight percent loss on the figure; percentage of theoretical density is also indicated in parentheses. A reaction and sintering scheme is also indicated on the figure.



a. ALON(SS) + liquid b. ALON(SS) + liquid c. Etched

**Figure 3. Microstructures of phase assemblages in the $Al_2O_3$-AlN system; 25 mole % AlN; T = 2025 C.**

19-066-1196/AMC-77

Reaction and Sintering → Sintering (No Liquid) →

Sintering (Liquid + Vapor Formation)

A preliminary apparent decrease in density occurs during formation of ALON, caused by the diminishing amount of the higher density $\alpha$-$Al_2O_3$. Optimum properties and liquidless sintering must be carried out below about 2000 C to prevent the formation and subsequent vaporization of a liquid phase. Residual, quenched liquid, either as a noncrystalline phase or as one of the "AlN" polytypes, will have a large effect on the properties of the material. Note the increase in density for the material sintered for three hours.

Formation of sintered, single-phase ALON ceramics free of second phases can be easily accomplished in the single-phase field. However, detailed reflected and transmitted light microscopy of sintered ALON samples do reveal occasional evidence of grain boundary phases - probably low melting phases due to unavoidable impurities in the starting powders, especially AlN. Figure 6 illustrates a typical

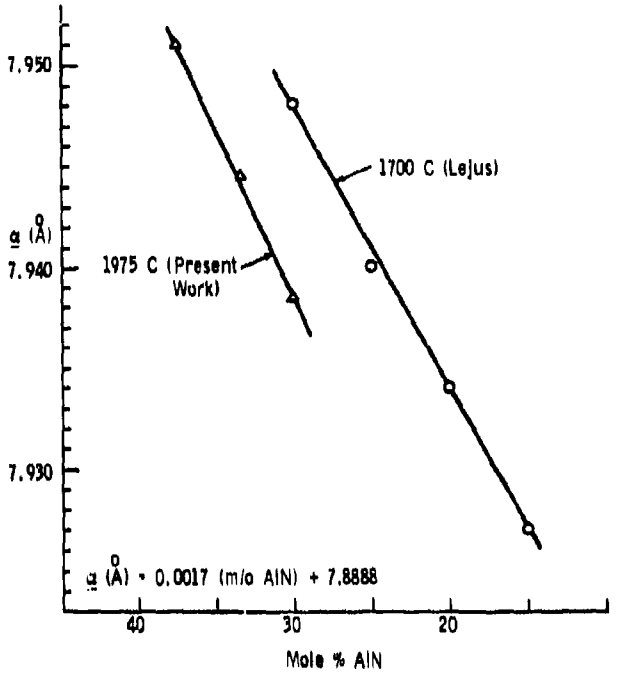



**Figure 4. Refined lattice parameters of cubic aluminum oxynitride spinel (ALON) as a function of mole % AlN at 1975 C.**

microstructure of material exhibiting a relatively large amount of remnant porosity. This material seems to react and sinter quite rapidly into a uniform grain size microstructure. For our experimental conditions, grain growth is quite rapid, with an average size somewhere between 50 and 100 μm. An abundance of apparent spinel-law ({111}) twins is also evident in Figure 6b, which seem to polish at different rates, since some are elevated and others are recessed. A small increase in sintering temperature (1975 C to 2025 C) results in much reduced residual porosity and a microstructure that only microscopic defocussing at low magnification (Figure 7a) will resolve the grain boundaries; scanning electron micrographs of fracture surfaces of this same material are illustrated in Figures 7b and c to further illustrate the microstructure.

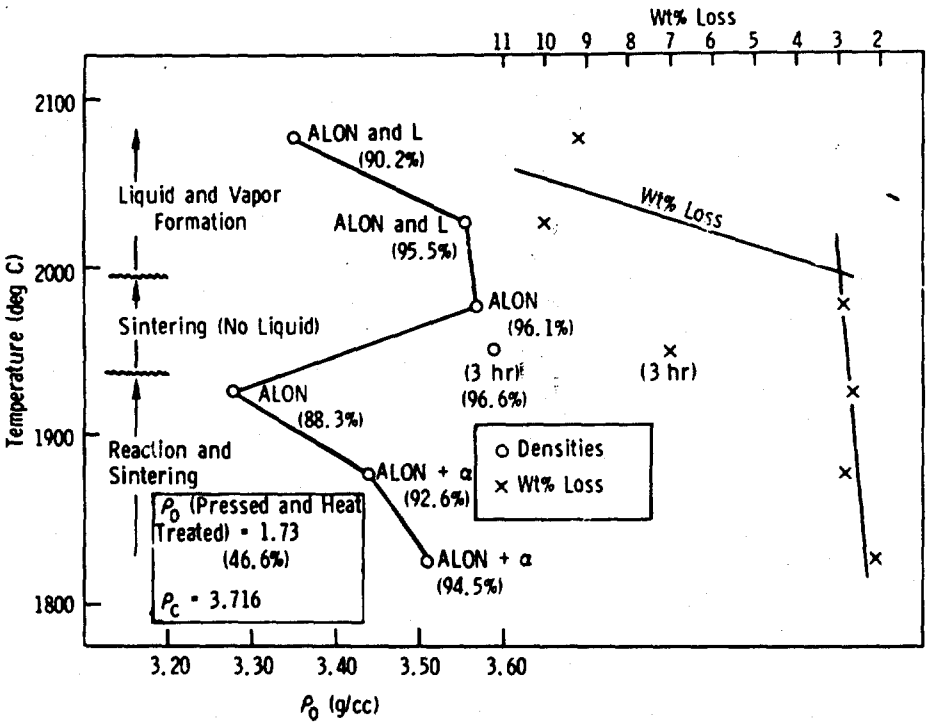



Figure 5. Variation of density and weight percent loss with temperature (1-hr runs) of 37.5 mole % AlN mixtures.

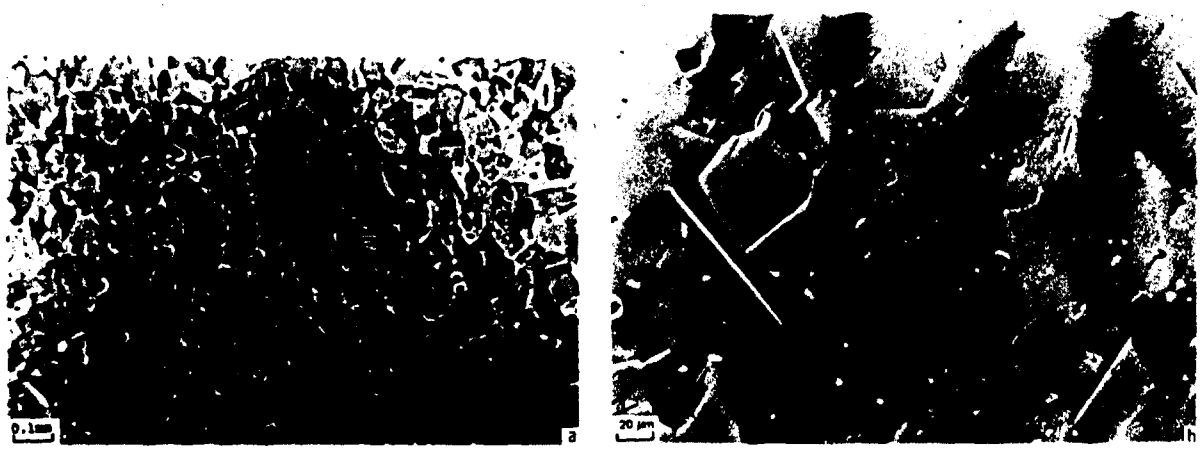


Figure 6. Microstructures of single-phase cubic aluminum oxynitride spinel; T = 1975 C; 30 mole % AlN; t = hour.

19-066-503/AMC-77

## IV. PROPERTY DETERMINATIONS

Preliminary properties are now being measured on sintered ALON materials. Figure 8 illustrates a polished disk (0.035" thick) of a 30 mole % AlN material exhibiting a high degree of transparency;




a. Reflected light - slightly defocussed to bring out microstructure



b. SEM - fracture surface




c. SEM - fracture surface

Figure 7. Microstructures of single-phase cubic aluminum oxynitride spinel; T = 2025 C; 30 mole % AlN; t = 1 hour.

19-066-504/AMC-77

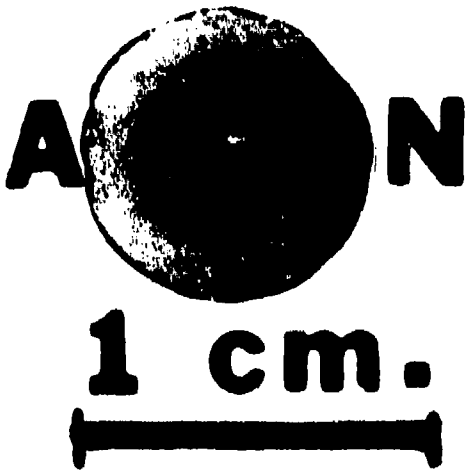



Figure 8. Transparent disk of single-phase cubic aluminum oxynitride spinel.

19-066-1874/AMC-77

the word ALON is behind the disk. Table 1 lists several of the properties of ALON which have been measured. Of particular interest in terms of a multimode sensor window is the high hardness, relatively high IR cutoff and the excellent dielectric properties. Table 2 is a more extensive listing of the dielectric properties and Figure 9 illustrates the dielectric constant and loss tangent as a function of temperature for measurements at 10 MHz frequency. These properties will significantly improve where 100% of theoretical density is achieved. Figure 10 illustrates the excellent oxidation resistance of ALON in air up to 1200 C. At this temperature a protective film is formed that serves to persist and inhibit additional oxide formation. Thermal expansion has also been measured and is illustrated in Figure 11; note that the sample returned to its original exact dimension. The average coefficient of thermal expansion ($\alpha$) calculated from 20 C to 980 C is $7\times10^{-6}$ $C^{\circ-1}$, significantly lower (22%) than

Table 1. PROPERTY DATA FOR CUBIC ALUMINUM OXYNITRIDE SPINELS

| Mole% AlN | Density g/cc | Max% Density | Sonic Modulus x$10^6$ psi | Hardness Knoop(100) kg/mm$^2$ | $K_{Ic}$ MN/m$^{3/2}$ | $\kappa'$ (10 MHz) (25 C) | Tan $\delta$ (10 MHz) (25 C) | IR Cutoff ($\mu$m) | Ref. Index ($\lambda$=0.55 $\mu$m) |
|---|---|---|---|---|---|---|---|---|---|
| 30 | 3.66 | 98.8 | 46.03 | 1788 | 0.48 | 8.56 | 0.0004 | 5.18 | 1.770 |
| 33.3 | 3.68 | 99.2 | | | | | | | |
| 37.5 | 3.57 | 96.1 | | 1624 | | | | | 1.785 |

$K_{Ic}$ = Critical stress intensity factor from Vickers hardness indent
$\kappa'$ = Dielectric constant
Tan $\delta$ = Dielectric loss tangent

Table 2. DIELECTRIC PROPERTIES OF ALON

| T°C | Freq., Hz | $10^2$ | $10^3$ | $10^4$ | $10^5$ | $10^6$ | $10^7$ |
|---|---|---|---|---|---|---|---|
| 25 | $\kappa'$ | 8.56 | 8.56 | 8.56 | 8.56 | 8.56 | 8.56 |
| | $\kappa''$ | 0.013 | | | | | |
| | tan $\delta$ | 0.0015 | 0.0011 | 0.0006 | 0.0005 | 0.0005 | 0.0004 |
| | $\sigma$ | $7.1 \times 10^{-13}$ | | | | | |
| 150 | $\kappa'$ | 8.62 | 8.60 | 8.60 | 8.60 | 8.60 | 8.60 |
| | $\kappa''$ | 0.028 | | | | | |
| | tan $\delta$ | 0.0033 | 0.0037 | 0.0024 | 0.0018 | 0.0010 | 0.0006 |
| | $\sigma$ | $1.6 \times 10^{-12}$ | | | | | |
| 300 | $\kappa'$ | 8.97 | 8.79 | 8.72 | 8.65 | 8.64 | 8.64 |
| | $\kappa''$ | 0.242 | | | | | |
| | tan $\delta$ | 0.0270 | 0.0108 | 0.0044 | 0.0029 | 0.0026 | 0.0021 |
| | $\sigma$ | $1.343 \times 10^{-11}$ | | | | | |
| 400 | $\kappa'$ | 10.0 | 9.40 | 9.13 | 8.95 | 8.76 | 8.72 |
| | $\kappa''$ | 0.709 | | | | | |
| | tan $\delta$ | 0.0709 | 0.0495 | 0.0229 | 0.0078 | 0.0037 | 0.0031 |
| | $\sigma$ | $3.94 \times 10^{-11}$ | | | | | |
| 500 | $\kappa'$ | 14.0 | 11.7 | 9.92 | 9.18 | 8.95 | 8.87 |
| | $\kappa''$ | 14.2 | 2.48 | | | | |
| | tan $\delta$ | 1.014 | 0.212 | 0.0941 | 0.0370 | 0.0136 | 0.0070 |
| | $\sigma$ | $7.89 \times 10^{-10}$ | $1.4 \times 10^{-9}$ | | | | |

$\kappa'$ = Dielectric constant
$\kappa''$ = Relative loss factor
tan $\delta$ = Loss tangent = $\kappa''/\kappa'$
$\sigma$ = Dielectric conductivity

$\alpha$-$Al_2O_3$ at $9 \times 10^{-6}$ $C^{\circ -1}$ at equivalent temperatures. This value suggests a much superior thermal shock resistance of this material over $\alpha$-$Al_2O_3$. Many preliminary shapes and sizes of ALON have been fabricated. These are illustrated in Figure 12. The thin disk covering the AMMRC logo was simply saw cut and was not polished at all.

## V. SUMMARY AND CONCLUSIONS

A refined high temperature phase diagram in the region of stability of cubic aluminum oxynitride spinel (ALON) along the $Al_2O_3$-AlN composition join has been determined. This material can also be described as nitrogen-stabilized cubic aluminum oxide. Using this newly determined diagram, single-phase ALON has been reactively sintered to nearly full density. Sintering is carried out quite easily and polished thin disks exhibit visible light transparency. The lattice parameter of ALON varies with composition from 7.938 Å for 30 mole % AlN to 7.951 Å for 37.5 mole % AlN sintered at 1975 C. At this temperature the limit of ALON crystalline solution is from 40 to about 27 mole % AlN.

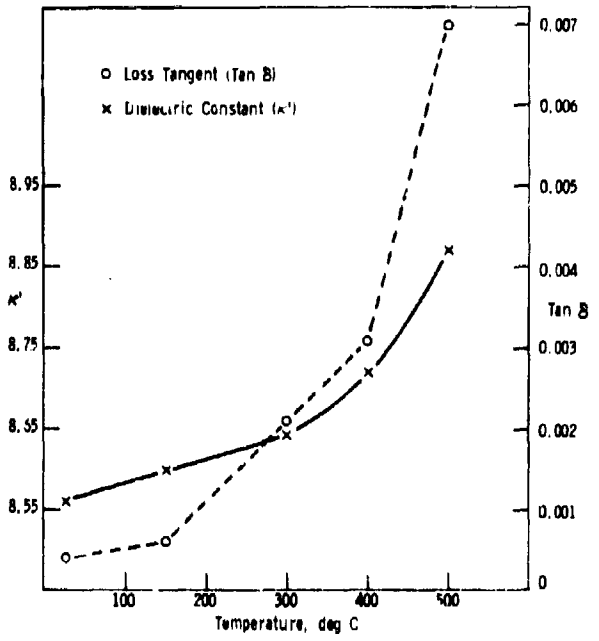



Figure 9. Dielectric properties of 30 mole % AlN ALON measured at 10 MHz.

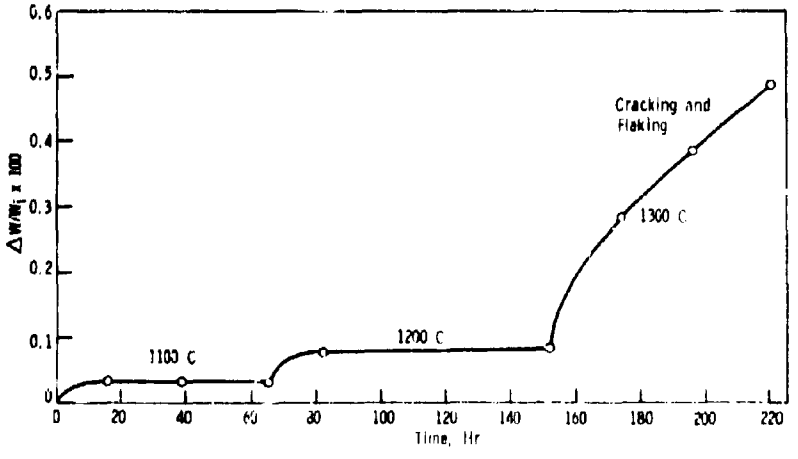



Figure 10. Oxidation of 30 mole % AlN ALON at various temperatures as a function of time in air.

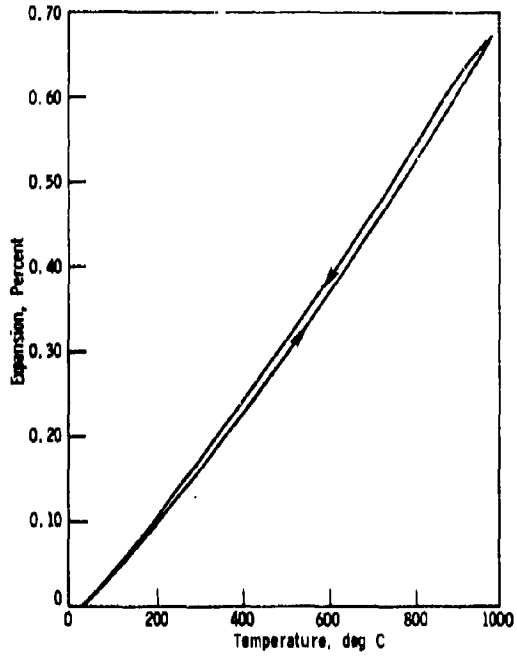



Figure 11. Thermal expansion of 30 mole % AlN ALON.

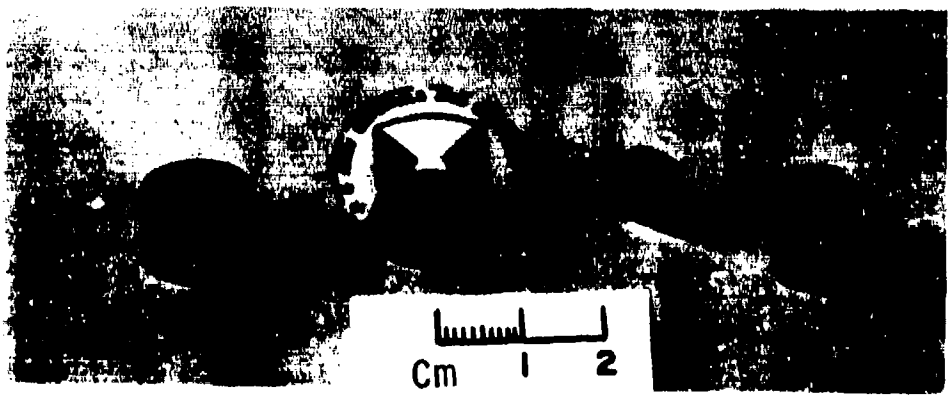



Figure 12. Various fabricated bodies of ALON.

19129-18J-1 DARCOM-79

Preliminary properties determined on ALON strongly suggest that it is a viable candidate for future multimode EM sensor window requirements and as an improved substitute for $Al_2O_3$ in many other applications.

## VI. ACKNOWLEDGMENT

The authors wish to thank Dr. D. J. Viechnicki for help in the design of the furnace and D. H. Corkum for help in various experimental parts of this work. We also would like to acknowledge the contributions of W. B. Westphal (dielectric constant), R. Brockelman (sonic modulus), J. H. Rosolowski (thermal expansion) and A. Zani, R. Middleton, and T. Sheridan for their help on the metallography and X-ray diffraction. Thanks also go to G. Gazza, D. Messier, and R. N. Katz for helpful discussions.

REFERENCES

1. BARATA, J., and GORNI, J. *Improvements in Mechanical and Electrical Properties of Alumina.* Powder Metallurgy International, v. 4, 1972, no. 3, p. 124-128 and no. 4, p. 201-203.

2. CAMPBELL, W. J., and GRAIN, C. *Thermal Expansion of Alpha Alumina.* Bureau of Mines Report of Investigation No. 5757, 1961, 16 pages.

3. WEFERS, K., and BELL, G. M. *Oxides and Hydroxides of Aluminum.* Technical Paper No. 19, Alcoa Research Laboratories, 1972, 51 pages.

4. JELLINEK, M. H., AND FANKUCHEN, I. *X-Ray Diffraction Examination of Gamma Alumina.* Ind. and Eng. Chem., v. 37, no. 2, 1945, p. 158-163.

5. STEINER, C. J. P., HASSELMAN, D. P. H., and SPRIGGS, R. H. *Kinetics of the Gamma-to-Alpha Alumina Phase Transformation.* J. Am. Ceram. Soc., v. 54, no. 8, 1971, p. 412-413.

6. ADAMS, I., AU COIN, T. R., and WOLFF, G. A. *Luminescence in the System $Al_2O_3$-AlN.* J. Electrochem. Soc., v. 109, 1962, p. 1050-1054.

7. LEJUS, A. *Formation at High Temperature of Nonstoichiometric Spinels and of Derived Phases in Several Oxide Systems Based on Alumina and in the System Alumina-Aluminum Nitride.* Rev. Int. Hautes Temp. Refract., v. 1, no. 1, 1964, p. 53-95.

8. IRENE, E. A., SILVESTRI, V. J., and WOOLHOUSE, G. R. *Some Properties of Chemically Vapor Deposited Films of $Al_xO_yN_z$ on Silicon.* J. Electron. Mat., v. 4, 1975, p. 409-427.

9. GAUCKLER, L. J., and PETZOW, G. *Representation of Multicomponent Silicon Nitride Based Systems.* Nitrogen Ceramics, F. L. Riley, ed., Noordhoff, Leyden, 1977, p. 41-62.

10. JACK, K. H. *Sialons and Related Nitrogen Ceramics.* Review, J. Materials Sci., v. 11, 1976, p. 1135-1158.

11. McCAULEY, J. W. *A Simple Model for Aluminum Oxynitride Spinels.* J. Am. Ceram. Soc., v. 61, nos. 7-8, 1978, p. 372-373.

# DISPOSAL OF SOLUBLE SALT WASTE FROM COAL GASIFICATION

CHARLES E. MC KNIGHT, PE
CHEMICAL SYSTEMS LABORATORY, USAARRADCOM
ABERDEEN PROVING GROUND, MD 21010

## INTRODUCTION

Energy predictions from various sources indicate a shortfall. Attention has, therefore, become focused on the use of coal as a partial solution, with subsequent realization that numerous aspects of coal utilization need to be studied. Among these aspects is the production of gas from coal. In such a study, restrictions on degrading the environment have a direct bearing on discharge of water from the plants, this largely by virtue of the solids content of the water. Increasing tightness of environmental restrictions leads to a probable requirement for zero discharge of liquid effluents. A suggested method of achieving this is examined in this paper.

For the waste disposal of salty water, ocean dumping may seem an obvious solution. However, ocean dumping is gradually being outlawed and is not at all available to most of the geographical areas of the world. For land areas (this applies to the bulk of the world's industry) the ocean dumping counterpart is stream or river dumping. This is restricted by regulation and becomes a matter of negotiation and controversy for all but the cleanest of salty waste. The sludges that invariably accompany waste soluble salts are an additional impediment to river dumping.

Public Law 92-500 of the United States reads, in part:

> "It is the National goal that the discharge of pollutants into the navigable waters be eliminated by 1985."



PRECEDING PAGE BLANK

Public Law 94-580 encourages resource recovery on a National scale.

This paper addresses pollutants in the form of soluble salts and resource recovery in the form of water and land.

River dumping cannot be considered a satisfactory way for industry to be environmentally clean, even for those industry owners who have access to a river. A method available to inland industry is deep-well dumping. This will handle both solutions and sludges, but it entails capital expense for the well, the pollution of an underground resource, and uncertainty as to where the pollutants will reappear. Various methods of disposal that have been considered are listed below:

1. Sale or reuse
2. Ocean dumping
3. Controlled or secure landfill
4. Under ocean floor burial
5. Deep well injection
6. Encapsulation:
   a. Pozzolanic landforming
   b. Deep formation grout injection
   c. Glass encapsulation
   d. Plastic encapsulation
7. Chemical fixation
8. Electrolytic conversion
9. Dry-land (desert) storage
10. In-door storage

The method in this paper is a highly-refined, expertly-arranged instance of pozzolanic landfill with essential previous preparation of the waste. Such preparation is part of the present paper. The method is comprised of a sequence of commercially-available processes to produce a material that is valuable for landform alteration. The sequence was conceived during chemical process planning for a coal conversion process. It applies to numerous other industries and applies world-wide, wherever a minimum amount of water is present to aid in consolidation of the landform.

## GASIFICATION OF COAL

The coal process, to convert coal to methane for use as pipeline gas, of the Lurgi type will be used as a basis to explain the disposal process. The purpose of the coal gasification plant is to

produce methane starting with coal, gasifying it and hydrogenating it. As supplies of low-sulfur coal diminish, coal gasification must operate on high-sulfur coal, and this gives rise to an important section of the complete process. To maintain the integrity of the environment and at the same time to be more efficient, energy-wise, is the challenge of coal gasification. In the Lurgi process, one example out of the many coal gasification processes, lump coal is introduced into the top of the gasifier (3). Steam and oxygen are injected. The resulting high temperature produces a gas which then undergoes a quench-cooling, water-gas shift, hydrogen sulfide and ammonia removal, methanation, compression and dehydration.

The slag or ash is cooled by quench water (including cooling tower and other blowdown water). The condensates from gas cooling and scrubbing lead to coal oils and naphthas as well as phenols, ammonia and aqueous wastes. The removal of acid gases (i.e., carbon dioxide and hydrogen sulfide) result in vented carbon dioxide and a sulfur product such as elemental sulfur, sodium sulfate, or sulfuric acid products that can be kept out of the environment.

## WATER USAGE

In this process the basic reaction for producing methane is the reaction of water (steam) with carbon (coal): $2C + 2H_2O = CO_2 + CH_4$. Water from the environment must be supplied for this and for boiler feed water. Rainfall naturally enters the plant site and must be accounted for before it leaves the site. During transit through the coal conversion process, the water is used and re-used several times, but finally it becomes a wastewater and must be treated for disposal to the environment. The proposed disposal process starts with the formation of a "combined feed" to the disposal process proper. To form the combined feed, site runoff (which may include leachate from coal and slag piles) is mixed with the waste water from gas purification. This mixture is wet-oxidized, then polished by filtration through activated carbon. The low-organic effluent which results is made up with softened raw water to become the cooling tower make-up water. The cooling tower blowdown now becomes the waste. The "combined feed" to waste disposal is now formed by adding sulfur recovery waste water, sanitary effluent, boiler blowdowns and ion exchanger blowdowns. A combined feed analysis calculated during the design study showed it would be approximately 1.3 percent total solids, a mixture mainly of sodium sulfate and chloride.

A full-size coal gasification plant is expected to produce 250 million standard cubic feet of methane per day. For this size plant, the combined feed to disposal is approximately 1 million gallons per day (1 MGD), depending upon geographic location and other factors. Larger quantities of water are recycled and are lost by evaporation at the cooling towers. This figure would apply separately to each of the many full-size coal gasification plants required to provide energy self-sufficiency in the United States, this being the objective of the federally-sponsored Project Independence.

## DISPOSAL-RECOVERY PROCESS

The proposed disposal-recovery process is shown in Figure 1. Use is made of electrodialysis to recover a large amount of water for reuse in the coal process and to prepare a strong brine. With a strong enough brine it appeared feasible to evaporate economically to a stronger brine or even to dryness in the next step, again recovering water for reuse. The final wet or dry soluble salts are then to be encapsulated by one of several techniques for use in forming a landfill suitable for agriculture, construction, or for any other land use. Pozzolanic encapsulation is preferred. The sequence of waste treatments proposed here has not been reduced to practice as a totality on a pilot plant or larger scale, although each separate step (such as wet oxidation, electrodialysis, evaporation and encapsulation) has been proved separately in commercial applications. A paper study was therefore made to explore the water treatment outlined above. The general results of the paper design study are presented in this paper.

## ELECTRODIALYSIS

Electrodialysis takes salts, the minor constituent, out of water, rather than water from the salts. The electrical energy required is primarily that required only by the salts, the minor constituent, and does not depend primarily on the total bulk of water. A brine results whose strength is theoretically limited only by the solubility of each salt. The method has a high energy efficiency and its proponents have been overcoming the economic limitation of high equipment (membrane and cell) cost. Electrodialysis generates only a minimum amount of wastes and these can be handled downstream. Membranes must be replaced; pumping costs and salt transfer energy are required.

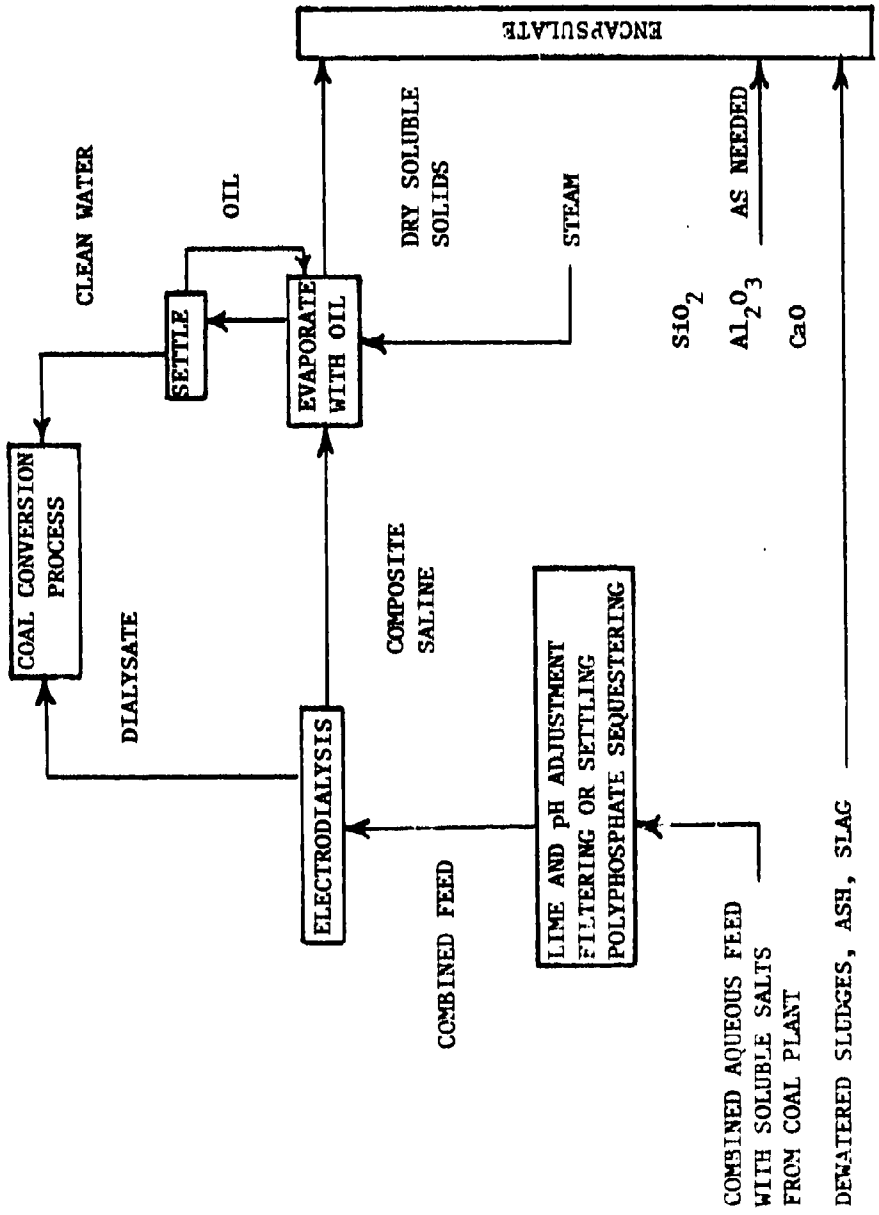


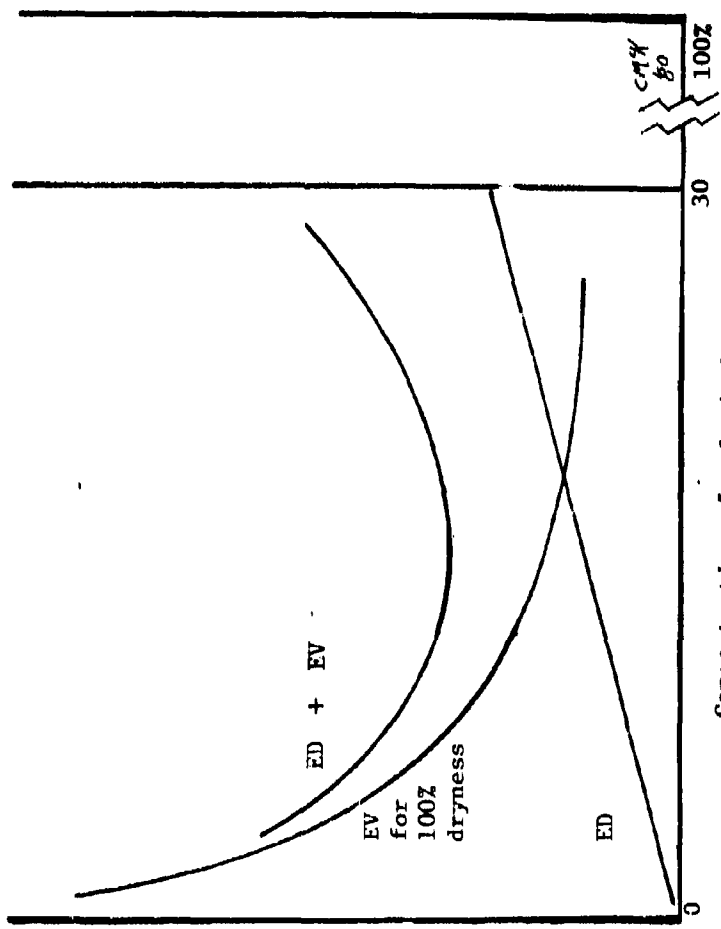

Figure 1. ED-EV-EN Recovery Process

A theoretical minimum work of separation can be calculated from the Gibbs free energy of mixing, resulting in 1 BTU per pound of reusable water for the conditions studied. For other conditions, mathematical optimization shows the energy required versus concentration of feed starts from zero for absolutely salt-free water and increases in proportion to the square root of concentration. These "other conditions" include localized parts of the electrodialysis equipment where strong brine is present and is acting as a "feed" (6) in the strong-brine end of the equipment, before being discharged to evaporation. Consequently, the actual energy required will be more than 1 BTU, but will be less than the 1000 BTU required by oridnary evaporation. It is, therefore, possible to recover most of the water by electrodialysis with the expenditure of upwards of 1 BTU per pound of water recovered and the remainder by evaporation, expending downwards of 1000 BTU per pound of water on the remaining few pounds. Increased exit concentration from electrodialysis requires more stages in the process design, with a moderate increase in cost per unit of water recovered. This effect is illustrated in Figure 2.

## EVAPORATION

For evaporation in the presence of solid salts, an oil-fluidized process (1) is desirable. By mixing oil with brine, the brine can be evaporated completely to dryness and pressed or centrifuged without thickening, scaling, fouling, or caking. The mixture passes through the evaporator, leaving the solids suspended in the dry oil as fluid slurry. The oil is then centrifuged off to be used again, and the dry solids are left. Water is recovered for reuse in the coal process; its oil is recycled, and the water-soluble solids can emerge in a dry powder form. This form is suitable for encapsulation to permit final disposition of the salts as insoluble forms that will not contaminate water in the environment. The purpose of the oil is to lower the vaporization temperature of the water, to improve heat transfer in the evaporator, to keep the salts suspended so they do not form scale, and to prevent especially scaling and fouling as the salts become completely dry. Water is recovered from condensates to be reused as a recovered resource. The oil evaporation technique is appropriate for a coal conversion plant, as a suitable oil is at hand in the form of one of the products of coal conversion.

For evaporation, the energy consumed is nearly proportional to the quantity of water recovered and is practically independent of feed concentration for a given quantity of water recovered.




$$\text{Evaporation fraction-toward-dryness, frd} = \frac{x_{EN} - x_B}{100 - x_B}$$

Figure 2. Cost Interactions

When evaporator exit concentration is markedly higher than its product water concentration, the boiling point elevation may permit heat economy gain by reusing the heat from one stage in other stages. A mathematical optimization showed that if this effect can be utilized, the optimum number of stages without electrodialysis varies inversely as the square root or less of the concentration level above that of the water condensate (7). Evaporation energy requirements can be explored by merely using an even 1000 BTU per pound for the latent heat of vaporization of water and dividing by the number of effects to incorporate the factor of steam economy. This quick method indicates greater than actual economy for evaporation since boiling point elevation must be added to the temperature drop for heat transfer surfaces to get the actual steam economy. The net result is a slight effect due to concentration. Evaporation remains energy-intensive regardless of concentration. As electrodialysis generates a stronger and stronger brine, the quantity of brine into evaporation lessens. Consequently, the energy and cost of evaporation diminish. The situation is illustrated by the falling evaporation curve on Figure 2, using evaporation to complete dryness for the illustration.

It is noted that Figure 2 shows a rising curve for electrodialysis and a falling curve for evaporation when these operations are operated in series as described above. This situation points to the existence of an optimum cost peculiar to the sequence of these two processes, and it has been verified by design calculations.

Figure 2 was drawn so that the minimum point is at 20 percent, a value selected based on engineering judgment. The exact value must be found by experiment or by a comprehensive design calculation. Regardless of the final correctness of the 20 percent figure, it may be used to illustrate the possible energy savings when electrodialysis preceeds evaporation to dryness (4). A material balance for this case, based on exactly 1 MGD, is shown in Figure 3. The theoretical minimum energy (basis: 1 hour) required is 1 BTU per pound of water times 326,242 pounds of water recovered by electrodialysis plus 1000 BTU per pound of water times 16,961 pounds of water for a total of 17,287,242 BTU per hour. Recovering the same amount of water by evaporation alone requires 1000 BTU per pound times 343,203 pounds of recovered water for a total of 343,203,000 BTU. The comparison is as 19.8:1. If the 1 BTU is increased to 10 BTU to allow for concentration effect, current efficiency, reflux load, pumping and other losses, the comparison is as 16.9:1. As far as a preliminary design study can go, the benefits of using electrodialysis with evaporation seem real.

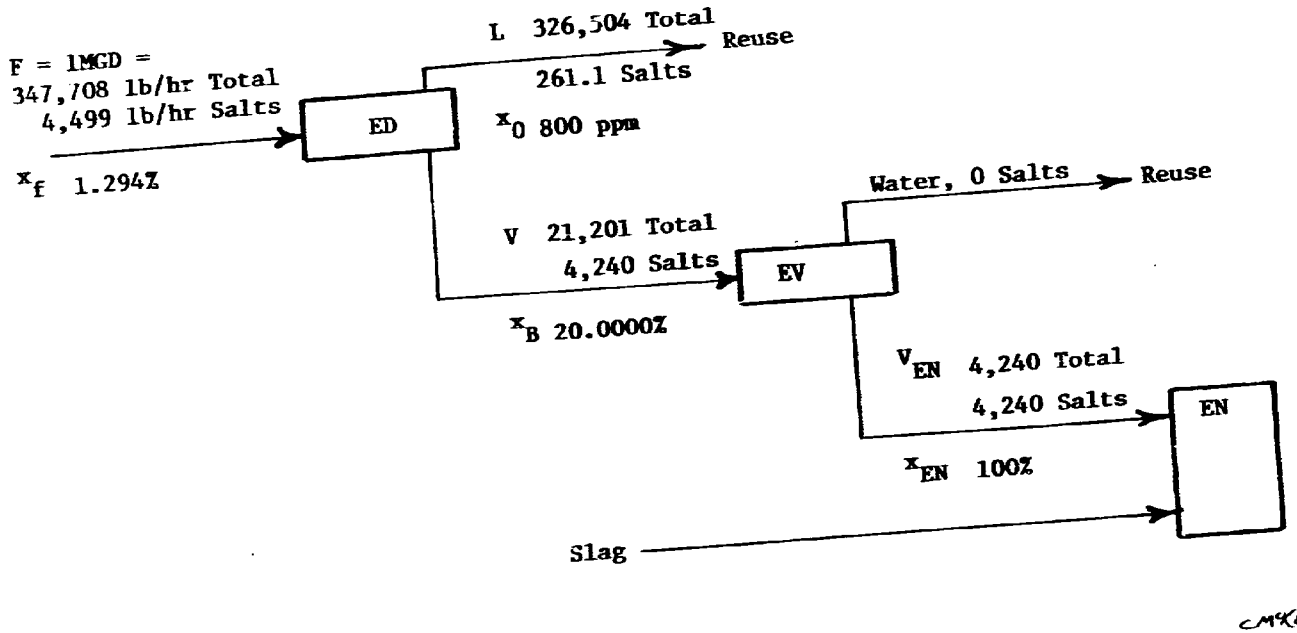



Figure 3. Recovery System Material Balance

## ENCAPSULATION

The dried soluble salts or strong brine from evaporation are to be encapsulated for permanent storage. Pozzolanic encapsulation (2) is a desirable method for accomplishing this. In this method the salts are blended with fly ash, alumina, silica and lime in the proper proportions, and caused to harden as a cement-like monolith of very high density. This can act as a paving landfill, can be used to underlay crop-yielding soil, artificial lakes, building foundations, landscaping forms, etc. The author believes that demilitarization residues can be encapsulated in this manner to prevent leaching if it is required to store them in a moist, non-desert, geographical area. The evaporated salts from the coal process can be considered, as a gross oversimplification, to be equal parts of sodium sulfate and sodium chloride. The sulfate is especially suited to encapsulation by means of the pozzolanic method, because one of the most important insoluble compounds formed when the mixture sets up is a calcium sulfo-aluminate. The entire mass is so impermeable that sodium chloride or other soluble salts distributed through it will not be leached out by water and, therefore, will not leach out. A nominal amount of leaching of the external surface occurs and diminishes rapidly with time. There is no leaching from the internal surfaces.

Water can be forced slowly into the fill for experimental purposes if a depth of 90 feet of water is used. This is not likely to happen in the natural environment. Thirty years would be required to fill the small amount of porosity of a 10 foot thick bed under conditions of rainfall of 1/4 inch per 24 hours and 30 inches per year, with no runoff or evaporation. The penetration rate under these conditions is approximately four inches per year of water going in and none coming out. The compacted material as actually used is graded, and drains are installed to permit runoff and to minimize accumulation of standing water. The area required for a commercial-size plant for 20 years (10 feet deep landfill) would be approximately 45 acres. Each section would be settled, compact and stabilized in a matter of months after each last addition. As...as merely dumped into water can take a long time to stabilize. Unstabilized and "secure" landfills are essentially destroyed land, for some uses.

Compressive strength develops rapidly, so that as a landfill, the completed fill is valuable for nearly every imaginable land use. Because of its strength it can be disposed of at a slope that depends on such considerations as final contour desired and

retention of cover material rather than on stacking stability. However, before hardening the fill will conform to a wide variety of disposal areas from low flat land along coastal and tidal areas to hilly natural ravines.

Other methods of encapsulation may be indicated in special situations. Sulfur in massive form from the coal plants may have a future use as in divider walls for divided highways. In this case dry soluble salts can be encapsulated in the sulfur. Bituminous road pavings are also capable of encapsulating dry soluble salts.

Encapsulation is considered to be an energy-unintensive operation, but does, nevertheless, have a certain energy consumption. As applied to a coal gasification plant it may encompass slag disposal as well as fly ash and soluble salt disposal. The process starts with slag of suitable particle size and a suitable reactivity, mixes it efficiently with the correct porportion of water and available fly ash and small amounts of additives if necessary. The water may include a certain amount of soluble salts and organics. The mix assumes an earth-like consistency and is then placed in the area to be filled. During several weeks the chemistry of cement formation proceeds, bonding the soluble salts into a stone-like mass, protecting them from extensive weathering. Energy is consumed by transport of slag and water to mixer, slag conditioning, mixing, transport to final emplacement, and consolidation or compaction. The mix should have approximately three parts of slag to one part of liquid (brine). Consequently, the quantity of mix to be handled will depend on the volume of brine from evaporation and not on salt content. All energy and costs in encapsulation will, therefore, depend on the final volume of brine produced. The actual energy required will reflect equipment requirements, in this case conveyors, trucks, and similar earth moving equipment, and most of the energy will go for acceleration and friction in machinery. Grinding slag and solids mixing may also be significant contributors to cost.

## THREE-PROCESS SEQUENCE

The three processes exhibit a cost interaction (5). Figure 2 showed an optimum brine concentration feed to evaporation, portrayed for one evaporator condition, viz, 100 percent evaporation of the brine. There is a corresponding optimum for each degree of evaporation. There is a locus of concentrations of effluents from electrodialysis and evaporation under the condition that the salt quantity per unit time is held constant, with the recovered water

being taken in differing proportions from electrodialysis and evaporation. For any given degree of evaporation and feed concentration, the cost of evaporation will depend on the feed rate to the evaporator. Because of rising electrodialysis cost with rising evaporator feed concentration (Figure 2), and rising evaporation cost with falling concentration, the combined cost, in the general case, shows the same cost at two different values of the feed concentration. The situation is shown schematically in Figure 4, which shows contours of equal cost for various degrees of dryness achieved by evaporation applied to each possible electrodialysis brine effluent concentration. Encapsulation cost is not shown on this schematic because of the possibility of encapsulation using some other source of water and being affected by large quantities of slag, fly ash, other sludges, etc. All contours shift to the left (higher cost) when cost of encapsulation is added. For greater clarity, Figure 4 is a schematic only, and does not show actual costs.

## CONCLUSION

In conclusion, a design for disposal of soluble salts has been produced. The interactions of its parameters have been shown by a process design study. The design will enable harmonious compliance with United States Public Laws 92-500 and 94-580, relating to water pollution and resource recovery. In the disposal of waste salt solutions, natural water resources need not be contaminated, because an encapsulation technique is available which will immobilize the salts. At the same time it will make useful landforms available, and water as a resource can be recovered. The bulk of the water can be recovered at a low cost, additional water can be recovered at a higher cost, and the water which must be discarded unavoidably is discarded at a low cost. The waste soluble salts can be provided with an environmentally satisfactory burial.

There is a cost minimum when electrodialysis and evaporation are combined, which is not realizable with evaporation alone, unless very low-cost thermal energy is available or unless very high-cost pretreatment for electrodialysis is required.

All the processes making up the proposed disposal process are commercially available, although they are nowhere operating commercially as one process. Because of the commercial availability of the processes, the proposed process may be a candidate "best commercially available treatment" for soluble salt disposal.

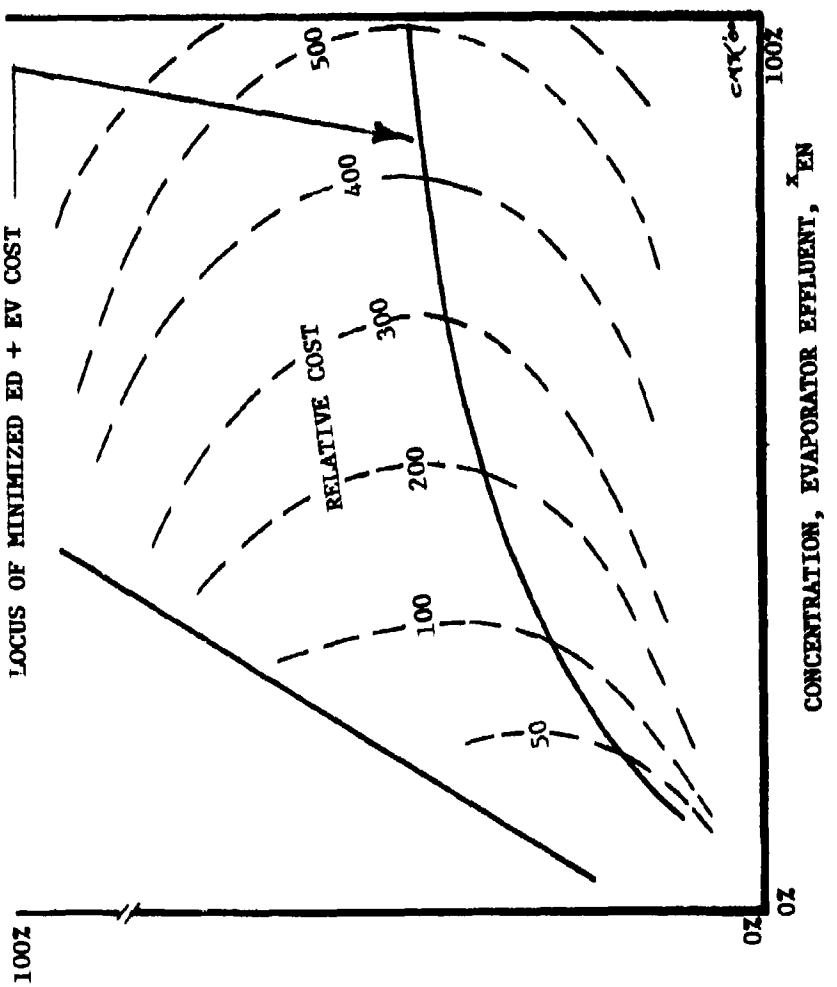



Figure 4. Cost Contours

## REFERENCES

1. Dehydro-Tech Corporation, East Hanover, New Jersey, personal communication.

2. IU Conversion Systems, Inc., Philadelphia, Pennsylvania, personal communication.

3. McKnight, C.E. "Treatment of Solid Residues from Coal Gasification." 1977 National Conference on Treatment and Disposal of Industrial Wastewaters and Residues. Houston, Texas. April 26-28, 1977.

4. McKnight, C.E. "Ultimate Environmental Disposal of Salts from Coal Gasification." Tenth Annual Chemical Systems Laboratory Technical Conference. Aberdeen Proving Ground, Maryland. June 8-9, 1977.

5. McKnight, C.E. "Water Recovery and Salt Disposal in Coal Gasification." 1978 Chemical Systems Laboratory Technical Conference. Aberdeen Proving Ground, Maryland. May 24-25, 1978.

6. McKnight, C.E. United States Patent Application 915,733, dated July 13, 1978, "Ultimate Disposal or Concentration of Salt Solutions and Suspensions."

7. Probstein, R.F., Sonin, A.A., and Spielman, L.A. "Water Purification." Short Course Notebook. AIAA, New York, NY, 1972.

# NON-LINEAR LEAST CHI-SQUARE ALGORITHM
# AN IMPROVEMENT ON NON-LINEAR LEAST SQUARES

RICHARD L. MOORE, PhD
ARMAMENT RESEARCH AND DEVELOPMENT COMMAND
DOVER, N.J. 07801

Introduction. Because of the high cost of testing, many large weapon systems cannot be tested over the full range of possible battlefield parameters. As a result, the developer and the reviewing authorities have come to rely on system simulation to demonstrate the system capability over the range of untested parameters. These simulations also are useful to investigate the change in performance resulting from possible subsystem modifications. In some important programs, the Government relies on simulations of competing systems to indicate the relative performance of these systems in situations for which no tests have been made, although of course, simulations such as these have been validated as much as possible by system tests. In these instances the procurement decision rests heavily on the validity of the system simulations. Consequently the need arises for a generally accepted procedure which is undoubtedly fair to each contractor and provides the maximum amount of objective judgment about the validity of the simulation. In any such procedure the Government must be able to rapidly evaluate simulations furnished from a variety of sources.

The procedure must be workable and economical -- that is it must apply a lot of leverage to the problem with regards to manpower, -- computer programmers and engineers -- the cost, -- computer running time and validation experiments -- and elapsed time. Implicit in this discussion is that planning for system simulation validation must be completed before the first system RFQ is issued.

Many methods are used for system simulations: Monte Carlo, analog, hybrid, and digital simulation of differential equations. A variety of special and general purpose programs are available for the simulator's use. Among then are "SPERT", "ACSL," and HIT PRO." The problem for the user of these simulations comes when he needs to compare theory with experiment and asks the questions: How good is the theory? Is

the agreement between theory and experiment good enough to validate the simulation? (As an example of these questions see Pastrick (1, 2).) Another question to be considered is: Could it be that the experiment was defective in any way?

Many simulations have not been prepared in such a way so that they can be used to answer these questions. In the first place, the simulations are not designed to adjust parameters to fit data. In the second place, the system itself may be so complex that the computing time for complete system simulation is so long that adjusting the parameters to achieve a better fit between simulation and experiment is not feasible. Thus a new procedure is needed to combine theory and experiment.

The procedure suggested by this paper is the use of the least chi square computer program to simulate the major subsystems of a system simulation and validate it against test data.

Criteria for comparison, and iteration procedure. In fitting data to non-linear models of system performance such as systems of differential equations, the usual criteria is the iterative minimization of the sum of the squares of the residuals. Other criteria, such as generalized least squares have also been considered and demonstrated (3). As Aitken (4) noted with respect to generalized least squares, the criteria to be used are a matter of choice. In other words, we are free to decide whether least squares is the best criteria for our purpose. A particular concern with the ordinary least squares procedure is whether the residuals are consistent with being drawn from a random sequence.

Many tests have been devised (5) for this purpose. One test of special interest in this paper is the Box-Pierce (6) test which is the sum of the squares of the autocorrelation coefficients divided by their variances. A typical term is $(r_i)^t v_i^{-1}$ .

Given all these tests, no way had been devised to adjust the parameters to better satisfy the data until it was proposed that this criteria be combined with least squares to obtain a new criteria: least chi-square (Moore, 7, 8, 9). By finding the parameters which minimize chi-square, the probability is maximized that the residuals come from a population with a given variance $\sigma_e^2$ , and from a random sequence. The variance must be independently determined from theory or measurement as the measurement error.

Thus, a probability can be generated from the computed chi-square which permits the statistician and decision-maker to compare the "goodness of fit" of the simulation of several quite different systems. In this way a direct comparison of the validity of the simulations can be mac

The figure of merit, $\chi^2_{TOT}$ (chi-square total) is the sum of $\chi_1^2$ and $\chi_2^2$ , the Box-Peirce number $\{ \sum_j v_j^{-1} (\sum_i^n d_i d_{i+j}) / \sum d_i^2 \}$

Derivation: We will follow the procedure and most of the notation of Aitken (4) for generalized least squares:

Let the representation of the vector of data:

$$u = (u(x_1), u(x_2), \ldots u(x_n))$$

by the theoretical vector, be:

$$y = (y(x_1), y(x_2), \ldots y(x_n))$$

$$u_i^* = u_i - y_i^* .$$

Let $\Theta^*$ denote a column vector of k + 1 coefficients independent of y such that:

$$\Theta^* = (\theta_1^*, \theta_2^*, \theta_3^*, \ldots \theta_{k+1}^*)$$

Define the matrix P* as the matrix whose ith row is

$$\frac{\partial y_i^*}{\partial \theta_1} , \frac{\partial y_i^*}{\partial \theta_2} \ldots \frac{\partial y_i^*}{\partial \theta_{r..}}$$

(The asterisk symbol * will be used to indicate an estimate of the indicated sysbol where convenient. However, it will not be used on complex expressions involving $\chi_r^2$, $\sigma^2$, and $r_j$ because of typographical difficulties).

In this expression $v_j^{-1}$ is defined as follows:

$$v_1^{-1} = \begin{bmatrix} 0\,1\,0\ldots 0 \\ 0\,0\,1\,0\ldots 0 \\ 0\,0\,0\,1\,0\ldots 0 \\ \ldots \end{bmatrix}; \quad v_2^{-1} = \begin{bmatrix} 0\,0\,1\,0\ldots 0 \\ 0\,0\,0\,1 \\ 0\,0\,0 \end{bmatrix} \qquad v_j^{-1} = \begin{bmatrix} 0\,0\ldots 0\,1\ldots \\ 0\,0\,0\ldots\ldots 0\,1\ldots \\ \ldots \\ \ldots \end{bmatrix}$$

In these, the subscript "j" indicates a unit value in each of the ith rows and (i + j)th column.

If $v_j$ is the variance of , then

$$\chi_T^2 = \sigma_0^{-1} d'd + \sum_{j=1}^{n} r_j^2 / v_j$$

and: $r_j = d' v_j^{-1} d / (dd')$

(Note the difference between $v_j^{-1}$ and $v_j^{-1}$.

On differentiating ( $\chi_\nu^2$ ) with respect to ( $\theta^*$ ) and substituting

$$(d^*) = P^* [\delta\theta^*] - u^*,$$

as an estimate of the increment of the residuals needed to minimize $\chi_\nu^2$ , the algorithm for $[\delta\theta^*]$ becomes:

$$[\delta\theta^*] = [P^{*\prime} \Gamma P^*]^{-1} P^{*\prime} \Gamma u^*,$$

WHERE:

$$a_j = \frac{2r_j V_j^{-1}}{(d)'(d)/\sigma_e^2 - 2\sum_{j=1}^{n} (r_j)^2 V_j^{-1}}$$

$$\Gamma = I + \sum_{j=1}^{n} a_j r_j V_j^{-1}$$

If $\Gamma$ equals I, the expression for $[\delta\theta^*]$ reduces to $[P^{*\prime} P^*]^{-1} P^{*\prime} u^*$, which is the same as the algorithm for ordinary non-linear least squares used in such computer programs as provided by both IBM and CDC libraries as well as in SAAM-27.

By inspection, $P^{*\prime}\Gamma$ replaces $P^{*\prime}$ in the ordinary expression. To modify the ordinary expression, $\Gamma$ is computed. P*' is postmultiplied by $\Gamma$ , and the product placed in the computer memory where P*' is normally stored. $\chi_\nu^2$ is substituted for $dd'$ whereever it occurs and no further change is need in the iteration procedure.

These expressions have been programmed into the Simulation And Analysis Modeling (SAAM-27) program of Berman et al, (10, 11) as indicated above, multipying P*', by $\Gamma$ , and letting the program proceed from that point. The usual iteration continues. The computer program resulting from this change has been designated for control purposes as SAACH, and has been tested on the CDC 6600 at ARRADCOM, Dover, to determine the following questions:

1. How much change is there in the final parameter estimates?

2. What change, if any, is there in the number of iterations?

3. What change is there in the time per iteration?

Four problems of different origin and which use different mathematical models have been run on the SAACH program to answer the above questions. In the first example: Gun Chamber Pressure Waves, the mathematical model used is the superposition of two pressure waves generated by analytic models in the program, with the adjustment of up

to eight parameters to obtain the best fit to observed data. In the second example, an aircraft control system simulation, the mathematical model is a set of four linear differential equations, simulating the Yaw Damper system on an aircraft. These equations were solved by a special procedure developed for SAAM-27 by Berman et al. (12), with up to four adjustable parameters. In the third example, a biomedical problem furnished as a test case by Miss Rita Straub of Brookhaven National Laboratory, the mathematical model was a set of seven coupled linear differential equations with five adjustable parameters; this was solved by the same method as used in the second case. In the fourth and final example: KEWB Kinetics, a simulation of the nuclear reactor transients of the Kinetic Experiment Water Boiler, the mathematical model was an extremely non-linear set of coupled differential equations as described by Hetrick and Gamble (13). These equations were integrated by the fourth order Runge-Kutte integration procedure of SAAM-27, with only one adjustable parameter. The results of these analyses were discussed in detail at the 1979 Design of Experiments Conference (9). Abbreviated discussions of the result of each problem follow.

Gun Chamber Pressure Waves. Unusual pressure waves suggestive of an acoustic wave superposed on the normal gun chamber pressure-time curve, have occurred in tests of the XM211 propellant charges at zone 3 for the M101 projectile in the 155mm gun, (Knutelski, (14)). The mathematical model used was:

$$P = P_1 \exp\{-(t-t_1)^2/2\sigma_1^2\} + P_2 \exp\{-(t-t_2)^2/2\sigma_2^2\} \times \sin\{2\pi f (t-t_3) + \pi/2\}$$

Three parallel cases were computed once the fit was good enough to permit iteration with different ranks of autocorrelation. Because of computing difficulties which arose when trying to converge on six or seven parameters, the iteration was initially restricted to four parameters: Once the fit was good and had converged using these four parameters, their final values were used as initial values for a six-parameter fit. Finally, all eight parameters were allowed to vary.

Two results of this series of analysis are plotted in Figs 1, and 2. The case numbers are BGK-3.30356301-0, and 3.30356511-5. The first has no autocorrelation coefficients; the second, 5, a third, (not shown), 10. The parameters for these cases are given in Table 1, (note that the last three digits only of the identifier are used here). Some parameters are quite different from case to case.

The apparent fit from the figures is best for the case of five autocorrelations given in Fig. 2. The fit of this case was also better than that for 10 autocorrelations which is not illustrated.

The value of the sum of the squares of the residuals as shown in Table 1 was actually much smaller for cases 511-10, and 301 than for 511-5 yet the fit as noted above was not as good.

The last row of Table 1, gives the values of $\sigma_s^2$, the experimental variances assumed for these cases. These were arbitrary numbers taken as example only, because the precision of the measurement system is probably much greater than the value given; i.e., the variances should be smaller. However, if smaller values were used, such as when case 511-10 is compared to 511-5, the weight on the sum of the squares is greater but the goodness of fit appears to decrease, thus illustrating the need for least chi-square iteration.

Table 1 also shows the effect of least chi-square in terms of number of iterations, and computing time. When five autocorrelations were used, as in case 511-5, only a small increase in number of iterations is found and a moderate increase in computing time as compared to 301. If ten autocorrelations were used, as in 511-10, the number of iterations increased, and the time increase was 1.8 times greater, giving about double the increase in time for double the number of autocorrelation coefficients.

Table 2 shows the autocorrelations up to order 20 for the three cases. The values of $X_1^2$, $X_2^2$, and $X_3^2$ for the number of autocorrelations used (0, 5, 10) is shown in the last rows of this table.

Aircraft Control Systems. A typical aircraft yaw damper design problem (15), was analyzed to illustrate the use of least chi-square. To optimize the design, four parameters may be adjusted to give the best fit to a desired response curve. These parameters are $K_\phi$, $K_r$, $\tau$, and $\delta$. These correspond to the parameters L(0,4), L(4,1), L(4,1), L(4,2) and L(4.3). A vector of a random sequence of normally distributed errors from a population with variance of $(.033)^2$ was added to the data vector to simulate the effects of sampling error; this may be considered to represent an allowable error or tolerance in fitting the curve.

The value of was set at $(.033)^2$, six autocorrelations were used for the problem which was identified as CONRLM 4.011-6. Another run was used on the same data with the standard least squares algorithm. Fig. 3 shows the fit obtained for the data and is typical of the results. Table 3 showns the number of iterations for each case. It took 4 iterations for the ordinary algoithm to converge, and only two for the least chi-sqare algorithm with six autocorrelation coefficients (CONRLM 4.011-6).

| | | CASE | | |
|---|---|---|---|---|
| UNITS | SYMBOL | 301 | 515-1 | 511-10 |
| 10 kpsi | $P_1$ | 20.04 | 19.7 | 19.95 |
| 10 kpsi | $P_2$ | 4.146 | 33.3 | 3.998 |
| sec | $t_1$ | .057 | .0568 | .0568 |
| sec | $t_2$ | .056 | .0564 | .0565 |
| sec | $t_3$ | .0557 | .0558 | .0558 |
| sec | $\sigma_1$ | .0027 | .00279 | .00270 |
| sec | $\sigma_2$ | .00051 | .000597 | .000519 |
| Hz | $f$ | 327.6 | 344.4 | 361.08 |
| No of Iterations | | 18 | 19 | 21 |
| Computing Time (sec) | | 51.2 | 58.0 | 63.2 |
| Experimental Variance ($\sigma_e^2$) | | NOT APPLICABLE | $(.45)^2$ | $(.317)^2$ |

TABLE 1. Parameters Fitting Pressure Curve.

| CASE / ORDER | 301 | 511-5 | 511-10 |
|---|---|---|---|
| 1 | .717 | .680 | .696 |
| 2 | .477 | .427 | .443 |
| 3 | .286 | .247 | .248 |
| 4 | .058 | .027 | .013 |
| 5 | -.113 | -.136 | -.160 |
| 6 | -.246 | -.258 | -.291 |
| 7 | -.303 | -.315 | -.346 |
| 8 | -.322 | -.334 | -.361 |
| 9 | -.307 | -.312 | -.342 |
| 10. | -.245 | -.240 | -.276 |
| 11. | -.130 | -.104 | -.155 |
| 12. | -.011 | .031 | -.035 |
| 13. | .069 | .130 | .048 |
| 14. | .137 | .217 | .123 |
| 15. | .110 | .203 | .106 |
| 16. | .075 | .168 | .085 |
| 17. | -.028 | .057 | .0004 |
| 18. | -.141 | -.068 | -.090 |
| 19. | -.202 | -.149 | -.130 |
| 20 | -.235 | -.205 | -.145 |
| Sum. Sqs. | 57.9 | 67.9 | 57.62 |
| $X_1^2$ | 57.9 | 62.96 | 116.2 |
| $X_2^2$ | - | 33.01 | 62.4 |
| $X_{TOT}$ | - | 95.97 | 178.6 |

Table 2. Autocorrelations and Chi-Square for final model of XM211 Pressure Oscillations. ($X_2^2$ based on the first 5 Autocorrelations for Case 511-5, and the first 10, for case 511-10)

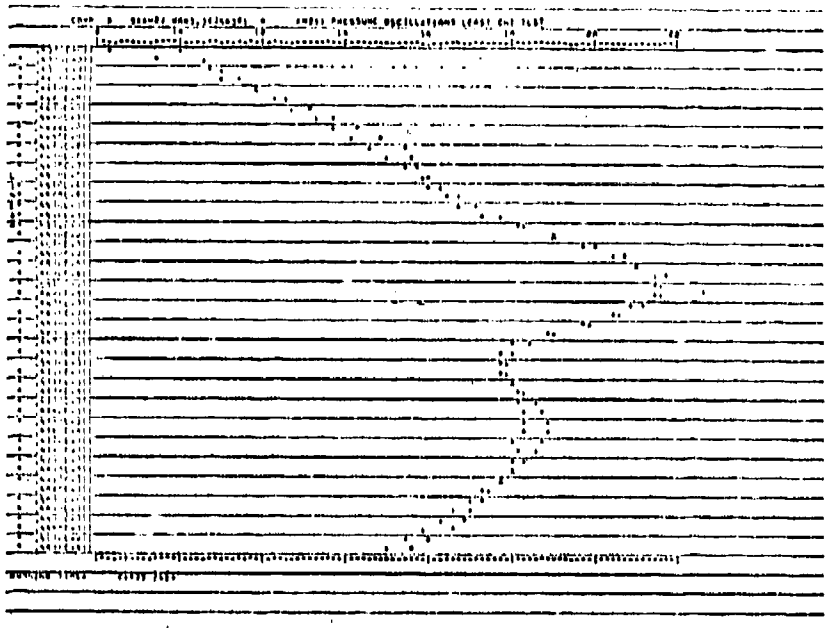


Figure 1 - Pressure as a function of time. "+"s indicate theory for case 301.

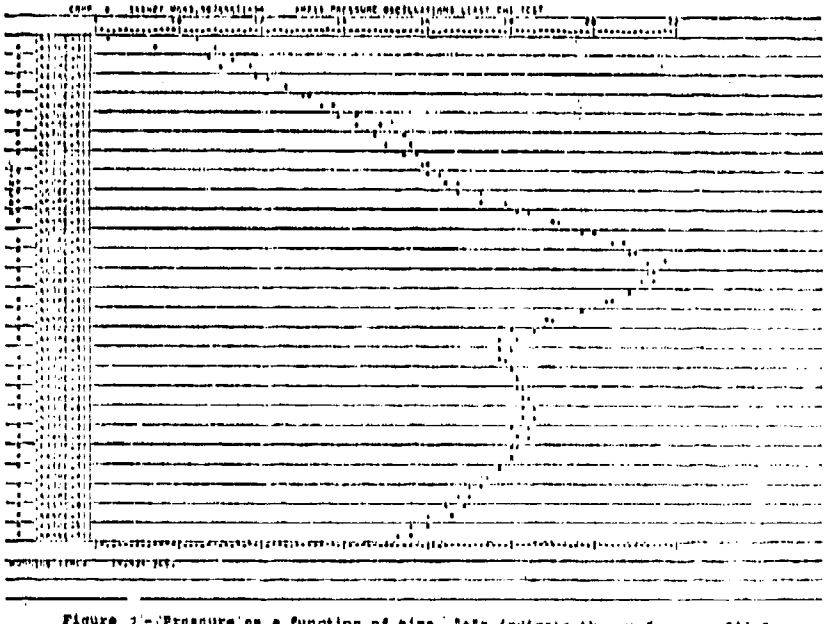


Figure 2 - Pressure as a function of time. "+"s indicate theory for case 511-5



THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

| CASE | | 4.012-0 | 4.011-6 |
|---|---|---|---|
| NO. of ITER. | | 4 | 2 |
| L(0,4) | | 17.70 | 18.58 |
| L(4,1) | | 53 :1 | 53.02 |
| L(4,?) | | 1.094 | 0.605 |
| L(4,.. | | 6.204 | 10.26 |
| R(I) | 1 | | -.110 |
| | 2 | | -.232 |
| | 3 | | -.014 |
| | 4 | | .068 |
| | 5 | | -.124 |
| | 6 | | .030 |
| Sum of sqs | | .03147 | .03138 |
| $X_1^2$ | | 28.90 | 28.82 |
| $X_2^2$ | | 2.77 | 2.49 |
| $X_T^2$ | | 31.67 | 31.31 |

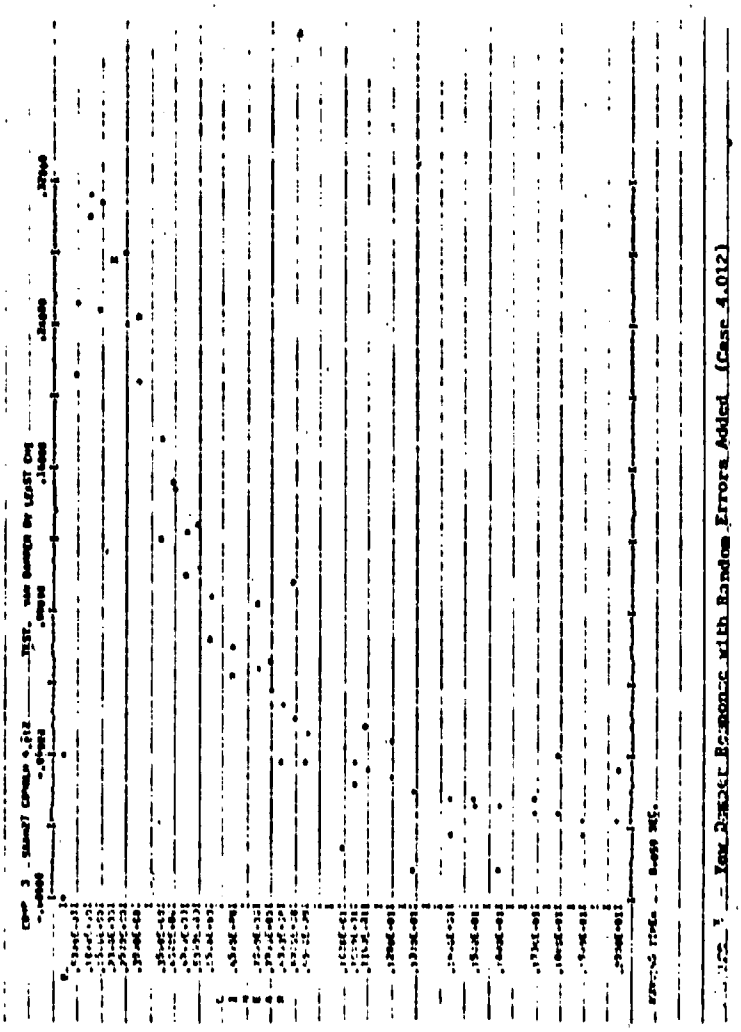
Table 3. Results of Yaw Damper Calculations

| ITEM/CASE | KJE 1.0021-0 | KJE 1.0023-5 |
|---|---|---|
| L(1,2) | .2153 | .2199 |
| L(2,1) | .4527 | .4442 |
| L(3,1) | .0431 | .0431 |
| L(4,1) | .0252 | .0251 |
| L(5,1) | .0743 | .0824 |
| R(1) | .122 | -.086 |
| R(2) | .065 | -.052 |
| R(3) | .009 | .034 |
| R(4) | .227 | -.223 |
| R(5) | .179 | -.168 |
| $X_1^2$ | 31.22 | 31.24 |
| $X_2^2$ | 3.99 | 3.04 |
| $X_{TOT}$ | 35.21 | 34.28 |

Table 4. Results of Brookhaven example calculation. Autocorrelation, $X_2$, and $X_T^2$ for case KJE 1.0021-0 computed for comparison.

| ITEM/CASE | KWB 1.005-0 | KWB 1.005-3 | KWB 1.0007-6 |
|---|---|---|---|
| L(11,1) | $5.318 \times 10^{-4}$ | $5.3183 \times 10^{-4}$ | $5.262 \times 10^{-4}$ |
| R(1) | .782 | .782 | .786 |
| R(2) | .44 | .44 | .453 |
| R(3) | .098 | .098 | -.122 |
| R(4) | -.204 | - | -.170 |
| R(5) | -.316 | - | -.274 |
| R(6) | -.235 | - | .290 |
| $X_1^2$ | 121.14 | $1.2 \times 10^9$ | 121.89 |
| $X_2^2$ | 36.71 | | 35.0 |
| $X_T^2$ | 157.85 | | 156.89 |

Table 5. Results of Kinetic Experiment Water Boiler Calculations



THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

The times for one iteration were 8.2 and 8.5 sec respectively. (Part of the increase in time for the least chi-square case was due to several attempts in both iterations to improve the fit by reducing the step size.) As shown in Table 3 the parameters L(0, 4), L(4,2) and L(4,3) appear to be different by significant amounts. (The autocorrelations for case 4.011-6 appear well within the random range.) The normalized sum of squares of the residuals is less, as expected, for the ordinary least squares, Case 4.012.

Brookhaven Example. A sample test case was received from Miss Rita Straub of Brookhaven National Laboratory. The exact nature of the probem was unspecified but from the form of the differential equations it appears to be a kinetic problem in which the material in component one decays into components two to five, and component two may change into component one. Component seven is composed of components three, four, and five. Although some coupling parameters may actually be unknown, they were assumed known, because the present version of the program will not iterate either type of linear coupling parameter with the least chi-square algorithm. The data were available for the amount of components 1, 3, 4, and 6 as a function of time, (where component 6 is the sum of components 1, 2, and 5).

Both the run with no autocorrelations and the run with 5 correlations (KJE 1.0023-5), took 7 iterations to converge. The results for the two cases are compared in Table 4. Since the value of $X_1^2$, (31.24) is large compared to $X_2^2$ (3.04), the major emphasis in this case was on reducing the sum of squares, and thus it is similar to the case run with no weight on the autocorrelations. As would be expected, there is only a small difference between the final values of the parameters of the two cases.

Reactor Kinetics Example. This example illustrates two things: First the use of the least chi-square algorithm, and second an apparently good fit between data and a physically incorrect model. Hetrick and Gamble (13) proposed a non-linear feed-back term proportional to the energy in the reactivity of the KEWB reactor to describe the fit. Later experiments (16) where the void amount was inferred from measurements and where the thermal effects on reactivity were also carefully measured, showed that shutdown was due to thermal, not void effects. In the simulation, the effect of the energy on void formation was simulated by the parameter L(11, 1). The functions correspond, in numerical order, to the functions used in the simulation: (1) Nuclear reactor power level, (2) Mean temperature, (3) Mean void volume, (4)-(9) Delayed neutron groups, (10) Not used, (11) Energy released to that time. The result of the iterations is shown in Figure 4, a logarithm plot of theoretical and experimental nuclear power. In Table 5, three different cases are shown:

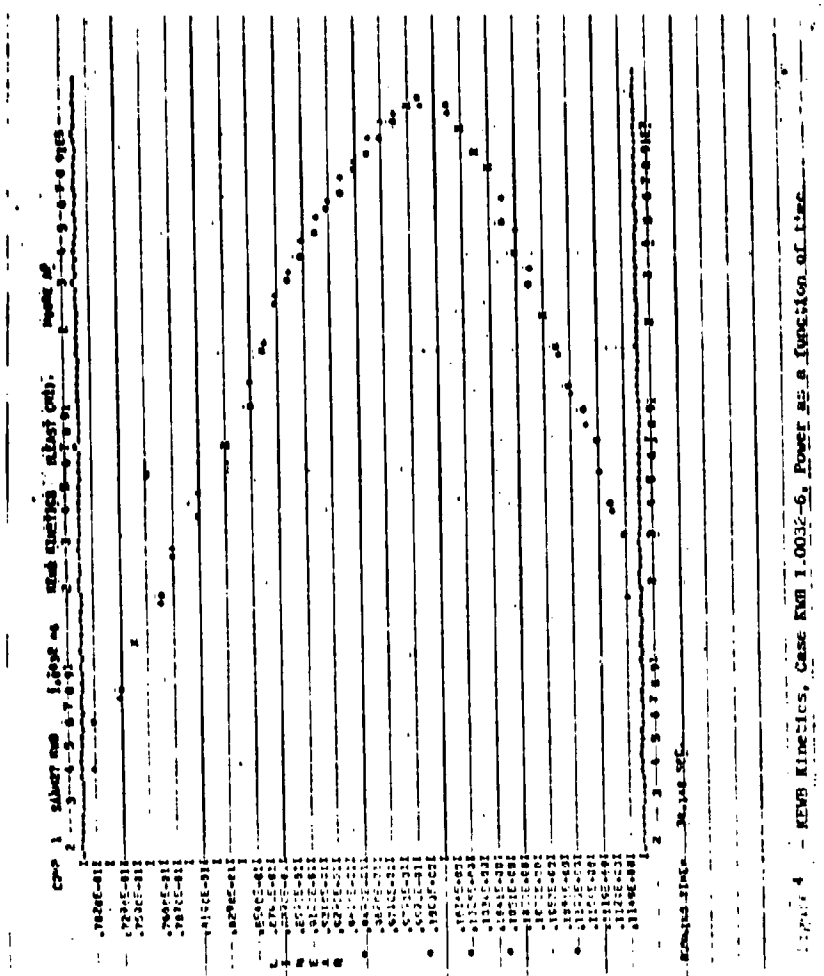


Figure 4 - KEVB Kinetics, Case KWB 1.0032-6, Power as a function of time

Case 1.003-0 was ordinary least-squares. The values of the autocorrelations and chi-squares are shown for comparison with the other two cases. Case 1.005-3 used three autocorrelations with a small value of the experimental variance thus resulting in a large value of $X_1^2$. Both case 1.007-6 and 1.003-0 use $1 \times 10^7$ for the experimental variance thereby reducing the emphasis on the sum of the squares of the errors. All of these runs took four iterations to converge.

Cases 1.003-0 and 1.005-3 give almost exactly the same results. On comparing 1.003-0 with 1.007-6, a difference is found in the value of the adjustable parameter L(11,1). The value of chi-square total is smaller for 1.007-6, and thus this result would be chosen over that of the other case.

The value of the chi-square for the Box-Pierce number is much smaller for case 1.007-6, although $X_1^2$ is slightly larger for the same case--thus illustrating the trade-off between getting the minimum as in ordinary least squares, and reducing the autocorrelations as in least chi-squares. The data for Case 1.003 show the values for R(1) to R(6) for comparison purposes. The data show that the sum of squares does not increase from one to the other appreciably, but $X_2^2$, the Box-Pearce statistic, does change appreciably. Each of the calculations give a total chi square which is too large to be consistent with the residuals being drawn from a random sample, and thus would have given support for the rejection of the Hetrick-Gamble model.

Comparison of Computing Time. Table 6 summarizes the comparison of the number of iterations to converge, and the computing time. The number of iterations was usually about the same. As seen in the last column the computing time is comparable, with a tendency for the computing time to be longer for least chi-square than for least squares. The relative difference is greater when the original total computing time is short. This just means that, as would be expected, it takes a larger fraction of the computing time to compute the matrix $\Gamma$ and post-multiply into P*' for cases where the time of iteration is short.

Conclusions. Based on four different types of non-linear theoretical models for data analysis, our results indicate that:

(1) Least chi-square is practicable for non-linear analysis.

(2) Least chi-square gives a better fit, and is a more reliable iteration procedure.

(3) The computing time for least chi-square is longer for the models which use less computing time, but because the convergence of this iterative procedure is somewhat better, the number of iterations (and particularly the number of "tries" per iteration) is reduced, thus keeping the total computing time about the same. Those models with longer integrating time would be expected to benefit more from least chi-square.

(4) In validation of simulations of future Army systems, the SAAM-27 computer program modified for least chi-square can be used at various stages in the system development. First, as a tool to simulate subsystems and compare the projected performance with the designer's simulation. Second, as subsystems are built and tested, they can be run as "hardware in the loop" and the test data used in the least chi-square program to validate the computer simulation and provide system parameter identification. Because no programming is needed to run SAAM-27 on a variety of problems, both the programming time and the elapsed time is greatly reduced.

By planning ahead to use SAAM in the validation of the subsystem modeling and providing the needed subsystem tests, a Program Manager can reduce the time and effort needed to validate the contractor's system simulation and will be able to give an impartial, knowledgeable, and timely evaluation of each system.

Acknowledgements: Mr J. Bay of ARRADCOM has capably performed the programming needed to modify SAAM-27. Discussions with Dr. Ray Boston of La Trobe University, Bundoora, Australia on details of the modification of SAAM-27 have been essential for its success.

References:

1. Pastrick, H. Proc., 23 Army Conference on Design of Experiments, ARO Report 1977.

2. Griner, Gary M., John Mango, and H. Pastrick, Validation of Simulation Models Using Slatistical Techniques, Proc. of Summer Computer Simulation Conference, Newport Beach, Ca., July, 1978.

3. J. M. Chambers, Biometrika, 60,, 1, (1973). A Celmins, A Manual for General Least Square Model Fitting, Ballistic Research Laboratory, TR-02167, June 79. Stewart, Warren E. and Jan P. Sorenson, Bayesian Estimation of Common Parameters from Irregular Multiple Response Data. Proc. 25th Army Conference on Design of Experiments, ARO Report, 1980.

4. Aitken, A, C,. Proc. Roy. Soc. Edinb. A 55, 42-47 (1934) and R.L. Plackett, Principles of Regression Analysis, Oxford, Clarendon Press, 1960.

5. Anscombe, F. J., and J. W. Tukey, Technometrics, 5, 141, 1963 and J. R. MacDonald, Rev. Mod. Phys. 41, 316, 1969.

6. Box, G. E. P. and D. A. Pierce, J. Amer. Stat. Assoc. 64, 1509, 1970.

7. Moore, R. L., Proc. 1975 Army Numerical Analysis Conference, ARO Report 75-3.

8. Moore, R. L., Proc. 23d Army Conference on Design of Experiments, ARO Report, 1977.

9. Moore, R. L. Proc. 25th Army Conference on Design of Experiments, ARO Report, 1980.

10. Berman M., E. Shahn, and M. F. Weiss, Biophys J. 2, 275, 1962.

11. Berman, Mones, and M. F. Weiss, USERS MANUAL FOR SAAM (Simulation, Analysis and Modeling) Version SAAM 27, April 15, 1977, Math Research Branch, National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, Bethesda, MD.

12. Model 10 of reference 11.

13. Hetrick, D., and J. Gamble, Transient Reactivity during Power Excursions in a Water Boiler Reactor, Atomics International, Feb 9, 1958.

14. Knutelski, B. G., private communication.

15. Data from Dr. Y. Hiroshige of Douglas Aircraft Co.

16. Dunenfeld, M. S., R. K. Stett, Summary Review of the Kinetics Experiments on Water Boilers, Atomic International, SR 7087, Feb 1, 1963.

| Case | No. of Adjustable Parameters | Rank of Audo-correlation | No of Iterations | Time (sec) |
|---|---|---|---|---|
| Gun Chamber Pressure Curve. | 8 | 0 | 18 | 51.2 |
| | 8 | 5 | 19 | 58.0 |
| | 8 | 10 | 21 | 63.7 |
| Yaw Damper. | 4 | 0 | 4 | 8.2 |
| | 4 | 6 | 2 | 8.5 |
| Biomedical Test Case. | 5 | 0 | 5 | 14.9 |
| | 5 | 5 | 5 | 15.2 |
| | 5 | 5 | 7 | 21.5 |
| Reactor Kinetics Experiment. | 1 | 0 | 4 | 84 |
| | 1 | 3 | 4 | 81 |
| | 1 | 6 | 4 | 89 |

Table 6. Comparison of computing time and number of iterations.

# ARMOR DESIGN BASED ON MATERIAL PROPERTIES (U)

GERALD L. MOSS, PhD
USA ARRADCOM, BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MARYLAND 21005

## I. INTRODUCTION

As a result of extensive research over the last 20 years, an analysis of high-rate loading and fracture has evolved that models the three main features of failure caused by stress waves. The nucleation and growth of voids are the governing processes on which the model is based. In the following, a description will be given of the characterization of a rolled homogeneous steel armor and a penetrator alloy in terms of this model. It will be shown by the results that the model is sensitive to the microstructural detail of the materials characterized. Finally, extensions of the model that are essential to the design of armor will be described along with an indication of how such a design is accomplished.

## II. FRACTURE WITH STRESS WAVES

The nucleation, growth and coalescence of large numbers of voids were first identified as characteristic features of fracture with stress waves (spallation) in 1960.(1,2) Large numbers of independent voids are a unique feature of spallation and occur because stresses much greater than the strength of a material can be applied before there is appreciable crack growth. This is possible because crack velocities cannot exceed approximately $0.38\sqrt{E/\rho}$ ($\sqrt{E/\rho}$ is the longitudinal elastic wave velocity $V_1$), and because disturbances associated with the opening of a void can only be transferred with the speed of stress waves. A high density of independent voids is, therefore, a general feature of spallation.



PRECEDING PAGE BLANK-NOT FILMED

Seaman, Barbee, Crewdson, and Curran subsequently developed an ingeneous technique that allows one to describe the spallation of a given material quantitatively.(3,4) The technique is based on an experimental determination of an expression that characterizes the size distribution N(R) (R is the radius of the void) of voids and rate equations, $\dot{N}$ and $\dot{R}$, for the nucleation and growth of voids. No assumption is made about the form of these equations. It is only assumed that the progression of failure can be described in terms of these variables.

The approach is to determine the relations for N(R), $\dot{N}$ and $\dot{R}$ experimentally with plate impact tests conducted over a range of impact velocities. Details of the method have been described in summary articles,(4,5) but it is emphasized that a unique and especially important feature of the analysis is the method of handling large numbers of voids. This has been possible through a continuum approach that describes void densities, N(R), at a point. Another unique feature of the model is that it includes a description of the effect of fracture on the loading history. It is, therefore, a truly "active" fracture model.

Characterizations of the spallation of several materials,(6) including Al, Cu, Fe, S-200 Be and Lexan have resulted in expressions for N(R), $\dot{N}$ and $\dot{R}$, and it has been found that equations with the same form approximate the behavior of each of these materials. These equations are

$$N(R) = N_o e^{-R/R_1} \tag{1}$$

$$\dot{N} = \dot{N}_o e^{(\sigma-\sigma_{no})/\sigma_1} \tag{2}$$

$$\dot{R} = (\sigma-\sigma_{go})R/4\eta \ . \tag{3}$$

The parameters $N_o$, $R_1$, $\dot{N}_o$, $\sigma_{no}$, $\sigma_1$, $\sigma_{go}$, and $\eta$ reflect the nature of a material, and the rate equations depend on the applied stress $\sigma$ explicitly which makes the progression of failure sensitive to the loading history. Once the Eqs. for N, $\dot{N}$ and $\dot{R}$ have been determined for a specific material, the approximate spallation of this material can be predicted for general stress histories bounded by the range of the calibration. Presumably, the range of applicability can be extended somewhat by an extrapolation of results.

## III. SPALLATION OF STEEL ARMOR

In a characterization of the spallation of a low alloy Ni-Cr

rolled steel armor (RHA) with the approach just described, it was found that Eqs. 1-3 are also approximately representative of the spallation behavior of this material. The material parameters determined are listed in Table I along with similar results for a W-7Ni-3Fe penetrator alloy and Cu.(6) The parameters for each of the materials are different as they should be in a realistic material-dependent model, but even though they are different for each material and allow one to compute how a previously characterized material will break under general loading conditions, it is still not possible to design materials with the nucleation and growth (NAG) model. This possibility has

TABLE I

VOID NUCLEATION AND GROWTH PARAMETERS FOR SPALLATION

| PARAMETER | UNITS | RHA* | 90W-7Ni-3Fe | Cu |
|---|---|---|---|---|
| $\sigma_{no}$ | $GPa/cm^2$ | 1.8 | -2.8 | -0.50 |
| $\dot{N}_o$ | $No/(cm^3 sec)$ | $5.0 \times 10^8$ | $2.5 \times 10^{14}$ | $2.8 \times 10^{12}$ |
| $\sigma_1$ | $GPa/cm^2$ | -0.25 | -2.53 | -0.20 |
| $\sigma_{go}$ | $GPa/cm^2$ | -0.2 | -1.0 | -0.50 |
| $1/4\ \eta$ | $cm^2/(GPa\ sec)$ | $-7.0 \times 10^5$ | $-5.0 \times 10^6$ | $-1.3 \times 10^7$ |
| $R_1$ | cm | $3.0 \times 10^{-3}$ | $1.4 \times 10^{-4}$ | $1.0 \times 10^{-4}$ |

*1.27cm rolled steel armor plate

eluded us because of the order in which we have learned about spallation. Initially, the goal was to quantitatively describe N(R), $\dot{N}$ and $\dot{R}$ with simple enough equations to be useful in treating complex ballistics problems. The approach was to use the data from plate impact-fracture tests exclusively to establish the equations.(4) Hence, the equations certainly apply to spallation, but the parameters are not necessarily related to material variables that are known to govern mechanical behavior and to be controllable with manufacturing methods. These connections still need to be made.

Apparently, more information is required about the NAG parameters and material behavior before material design can be attempted with the nucleation and growth equations. Some information

along these lines is implied by the crack-size distributions that were used in determining the parameters listed in Table I. A typical result is shown in Fig. 1 where it can be seen how the cumulative number of cracks varied with crack radius. Clearly, these results cannot be completely described with Eq. 1, although it is made up of straight-line segments.

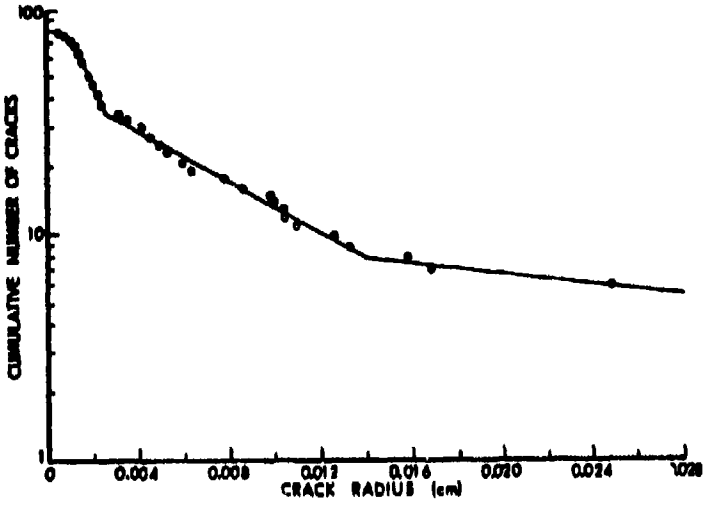



Figure 1. Size distribution of cracks formed in steel armor (1.27cm RHA) with the stress waves from a plate impact. Impact velocity was 0.1178 mm/μsec.

Without additional information, it is not obvious why the curve is segmented, but when the cracks were observed on cross sections of partially broken samples, the reasons for the two changes in slope became clear. Three distinct stages of cracking were observed, and these are illustrated with Figs. 2 through 4.

First, there was crack nucleation at MnS inclusions during which the inclusions were cracked as well as separated from the matrix. At intermediate stages of cracking, short irregularly oriented "hair-line" cracks extended from inclusion to inclusion in planes approximately parallel to the plane of the plate. Most of the MnS inclusions were found in these planes which were the weak planes of a banded microstructure that is typical of rolled steel armor. In the last stage of failure, long non-coplanar cracks opened and coalesced by plastic shear on connecting surfaces.

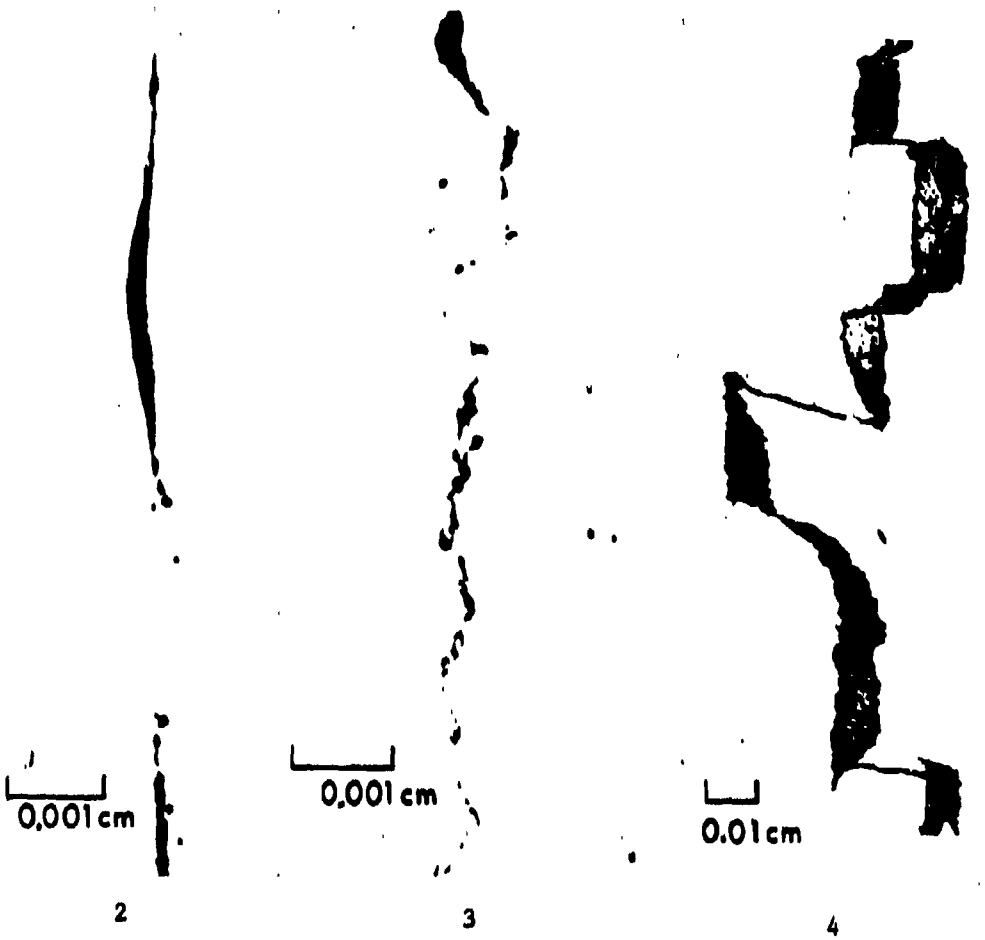



Figure 2. Early stage of fracture in steel armor showing crack nucleation at sulfide inclusions.

Figure 3. Intermediate stage of fracture in steel armor showing cracks extending from inclusion to inclusion in approximately a single plane.

Figure 4. Late stage of fracture of steel armor showing the coalescense of non-coplanar cracks by plastic shear on connecting surfaces.

These distinct stages of cracking are exactly what is reflected in the crack-size distribution shown in Fig. 1. The part of the curve bounded by R ≤ 0.0025cm is the region where there was primarily nucleation at MnS inclusions. Inclusion diameters observed with the optical microscope ranged from the resolution limit to 0.005cm. The change in slope at 0.0025cm corresponds to the beginning of the intermediate stage of cracking. Finally, the change in slope at R = 0.0135cm is an indication of the onset of coalescence of non-coplanar cracks. This is in agreement with the crack lengths in Fig. 4 which range from 0.01 to 0.04cm.

The coalescence of the non-coplanar cracks is the mechanism by which the material finally comes apart, and an independent model of this phenomenon has been presented by Seaman et al.(5) Hence, the consideration will be limited here to the first two stages of failure. These can be reasonably approximated with the NAG analysis because the volume of the cracks with radii less than 0.0025cm is small compared with the total crack volume. They contribute little to the total separation, and it is sufficient to describe the size distribution of cracks with radii in the range 0.0025 ≤ R ≤ 0.0135cm. Hence, even though the entire curve in Fig. 1 is too complex to be described with Eqs. 1-3, they reasonably approximate the intermediate stage of cracking. This is the important stage to characterize because this is when the greatest void volume, i.e., damage, develops.

An important conclusion to be drawn from these observations is that the crack measurements displayed in Fig. 1 are sensitive to the microstructure of the material. Clearly, banding, inclusion size and inclusion distribution are important microstructural features that govern the failure of RHA, and if a method were to be derived to allow steel armor to be tailored to resist spallation, it would be important to include the effects of banding and inclusions in the analysis. Such an analysis will be presented next.

## IV. MATERIAL DEPENDENCE OF FRACTURE

Conceivably, the material dependence of the parameters in Eq. 3 could be established by deriving an independent relation for $\dot{R}$ in terms of the governing material features and equating it to Eq. 3. Such an attempt will be described because it has yielded useful results.

The velocity of an elastic crack $\dot{C}_e$ can be determined from an expression of energy conservation during fracture

$$\frac{d}{dC}\left(-\frac{\pi\sigma^2C^2}{E} + 4\gamma C + \tfrac{1}{2}k\rho\dot{C}_e^2C^2\frac{\sigma^2}{E^2}\right) = 0, \qquad (4)$$

where the first term in parenthesis is the elastic energy, the second term is the surface energy and the last term is the kinetic energy. The applied stress is $\sigma$, C is the crack radius, $\gamma$ is the surface energy, $\rho$ is the material density and E is Young's modulus.

Differentiating Eq. 4 and solving for $\dot{C}_e^2$ leaves

$$\dot{C}_e^2 = \left(\frac{2\pi\sigma^2C}{E} - 4\gamma\right)\left(\frac{1}{k\rho C}\right)\left(\frac{E}{\sigma}\right)^2. \qquad (5)$$

This can be expressed in terms of the critical crack size $C_o$ at which crack growth becomes unstable. This is given by the Griffith condition as

$$C_o = \frac{2\gamma E}{\pi\sigma_c^2}, \qquad (6)$$

where $\sigma_c$ is the critical stress corresponding to the instability condition. Substituting Eq. 6 into Eq. 5 and solving for $\dot{C}_e$ leaves

$$\dot{C}_e = \sqrt{\frac{2\pi}{k}}\sqrt{\frac{E}{\rho}}\left(1 - \frac{C_o\sigma_c^2}{C\sigma^2}\right)^{\frac{1}{2}}. \qquad (7)$$

Usually, stresses are applied gradually until $\sigma$ equals $\sigma_c$, and it is assumed that subsequent failure proceeds at this stress level. Then,

$$\dot{C}_e = \sqrt{\frac{2\pi}{k}}\,V_1\left(1 - \frac{C_o}{C}\right)^{\frac{1}{2}}. \qquad (8)$$

This is the equation commonly used for $\dot{C}_e$. However, as explained in §2, it is possible to impose stresses during fracture that are much greater than $\sigma_c$ when the loads are applied with stress waves. Then, Eq. 7 is the correct expression for $\dot{C}_e$.

Steel armor is not a purely elastic material, but also deforms plastically as it fails. It is well known that for such materials, the expression for the surface energy should include the effects of plastic deformation. Gilman(7) has described such a result, and if it is combined with his relation for the surface energy of elastic cleavage, an expression is obtained for the energy of a

cleaved surface formed with plastic deformation. The result is

$$\sigma_p = \frac{9E_{hkl}}{y_{hkl}} \left(\frac{a_o}{\pi}\right)^2 \ln\left(\frac{2G}{\pi\sigma_Y}\right) , \qquad (9)$$

where the indeces hkl define the cleavage plane, $y_{hkl}$ is the distance between cleavage planes, $a_o$ is the atomic radius, G is the shear modulus and $\sigma_Y$ is the yield strength.

There is an additional plastic effect on the crack velocity that has previously been ignored. The effect is to shift the elastic stress field beyond the crack by the range of the plastic strain field. For plane-strain conditions, applicable when loading is with one-dimensional stress waves, the length of the plastic zone in front of the crack tip is given by(8)

$$R_m = \rho_t \left[e^{(\pi-\alpha)/2} - 1\right] , \qquad (10)$$

where $\rho_t$ is the crack-tip radius, and $\alpha$ is the wedge angle of the crack.

The factor $\sqrt{2\pi/k}$ in Eq. 7 equals the range of the stress field(7) normalized to the crack length. It is limited to $\dot{C}_m/V_1$ ($\dot{C}_m$ is the maximum crack velocity) for an elastic material, but when there is plastic deformation, this must be increased by $R_m/C$. Hence, when there is a plastic zone leading the crack,

$$\sqrt{\frac{2\pi}{k}} = \frac{\rho_t\left[e^{(\pi-\alpha)/2} - 1\right]}{C} + \frac{\dot{C}_m}{V_1} . \qquad (11)$$

The radius of the crack tip is approximately $\rho_t = C\sigma^2/(2\sigma_Y G)$.(7,9)

Substituting Eq. 11 for $\sqrt{2\pi/k}$ and Eq. 9 for $\gamma$ in Eq. 7, indicates that the crack velocity $\dot{C}_p$ in an elastic-plastic material is given by

$$\dot{C}_p = \left\{\frac{\sigma^2}{2\sigma_Y G}\left[e^{(\pi-\alpha)/2} - 1\right] + \frac{\dot{C}_m}{V_1}\right\} V_1 \left\{1 - \frac{\frac{\sigma_c^2}{\sigma_Y^2}\frac{18}{\pi}E_{hkl}\left[\frac{E_{hkl}}{Y_{hkl}}\left(\frac{a_o}{\pi}\right)^2\right]\ln\left(\frac{2G}{\pi\sigma_Y}\right)}{C\sigma^2}\right\}^{1/2} . \qquad (12)$$

It was initially assumed that cracking would propagate through the weakest part of a material, and in a bainitic steel such

as steel armor this would be in the ferrite. With this assumption and the properties of ferrite, which cleaves on {100} planes, it was estimated with Eq. 12 that the radius of the crack nucleus activated with an applied stress of 1.65 GPa should be 3.9μm in 1.27cm thick RHA plate. A radius of 1.3μm is predicted for a 2.88 GPa load. This is in approximate agreement with the smallest cracks that have been observed, but the crack velocity computed with the same assumptions was found to be greater than has been observed.

The possibility that cracks might pass the strongest barriers was tested by expressing $\gamma$ in terms of the fracture toughness of RHA. At an applied stress of 2.88 GPa, this procedure predicted a nucleus radius of 0.019cm. This is larger than most of the observed cracks as documented in Fig. 1. The conclusion from this must be that for stresses up to at least 2.88 GPa cracks do not propagate straight through rolled steel armor without regard for the microstructure.

The results for the two conditions investigated above indicate that initial crack growth is in the ferrite since the critical nucleus size computed is about the size of the smallest cracks observed, but these cracks must stop when they intersect strong barriers. This is why the average crack velocity is less than computed. The computed velocity applies only up to the time the crack is stopped. Continued growth can be by either coalescence of cracks or reactivation of the stopped crack. If continuation were by coalescence, there should be a sea of microcracks surrounding the main crack. There is some evidence of this where there are clusters of inclusions, but this is not typical. Continued growth by reactivation is the alternative. Irregular continuous cracks should be expected when barriers are important and overcome by reactivation of a stopped crack, for if barriers are important, the crack will continue to propagate in the weakest phase. Hence, it will change direction as it moves into material with a different crystallographic orientation. An irregular crack path is exactly what is seen in Fig. 3. Why is a delay time required if the applied stress is greater than the nucleation threshold stress $\sigma_{no}$? It is proposed that this time is required for the crack to open sufficiently to activate the blunted crack at the barrier.

At this point, several features of the breaking process have been clarified, but it is apparent that $\dot{C}_p$ cannot be equated to $\dot{R}$ for all C to establish fracture parameters because crack growth is intermittent.

Presumably, Eq. 12 gives the correct crack velocities when growth is in progress, but part of the time the cracks are blocked.

The effective velocity is given by

$$\dot{C}(C)_E = \frac{l_i \cos\theta_i}{\Delta t_{ci} + \Delta t_{oi}} = \frac{\dot{C}_p(C)\ \Delta t_{ci} \cos\theta_i}{\Delta t_{ci} + \Delta t_{oi}}, \qquad (13)$$

where the subscript i identifies the segment of the crack to open as the crack grows from C to C+ΔC, $l_i$ is the length of the crack segment breaking, $\cos\theta_i$ is the projection of $l_i$ onto the main crack path, $\Delta t_{ci}$ is the time during which the crack propagates and $\Delta t_{oi}$ is the time during which the crack is blocked at a barrier before cracking starts along $l_i$. As the crack lengthens, segments break more quickly according to Eq. 12, which means $\Delta t_{ci}$ decreases with C. The dwell time at barriers should also decrease since the time to attain the crack opening sufficient to overcome a barrier must decrease as the rate at which the crack opens increases. It is proposed that the dependence of $\Delta t_{oi}$ on C is approximated by letting $\Delta t_{oi}$ be proportional to $\Delta t_{ci}$. Then $\Delta t_{oi} = F\Delta t_{ci}$, and

$$\dot{C}(C)_E = \dot{C}_p \cos\theta_i/(1+F). \qquad (14)$$

As indicated in §III, cracks nucleate in RHA at inclusions and grow in the weak bands of the microstructure, but the MnS inclusions are concentrated in the weak bands. This leads to the coalescence of major coplanar cracks after growth dependent on the inclusion spacing. Since the parameters listed in Table I were determined to represent the intermediate stage of cracking, $\dot{R}$ computed with these parameters and Eq. 3 should reflect crack growth and coalescence except for crack lengths less than the inclusion spacing. Then, $\dot{R}$ equals $\dot{C}_E$. This equality at sufficiently small crack sizes allows $\cos\theta_i/(1+F)$ to be determined from experimental measurements for an average angle $\theta_i$ characteristic of the steel. Now, Eqs. 12 and 14 should reasonably express the magnitude and material dependence of the crack velocity in sound material.

Eq. 12 is based on the plastic work at the new surface created as a crack propagates and on the effect of plastic deformation on the stress field bounding the crack tip. Hence, the form of the equation and the yield strength, which is included in the result, reflect the plastic nature of the material. Although nucleation and growth have been related to the banding and inclusions of steel armor, the quantitative effects of these microstructural features on $\dot{C}_E$ are not explicitly stated, and as far as the author knows, they have escaped previous description. However, one effect of the inclusions on the crack velocity is now apparent. They limit the maximum crack length on which the crack velocity is directly dependent. The crack

length is bounded because the MnS inclusions exist in planes of the banded microstructure and because cracks are nucleated at the inclusions independently, as described in §II. The maximum crack length is approximately half the inclusion spacing since coalescence must occur when the cracks reach this length. The width and spacing of the bands of the microstructure, in which the sulfides occur, establishes the actual inclusion spacing for a given sulfide volume and inclusion size. In this way, the banding of the steel indirectly affects the crack velocity.

The new point stressed is that quantitative predictions of the crack velocity of steel armor can be made with Eqs. 12 and 14 as a function of the chemistry and processing of the steel. This is possible because the spacing and width of the bands in which the MnS inclusions reside, the sulfur content of the steel and the inclusion size establishes the distance between the inclusions and the crack length on which the crack velocity is directly dependent.

Manufacturing techniques allow substantial control over the sulfide inclusions in a steel. For example, by varying the sulfur content from $10^2$ to $10^4$ ppm and the inclusion diameter from 0.1 to 10μm, the distance between inclusions can be changed from 0.2 to 200μm. This is equivalent to limiting the maximum crack size and, therefore, the crack velocity.

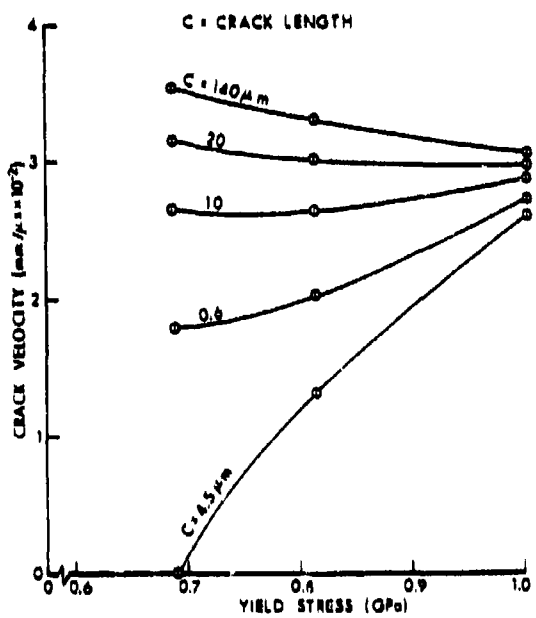



Figure 5. Crack Velocity under an applied stress of 2.88 GPa.

The effect of the crack size and the yield strength on the crack velocity predicted with Eq. 14 is illustrated in Fig. 5 for an applied stress of 2.88 GPa. It is apparent that the crack velocity can either increase or decrease as the yield strength is decreased. Whichever occurs, depends on the crack size. This is a new result that occurs because of the effect of the plastic zone on the stress field at the crack tip.

When cracks are long, it is clear that the crack velocity increases to infinity as $\sigma_Y$ goes to zero. This occurs because $\rho_t$ is proportional to $\sigma_Y^{-1}$, but there is a degree of realism to the result. It can be interpreted as though the entire cross section of the sample is the process zone of a crack with complete certainty that microcracking will occur everywhere in this zone.

The material parameter $N_o$ is the total number of crack nuclei. The total number available equals the number of inclusions larger than the critical nucleus size. The parameter $\dot{N}_o$ is the nucleation rate when $\sigma$ equals $\sigma_{no}$. This should be proportional to the density of available nucleation sites and, therefore, to the inclusion density. Hence, $\dot{N}_o = KN$. The proportionality constant was estimated with results from the program in which the parameters in Table I were determined. The parameter $N_o$ and the relation for $\dot{N}_o$ are two quantitative ways of expressing the effects of inclusions on spallation.

Methods of predicting the void nucleation threshold stress $\sigma_{no}$ have been proposed, but in each of these, the shape of the void must be known, or assumed. Bounds can be placed on $\sigma_{no}$ as a function of the yield strength, but these allow widely different values of $\sigma_{no}$ for a given yield stress. At this point, an experimental determination of the dependence of $\sigma_{no}$ on the yield strength is more reliable. Such observations for steel armor reveal that in the yield strength range from $0.75 \leq \sigma_Y \leq 1.0$ GPa, $\sigma_{no}$ increases as $\sigma_Y$ decreases. These experimental results were used in the material computations described next.

## V. ARMOR DESIGN

A new method of designing a material to resist spallation is proposed that is based on the crack nucleation and growth-rate equations, the dependence of these equations on the governing material properties and a method of computing the damage for a given impulse. In such a design, the anticipated stress history is an important design condition. Stresses can be applied that will certainly eventually destroy an armor plate. The general question is, How resistant is a material to a stress for a specific time? The design and performance of a material are closely tied to the impulse it must withstand.

The rate equations and the governing material properties have been described, but these are not sufficient to fully assess the changes in spallation resistance with material changes. A description is still required of the in-plane crack coalescence that is related to the planar arrays of MnS inclusions. The inclusion spacing governs

this coalescense rate directly, and it is a major effect of the inclusions since coalescence is an especially abrupt unzipping mechanism by which a material comes apart. The approach taken here was to selectively treat the damage of the weak bands of the banded microstructure, which is where failure actually occurs, rather than to assume uniformly distributed damage. Coalescence was computed with a random linear coalescence model(10) in which a new crack is created whenever two cracks collide. The new crack length is the sum of the lengths of the colliding cracks, and the number of cracks is reduced by one each time there is a collision. Damage was computed with this model by determining the fraction f of a sample covered by cracks. The smaller f, the more resistant a material is to spallation.

Average crack velocities were estimated with Eq. 14 and a crack length limited by half the inclusion spacing which was specified as part of the material design.

The result of these considerations is that the approximate effect of yield strength, band thickness, band separation and inclusion size and density on the nucleation, growth and coalescence of cracks and the damage of rolled steel armor can be computed. Hence, it is possible to design the spallation resistance of such a material as a function of the governing material variables. Computations to demonstrate this have been completed, and a representative result is shown in Fig. 6.

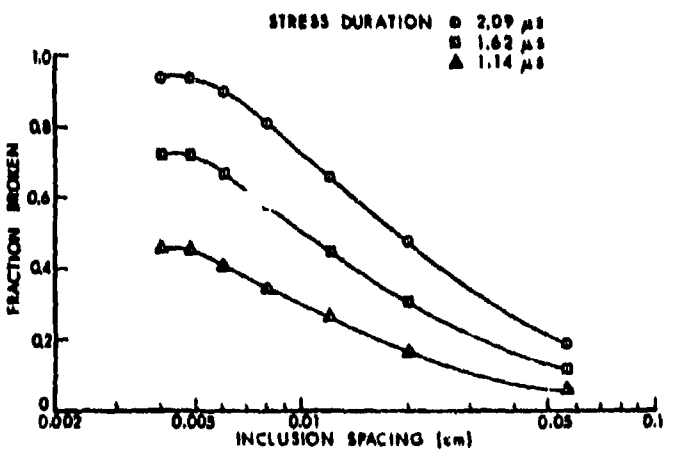



Figure 6. Extent of failure created with stress waves.

Results from this approach to material design should allow quantitative material considerations that have been impossible in the past. The role of inclusions and banding are specific examples. Furthermore, it should allow many new types of decisions to be made about material selection prior to expensive procurements and field

tests. For example, it may be possible to select material conditions to allow a prespecified degree of damage.

New conclusions about the role of inclusions can also be derived from Fig. 6. It appears that for the loads considered an inclusion spacing of from 40 to 50μm is particularly undesirable. It is noted that cleaning a material moderately, i.e., increasing the inclusion spacing from 40 to 60μm does not improve the spallation resistance much. Finally, there is an inclusion spacing beyond which the effect of cleaning the material becomes less pronounced.

## VI. CONCLUSIONS

a. Crack nucleation occurs at inclusions in rolled steel, and cracks propagate in the weak bands of the banded microstructure.

b. A comparison of the microstructural details of fracture in rolled steel with the cumulative crack-size distribution reveals that planar crack coalescence accompanies an intermediate stage of failure. Crack velocities determined for this stage of cracking without regard for coalescence do not represent true crack velocities.

c. Late stage coalescence of non-coplanar cracks occurs by plastic shear on connecting surfaces.

d. The method proposed for material design to resist spallation can be successfully applied when the governing material characteristics can be related to the crack nucleation and growth-rate equations. The inclusion spacing, banded microstructure and flow stress of rolled steel relate to these rate equations. Hence, material design considerations are possible for rolled steel.

e. Minor increases in inclusion spacing may not improve spallation resistance appreciably. There is an inclusion spacing beyond which only small gains in spallation resistance are possible.

## VII. ACKNOWLEDGEMENT

It is a pleasure to acknowledge helpful technical discussions with Dr. Donald Curran in which he indicated he had independently concluded that cracks must propagate intermittently under certain conditions.

## REFERENCES


1. G. Moss and C. M. Glass, "Some Microscopic Observations of Cracks Developed in Metals by Very Intense Stress Waves," Ballistic Research Laboratories Technical Note No. 1312, April 1960.

2. C. M. Glass, G. L. Moss, and S. K. Golaski, "Effects of Explosive Loading on Single Crystals and Polycrystalline Aggregates," Response of Metals to High Velocity Deformation, Ed. Shewman and Zackay, Interscience Publishers, New York, 1961, pp. 115-143.

3. T. Barbee, L. Seaman, and R. C. Crewdson, "Dynamic Fracture Criteria of Homogeneous Materials," Technical Report No. AFWL-TR-70-99, Air Force Weapons Laboratory, Kirtland Air Force Base, New Mexico, Nov. 1970.

4. T. W. Barbee, Jr., L. Seaman, R. Crewdson, and D. Curran, "Dynamic Fracture Criteria for Ductile and Brittle Metals," J. Materials, Vol. 7, No. 3, 1972, pp. 393-401.

5. Lynn Seaman, Donald R. Curran, and Donald A. Shockey, "Computational Models for Ductile and Brittle Fracture," J. Appl. Phys., Vol. 47, 11, 1976, pp. 4814-4826.

6. L. Seaman, T. W. Barbee, Jr., and D. R. Curran, Technical Report No. AFWL-TR-71-156, Air Force Weapons Laboratory, Kirtland Air Force Base, New Mexico, Feb., 1972.

7. B. L. Averbach, D. K. Felbeck, G. T. Hahn, and D. A. Thomas, Fracture, John Wiley & Sons, Inc., New York, 1959.

8. A. S. Tetelman and A. J. McEvily, Jr., Fracture of Structural Materials, John Wiley & Sons, Inc., New York, 1967, p. 296.

9. Frank A. McClintock, "On the Plasticity of the Growth of Fatigue Cracks," Fracture of Solids, Ed. Drucker and Gilman, Interscience Publishers, John Wiley & Sons, New York, 1963, pp. 65-102.

10. W. deRosset, "Coalescence of Cavities During Dynamic Fracture," BRL R 1631, Ballistic Research Laboratory, Aberdeen Proving Ground, MD., 1973.