ADA278694

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1993

SUMMER RESEARCH PROGRAM FINAL REPORTS

VOLUME 3

PHILLIPS LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Col. Hal Rhoades

Program Manager, RDL
Scott Licoscos

Program Administrator, RDL
Gwendolyn Smith

Program Administrator
Johnetta Thompson

Accesion For

NTIS CRA&I
DTIC TAB
Unannounced
Justification

By
Distribution /

Availability Codes

Dist	Avail and / or Special
A-1	

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1993

DTIC QUALITY INSPECTED 3

Best
Available
Copy

MEMORANDUM FOR DTIC (Acquisition)
(Attn: Pat Mauby)

SUBJECT: Distribution of USAF (AFOSR Summer Research Program (Air Force Laboratories) and Universal Energy Systems, Inc., and the Research Initiation Program

FROM: AFOSR/XPT
Joan M. Boggs
110 Duncan Avenue, Suite B115
Bolling AFB DC 20332-0001

1. All of the books forwarded to DTIC on the subjects above should be considered Approved for Public Release, distribution is unlimited (Distribution Statement A).

2. Thank you for processing the attached information.

Joan M. Boggs

JOAN M. BOGGS
Chief, Technical Information Division

1993 SRP Vol 1, 2, 3, 4, 5A, 5B, 6, 7, 8, 9
(10 Books)

Master Index for Faculty Members

Abbott, Ben
Research, MS
Box 1649 Station B
Vanderbilt University
Nashville, TN 37235-0000

Field: Electrical Engineering
Laboratory: AEDC/

Vol-Page No: 6- 1

Abrate, Serge
Assistant Professor, PhD
Mechanical & Aerospace En
University of Missouri - Rolla
Rolla, MO 65401-0249

Field: Aeronautical Engineering
Laboratory: WL/FI

Vol-Page No: 5-15

Almallahi, Hussein
Instructor, MS
P.O. Box 308
Prairie View A&M University
Prairie View, TX 77446-0000

Field: Electrical Engineering
Laboratory: AL/HR

Vol-Page No: 2-25

Anderson, James
Associate Professor, PhD
Chemistry
University of Georgia
Athens, GA 30602-2556

Field: Analytical Chemistry
Laboratory: AL/EQ

Vol-Page No: 2-18

Anderson, Richard
Professor, PhD
Physics
University of Missouri, Rolla
Rolla, MO 65401-0000

Field: Physics
Laboratory: PL/LI

Vol-Page No: 3- 7

Ashrafiuon, Hashem
Assistant Professor, PhD
Mechanical Engineering
Villanova University
Villanova, PA 19085-0000

Field: Mechanical Engineering
Laboratory: AL/CF

Vol-Page No: 2- 6

Backs, Richard
Assistant Professor, PhD
Dept. of Psychology
Wright State University
Dayton, OH 45435-0001

Field: Experimental Psychology
Laboratory: AL/CF

Vol-Page No: 2- 7

Baginski, Thomas
Assoc Professor, PhD
200 Broun Hall
Auburn University
Auburn, AL 36849-5201

Field: Electrical Engineering
Laboratory: WL/MN

Vol-Page No: 5-40

Baker, Suzanne
Assistant Professor, PhD
Dept. of Psychology
James Madison University
Harrisonburg, VA 22807-0000

Field:
Laboratory: AL/OE

Vol-Page No: 2-36

Baker, Albert
Assistant Professor, PhD

University of Cincinnati
, - 0

Field: Electrical Engineering
Laboratory: WL/MT

Vol-Page No: 5-53

Balakrishnan, Sivasubramanya
Associate Professor, PhD

University of Missouri, Rolla
, - 0

Field: Aerospace Engineering
Laboratory: WL/MN

Vol-Page No: 5-41

Bannister, William
Professor, PhD

Univ of Mass.-Lowell
Lowell, MA 1854-0000

Field: Organic Chemistry
Laboratory: WL/FI

Vol-Page No: 5-16

Barnard, Kenneth
Assistant Professor, PhD

Memphis State University
, - 0

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 1

Bayard, Jean-Pierre
Associate Professor, PhD
6000 J Street
California State Univ-Sacramen
Sacramento, CA 95819-6019

Field: Electrical/Electronic Eng
Laboratory: RL/ER

Vol-Page No: 4- 7

Beardsley, Larry
Research Professor, MS

Athens State College
, - 0

Field: Mathematics
Laboratory: WL/MN

Vol-Page No: 5-42

Beecken, Brian
Associcate Professor, PhD
3900 Bethel Dr.
Bethel College
St. Paul, MN 55112-0000

Field: Dept. of Physics
Laboratory: PL/VT

Vol-Page No: 3-23

Bellem, Raymond
Dept, CHM. EE cs, PhD
3200 Willow Creek Road
Embry-Riddle Aeronautical Univ
Prescott, AZ 86301-0000

Field: Dept. of Computer Science
Laboratory: PL/VT

Vol-Page No: 3-24

Bellem, Raymond
Dept, CHM. EE cs, PhD
3200 Willow Creek Road
Embry-Riddle Aeronautical Univ
Prescott, AZ 86301-0000

Field: Dept. of Computer Science
Laboratory: /

Vol-Page No: 0- 0

Bhuyan, Jay
Assistant Professor, PhD
Dept. ofComputer Science
Tuskegee University
Tuskegee, AL 36088-0000

Field: Computer Science
Laboratory: PL/WS

Vol-Page No: 3-33

Biegl, Csaba
Assistant Professor, PhD
Box 1649 Station B
Vanderbilt University
Nashville, TN 37235-0000

Field: Electrical Engineering
Laboratory: AEDC/

Vol-Page No: 6- 2

Biggs, Albert
Professor, PhD
Electrical Engineering
Univ. of Alabama, Huntsville
Huntsville, AL 35899-0000

Field:
Laboratory: PL/WS

Vol-Page No: 3-34

Blystone, Robert
Professor, PhD
Trinity University
715 Stadium Drive
San Antonio, TX 78212-7200

Field: Dept of Biology
Laboratory: AL/OE

Vol-Page No: 2-37

Branting, Luther
Assistant Professor, PhD
PO Box 3682
University of Wyoming
Laramie, WY 82071-0000

Field: Dept of Computer Science
Laboratory: AL/HR

Vol-Page No: 2-26

Bryant, Barrett
Associate Professor, PhD
115A Campbell Hall
University of Alabama, Birming
Birmingham, AL 35294-1170

Field: Computer Science
Laboratory: RL/C3

Vol-Page No: 4- 1

Callens, Jr., Eugene
Assocition Professor, PhD
Industrial
Louisiana Technical University
Ruston, LA 71270-0000

Field: Aerospace Engineering
Laboratory: WL/MN

Vol-Page No: 5-43

Cannon, Scott
Associate Professor, PhD
Computer Science
Utah State University
Logan, UT 84322-0000

Field: Computer Science/Biophys.
Laboratory: PL/VT

Vol-Page No: 3-25

Carlisle, Gene
Professor, PhD
Dept. of Physics
West Texas State University
Canyon, TX 79016-0000

Field: Killgore Research Center
Laboratory: PL/LI

Vol-Page No: 3- 8

Catalano, George
Associate Professor, PhD
Mechanical Engineering
United States Military Academy
West Point, NY 10996-1792

Field: Department of Civil &
Laboratory: AEDC/

Vol-Page No: 6- 3

Chang, Ching
Associate Professor, PhD
Euclid Ave at E. 24th St
Cleveland State University
Cleveland, OH 44115-0000

Field: Dept. of Mathematics
Laboratory: WL/FI

Vol-Page No: 5-17

Chattopadhyay, Somnath
Assistant Professor, PhD

University of Vermont
Burlington, VT 5405-0156

Field: Mechanical Engineering
Laboratory: PL/RK

Vol-Page No: 3-14

Chen, C. L. Philip
Assistant Professor, PhD
Computer Science Engineer
Wright State University
Dayton, OH 45435-0000

Field: Electrical Engineering
Laboratory: WL/ML

Vol-Page No: 5-26

Choate, David
Assoc Professor, PhD
Dept. of Mathematics
Transylvania University
Lexington, KY 40505-0000

Field: Mathematics
Laboratory: PL/LI

Vol-Page No: 3- 9

Chubb, Gerald
Assistant Professor, PhD
164 W. 19th Ave.
Ohio State University
Columbus, OH 43210-0000

Field: Dept. of Aviation
Laboratory: AL/HR

Vol-Page No: 2-27

Chuong, Cheng-Jen
Associtae Professor, PhD
501 W. 1st Street
University of Texas, Arlington
Arlington, TX 76019-0000

Field: Biomedical Engineering
Laboratory: AL/CF

Vol-Page No: 2- 8

Citera, Maryalice
Assistant Professor, PhD
Department of Psychology
Wright State University
Dayton, OH 4-5435

Field: Industrial Psychology
Laboratory: AL/CF

Vol-Page No: 2- 9

Collard, Jr., Sneed
Professor, PhD
Ecology & Evolutionary Bi
University of West Florida
Pensacola, FL 32514-0000

Field: Biology
Laboratory: AL/EQ

Vol-Page No: 2-19

Collier, Geoffrey
Assistant Professor, PhD
300 College St., NE
South Carolina State College
Orangeburg,, SC 29117-0000

Field: Dept of Psychology
Laboratory: AL/CF

Vol-Page No: 2-10

Cone, Milton
Assistat Professor, PhD
3200 Willow Creek Road
Embry-Riddel Aeronautical Univ
Prescott, AZ 86301-3720

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 2

Cundari, Thomas
Assistant Professor, PhD
Jim Smith Building
Memphis State University
Memphis, TN 38152-0000

Field: Department of Chemistry
Laboratory: PL/RK

Vol-Page No: 3-15

D'Agostino, Alfred
Assistant Professor, PhD
4202 E Fowler Ave/SCA-240
University of South Florida
Tampa, FL 33620-5250

Field: Dept of Chemistry
Laboratory: WL/ML

Vol-Page No: 5-27

Das, Asesh
Assistant Professor, PhD
Research Center
West Virginia University
Morgantown, WV 26505-0000

Field: Concurrent Engineering
Laboratory: AL/HR

Vol-Page No: 2-28

DeLyser, Ronald
Assistant Professor, PhD
2390 S. York Street
University of Denver
Denver, CO 80208-0177

Field: Electrical Engineering
Laboratory: PL/WS

Vol-Page No: 3-35

DelVecchio, Vito
Professor, PhD
Biology
University of Scranton
Scranton, PA 18510-4625

Field: Biochemical Genetics
Laboratory: AL/AO

Vol-Page No: 2- 1

Dey, Pradip
Associate Professor, PhD

Hampton University
, - 0

Field: Computer Science
Laboratory: RL/IR

Vol-Page No: 4-16

Ding, Zhi
Assistant Professor, PhD
200 Broun Hall
Auburn University
Auburn, AL 36849-5201

Field: Electrical Engineering
Laboratory: WL/MN

Vol-Page No: 5-44

Doherty, John
Assistant Professor, PhD
201 Coover Hall
Iowa State University
Ames, IA 50011-1045

Field: Electrical Engineering
Laboratory: RL/OC

Vol-Page No: 4-21

Dolson, David
Assistant Professor, PhD

Wright State University
, - 0

Field: Chemistry
Laboratory: WL/PO

Vol-Page No: 5-56

Dominic, Vincent
Assustant professor, MS
300 College Park
University of Dayton
Dayton, OH 45469-0227

Field: Electro Optics Program
Laboratory: WL/ML

Vol-Page No: 5-28

Donkor, Eric
Assistant Professor, PhD
Engineering
University of Connecticut
Stroes, CT 6269-1133

Field: Electrical Engineering
Laboratory: RL/OC

Vol-Page No: 4-22

Driscoll, James
Associate Professor, PhD
3004 FXB Bldg 2118
University of Michigan
Ann Arbor, MI 48109-0000

Field: Aerospace Engineering
Laboratory: WL/PO

Vol-Page No: 5-57

Duncan, Bradley
Assistant Professor, PhD
300 College Park
University of Dayton
Dayton, OH 45469-0226

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 3

Ehrhart, Lee
Instructor, MS
Communications & Intellig
George Mason University
Fairfax, VA 22015-1520

Field: Electrical Engineering
Laboratory: RL/C3

Vol-Page No: 4- 2

Ewert, Daniel
Assistant Professor, PhD
Electrical Engineering
North Dakota State University
Fargo, IN 58105-0000

Field: Physiology
Laboratory: AL/AO

Vol-Page No: 2- 2

Ewing, Mark
Associate Professor, PhD
2004 Learned Hall
University of Kansas
Lawrence, KS 66045-2969

Field: Engineering Mechanics
Laboratory: PL/SX

Vol-Page No: 3-22

Foo, Simon
Assistant Professor, PhD
College of Engineering
Florida State University
Tallahessee, FL 32306-0000

Field: Electrical Engineering
Laboratory: WL/MN

Vol-Page No: 5-45

Frantziskonis, George
Assistant Professor, PhD
Dept of Civil Engrng/Mech
University of Arizona
Tuson, AZ 85721-1334

Field: College of Engrng/Mines
Laboratory: WL/ML

Vol-Page No: 5-29

Frenzel III, James
Assistant Professor, PhD
Dept of Electrical Engnr
University of Idaho
Moscow, ID 83844-1023

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 4

Fried, Joel
Professor, PhD
Chemical Engineering
University of Cincinnati
Cincinnati, OH 45221-0171

Field: Polymer Science
Laboratory: WL/PO

Vol-Page No: 5-58

Friedman, Jeffrey
Assistant Professor, PhD
Physics
University of Puerto Rico
Mayaguez, PR 681-0000

Field: Physics/Astrophysics
Laboratory: PL/GP

Vol-Page No: 3- 1

Fuller, Daniel
Dept. Chairman, PhD
Chemistry & Physics
Nicholls State University
Thibodaux, LA 70310-0000

Field: Chemistry
Laboratory: PL/RK

Vol-Page No: 3-16

Gao, Zhanjun
Assistant Professor, PhD
203 W. Old Main, Box 5725
Clarkson University
Potsdam, NY 13699-5725

Field: Mechanical/Aeronautical E
Laboratory: WL/ML

Vol-Page No: 5-30

Gavankar, Prasad
Asst Professor, PhD
Campus Box 191
Texas A&I University
Kingsville, TX 78363-0000

Field: Mech & Indust Engineering
Laboratory: WL/MT

Vol-Page No: 5-54

Gebert, Glenn
Assistant Professor, PhD
Mechanical
Utah State University
Logan, UT 84339-0000

Field: Aerospace Engineering
Laboratory: WL/MN

Vol-Page No: 5-46

Gerdom, Larry
Professor, PhD
Natural Science
Mobile College
Mobil, AL 36663-0220

Field: Chemistry
Laboratory: AL/EQ

Vol-Page No: 2-20

Ghajar, Afshin
Professor, PhD
Mech. & Aerospace Enginee
Oklahoma State University
Stillwater, OK 74078-0533

Field: Mechanical Engineering
Laboratory: WL/PO

Vol-Page No: 5-59

Gopalan, Kaliappan
Associate Professor, PhD
Dept of Engineering
Purdue University, Calumet
Hammond, IN 46323-0000

Field:
Laboratory: AL/CF

Vol-Page No: 2-11

Gould, Richard
Assistant Professor, PhD
Mechanical & Aerospace En
N.Carolina State University
Raleigh, NC 27695-7910

Field: Mechanical Engineering
Laboratory: WL/PO

Vol-Page No: 5-60

Gowda, Raghava
Assistant Professor, PhD
Dept of Computer Science
University of Dayton
Dayton, OH 45469-2160

Field: Computer Information Sys.
Laboratory: WL/AA

Vol-Page No: 5- 5

Graetz, Kenneth
Assistant Professor, PhD
300 College Park
University of Dayton
Dayton, OH 45469-1430

Field: Department of Psychology
Laboratory: AL/HR

Vol-Page No: 2-29

Gray, Donald
Associate Professor, PhD
PO Box 6101
West Virginia Unicersity
Morgantown, WV 20506-6101

Field: Dept of Civil Engineering
Laboratory: AL/EQ

Vol-Page No: 2-21

Green, Bobby
Assistant Professor, MS
Box 43107
Texas Tech University
Lubbock, TX 79409-3107

Field: Electrical Engineering
Laboratory: WL/FI

Vol-Page No: 5-18

Grubbs, Elmer
Assistant Professor, MS
Engineering
New Mexico Highland University
Las Vegas, NM 87701-0000

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 6

Guest, Joyce
Associate, PhD
Department of Chemistry
University of Cincinnati
Cincinnati, OH 45221-0172

Field: Physical ChemIstry
Laboratory: WL/ML

Vol-Page No: 5-31

Gumbs, Godfrey
Professor, PhD
Physics & Astronomy
University New York Hunters Co
New York, NY 10021-0000

Field: Condensed Matter Physics
Laboratory: WL/EL

Vol-Page No: 5-12

Hakkinen, Raimo
Professor, PhD
207 Jolley Hall
Washington University
St. Louis, MO 63130-0000

Field: Mechanical Engineering
Laboratory: WL/FI

Vol-Page No: 5-19

Hall, Jr., Charles
Assistant Professor, PhD
Mech & Aerospace Engr.
North Carolina Univ.
Raleigh, NC 27695-7910

Field:
Laboratory: WL/FI

Vol-Page No: 5-20

Hancock, Thomas
Assistant Professor, PhD

Grand Canyon University
, - 0

Field: Educational Psychology
Laboratory: AL/HR

Vol-Page No: 2-30

Hannafin, Michael
Visiting Professor, PhD
305-D Stone Building,3030
Florida State University
Tallahassee, FL 3-2306

Field: Educational Technology
Laboratory: AL/HR

Vol-Page No: 2-31

Helbig, Herbert
Professor, PhD
Physics
Clarkson University
Potsdam, NY 13699-0000

Field: Physics
Laboratory: RL/ER

Vol-Page No: 4- 8

Henry, Robert
Professor, PhD
Electrical Engineering
University of Southwestern Lou
Lafayette, LA 70504-3890

Field: Electrical Engineering
Laboratory: RL/C3

Vol-Page No: 4- 3

Hong, Lang
Assistant Professor, PhD
Dept of Electrical Engin
Wright State University
Dayton, OH 45435-0000

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 7

Hsu, Lifang
Assistant Professor, PhD

Field: Mathematical Statistics
Laboratory: RL/ER

Le Moyne College
, - 0

Vol-Page No: 4- 9

Huang, Ming
Assistant Professor, PhD
500 NW 20th Street
Florida Atlantic University
Boca Raton, FL 33431-0991

Field: Mechanical Engineering
Laboratory: AL/CF

Vol-Page No: 2-12

Humi, Mayer
Professor, PhD
Mathematics
Worchester Polytechnic Institu
Worchester, MA 1609-2280

Field: Applied Mathematics
Laboratory: PL/GP

Vol-Page No: 3- 2

Humi, Mayer
Professor, PhD
Mathematics
Worchester Polytechnic Institu
Worchester, MA 1609-2280

Field: Applied Mathematics
Laboratory: /

Vol-Page No: 0- 0

Jabbour, Kamal
Associate Professor, PhD
121 Link hall
Syracuse University
Syracuse, NY 13244-1240

Field: Electrical Engineering
Laboratory: RL/C3

Vol-Page No: 4- 4

Jaszczak, John
Assistant Professor, PhD
Dept. of Physics
Michigan Technological Univers
Houghton, MI 49931-1295

Field:
Laboratory: WL/ML

Vol-Page No: 5-32

Jeng, San-Mou
Associte, PhD
Mail Location #70
University of Cincinnati
Cincinnati, OH 45221-0070

Field: Aerospace Engineering
Laboratory: PL/RK

Vol-Page No: 3-17

Johnson, David
Associate Professor, PhD
Dept of Chemistry
University of Dayton
Dayton, OH 45469-2357

Field: Chemistry
Laboratory: WL/ML

Vol-Page No: 5-33

Karimi, Amir
Associate, PhD
Division Engineering
University of Texas, San Anton
San Antonio, TX 7824-9065

Field: Mechanical Engineering
Laboratory: PL/VT

Vol-Page No: 3-26

Kheyfets, Arkady
Assistant Professor, PhD
Dept. of Mathematics
North Carolina State Univ.
Raleigh, NC 27695-7003

Field:
Laboratory: PL/VT

Vol-Page No: 3-27

Koblasz, Arthur
Associate, PhD
Civil Engineeing
Georgia State University
Atlanta, GA 30332-0000

Field: Engineering Science
Laboratory: AL/AO

Vol-Page No: 2- 3

Kraft, Donald
Professor, PhD
Dept. of Computer Science
Louisiana State University
Baton Rouge, LA 70803-4020

Field:
Laboratory: AL/CF

Vol-Page No: 2-13

Kumar, Rajendra
Professor, PhD
1250 Bellflower Blvd
California State University
Long Beach, CA 90840-0000

Field: Electrical Engineering
Laboratory: RL/C3

Vol-Page No: 4- 5

Kumta, Prashant
Assistant Professor, PhD
Dept of Materials Science
Carnegie-Mellon University
Pittsburgh, PA 15213-3890

Field: Materiels Science
Laboratory: WL/ML

Vol-Page No: 5-34

Kuo, Spencer
Professor, PhD
Route 110
Polytechnic University
Farmingdale, NY 11735-0000

Field: Electrophysics
Laboratory: PL/GP

Vol-Page No: 3- 3

Lakeou, Samuel
Professor, PhD
Electrical Engineering
University of the District of
Washnington, DC 20008-0000

Field: Electrical Engineering
Laboratory: PL/VT

Vol-Page No: 3-28

Langhoff, Peter
Professor, PhD

Indiana University
Bloomington, IN 47405-4001

Field: Dept. of Chemistry
Laboratory: PL/RK

Vol-Page No: 3-18

Lawless, Brother
Assoc Professor, PhD
Dept. Science /Mathematic
Fordham University
New York, NY 10021-0000

Field: Box 280
Laboratory: AL/OE

Vol-Page No: 2-38

Lee, Tzesan
Associate Professor, PhD
Dept. of Mathematics
Western Illinois University
Macomb, IL 61455-0000

Field:
Laboratory: AL/OE

Vol-Page No: 2-39

Lee, Min-Chang
Professor, PhD
167 Albany Street
Massachusetts Institute
Cambridge, MA 2139-0000

Field: Plasma Fusion Center
Laboratory: PL/GP

Vol-Page No: 3- 4

Lee, Byung-Lip
Associate Professor, PhD
Engineering Sci. & Mechan
Pennsylvania State University
University Park, PA 16802-0000

Field: Materials Engineering
Laboratory: WL/ML

Vol-Page No: 5-35

Leigh, Wallace
Assistant Professor, PhD
26 N. Main St.
Alfred University
Alfred, NY 14802-0000

Field: Electrical Engineering
Laboratory: RL/ER

Vol-Page No: 4-10

Levin, Rick
Research Engineer II, MS
EM Effects Laboratory
Georgia Institute of Technolog
Atlanta, GA 30332-0800

Field: Electrical Engineering
Laboratory: RL/ER

Vol-Page No: 4-11

Li, Jian
Asst Professor, PhD
216 Larsen Hall
University of Florida
Gainesville, FL 32611-2044

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 8

Lilienfield, Lawrence
Professor, PhD
3900 Reservoir Rd., NW
Georgetown University
Washington, DC 20007-0000

Field: Physiology & Biophysics
Laboratory: WHMC/

Vol-Page No: 6-14

Lim, Tae
Assistant Professor, PhD
2004 Learned Hall
University of Kansas
Lawrence, KA 66045-0000

Field: Mechanical/Aerospace Engr
Laboratory: FJSRL/

Vol-Page No: 6- 8

Lin, Paul
Associate Professor, PhD
Mechanical Engineering
Cleveland State University
Cleveland, OH 4-4115

Field: Associate Professor
Laboratory: WL/FI

Vol-Page No: 5-21

Liou, Juin
Associate Professor, PhD
Electrical & Computer Eng
University of Central Florida
Orlando, FL 32816-2450

Field: Electrical Engineering
Laboratory: WL/EL

Vol-Page No: 5-13

Liu, David
Assistant Professor, PhD
100 Institute Rd.
Worcester Polytechnic Inst.
Worcester, MA 1609-0000

Field: Department of Physics
Laboratory: RL/ER

Vol-Page No: 4-12

Losiewicz, Beth
Assistant Professor, PhD
Experimental Psychology
Colorado State University
Fort Collins, CO 80523-0000

Field: Psycholinguistics
Laboratory: RL/IR

Vol-Page No: 4-17

Loth, Eric
Assistant Professor, PhD
104 S. Wright St, 321C
University of Illinois-Urbana
Urbana, IL 61801-0000

Field: Aeronaut/Astronaut Engr
Laboratory: AEDC/

Vol-Page No: 6- 4

Lu, Christopher
Associate Professor, PhD
300 College Park
University of Dayton
Dayton, OH 45469-0246

Field: Dept Chemical Engineering
Laboratory: WL/PO

Vol-Page No: 5-61

Manoranjan, Valipuram
Associate Professor, PhD
Neill Hall
Washington State University
Pullman, WA 99164-3113

Field: Pure & AppliedMathematics
Laboratory: AL/EQ

Vol-Page No: 2-22

Marsh, James
Professor, PhD
Physics
University of West Florida
Pensacola, FL 32514-0000

Field: Physics
Laboratory: WL/MN

Vol-Page No: 5-47

Massopust, Peter
Assistant Professor, PhD

Sam Houston State University
Huntsville, TX 77341-0000

Field: Dept. of Mathematics
Laboratory: AEDC/

Vol-Page No: 6- 5

Miller, Arnold
Senior Instructor, PhD
Chemistry & Geochemistry
Colorado School of Mines
Golden, CO 80401-0000

Field:
Laboratory: FJSRL/

Vol-Page No: 6- 9

Misra, Pradeep
Associate Professor, PhD

University of St. Thomas
, - 0

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 9

Monsay, Evelyn
Associate Professor, PhD
1419 Salt Springs Rd
Le Moyne College
Syracuse, NY 13214-1399

Field: Physics
Laboratory: RL/OC

Vol-Page No: 4-23

Morris, Augustus
Assistant Professor, PhD

Central State University
, - 0

Field: Biomedical Science
Laboratory: AL/CF

Vol-Page No: 2-14

Mueller, Charles
Professor, PhD
W140 Seashore Hall
University of Iowa
Iowa City, IA 52242-0000

Field: Dept of Sociology
Laboratory: AL/HR

Vol-Page No: 2-32

Murty, Vedula
Associate Professor, MS

Texas Southern University
, - 0

Field: Physics
Laboratory: PL/VT

Vol-Page No: 3-29

Musavi, Mohamad
Assoc Professor, PhD
5708 Barrows Hall
University of Maine
Orono, ME 4469-5708

Field: Elect/Comp. Engineering
Laboratory: RL/IR

Vol-Page No: 4-18

Naishadham, Krishna
Assistant Professor, PhD
Dept. of Electrical Eng.
Wright State University
Dayton, OH 45435-0000

Field: Electrical Engineering
Laboratory: WL/EL

Vol-Page No: 5-14

Noel, Charles
Associate Professor, PhD
151A Campbell Hall
Ohio State University
Columbus, OH 43210-1295

Field: Dept of Textiles & Cloth
Laboratory: PL/RK

Vol-Page No: 3-19

Norton, Grant
Asst Professor, PhD
Mechanical & Materials En
Washington State University
Pullman, WA 99164-2920

Field: Materials Science
Laboratory: WL/ML

Vol-Page No: 5-36

Noyes, James
Professor, PhD
Mathematics & Computer Sc
Wittenberg University
Springfield, OH 45501-0720

Field: Computer Science
Laboratory: WL/FI

Vol-Page No: 5-22

Nurre, Joseph
Assistant Professor, PhD
Elec. & Computer Engineer
Ohio University
Athens, OH 45701-0000

Field: Mechanical Engineering
Laboratory: AL/CF

Vol-Page No: 2-15

Nygren, Thomas
Associate Professor, PhD
1885 Neil Ave. Mail
Ohio State University
Columbus, OH 43210-1222

Field: Department of Psychology
Laboratory: AL/CF

Vol-Page No: 2-16

Osterberg, Ulf
Assistant Professor, PhD
Thayer School of Engrg.
Dartmouth College
Hanover, NH 3755-0000

Field:
Laboratory: FJSRL/

Vol-Page No: 6-10

Pan, Ching-Yan
Associate Professor, PhD
Physics
Utah State University
Logan, UT 84322-4415

Field: Condensed Matter Physics
Laboratory: PL/WS

Vol-Page No: 3-36

Pandey, Ravindra
Assistant Professor, PhD
1400 Townsend Dr
Michigan Technological Univers
Houghton, MI 49931-1295

Field: Physics
Laboratory: FJSRL/

Vol-Page No: 6-11

Patton, Richard
Assistant Professor, PhD
Mechanical&Nuclear Engine
Mississippi State University
Mississippi State, MS 39762-0000

Field: Mechanical Engineering
Laboratory: PL/VT

Vol-Page No: 3-30

Peretti, Steven
Assistant Professor, PhD
Chemical Engineering
North Carolina State Univ.
Raleigh, NC 27695-7905

Field:
Laboratory: AL/EQ

Vol-Page No: 2-23

Petschek, Rolfe
Associate Professor, PhD
Department of Physics
Case Western Reserve Universit
Cleveland, OH 44106-7970

Field: Physics
Laboratory: WL/ML

Vol-Page No: 5-37

Pezeshki, Charles
Assistant Professor, PhD

Washington State University
Pullman, WA 99164-2920

Field: Mechanical Engineering
Laboratory: FJSRL/

Vol-Page No: 6-12

Piepmeier, Edward
Assistant Professor, PhD
College of Pharmacy
University of South Carolina
Columbia, SC 29208-0000

Field:
Laboratory: AL/AO

Vol-Page No: 2- 4

Pittarelli, Michael
Associate Professor, PhD
PO Box 3050, Marcy Campus
SUNY, Institute of Technology
Utica, NY 13504-3050

Field: Information Sys & Engr.
Laboratory: RL/C3

Vol-Page No: 4- 6

Potasek, Mary
Research Professor, PhD

Columbia University
, - 0

Field: Physics
Laboratory: WL/ML

Vol-Page No: 5-38

Prasad, Vishwanath
Professor, PhD

SUNY, Stony Brook
Stony Brook, NY 11794-2300

Field: Mechanical Engineering
Laboratory: RL/ER

Vol-Page No: 4-13

Priestley, Keith
Research Scientist, PhD

University of Nevada, Reno
, - 0

Field: Geophysics
Laboratory: PL/GP

Vol-Page No: 3- 5

Purasinghe, Rupasiri
Professor, PhD
5151 State Univ. Dr.
California State Univ.-LA
Los Angeles, CA 90032-0000

Field: Dept of Civil Engineering
Laboratory: PL/RK

Vol-Page No: 3-20

Raghu, Surya
Assistant Professor, PhD
Mechanical Engineering
SUNY, Stony Brook
Stony Brook, NY 11794-2300

Field: Mechanical Engineering
Laboratory: WL/PO

Vol-Page No: 5-62

Ramesh, Ramaswamy
Associate Professor, PhD
School of Management
SUNY, Buffalo
Buffalo, NY 14260-0000

Field: Magement Science/Systems
Laboratory: AL/HR

Vol-Page No: 2-33

Ramm, Alexander
Professor, PhD
Mathematics
Kansas State University
Manhattan, KS 66506-2602

Field:
Laboratory: AL/CF

Vol-Page No: 2-17

Ray, Paul
Assistant Professor, PhD
Box 870288
University of Alabama
Tuscaloosa, AL 35487-0288

Field: Industrial Engineering
Laboratory: AL/OE

Vol-Page No: 2-40

Reimann, Michael
Assistant Instructor, MS
Information Systems
The University of Texas-Arling
Arlington, TX 76019-0437

Field: Computer Science
Laboratory: WL/MT

Vol-Page No: 5-55

Rodriguez, Armando
Assistant Professor, PhD

Arizona State University
Tempe, AZ 85287-7606

Field: Electrical Engineering
Laboratory: WL/MN

Vol-Page No: 5-48

Rohrbaugh, John
Research Engineer, PhD
347 Ferst St
Georgia Institute of Technolog
Atlanta, GA 30332-0800

Field: Sensors & Applied Electro
Laboratory: RL/ER

Vol-Page No: 4-14

Roppel, Thaddeus
Associate Professor, PhD
200 Broun Hall
Auburn University
Auburn, AL 36849-5201

Field: Electrical Engineering
Laboratory: WL/MN

Vol-Page No: 5-49

Rosenthal, Paul
Professor, PhD
Mathematics
Los Angeles City College
Los Angeles, CA 90027-0000

Field: Mathematics
Laboratory: PL/RK

Vol-Page No: 3-21

Rotz, Christopher
Associate Professor, PhD

Brigham Young University
Provo, UT 84602-0000

Field: Mechanical Engineering
Laboratory: PL/VT

Vol-Page No: 3-31

Rudolph, Wolfgang
Associate Professor, PhD
Dept of Physics and Astro
University of New Mexico
Albuquerque, NM 84131-0000

Field: Physics
Laboratory: PL/LI

Vol-Page No: 3- 0

Rudzinski, Walter
Professor, PhD
Dept. of Chemistry
Southwest Texas State Universi
San Marcos, TX 78610-0000

Field: Professor
Laboratory: AL/OE

Vol-Page No: 2-41

Rule, William
Asst Professor, PhD
Mechanical Engineering
University of Alabama
Tuscaloosa, AL 35487-0278

Field: Engineering Mechanics
Laboratory: WL/MN

Vol-Page No: 5-50

Ryan, Patricia
Research Associate, MS
Georgia Tech Research Ins
Georgia Institute of Tech
Atlanta, GA 30332-0000

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5-10

Saiduddin, Syed
Professor, PhD
1900 Coffey Rd
Ohio State University
Columbus, OH 43210-1092

Field: Physiology/Pharmacology
Laboratory: AL/OE

Vol-Page No: 2-42

Schonberg, William
Assoc Professor, PhD
Engineering Dept.
University of Alabama, Huntsvi
Huntsville, AL 35899-0000

Field: Civil & Environmental
Laboratory: WL/MN

Vol-Page No: 5-51

Schulz, Timothy
Assistant Professor, PhD
1400 Townsend Dr
Michigan Technological Univers
Houghton, MI 49931-1295

Field: Electrical Engineering
Laboratory: PL/LI

Vol-Page No: 3-11

Shen, Mo-How
Assistant Professor, PhD
2036 Neil Ave.
Ohio State University
Columbus,, OH 43210-1276

Field: Aerospace Engineering
Laboratory: WL/FI

Vol-Page No: 5-23

Sherman, Larry
Professor, PhD
Dept. of Chemistry
University of Scranton
Scranton, PA 18510-4626

Field: Analytical Chemistry
Laboratory: AL/OE

Vol-Page No: 2-43

Shively, Jon
Professor, PhD
Civil & Industrial Eng.
California State University, N
Northridge, CA 91330-0000

Field: Metallurgy
Laboratory: PL/VT

Vol-Page No: 3-32

Snapp, Robert
Assistant Professor, PhD
Dept of Computer Science
University of Vermont
Burlington, VT 5405-0000

Field: Physics
Laboratory: RL/IR

Vol-Page No: 4-19

Soumekh, Mehrdad
Associate Professor, PhD
201 Bell Hall
SUNY, Buffalo
Amherst, NY 14260-0000

Field: Elec/Computer Engineering
Laboratory: PL/LI

Vol-Page No: 3-12

Spetka, Scott
Assistant Professor, PhD
PO Box 3050, Marcy Campus
SUNY, Institute of Technology
Utica, NY 13504-3050

Field: Information Sys & Engrg
Laboratory: RL/XP

Vol-Page No: 4-26

Springer, John
Associate Professor, PhD

Field: Physics
Laboratory: AEDC/

Fisk University
, - 0

Vol-Page No: 6- 6

Stevenson, Robert
Assistant Professor, PhD
Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556-0000

Field: Electrical Engineering
Laboratory: RL/IR

Vol-Page No: 4-20

Stone, Alexander
Professor, PhD
Mathematics & Statistics
University of New Mexico
Alburquerque, NM 87131-1141

Field:
Laboratory: PL/WS

Vol-Page No: 3-37

Sveum, Myron
Assistant Professor, MS
Electronic Engineering Te
Metropolitan State College
Denver, CO 80217-3362

Field: Electrical Engineering
Laboratory: RL/OC

Vol-Page No: 4-24

Swanson, Paul
Research Associate, PhD
Electrical Engineering
Cornell University
Ithaca, NY 14853-0000

Field: Electrical Engineering
Laboratory: RL/OC

Vol-Page No: 4-25

Swope, Richard
Professor, PhD
Engineering Science
Trinity University
San Antonio, TX 78212-0000

Field: Mechanical Engineering
Laboratory: AL/AO

Vol-Page No: 2- 5

Tan, Arjun
Professor, PhD
Physics
Alabama A&M University
Normal, AL 35762-0000

Field: Physics
Laboratory: PL/WS

Vol-Page No: 3-38

Tarvin, John
Associate Professor, PhD
800 Lakeshore Drive
Samford University
Birmingham, AL 35229-0000

Field: Department of Physics
Laboratory: AEDC/

Vol-Page No: 6- 7

Taylor, Barney
Visiting Assist Professor, PhD
1601 Peck Rd.
Miami Univ. - Hamilton
Hamilton, OH 4-5011

Field: Dept. of Physics
Laboratory: WL/ML

Vol-Page No: 5-39

Thio, Y.
Associate Professor, PhD

University of Miami
Coral Gables, FL 33124-0530

Field: Physics Dept.
Laboratory: PL/WS

Vol-Page No: 3-39

Tong, Carol
Assistant Professor, PhD
Electrical Engineering
Colorado State University
Fort Collins, CO 80523-0000

Field:
Laboratory: WL/AA

Vol-Page No: 5-11

Truhon, Stephen
Associate Professor, PhD
Social Sciences
Winston-Salem State University
Winston-Salem, NC 27110-0000

Field: Psychology
Laboratory: AL/HR

Vol-Page No: 2-34

Tzou, Horn-Sen
Associate Professor, PhD
Mechanical Engineering
University of Kentucky
Lexington, KY 40506-0046

Field: Mechanical Engineering
Laboratory: WL/FI

Vol-Page No: 5-24

Vogt, Brian
Professor, PhD

Bob Jones University
, - 0

Field: Pharmaceutical Sciences
Laboratory: AL/EQ

Vol-Page No: 2-24

Wang, Xingwu
Asst Professor, PhD
Dept. of Electrical Eng.
Alfred University
Alfred, NY 14802-0000

Field: Physics
Laboratory: WL/FI

Vol-Page No: 5-25

Whitefield, Philip
Research Assoc Professor, PhD
Cloud & Aerosol Sciences
University of Missouri-Rolla
Rolla, MO 65401-0000

Field: Chemistry
Laboratory: PL/LI

Vol-Page No: 3-13

Willson, Robert
Research Assoc Professor, PhD
Robinson Hall
Tufts University
Medford, MA 2155-0000

Field: Physics and Astronomy
Laboratory: PL/GP

Vol-Page No: 3- 6

Witanachchi, Sarath
Assistant Professor, PhD
4202 East Fowler Avenue
University of South Florida
Tampa, FL 33620-7900

Field: Department of Physics
Laboratory: FJSRL/

Vol-Page No: 6-13

Woehr, David
Assistant Professor, PhD
Psychology
Texas A&M University
College Station, TX 77845-0000

Field: Psychology
Laboratory: AL/HR

Vol-Page No: 2-35

Xu, Longya
Assistant Professor, PhD
Electrical Engineering
The Ohio State University
Columbus, OH 43210-0000

Field: Electrical Engineering
Laboratory: WL/PO

Vol-Page No: 5-63

Yavuzkurt, Savas
Associate Professor, PhD

Pennsylvania State University
University Park, PA 16802-0000

Field: Mechanical Engineering
Laboratory: WL/PO

Vol-Page No: 5-64

Zhang, Xi-Cheng
Associate Professor, PhD
Physics Department
Rensselaer Polytechnic Institu
Troy, NY 12180-3590

Field: Physics
Laboratory: RL/ER

Vol-Page No: 4-15

Zhou, Kemin
Assistant Professor, PhD
Dept. of Elec & Comp. Eng
Louisiana State University
Baton Rouge, LA 70803-0000

Field:
Laboratory: WL/MN

Vol-Page No: 5-52

Zimmermann, Wayne
Associate Professor, PhD
P.O. Box 22865
Texas Woman's University
Denton, TX 76205-0865

Field: Dept Mathematics/Computer
Laboratory: PL/WS

Vol-Page No: 3-40

ION-MOLECULE REACTIONS AT HIGH TEMPERATURES

Jeffrey Friedman
Assistant Professor
Department of Physics
University of Puerto Rico at Mayaguez
Mayaguez, P.R. 00680

Melani Menendez-Barreto
Graduate Student
Department of Physics Astronomy
University of Oklahoma
Norman, OK 73019

Final Report for:
Summer Research Program
Phillips Laboratory, Geophysics Directorate

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, DC

August 1993

ION-MOLEC' : REACTIONS AT HIGH TEMPERATURES

Jeffrey Friedman
Assistant Professor
Department of Physics
University of Puerto Rico at Mayaguez
Mayaguez, P.R. 00680

Melani Menendez -Barreto
Graduate Student
Department of Physics and Astronomy
University of Oklahoma
Norman, OK 73019

Abstract

A High Temperature Flowing Afterglow was utilized in the study of ion-molecule reactions. Reaction rate coefficients were measured for the following reactions in the temperature range 300-1300 K: O^- + H_2, D_2, and CH_4; O_2^+ + CH_4.

The system performance was evaluated in terms of changes needed to effect a more durable and reliable instrument. The effects of the high temperatures and cycling to room temperature dictate further design modifications that will be the subject of another paper (in preparation)[1].

ION-MOLECULE REACTIONS AT HIGH TEMPERATURES

Jeffrey F. Friedman, and Melani Menendez-Barreto

INTRODUCTION

Information on ion-molecule reactions is desirable for both scientific and practical reasons.[2] At high temperatures rotational and vibrational states are excited in both the ions and target molecule, making it possible for new reaction channels to open up. The controversial topic of "entropy-driven" reactions could more readily be investigated at high temperatures, as emphasized by Meot-Ner.[3]

Because of the need to understand and influence plasmas which occur naturally (e.g., auroras and the ionosphere) or man-made (e.g., around transonic aircraft, re-entry vehicles, or engine exhausts), the U.S. Air Force sponsors a great deal of research on low-energy interactions of electrons, ions, atoms, and molecules.

Laboratory studies of ion-molecule reactions are generally carried out within 200-300 Celsius degrees of room temperature. This represents a difficulty to the Air Force given many of the plasmas of their interest are at much higher temperatures (say, 1000-2000 K). Modelers are forced to make educated guesses as to the high temperature behavior of reaction rates and products.

Drift tube experiments have revealed the translational energy dependence of the reaction rate coefficients for some reactions, up to energies corresponding to thousands of degrees. However translational energy data are not directly applicable to high temperature plasmas because the target molecules are not at equivalently high

temperatures, and the projectile ions may or may not be excited by successive collisions with a buffer gas.

Only a few reaction rate coefficients had been measured as a function of temperature to prior our work on this high temperature flowing afterglow apparatus.[4] Those measurements typically extend into the 500-600 K range, limited by the materials and heaters used in the apparatus. There had only been one experiment that went beyond this conventional temperature limit before our last summer's experiments on the HTFA. In 1974, a group at NOAA modified the heaters in a flowing afterglow apparatus and measured reaction rate coefficients for several positive-ion reactions involving atmospheric species, up to 900 K.[5] Our experiments last year led to reaction rate coefficient measurements for a variety of both positive and negative reactions in the 300-1200 K range.

We have new measurements for some other ion-molecule systems of known importance in the ionospheric composition and have made a modest temperature advance to 1300K. We have studied O^- on H_2, D_2, and CH_4, as well as O_2^+ on CH_4.

EXPERIMENTAL METHOD

We used a newly constructed and now extensively modified flowing afterglow apparatus at Hanscom AFB that was specially designed for high temperatures. The flowing afterglow technique was developed in the 1960s for the study of ion-molecule reactions by a group at NOAA laboratories (then part of the National Bureau of Standards). The flowing afterglow method is now a well-established technique used in 20-30 laboratories around the world.

The experimental method involves interacting a swarm of ions with a known number density (n) of reactant molecules for a specific time (t). Exponential attenuation of the primary ion signal (I, which is proportional to the primary ion density) will occur

if reactions take place which generate other species:

$$I=I_0 \exp(-knt),$$

where the signal at t=0 is I_0. The ion-molecule reaction rate coefficient (k) is given by:

$$k=-n^{-1}t^{-1} \ln (I/I_0).$$

The experiment was carried out in a flowing buffer gas to effectively transform the time measurement into one of distance.

The measurement procedure involved recording the primary ion count as the neutral reactant concentration was stepped from zero to an amount that affected no less than a factor of two attenuation in the primary ion signal. The (negative) logarithmic slope of the primary ion count rate versus reactant concentration is equal to -kt. The reaction time was readily determined from a measurement of the ion velocity and the known reaction distance.

The ionic products of the reaction under study were identified by the mass spectra. The lack of ionic products indicates electron attachment. Neutral products are now observable because of the addition of a residual gas analyzer (an ionizer in the lens region). We hope to make extensive use of this new feature in the future.

APPARATUS and PROCEDURES

The ion swarm is created in the upstream end of the flow tube by electron impact on a suitable source gas (N_2O for O^- and O_2 or CO_2 for O_2^+) , and in the case of O_2^+ from CO_2 a rapid secondary reaction that must be essentially complete before the neutral reactant gases are added approximately 40 cm. downstream. The ions and neutrals can interact in a 63 cm. reaction region before sampling through a pinhole aperture to a mass spectrometer and detector. The time in the reaction region is determined by the gas flow. The ion velocity can be determined directly by applying an

electrical pulse to the internal thermocouples (each in turn) and timing the arrival of the ion attenuation minimum at the detector with a multichannel analyzer. The ion velocity is just the distance between the thermocouples divided by the arrival time difference. Velocities were generally in the range of 7000 to 35000 cm/s. Comparing this velocity with the neutral gas velocity determined by the temperature, pressure, and flow rate yields an almost constant ratio of Vion / $V_{neutrals}$ in the range of 1.6 to 1.9. The ratio has some dependence on the temperature and the pressure. The flows of the reactant and source gases can generally be ignored for this calculation compared to the He buffer, but this is not always the case as for slow rates when the reactant concentration could be a few percent. Occasionally high flows of source gas and Ar (to remove He metastables created in the electron beam) were also used. The total correction could run as high as 10% in some cases.

The concentration of neutral reactant in the flow tube was determined by comparing the flow rates of the He buffer gas and the reactant. Typical concentrations were .001 times the He buffer (except as noted above). The total gas pressure ranged from .3 to 1.2 torr.

The temperature of the flow tube is determined by averaging the readings from two thermocouples spaced about 30 cm. apart and roughly centered in the reaction region. Above ambient temperatures were achieved by 7 high temperature commercial heating tapes in four heating zones covering the flow tube from ion source to sampling port. For more information on special modifications made to the flow tube to allow operation at high temperatures see Menendez-Barreto, Friedman, and Miller [6] (and 1).

As a result of the heating-cooling cycle cracks developed at the downstream end of the flow tube where it is welded to the thin end flange. At first the resultant leaks could be sealed with conventional leak sealants, and generally disappeared at high temperature because of thermal expansion of the flow tube. This solution was deemed

expedient because a more permanent solution would require the removal of the flow tube, and the crystallized insulation on the heating tapes rendered this impracticable without the risk of destroying the tapes. However this proved only temporary as the cracks worsened. The use of the large vacuum box that surrounds the entire flow tube solved the problem, but is generally inconvenient, and we have undertaken a redesign of the end of the flow tube.[1]

In practice the lifetimes of the channel multipliers are very short at these elevated temperatures. It is not known if this is a purely temperature effect, which seems unlikely because of the low pressure in the detector region, or the result of highly reactive contaminants.

All of these issues will addressed more fully in a paper in preparation for "Review of Scientific Instruments".[1]

RESULTS AND DISCUSSION

In order to avoid thermal dissociation problems it was decided that only stable, but atmospherically important reactions would be studied. Figures 1-4 show reaction coefficient, k, versus temperature for the reactions O_2^+ on CH_4, O^- on CH_4, O^- on H_2 and O^- on D_2 respectively. It is clearly obvious that the methane data are very different in nature than the hydrogen data. Presumably this reflects the opening of vibrational channels in methane at high temperature that are not yet available to the hydrogen and not seen in drift tube data at equivalent kinetic temperatures. The internal excitation of methane is the subject of a paper by Viggiano et al.[7] that will include these data and is currently in preparation. The minimum in the O_2^+ curve is at the limit of the HTFA data, but is still consistent with them. You will notice a large amount of scatter in the molecular oxygen ion on methane plot, and this reflects the great difficulty in doing this experiment. However the trend of the data is clearly evident and represents results from

widely varying flow conditions in pressure and velocity as well as source and reactant concentrations and the presence or lack of Argon in the buffer. The data are also consistent with the VT-SIFT data with which they overlap (fig. 1). We are presently analyzing the ion-neutral chemistry in the flow tube and the possibility of new channels at elevated temperatures. There seems to be some, as yet unknown, mechanism for regenerating the molecular oxygen ion downstream, at least in the 600-800K range. The ratio of the ionic products is not constant in the data set as a function of temperature but it is not yet clear if this is a temperature effect, or the result of incomplete chemistry upstream of the reaction region because of different plasma velocities and/or changing reaction rates. The data are in fairly good agreement with the VT-SIFT data of Viggiano et al.[7] over the temperature range common to both. This is also true of the hydrogen data[8] and consistent with our 1992 data.[6] Comparison of the O^- + CH_4 at high temperature with equivalent kinetic temperature SIFT drift tube data shows that they are not equivalent. Further collaboration is indicated to sort out the ion/neutral chemistry as a function of temperature, and the openning of new reaction channels.

BIBLIOGRAPHY:

[1]Friedman, JF.., Miller, T.M., Menendez-Barreto, M., Paulson, J.F., Williamson, J., Dale, F., Viggiano, A.A., Morris, R.A., A Functioning 1300 K (1500K) Flowing Afterglow Apparatus, in preparation for *Review of Scientific Instruments.*

[2]*Gas Phase Ion Chemistry,* Vol. 2, Ed. Bowers, M.T. (Academic Press, New York, 1979)

[3]Moet-Ner, M., *J. Phys. Chem.* **95**, 6580, (1991).

[4]Ikezoe, Y., Matsuoka, S., Takebe, M., Viggiano, A.A., *Gas Phase Ion-Molecule Reaction Rate Constants Through 1986* (Maruzen Co. Ltd., Tokyo, 1987).

[5]Lindinger, W.L., Fehsenfeld, F.C., Schmeltekopf, A.L., and Ferguson, E.E., *J. Geophys. Res.* **79**, 4753 (1974).

[6]Menendez-Barreto, M., Friedman, J.F., Miller, T.M., *Final Reports AFOSR Summer Research Program,* Bolling AFB.

[7]Viggiano, A.A., Morris, R.A., Miller, T.M., Friedman, J.F., Menendez-Barreto, M., Paulson, J.F., Michels, H.H., and Montgomery Jr., J.A.,in preparation for J. Chem. Phys.

[8]Viggiano, A.A., Morris, R.A., Deakyne, C.A., Dale, F., Paulson, J.F., *J. Phys. Chem.* **95**, 3644 (1991).

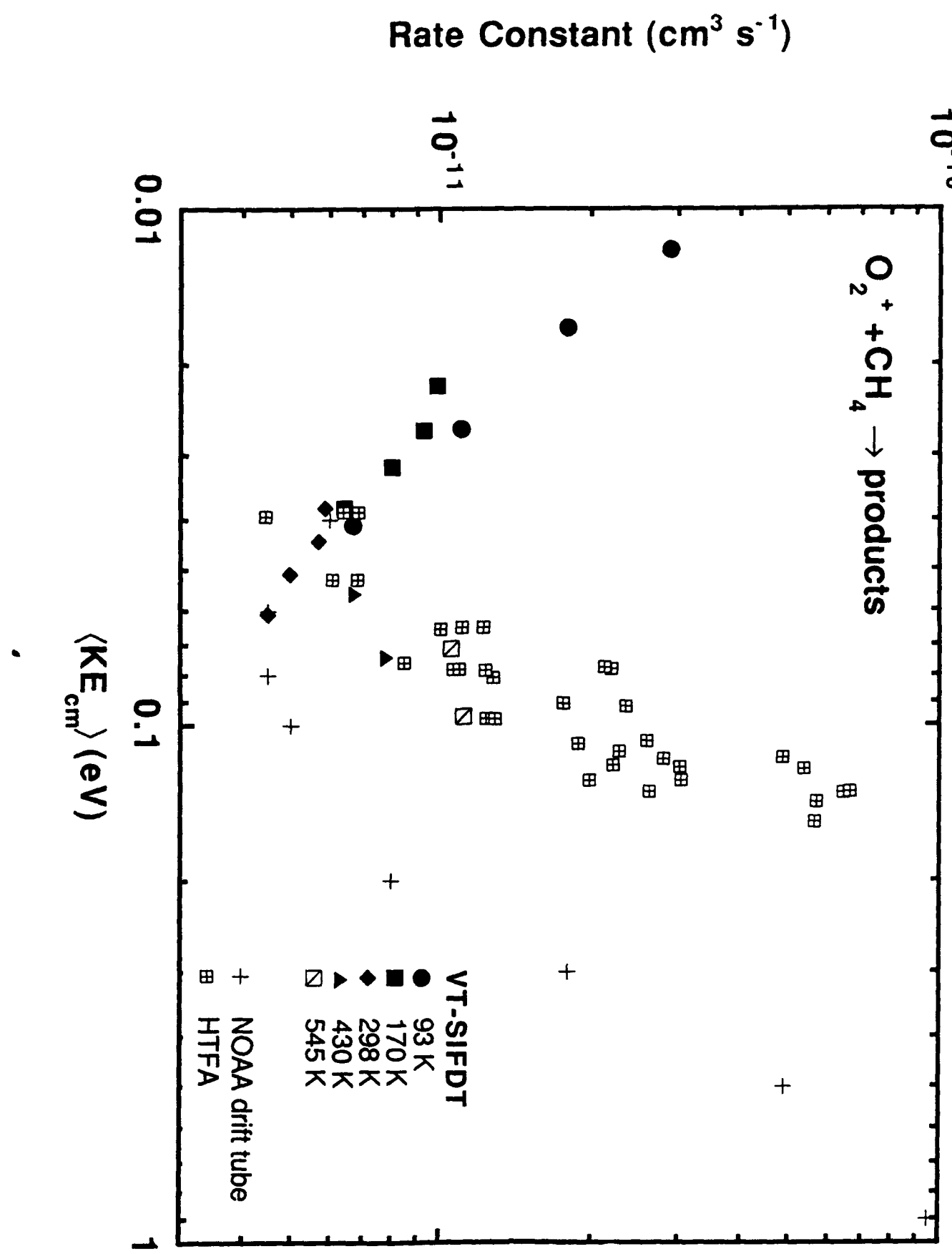

Rate Constant (cm^3 s^{-1})
10^{-11}
10^{-10}
0.01
0.1
1
$\langle KE_{cm} \rangle$ (eV)
$O_2^+ + CH_4 \rightarrow$ products
VT-SIFDT
93 K
170 K
298 K
430 K
545 K
NOAA drift tube
HTFA

Figure 1

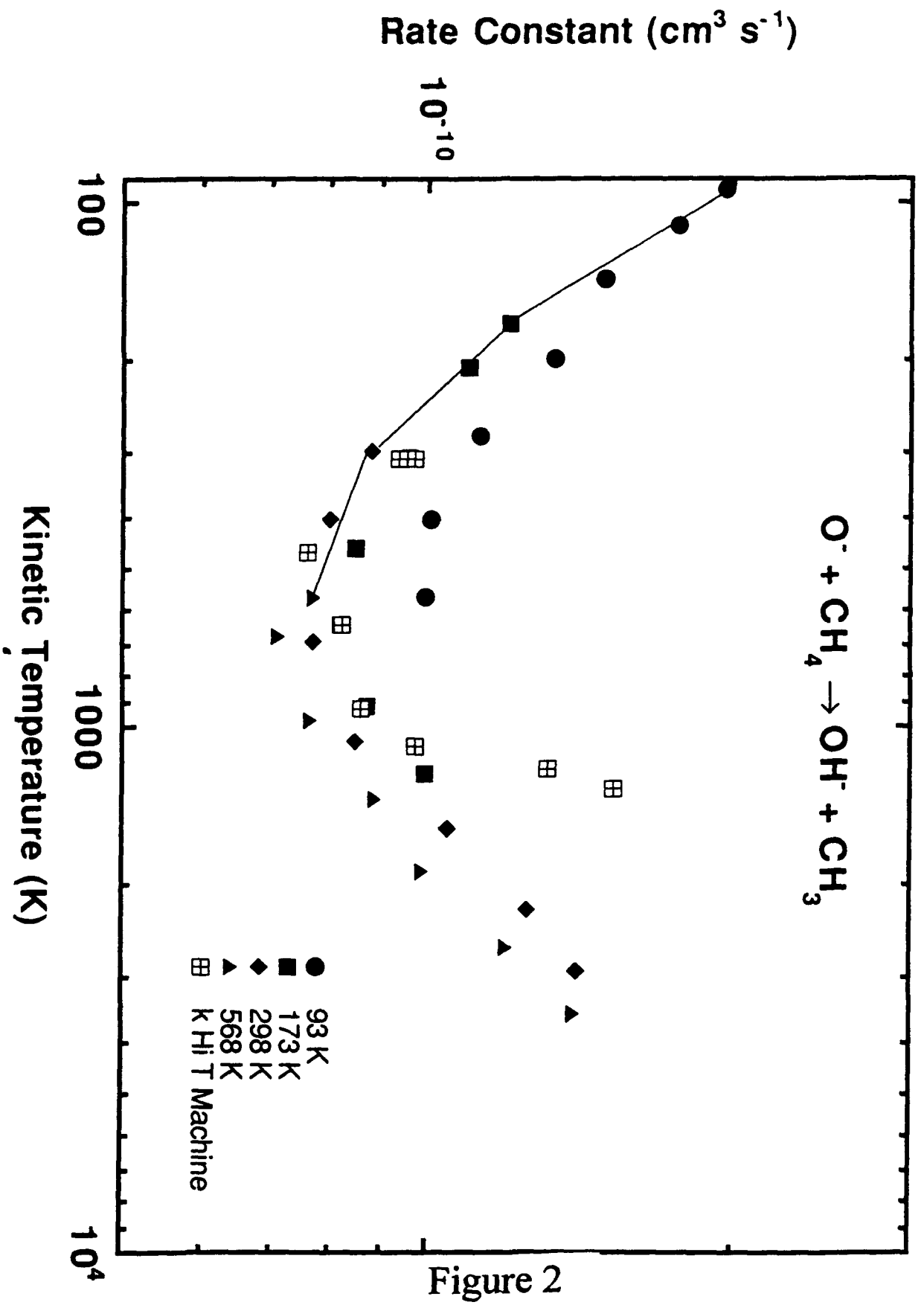

Rate Constant (cm^3 s^{-1})
10^{-10}
100
1000
10^4
Kinetic Temperature (K)
$O^- + CH_4 \rightarrow OH^- + CH_3$
93 K
173 K
298 K
568 K
k Hi T Machine

Figure 2

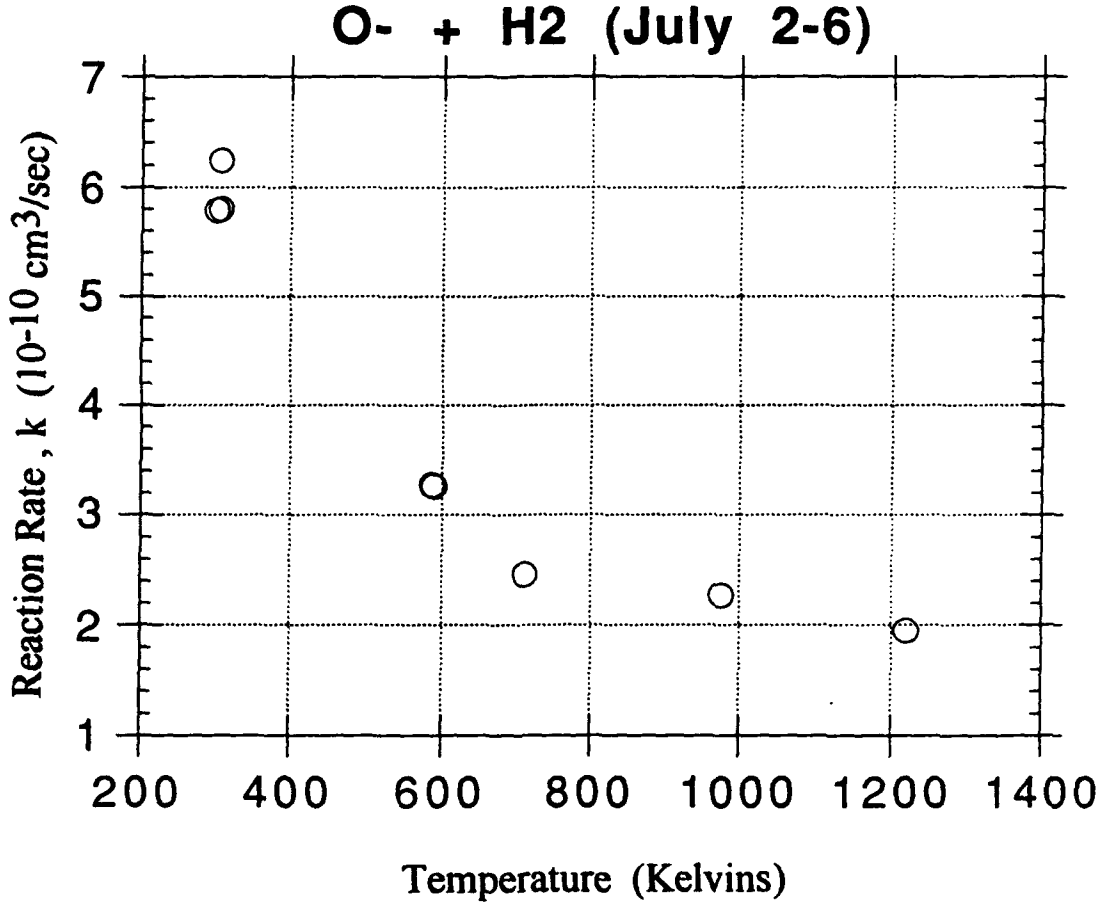

O- + H2 (July 2-6)
Reaction Rate , k (10-10 cm3/sec)
7
6
5
4
3
2
1
200
400
600
800
1000
1200
1400
Temperature (Kelvins)

Figure 3

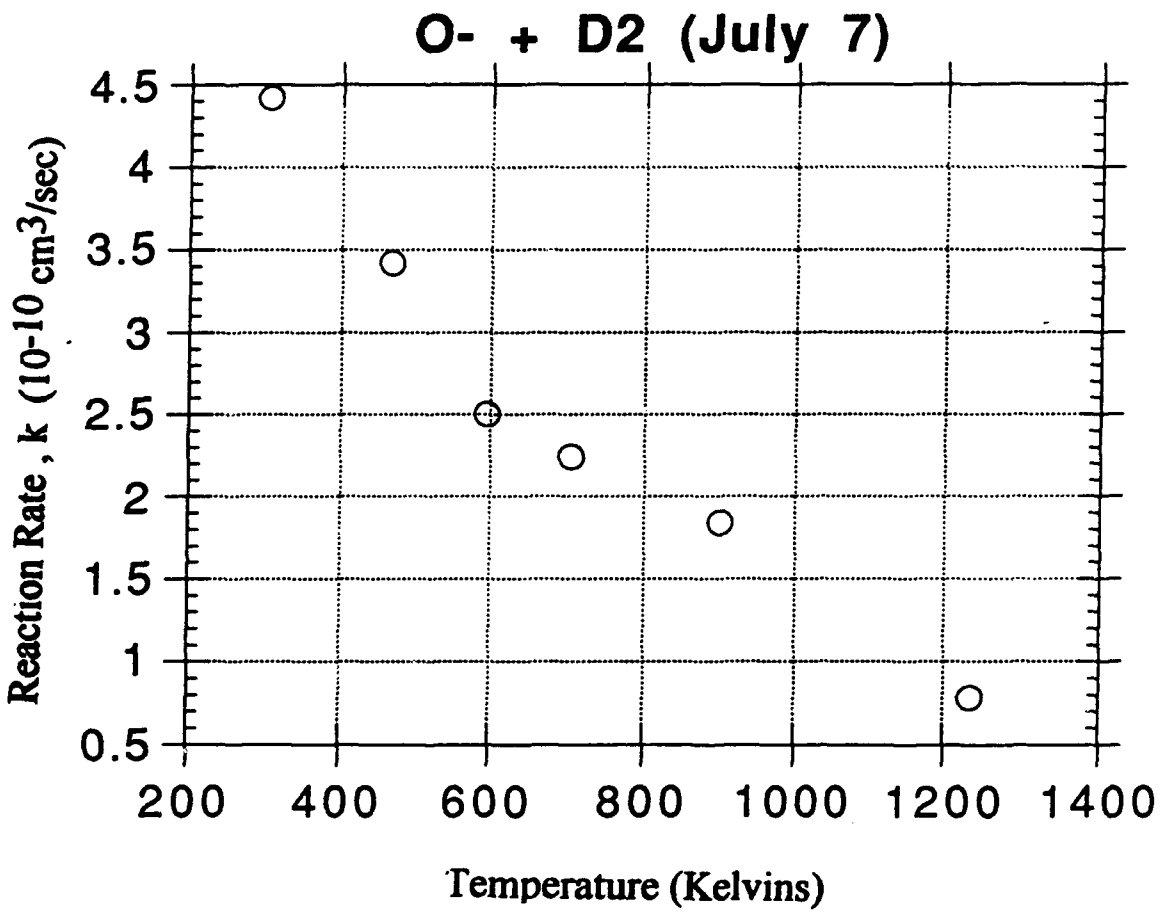

O- + D2 (July 7)
Reaction Rate , k (10-10 cm3/sec)
4.5
4
3.5
3
2.5
2
1.5
1
0.5
200
400
600
800
1000
1200
1400
Temperature (Kelvins)

Figure 4

OPTICAL AND ATMOSPHERIC TURBULENCE

submitted by

MAYER HUMI, Professor
DEPARTMENT OF MATHEMATICAL SCIENCES
Worcester Polytechnic Institute
Worcester, MA 01609–2280

Final Report for
Summer Research Program
Phillips Laboratory

Sponsored by
AFOSR, Bolling Air Force Base
Washington, D.C.

September 1993

OPTICAL AND ATMOSPHERIC TURBULENCE

MAYER HUMI, Professor
DEPARTMENT OF MATHEMATICAL SCIENCES
Worcester Polytechnic Institute

Abstract

For many important applications it is imperative to estimate correctly the spectral density of atmospheric turbulence. To this end one must collect meteorological data and "detrend" it to obtain the turbulent residual. The objective of this project was to experiment with different detrending (= filtering) strategies and gauge their impact on the computed spectral densities of the flow variables. To help accomplish this objective a general purpose software package was written and used.

I. Introduction.

It is a well known fact that the twinkling of the stars is caused by the fluctuations in the index of refraction of the earth atmosphere. These are due to atmospheric turbulence which in turn cause random fluctuations in the effective optical path of the electromagnetic signal. Due to these fluctuations there will be a spreading of the light beam, scintillations and a decrease in the spatial and temporal coherence. The same phenomena affect laser and radar systems and might degrade severely their performance. To compensate for these effects a careful study must be made on the relationship between the meteorological state of the atmosphere and the fluctuations in the index of refraction.

II. The structure constant C_n^2.

To study the effects of the fluctuations of the index of refraction in a clear atmosphere let $n(r,t)$ denote the index of refraction. This can be written as

$$n(\mathbf{r},t) = \langle n(\mathbf{r},t)\rangle + n_1(\mathbf{r},t) \tag{2.1}$$

where $\langle n(\mathbf{r},t)\rangle$ represents the average part and n_1 the fluctuation. For clear dry air $\langle n\rangle \equiv 1$. The fluctuating part n_1 consists of real and imaginary components

$$n_1 = n_{1r} + in_{1i} \tag{2.2}$$

From a practical point of view [1] the most important quantity affecting the signal propagation is the spatial correlation function for n_{1r}, i.e.

$$B_n(\mathbf{r},\rho) = \langle n_{1r}(\mathbf{r})n_{1r}(\mathbf{r}+\rho)\rangle \tag{2.3}$$

Here we neglected the dependence of n_{1r} on t since the time of propagation of an

electromagnetic signal in the atmosphere is short. If we assume further that the atmosphere is spatially homogeneous then $B_n(r,\rho)$ is independent of r. Using first order differentials to approximate n_{1r} we have (remember $\langle n_{1r} \rangle = 0$)

$$n_{1r} \cong \frac{\partial n_{1r}}{\partial T} dT + \frac{\partial n_{1r}}{\partial Q} dQ + \frac{\partial n_{1r}}{\partial P} dP \tag{2.4}$$

where T,Q,P stand for the temperature, humidity and pressure. However usually the pressure term is negligible and therefore we can write

$$n_{1r} \cong A_T \frac{dT}{\langle T \rangle} + A_Q \frac{dQ}{\langle Q \rangle}. \tag{2.5}$$

Substituting this in eq. (2.3) leads to

$$B_n(p) = \frac{A_T^2}{\langle T \rangle^2} B_T(p) + \frac{A_Q^2}{\langle Q \rangle^2} B_Q(p) + \frac{2A_T A_Q}{\langle T \rangle \langle Q \rangle} B_{TQ}(p) \tag{2.6}$$

where

$$B_T(p) = \langle \delta T(r) \delta T(r+p) \rangle \tag{2.7}$$

etc. Applying the Fourier transform to eq. (2.6) leads to

$$\varphi_n(k) = \frac{A_T^2}{\langle T \rangle^2} \varphi_T(k) + \frac{A_Q^2}{\langle Q \rangle^2} \varphi_Q(k) + \frac{2A_T A_Q}{\langle T \rangle \langle Q \rangle} \varphi_{TQ}(k) \tag{2.8}$$

where

$$f_n(\mathbf{p}) = \int_{R^3} \varphi_n(\mathbf{k}) \exp(i\mathbf{k}\cdot\mathbf{p}) d^3\mathbf{k} \qquad (2.9)$$

For well developed (clear) atmospheric turbulence it can be shown that the spectral density function $\varphi_n(\mathbf{k})$ can be approximated with proper assumptions by [1,2]

$$\varphi_n(\mathbf{k}) = \frac{0.33 C_n^2 \exp(-\mathbf{k}^2/k_m^2)}{(\mathbf{k}^2 + 1/L_0)^{11/6}}$$

where L_0 is the largest turbulence scale (typically on the order of 100 meters) C_n^2 is known as the "index of refraction structure constant". It has to be estimated from the spectral density of the turbulent fluctuations in the atmosphere.

A long mathematical derivation based on the works of Tatarski and others [5,6] then yields that $C_n^2 \approx k_\theta / k_m$ where

$$k_\theta = -\frac{\langle \theta' v' \rangle}{\langle \mathrm{grad}\,\overline{\theta} \rangle}$$

and K_M is the eddy diffusivity for momentum transfer which is related to the turbulent energy dissipation ϵ by

$$\epsilon = K_M (\mathrm{grad}\,\overline{U})^2$$

Here $\overline{\theta}, \overline{U}$ are the average potential temperature $\overline{U}$ the mean horizontal velocity and θ', v' represent the turbulent fluctuations.

To use this relations one must apply a proper algorithm to separate the mean part of the meteorological data from the turbulent fluctuations. The main objective of this

project was to use and compare different filtering methodologies in order to find the most appropriate for our purpose. On the other hand when n is large (e.g. n = 128) the influence of the data on the filter coefficients is strong but we have to solve a large number of nonlinear equations. Furthermore the set of equations becomes ill posed (due to the nature of eq. (3.1)) especially if the data contains large errors. Nevertheless in "mid–range" we were able to solve the resulting set of equations to obtain appropriate filter coefficients.

III PRINCIPAL COMPONENT ANALYSIS

The measurement and collection of meteorological data is subject to various types of errors. The goal of principal component analysis (PCA) is to identify and subtract these errors by the use of statistical analysis. We give here an overview of the algorithm used in this analysis and refer the reader to the literature for more details [15].

Suppose we have a set of (atmospheric) measurements $z(i,j)$, $i = 1,...,n$, $j = 1,...,nvar$. Here i stands (typically) for the time variable while j may refer to different atmospheric variables (pressure, temperature, etc.) or different sets of measurements of the same variable.

To apply PCA to this data set we first arrange it in a matrix of n rows and $nvar$ columns. The algorithm then consists of the following steps.

1. Center and normalize the data

$$y(i,j) = \frac{z(i,j) - \bar{z}(j)}{\sqrt{n}} \tag{3.1}$$

where $\bar{z}(j)$ represents the average of the measurements $z(i,j)$, $i = 1,...,n$.

2. Generate the covariance matrix of the data.

$$S = y^T \cdot y \tag{3.2}$$

(Note that S is nvar × nvar matrix.)

3. Find the eigenvalues and eigenvectors of S (λ_k, e_k), k = 1,...,nvar.

$$e_i \cdot e_j = \delta_{ij}. \tag{3.3}$$

The eigenvectors e_i thus provide an intrinsic frame for the decomposition of the measurements y(i,j)

$$y(i,*) = \sum_{k=1}^{nvar} x(i,k)e_k \tag{3.4}$$

where y(i,*) = (y(i,1), ... , y(i,nvar)).

In many cases some of the eigenvalues of S are small as compared to the leading one e.g.

$$|\lambda_1| < |\lambda_2| < \ldots < |\lambda_{nvar}| \tag{3.5}$$

and

$$\left| \lambda_k / \lambda_{nvar} \right| < \epsilon << 1 \ , \ k = 1, \ldots , m-1 \tag{3.6}$$

The basic premise of PCA is that one can attribute the components of y(i,x) along k = 1, ..., m − 1 to measurement errors rather than physical causes. We can therefore reconstruct the "error free" scaled measurements in the form

$$y(i,j) = \sum_{k=m}^{nvar} x(i,k)e_k(j). \qquad (3.7)$$

PCA can be applied to the meteorological data set obtained from the ER2 mission in two ways. The first is to use this analysis on the data obtained in one flight. (Thus we attempt to eliminate errors due to airplane vibrations and similar in flight errors). In this context the index j in the analysis refers to the six atmospheric variables which were measured viz. pressure, temperature, true air speed and the three velocity components (u,v,w). We found that the ratio of the first two eigenvalues of S to the leading one was 10^{-3} and the data was reconstructed accordingly. As a second application of PCA we used the data collected for the temperature on eight daily missions. (The hope was to eliminate in this way possible instrument errors). We found that the ratio of the first three eigenvalues to the leading one was 10^{-2} and the data was reconstructed accordingly.

IV. Data Filters.

The raw meteorological data regarding the wind velocity, pressure and temperature have to be decomposed into the mean and fluctuating parts to derive a value for C_n^2. This can be accomplished by different methods which lead in turn to different estimates for C_n^2.

In the following we describe briefly the various techniques which were used in this project but due to space limitations we shall not be able to describe their technical aspects. To this end the reader is referred to the bibliography.

1. Time series approach [8]

In this method we fit a local low order polynomial to the data. The algorithm which is based on Box–Jenkins approach treat the meteorological data as due to a stochastic process. The decomposition, used properly, leads to residuals which have zero mean and variance. However the global ("smooth") structure of the mean quantities

remains obscured by this process. One can refine this algorithm by the computation of the autocorrelation function for the data to determine the nature of the local filtering polynomial.

2. Low pass filter based on the FFT.[7]

In this approach we apply the fast Fourier transform to the data and by (inverse transform) remove from it the low frequency components. This leads to a smooth trigonometric approximation to the mean quantities which captures its global structure. However as the meteorological data contains some discontinuities the approximation fails at these points. As a refinement one can divide the data into "windows" of predetermined length (e.g. 2k to 10k) and apply the algorithm separately on each of these windows.

3. "Global" least squares algorithm.

In this approach we divide the data into subsets (≡ windows) of 1000–5000 data points. For each window we compute the least squares curve $f(t)$ of appropriate order (usually 1 or 2) to the data. The turbulent residuals are then given by

$$r_i = x_i - f(t_i) \qquad (4.1).$$

To avoid sharp discontinuities on the edges of these windows we allowed an overlap of 256 data points on each side of these subsets.

4. Local least squares algorithm.

In this algorithm (which was inspired by the Kalman filter) we open a window of 50–100 points on each side of x_i and compute the least squares curve (of appropriate order) for this data window. The turbulent residual at x_i (only) is then given by (5.1). The difference between this and the previous algorithm is that here a new least squares curve is computed for each data point. This requires a fairly extensive resource.

5. Spline Smoothing Filter

To produce this filter we use a natural cubic spline $S(t)$ for the data set (t_i,x_i) with knots at al t_i. However the smoothing spline does not interpolate for the x_i's. Instead it minimizes

$$\int_a^b S''(t)^2 dt \tag{4.2}$$

subject to the constraint

$$\sum \left| \frac{S(t_i) - x_i}{\omega_i} \right| < \sigma \tag{4.3}$$

where the ω_i's are proper "data weights" (we used $\omega_i \equiv 1$) and σ a "smoothing parameter". As a refinement one can use various strategies to estimate the value of σ that is appropriate for a given data set. As before the turbulent residuals are given by $r_i = x_i - S(t_i)$.

6. Convolution Transform.

It is well known that

$$(f * x)\hat{}\,(s)) = \hat{f}(s)\hat{x}(s) \tag{4.4}$$

where $*$ is the convolution operator and $\hat{}$ stands for the Fourier transform. It follows then that if one wanted to eliminate from the spectrum of a signal $x(t)$ all the frequencies greater than ω then $\hat{f}(s)$ must be of the form

$$\hat{f}(s) = \begin{cases} 1 & 0 < s < \omega \\ 0 & s \geq w \end{cases} \tag{4.5}.$$

The filtered data will be given then by the convolution

$$(f * x)(s) = \int_{-\infty}^{\infty} f(s-t)x(t)dt. \qquad (4.6)$$

However as explained earlier filters of the form (4.5) are not appropriate in the context of meteorological analysis. To motivate the choice of an appropriate filter we observe that the simulation of fluid flow using large eddy simulation methodology [16] requires a subgrid filter (to eliminate and model the fast modes in the flow). An optimal choice for such a filter was shown to have the form of a normal distribution [17]. We used therefore such a filter (with different standard deviations) to detrend the meteorological data using (4.4), (4.6).

7. Wavelet transform [9]

Wavelets represent a new development in applied mathematics which is used principally for signal processing. The basic idea is to use a function $\varphi(x)$ with compact support (called the mother wavelet) and apply to it a set of translations and dilatations to produce a basis of $L^2(R)$. This basis can be used then to expand some sets of functions "inexpensively" (i.e. with few terms). The hope is that in this manner one can capture both the local and global structure of the original signal.

In this project the idea was to use wavelets to capture the global structure of the meteorological signal and at the same time identify any local "jumps". To this end we used a wavelet transform package which was written by the "wavelet lab" at Yale University. It turned out that the choice of $\varphi(x)$ was crucial for this transform to work. In fact the package contains four types of wavelets functions but only one detrended the data correctly, i.e. produced residuals with mean zero. To overcome these difficulties we

devised and implemented a new algorithm for "data filtered wavelets" which was applied to the data [21].

V. Implementation.

In the summer of '92 we wrote an initial software package to detrend the meteorological data using some of the algorithms and techniques described above. In the present assignment many of the programs in the original package were enhanced and new ones to detrend the data were added. In addition we implemented algorithms to compute:

1. the Lyaponouv exponents for the data.
2. the fractal dimension.
3. the autocorrelation of the meteorological variables.
4. the cospectrum of several variables.
5. perform vertical and lateral principal component analysis on the data.

The package is written in Fortran and uses the IMSL and TEKSIM libraries. It uses the "wavelet lab package" that was obtained from Yale University and implements some of the wavelet bases that appear in [12].

In the following we enumerate and describe briefly ONLY THE NEW programs that were added to this package this summer.

conv05.f -- Detrend the data by the convolution method. The filter function here corresponds to the normal distribution however the program is modular and this other functions can be implemented easily if so desired.

convsp.f – Detrend the data as in conv05.f but also compute and plots the spectrum of the detrended data and some basic statistics for the turbulent residuals.

cor02.f – Computes the autocorrelation of each of the six meteorological variable and plots graphs of these quantities. The program is used to find an independent estimate of the large scales in the flow.

fd3 – Executable to compute the fractal dimension of the data. To run this program one must convert the data to an appropriate format using the convrt.f. The source code and documentation for this program are in ~/fractal/src/ .

convrt.f – program to read the data from mm*.bin file and convert it into a form appropriate for fd3.

lce02.f – Program to compute the Lyaponouv exponents for the data in each day using linear approximations to compute the Jacobians of the flow.

pca03.f – Performs PCA analysis on the data collected in one day and plot the spectrum of the 'clean' data.

pca05.f – Performs PCA analysis on the collected for one variable in several ER2 missions. The program also plots the spectrum of the residuals after detrending by global least squares.

lsp02.f – Detrend data using local least squares algorithm and plots the spectrum of the detrended data.

lsq02.f – Detrend the data using global least squares algorithm and plots the spectrum of the detrended data.

lsq03.f – Detrend the data using linear least squares in two windows. The dividing point between the the windows is determined by the minimum value of the temperature.

lscos.f – computes and plots the cospectrum of several meteorological variable pairs using the same algorithm for data detrending in lsq03.f

daub04.f – Detrend and plots the spectrum using various wavelets from Daubechies book [8].

intr03.f – Detrend the data using data fitted wavelets where the global data structure is inferred by using least squares on a sample of data points.

intr08.f – Detrend the data by using data fitted wavelets where the global structure of smooth part of the data is obtained by using another wavelet basis.

The meteorological data base used by theses program is in ~/nasa5/data. Each of the mm*.bin files represents the data collected in one day of the ER2 missions. The file mm890209.spe contains the data for flight take off on that day. This deemed important since airborne lasers direct their beams in this way from the lower altitudes. Other programs in this directory clean and plot the data as it appears on the CD–ROM received from Nasa and convert it into a binary form.

In total the source code for this package occupies over 1Mb. A sample from the output of these programs is given in the appendix.

VI CONCLUSION

This work explored different paradigms for the detrending of meteorological data. The most successful of these were found to be based on least squares and wavelet transform. Using these techniques we carried out statistical and spectral analysis of the turbulent residuals.

As to the characterization of the "atmospheric signal" further research is needed. In particular we wish to point out that the meteorological data that was collected on the ER2 mission does not constitute a time series per se. In fact it represents a space–time series of measurements. As such it provides only a one–dimensional cross section of a 4–dimensional set of data points. To ameliorate this situation accurate modeling and simulations of the atmospheric conditions in the stratosphere are needed. In these efforts proper attention must be paid to the mechanisms that underlie the creation of turbulence in these heights. Work related to these issues is in progress.

REFERENCES

1. E. F. Danielsen et al – Irreversible Transport in the Stratosphere by Internal Waves of Short Vertical Wavelength, J. Geo Res. 96D p. 17433–17452 (1991).

2. P. J. Kennedy and M. A. Shapiro – Further Encounters with Clear Air Turbulence in Research Aircraft, J. Atmos. Sci. 37 p. 986–993 (1980).

3. C. Cot and J. Barat – Wave–Turbulence Interaction in the Stratosphere, J. Geo Res. 91D p. 2749–2756 (1986).

4. L. Kristensen and P. Kirkegraard – Comments on effects of finite sampling on atmospheric spectra, Boundary Layer Metes 48 p. 205–210 (1989).

5. R. L. Fante – Wave Propagation in Random Media, Prog. In Optics 21 p. 341–401 (1985).

6. E. M. Dewan – Optical [illegible] casting, AFGL–TR–80–0300 (1980).

7. C. F. Gerald – Applied Numerical Analysis, Addison Wesley, Reading, MA 1989.

8. G. Box and G. M. Jenkins – Time Series Analysis, Holden Day 1976.

9. J. Morlet et al – Wave Propagation and Sampling Theory, Geophysics 47 p. 202 (1982).

10. R. Coifman et al – Wavelet Analysis and Signal Analysis, IEEE Trans. Inform. Theory 38 p. 713–718 (1990).

11. F. C. Moon – Chaotic and Fractal Dynamics, J. Wiley, New York, 1992.

12. I. Daubechies – Ten Lectures on Wavelets, SIAM, Philadelphia, PA 1992.

13. G. Strang – Wavelets and Dilation Equations, SIAM Rev. 31 p. 614–622 (1989).

14. M. Farge – Wavelet Transforms and Their Applications to Turbulence, Ann. Rev. Fluid Mech. p. 395–457 (1992).

15. J. F. Jackson – A User's Guide [illegible] Components, J. Wiley, New York, 1991.

16. M. Lesieur and R. Rogallo – [illegible]udy Simulation of Passive Scalar Diffusion in Isotropic Turbulence, Phys. Fluids A [illegible]8–731 (1989).

17. M. Humi – Optimal Large Eddy Simulation in One Dimension, Phys. Fluids A 2 p. 1046–1048 (1990).

18. Th. M. Kruel – Computation of Lyaponouv Spectra, Physicon D p. 117–131 (1993).

19. Liebovitch and Toth – A Fast Algorithm to Determine Fractal Dimensions by Box Counting, Phys. Letts. A 141 p. 386–390 (1989).

20. R. Stoop and R. F. Meier – Evaluation of Lyaponouv Exponents and Scaling Functions From Time Series, J. Opt. Soc. Am. B5 p. 1037 (1987).

21. M. Humi and O. Coté – Data fitted for meteorological analysis (submitted for publication).

Appendix

STATISTICS OF ORIGINAL DATA

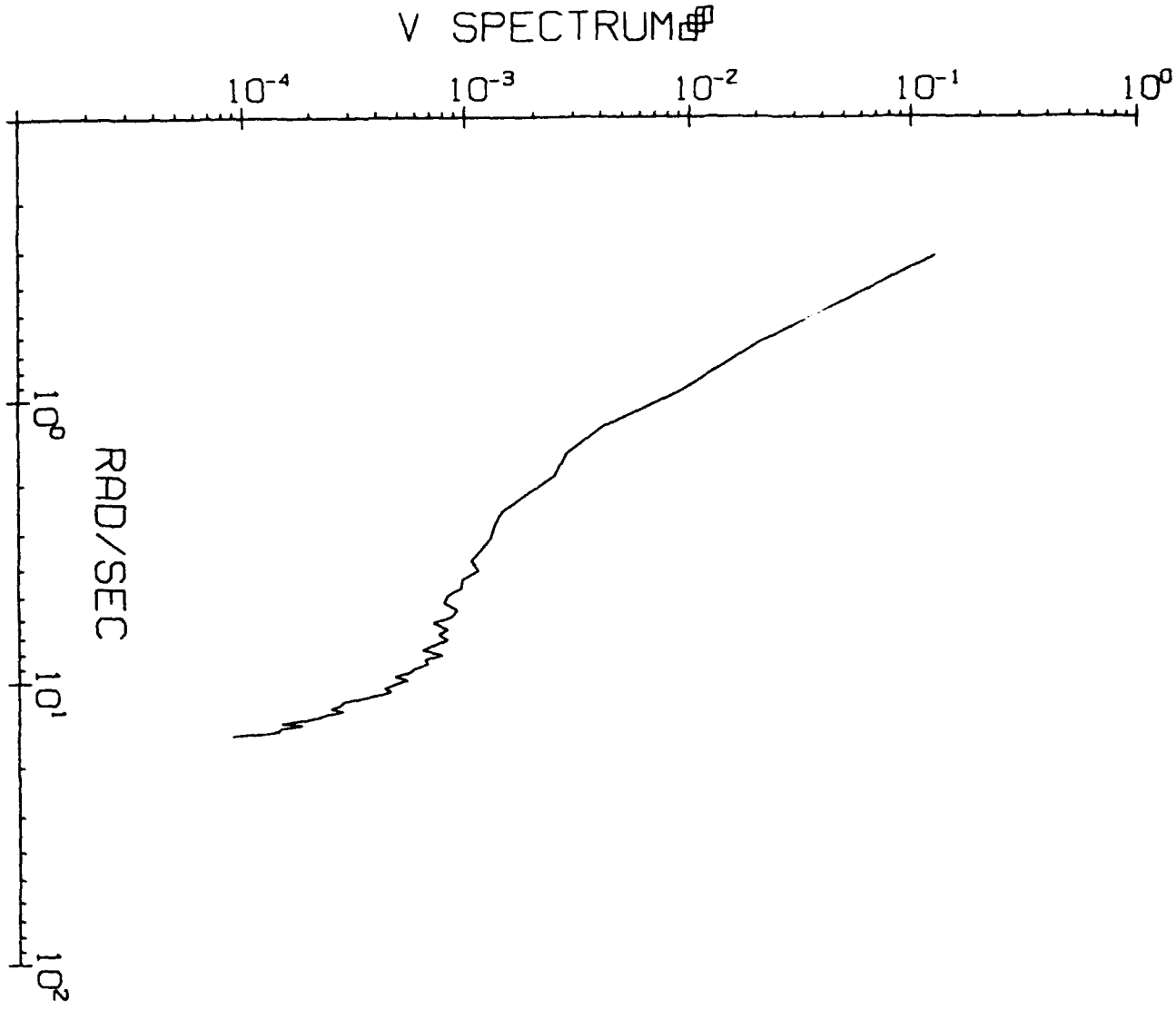
	psta	tsta	tas	u	v	w
mean	70.759	210.892	212.375	47.811	20.966	-0.877
variance	23.267	36.703	7.895	99.282	159.630	0.307
std. dev.	4.824	6.058	2.810	9.964	12.634	0.554
skewness	0.915	0.679	-0.336	-0.127	-0.275	0.364
kurtosis	0.103	-0.904	1.619	-1.160	-1.364	0.830
max	85.400	227.040	223.500	65.410	41.740	2.118
min	63.200	202.260	200.510	24.080	-3.000	-3.004
range	22.200	24.780	22.990	41.330	44.740	5.122
coef. var	0.068	0.029	0.013	0.208	0.603	-0.631
count	167199.	167199.	167199.	167199.	167199.	167199.

STATISTICS OF DETRENDED DATA
DIFFERENCING ORDER = 2

	psta	tsta	tas	u	v	w
mean	0.000	0.000	0.000	0.000	0.000	0.000
variance	0.022	0.019	0.090	0.074	0.033	0.009
std. dev.	0.150	0.137	0.300	0.272	0.182	0.097
skewness	0.010	-0.016	-0.014	-0.017	0.055	-0.058
kurtosis	0.981	0.093	0.451	0.635	3.617	6.178
max	0.800	0.640	1.880	1.910	2.500	1.228
min	-0.800	-0.700	-2.150	-2.280	-1.890	-1.715
range	1.600	1.340	4.030	4.190	4.390	2.943
count	167199.	167199.	167199.	167199.	167199.	167199.

SPECTRAL ANALYSIS FOR PSTA
using the window of Parzen

frequency	period	m=512	m=1024	m=2048
0.3132	20.06	0.1606E-06	0.1519E-06	0.1393E-06
0.6273	10.02	0.1060E-05	0.1064E-05	0.1107E-05
0.9415	6.67	0.3802E-05	0.3840E-05	0.3987E-05
1.2556	5.00	0.1055E-04	0.1065E-04	0.1110E-04
1.5698	4.00	0.2495E-04	0.2530E-04	0.2572E-04
1.8840	3.34	0.4489E-04	0.4340E-04	0.4036E-04
2.1981	2.86	0.8145E-04	0.7832E-04	0.7411E-04
2.5123	2.50	0.1334E-03	0.1310E-03	0.1253E-03
2.8264	2.22	0.1992E-03	0.1959E-03	0.1969E-03
3.1406	2.00	0.3031E-03	0.3169E-03	0.3277E-03
3.4548	1.82	0.3858E-03	0.3821E-03	0.3725E-03
3.7689	1.67	0.5640E-03	0.5695E-03	0.5559E-03
4.0831	1.54	0.6991E-03	0.6936E-03	0.6915E-03
4.3972	1.43	0.7672E-03	0.7684E-03	0.8028E-03
4.7114	1.33	0.9856E-03	0.9886E-03	0.1003E-02
5.0255	1.25	0.1164E-02	0.1156E-02	0.1227E-02
5.3397	1.18	0.1372E-02	0.1412E-02	0.1492E-02
5.6539	1.11	0.1537E-02	0.1555E-02	0.1539E-02
5.9680	1.05	0.1617E-02	0.1666E-02	0.1726E-02
6.2822	1.00	0.1939E-02	0.2037E-02	0.2136E-02
6.5963	0.95	0.1880E-02	0.1923E-02	0.1926E-02
6.9105	0.91	0.1970E-02	0.1978E-02	0.1986E-02
7.2247	0.87	0.2137E-02	0.2183E-02	0.2324E-02
7.5388	0.83	0.1918E-02	0.1888E-02	0.1861E-02
7.8530	0.80	0.2131E-02	0.2242E-02	0.2333E-02
8.1671	0.77	0.1893E-02	0.1882E-02	0.1878E-02
8.4813	0.74	0.1712E-02	0.1751E-02	0.1747E-02
8.7955	0.71	0.1874E-02	0.1950E-02	0.2007E-02
9.1096	0.69	0.1524E-02	0.1491E-02	0.1507E-02
9.4238	0.67	0.1461E-02	0.1503E-02	0.1602E-02


V SPECTRUM
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{0}
10^{1}
10^{2}
RAD/SEC

Generation of ELF and VLF Waves in the HF Heater-Modulated Polar Electrojet by Two Heater-Modulation Schemes

S. P. Kuo

Professor

Department of Electrical Engineering

Polytechnic University

Route 110

Farmingdale, NY 11735

Final Report for:

Summer Research Project

Geophysics Directorate of the Phillips Laboratory

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D.C.

September, 1993

Generation of ELF and VLF Waves in the HF Heater-Modulated Polar Electrojet by Two Heater-Modulation Schemes

S. P. Kuo

Professor

Department of Electrical Engineering

Polytechnic University

Abstract

Generation of ELF/VLF waves in the HF heater-modulated polar electrojet is studied. The electron temperature of the electrojet is modulated by an amplitude modulated HF heater. It in turn causes the modulation of the conductivity and, thus, the current of the electrojet. Emissions are then produced at the modulation frequency and its harmonics. Two heater-modulation schemes are considered. On, modulates the heater by a periodic rectangular pulse. The other one uses two intersecting heaters (beat wave) having a frequency difference equal to the desired modulation frequency. The nonlinear evolutions of the generated ELF/VLF waves are determined numerically. Their spectra are also evaluated. The results show that the signal quality of the second (beat wave) scheme is better. The field intensity of the emission at the fundamental modulation frequency is found to increase with the modulation frequency, consistent with the Tromso observations.

INTRODUCTION

Using the HF heater to modulate the ionospheric electrojet current, ELF/VLF have been generated as observed in many ionospheric heating experiments. [Getmantsev et al. (1974); Kapustin et al. (1977); Stubbe, et al., 1981, 1982 a&b; Stubbe and Kopka, 1981; Barr and Stubbe, 1984a&b, 1991a&b; Ferraro et al., 1982, 1984; James, et al., 1984; James, 1985; Rietveld et al., 1986; Lee, et al., 1990; McCarrick, et al., 1990; Wong, et al 1990; Barr, et al., 1991]. Two approaches were adopted in the previous heating experiments to modulate the electrojet current at a desired frequency. However, the plasma is a nonlinear medium, the harmonics of the modulation are also produced, in general. In one approach, e.g. adapted in the Tromso heating experiments [Stubbe et al., 1981], the wave field of the HF heater was modulated with periodic pulses in the ELF/VLF frequency range and given by

$$\vec{E}_{p1} = (\hat{x} \pm i\hat{y})\,(\varepsilon_p/2) \sum_{n=-\infty}^{\infty} P_{T/2}(t - nT_1)\, e^{i\,(k_0 z - \omega_0 t)} + \text{c.c.} \qquad (1)$$

where the rectangular pulse $P_{T/2}(t\text{-}nT_1)$ is defined to be $P_a(x)=1$ for $|x| < a$, and 0 for $|x| > a$; T_1 is the modulation period and T/T_1 is the percentage of the duty cycle.

In this approach, the modulation is not sinusoidal, only a small fraction of the total modulation contributes to the modulation at the fundamental frequency ω_1,. The harmonics are caused not only by the nonlinearity of the plasma by also by the modulation source itself.

It is shown in the Arecibo heating experiments, the electrojet modulation can also be achieved by using two HF heaters. Instead of operating the heater's entire antenna system as a whole, it is separated into two sets of antenna array. Each one radiates at the same power, but at slightly different frequency. Therefore, the heater power delivered to the ionosphere varies temporarily at the difference frequency w1 of the two antenna arrays and the resulting wave field of the HF heater can be expressed as

$$\vec{E}_{p2} = (\hat{x} \pm i\hat{y})\,(\varepsilon_p/2\sqrt{2})\,[1+e^{-i\,(\omega_1 t + \phi)}]\, e^{i\,(k_0 z - \omega_0 t)} + \text{c.c.} \qquad (2)$$

where $\omega_1 = 2\pi/T_1$ and ϕ is an arbitrary phase factor.

Comparing (2) to (1), the fractional power modulation at the fundamental frequency ω_1 is shown to increase by a factor $\pi/2\sin(\omega_1 T/2)$. Hence, one may expect that the second approach is more effective than the first approach in the electrojet modulation at the preferred fundamental frequency ω_1. This point will be clarified in the following analysis. Moreover, in many practical applications [Kossey and Brandt, 1992], it is not desirable to generate signals accompanying with harmonics. However, as mentioned before, the plasma is a nonlinear medium, the side effect of harmonic generation is not avoidable. Therefore, the following analysis is aimed to demonstrate that the second approach generates a better quality signal as shown by its Fourier spectrum (i.e. less harmonic content). This work is also aimed to extend the previous work of thermal instability [Kuo and Lee, 1993; Kuo, 1993] to the nonlinear saturation region. In other words, the electron thermal diffusion equation will be solved numerically without perturbation expansion. The solution describes the temporal evolution of the electron temperature starting from the linear instability stage and evolving into the saturated steady state stage. It is also used to determine the temporal evolution of the induced electromagnetic emissions

FORMULATION

ELF/VLF waves can be generated by the modulation of the ionospheric electrojet current at the corresponding frequency. The modulation is achieved by using a modulated HF heater and described by the electron thermal energy equation [Gurevich, 1978]:

$$\frac{d}{dt}T_e + \delta(T_e)\nu_e(T_e)(T_e - T_n) = \frac{2Q}{3n_0} \qquad (3)$$

where $\nu_e(T_e)$ is the effective collision frequency of the electrons with neutrals and $\delta(T_e)$ is the average relative energy fraction lost in [illegible] collision; T_n is the temperature of the background neutrals; Q is the total Ohmic heating source density in the background plasma and contributed by the flowing electrojet current and the induced a.c. current driven by the wave field of the HF heater.

The heat source density is given by

$$Q = \langle J^2/\sigma \rangle \cong \nu_{en} n_0 m \, (u^2 + \langle |\vec{v}_{pe}|^2 \rangle) \tag{4}$$

where $\vec{J} = -en_0 (\vec{u} + \vec{v}_{pe})$ is the total current density in the background plasma; $\langle\rangle$ stands for the time average over the heater wave period; $\sigma = n_0 e^2/m\nu_{en}$ is the conductivity of the plasma responsible for the Ohmic loss.

Considering that the electrojet current is driven by a d.c. field $\vec{E}_0 = \hat{x} E_0$ in a collisional plasma embedded in a d.c. magnetic field $\vec{B}_0 = -\hat{z} B_0$, then [Kuo and Lee, 1988],

$$\vec{u} \cong -(eE_0/m)(\hat{x}\nu_{en} - \hat{y}\Omega_e)(\nu_{en}^2+\Omega_e^2) \tag{5}$$

From the electron momentum equation, the linear velocity responses of electrons to the fields (1) and (2) are determined respectively to be

$$\vec{v}_{pe1} = -i(\hat{x} \pm i\hat{y})[e\varepsilon_p / 2m(\omega_0 \pm \Omega_e + i\nu_{en})] \sum_n P_{T/2}(t-nT_1) e^{i(k_0 z - \omega_0 t)} + \text{c.c.} \tag{6}$$

and

$$\vec{v}_{pe2} = -i(\hat{x} \pm i\hat{y})[e\varepsilon_p / 2\sqrt{2}m(\omega_0 \pm \Omega_e + i\nu_{en})][1+e^{-i(\omega_1 t + \phi)}] e^{i(k_0 z - \omega_0 t)} + \text{c.c.} \tag{7}$$

Thus,

$$u \cong eE_0/m(\nu_{en}^2+\Omega_e^2)^{1/2} \tag{8}$$

$$\langle |\vec{v}_{pe1}| \rangle = \nu_{q_\pm}^2 \sum_n P_{T/2}(t-nT_1) \tag{9}$$

$$\langle |\vec{v}_{pe2}| \rangle = \nu_{q_\pm}^2 [1+\cos(\omega_1 t + \phi)] \tag{10}$$

where $\nu_{q_\pm}^2 = (e\varepsilon_p / m)^2/[(\omega_0 \pm \Omega_e)^2 + \nu_{en}^2)]$

Both elastic and inelastic collisions contribute to the heat loss. The main processes involved in the inelastic collisions are the rotational and vibrational

excitation of N_2 and O_2. In the lower ionosphere, the total collision frequency can be expressed in terms of the unperturbed elastic collision frequency ν_{e0} as

$$\delta(T_e)\nu_e(T_e) \equiv 2(m/M_n)\nu_{eo}\{(\frac{T_e}{T_{eo}})^{5/6} + 8.6(\frac{T_e}{T_{eo}})^{-1/2} + (\frac{T_e}{T_{eo}})^{-5/2}\,[0.25e^{15.7(1-T_{eo}/T_e)}+27e^{3(1-T_{eo}/T_e)}]\} \tag{11}$$

where M_n is the mass of the neutral (N_2); the two terms inside the bracket are associated with rotational and vibrational excitation of neutrals (N_2 and O_2) respectively.

Since $\nu_{en} << \Omega_e << \omega_0$, the heat source density (4) increases with ν_{en}. On the other hand, $\nu_{en} \alpha T_e^{5/6}$ also increases with the heating. Such a dependence provides a feedback channel for the excitation of thermal instability oscillating at the modulating frequency of the HF heater and its harmonics. However, the heat loss in the electrojet also increases with heating, in particular, that caused by inelastic collisions increases exponentially. Thus, the instability is expected to saturate at a steady state level. In the following, (3) is solved numerically to obtain the steady state electron temperature modulation responding to the modulated heating by the incident HF heater. Since the electron temperature modulation leads to the modulation of the plasma conductivity tensor and thus the electrojet current, this induced a.c. current density $\delta\vec{J} = -\,en_0\delta\vec{u}$, where $\delta\vec{u} = \vec{u} - \vec{u}_0$, and $\vec{u}_0 = \vec{u}|_{\nu_{en}=\nu_{e0}}$, is the source of the ELF/VLF radiations. If the dimension of the source region is small comparing with the distance r to the receiver of the radiations, then the radiation field $\vec{E}$ at the receiver position can be determined approximately by

$$\vec{E}(\vec{r},t) = -\,(V/c^2r)\frac{\partial}{\partial t}\delta\vec{J}(t-r/c) = (en_0V/c^2r)\frac{\partial}{\partial t}\vec{u}(t-r/c) \tag{12}$$

where V is the volume of the modulated electrojet region.

Substituting (5) into (12), yields

$$\vec{E}(\vec{r},t) = -\,(\omega_{pe}^2VE_0/4\pi c^2r)[\hat{x}\,(\Omega_e^2 - \nu_{en}^2) + 2\hat{y}\Omega_e\nu_{en}](\frac{\partial}{\partial t}\nu_{en})/(\nu_{en}^2 + \Omega_e^2)^2$$

$$= -\frac{5}{6}\nu_{e0}(\omega_{pe}^2 VE_0/4\pi c^2 r)[\hat{x}(\Omega_e^2 - \nu_{en}^2) + 2\hat{y}\Omega_e\nu_{en}](\frac{T_e}{T_{e0}})^{-1/6}\frac{\partial}{\partial t}(\frac{T_e}{T_{e0}})/(\nu_{en}^2 + \Omega_e^2)^2 \quad (13)$$

It shows that the radiation field intensity received on the ground is proportional to the electron temperature modulation in the electrojet. This , in turn, leads to a modulation of the background electrojet current which becomes the source of radiations at the modulation frequency and its harmonics.

NUMERICAL ANALYSIS AND RESULTS

Equation (3) is first normalized into dimensionless form. It is done by introducing the following dimensionless functions variables and parameters: $\chi = T_e/T_{e0}$, $\tau = \nu_{e0}t/100$, $\omega = 100\,\omega_1/\nu_{e0}$, $T_n/T_{e0} = \alpha$, $v_{te0}^2 = T_{e0}/m$, and $(eE_0/m\Omega_e v_{te0}^2)^2 = \eta$, $v_{q\pm} = 6.94\times10^3\,\varepsilon_{p0}(1\pm1/3)$ m/s and $(v_{q\pm}/v_{te0})^2 = c_\pm$, and $(\nu_{e0}/\Omega_e)^2 = 6$. Thus, $\nu_{en} = \nu_{eo}\chi^{5/6}$, and (3) becomes

$$\frac{d}{d\tau}\chi_{a\pm}+200(m/M_n)\{\chi_{a\pm}^{5/6}+8.6\chi_{a\pm}^{-1/2}+\chi_{a\pm}^{-5/2}[0.25e^{1.57(1-\chi_{a\pm}^{-1})}+27e^{3(1-\chi_{a\pm}^{-1})}]\}(\chi_{a\pm}-\alpha)$$

$$= (200/3)\chi_{a\pm}^{5/6}[\eta/(1+6\chi_{a\pm}^{5/3}) + q_{a\pm}] \quad (14),$$

where a = 1, 2 corresponding to (9), (10) respectively; "±" corresponding to 0-mode and x-mode heater respectively; $q_{1\pm}=(c_\pm/2)[1+4\sum_{n=0}^{\infty}(-1)^n\cos(2n+1)\omega t/(2n+1)\pi]$ and $q_{2\pm}=c_\pm[1+\cos(\omega\tau+\phi)]$. In the case a = 1 of periodic pulse modulated heater, a 50% duty cycle $T/T_1 = 1/2$ is assumed and the relation of the Fourier series expansion

$$\sum_{n=-\infty}^{\infty} P_{T/2}(t-nT_1) = \frac{1}{2}[1+4\sum_{n=0}^{\infty}(-1)^n\cos(2n+1)\omega_1 t/(2n+1)\pi]$$

is used to express $q_{1\pm}$.

Let $\varepsilon(\vec{r},t)=|\vec{E}(\vec{r},t)/E_0|$, $R = (\nu_{e0}/100c)\, r$, $\xi = \tau - R$, and $c_0 = (5/6)\, 10^{-4}(\omega_{pe}^2 \nu_{e0}^3 V/4\pi c^3 \Omega_e^2)$, (13) reduces to

$$\varepsilon_{a\pm}(R, \tau) = (c_0/R)\chi_{a\pm}^{-1/6}(\xi)\frac{d}{d\tau}\chi_{a\pm}(\xi)/[1+b\chi_{a\pm}^{5/3}(\xi)] \qquad (15)$$

For the quantitative analysis, the following E region parameters are used: $T_n \equiv T_i \equiv 300°K$, $T_{e0} = 1500°K$, $\nu_{e0} = 5 \times 10^4$ sec^{-1}, $E_0 = 50$ mV/m, $M_n/m = 5.52 \times 10^4$, $\Omega_e/2\pi = 1.35$ MHz, $\omega_0/2\pi = 4.04$ MHz, and $v_{te0} = 1.5 \times 10^5$ m/s. Thus, $\alpha = 0.2$, $b = 3.48 \times 10^{-5}$, $\eta = 4.9 \times 10^{-5}$, and $m/M_n = 1.8 \times 10^{-5}$. [illegible] assuming that $\varepsilon_{p0} = 2.5$ V/m, then, $v_{q\pm} = 1.735 \times 10^4\,(1 \pm 1/3)$ m/s, $c_+ = 3.0157 \times 10^{-3}$ [illegible] $c_- = 7.5375 \times 10^{-3}$. We now solve (14) numerically subject to the initial condition $\chi_{a\pm}(0)=1$. In the beat wave case, the phase angle ϕ affects the initial growth rate of the excited perturbations. However, the steady state spectrum of the temperature perturbation or the emission is independent of ϕ, hence, $\phi = 0$ is set without losing generality. Considering three modulation frequencies: $\omega_1/2\pi$ = (10, 100, and 1000) Hz, thus $\omega/2\pi = 2 \times (10^{-2}, 10^{-1}, 1)$ respectively. Presented in Fig. 1 are the results of the electron temperature modulations and their corresponding Fourier spectra excited by an o-mode heater wave whose amplitude is modulated (by a periodic rectangular pulse) for the above-mentioned three modulation frequency cases. The similar results for the x-mode heater case are presented in Fig. 2. For comparison, the counterparts excited by using the beat wave scheme are presented in Figs. 3 and 4. In both cases, the perturbations produced by the x-mode heater (-) are stronger than those produced by the o-mode heater (+) as expected. More harmonic lines are generated by pulse modulation scheme, thus, it is better to use the beat wave scheme to modulated the electrojet electron temperature at the desired fundamental modulation frequency. The results also show that the higher the modulation frequency, the less the harmonic lines. Though the amplitude of the electron temperature modulation decreases with the modulation frequency, it is not necessary to imply a same trend to the field amplitude oscillating at the fundamental modulation frequency. Using the results presented in Figs. 1-4, (15) is evaluated and the corresponding results are presented in Figs. 5-8. In Figs. 5 and 6, the surges in the amplitudes of the EM emissions are caused by the discontinuity in the rectangular pulse modulation scheme. It increases the number of harmonic lines and, thus, the spectral bandwidth of the emission.

SUMMARY AND CONCLUSION

As shown in the previous works [Kuo and Lee, 1993; Kuo, 1993], a thermal instability can be excited in the polar ionospheric E region electrojet current by an amplitude-modulated HF heater. This instability process generates the ELF/VLF radiations more effectively than other known processes. In one work [Kuo and Lee, 1993], the heater is modulated by a periodic rectangular pulse. The intensity threshold is found to be higher than that excited by two intersecting heaters having a small frequency difference [Kuo, 1993].

The previous works have demonstrated the feasibility of the instability process, however, the steady state results are of practical interest. The nonlinearity of the plasma causes the saturation of the excited instability. The temporal evolutions of the excited temperature modulations in the electrojet and the associated EM emissions are evaluated numerically. The results show that x-mode heater is more effective than o-mode heater. The quality of the signal generated by the beat wave scheme is better than that generated by the rectangular pulse-modulation scheme. The percentage of modulation of the electrojet electron temperature decreases with the modulation frequency, however, the field intensity of the emission at the fundamental modulation frequency increases with the modulation frequency as shown in Figs 5-8. This increasing trend is consistent with the Tromso observations

ACKNOWLEDGMENTS

This work was performed at the Geophysics Directorate of the U.S. Air Force's Phillips Laboratory (PL/GP), Hanscom AFB, Massachusetts. The author would like to thank PL for its hospitality. He also wishes to acknowledge the frequent and useful discussions with Dr. Paul Kossey, Dr. Keith Groves, and Mr. John Heckscher of the PL/GP, and with Professor M.C. Lee of the Massachusetts Institute of Technology. The numerical computations were carried out by Mr. Antony Ho of Polytechnic University.

This work was sponsored in part by the summer faculty research program of the Air Force office of Scientific Research (AFOSR) Bolling AFB, Washington through the Research and Development Laboratory, Culver City, California.

REFERENCES

Barr, R. and P. Stubbe, "The Polar Electrojet Antennas as a Source of ELF Radiation in the Earth-Ionosphere Waveguide", J. Atmos. Terr. Phys., 46, 315, 1984a.

Barr, R. and P. Stubbe, "ELF and VLF Radiation from the Polar Electrojet Antenna", Radio Sci., 19,1111, 1984b.

Barr, R., P. Stubbe and H. Kopka, "Long Range Detection of VLF Radiation Produced by Heating the Auroral Electrojet", Radio Sci., 26,871,1991.

Barr, R. and P. Stubbe, "ELF Radiation from the Tromso Super Heater Facility", Geophys. Res. Lett., 18, 1971-1974, 1991a.

Barr, R. and P. Stubbe, "On the ELF Generation Efficiency of the Tromso Heater Facility", Geophys. Res. Lett., 18, 1035, 1991b.

Ferraro, A.J., H.S. Lee, R. Allshouse, K. Carroll, A.A. Tomko, F.J. Kelly, and R.G. Joiner, "VLF/ ELF radiation from the Ionospheric Dynamo Current System Modulated by Powerful HF Signals", J. Atmos. Terr. Phys., 44, 1113, 1982.

Ferraro, A.J., H.S. Lee, R. Allshouse, K. Carroll, R. Lunnen, and T. Collins, "Characteristics of Ionospheric ELF Radiation Generated by HF Heating", J. Atmos. Terr. Phys., 46, 855, 1984.

Getmantsev, G.G., N.A. Zuikov, D.S. Kotlk, L.F. Mirononko, N.A. Mityakov, V.O. Rapoport, Yu.A. Sazonov, V.Yu. Trakhtengerts, and V.Ya. Eidman, "Combination Frequencies in the Interaction Between High-Power Short-Wave Radiation and Ionospheric Plasma", JETP Lett., 20, 101, 1974.

Gurevich, V.A., "Nonlinear Phenomena in the Ionosphere", Chap. 2, Springer-Verlag publisher, New York, 1978.

James, H.G., R.L. Dowden, M.T. Rietveld, P. Stubbe, and H. Kopka, "Simultaneous Observations of ELF Waves from an Artificially Modulated Aurora Electrojet in Space and On the Ground", J. Geophys. Res., 89, 1655, 1984.

James, H.G., "The ELF Spectrum of Artificially Modulated D/E Region Conductivity", J. Atmos. Terr. Phys., 47, 1129, 1985.

Kapustin, I.N., R.A. Pertsovskii, A.N. Vasilev, V.S. Smirnov, O.M. Raspopov, L.E. Soloveva, A.A. Ulyachenko, A.A. Arykov, and N.V. Galakhova, "Generation of Radiation at Combination Frequencies in the Region of the Auroral Electric Jet", JETP Lett., 25,229,1977.

Kossey, P.A. and R. B. Brandt, "High Frequency Active Auroral Research Program (HAARP)", presented at National Radio Science Meeting, University of Colorado, Boulder, Colorado, January 7-10, 1992.

Kuo, S.P. and M.C. Lee, "Thermal Filamentation Instability Driven by the Auroral Electrojet Current", J. Geophys. Res., 93, 265, 1988.

Kuo, S.P. and M.C. Lee, "Generation of ELF and VLF Waves by HF Heater-Modulated Polar Electriojet via a Thermal Instability Process" Geophys. Res. Lett., 20, 189, 1993.

Kuo, S. P., "Generation of ELF and VLF Waves by a Thermal Instability excited in the HF Heater-Modulated Polar Electrojet", Radio Sci., in press, 1993.

Lee, H.S., A. J. Ferraro, and J. [illegible] Olson, "Detection and Characterization of Geomagnetic Pulsations Using [illegible]nospheric Heating", Radio Sci., 25, 1429, 1990.

McCarrick, M.J., D.D. Sentman, A.Y. Wong, R.F. Wuerker, and B. Chouinard, "Excitation of ELF Waves in the Schumann Resonance Range by Modulated HF Heating of the Polar Electrojet", Radio Sci. 25, 1291, 1990.

Rietveld, M.T., H. Kopka, and P. Stubbe, "D-region Characteristics Deduced from Pulsed Ionospheric Heating Under Auroral Electrojet Conditions", J. Atmos. Terr. Phys. 48, 311, 1986.

Stubbe, P., H. Kopka, and R.L. Dowen, "Generation of ELF and VLF Waves by Polar Electrojet Modulation: Experimental Results", J. Geophys. Res., 86, 9073, 1981.

Stubbe, P. and H. Kopka, "Generation of Pc5 Pulsations by Polar Electrojet Modulation: First Experimental Evidence", J. Geophys. Res., 86, 1606, 1981.

Stubbe, P., H. Kopka, H. Lauche, M.T. Rietveld, A. Brebbe, O. Holt, and R.L. Dowen, "Ionospheric Modification Experiment in Northern Scandinavia", J. Atmos. Terr. Phys., 44, 1025, 1982a.

Stubbe, P., H. Kopka, M.T. Rietveld, and R.L. Dowen, "ELF and VLF Wave Generation by Modulated HF Heating of the Current Carrying Lower Ionosphere", J. Atmos. Terr. Phys., 44,1173, 1982b.

Wong, A. Y., J. Carroll, R. Dickman, W. Harrison, W. Huhn, B. Lum, M. McCarrick, J. Santoru, C. Schock, G. Wong, and R.F. Wuerker, "High-Power Radiating Facility at the HLPAS Observatory", Radio Sci., 25, 1269, 1990.

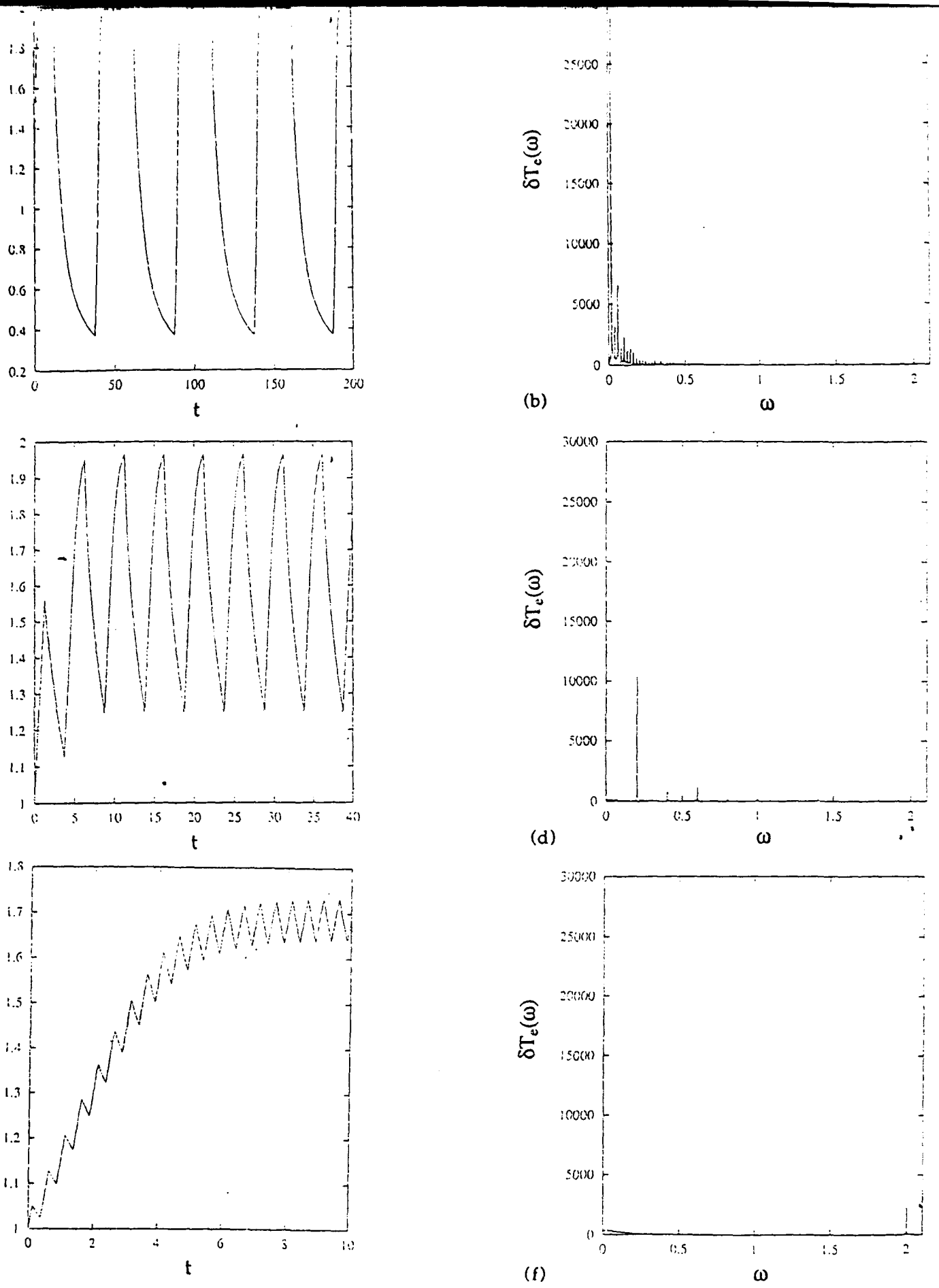


Fig. 1. Electron temperature modulation induced by a rectangular pulse modulated o-mode HF heater for three modulation frequencies (a) 10 Hz, (c) 100 Hz and (e) 1000 Hz; (b), (d) and (f) are the Fourier spectra of (a), (c) and (e) respectively.

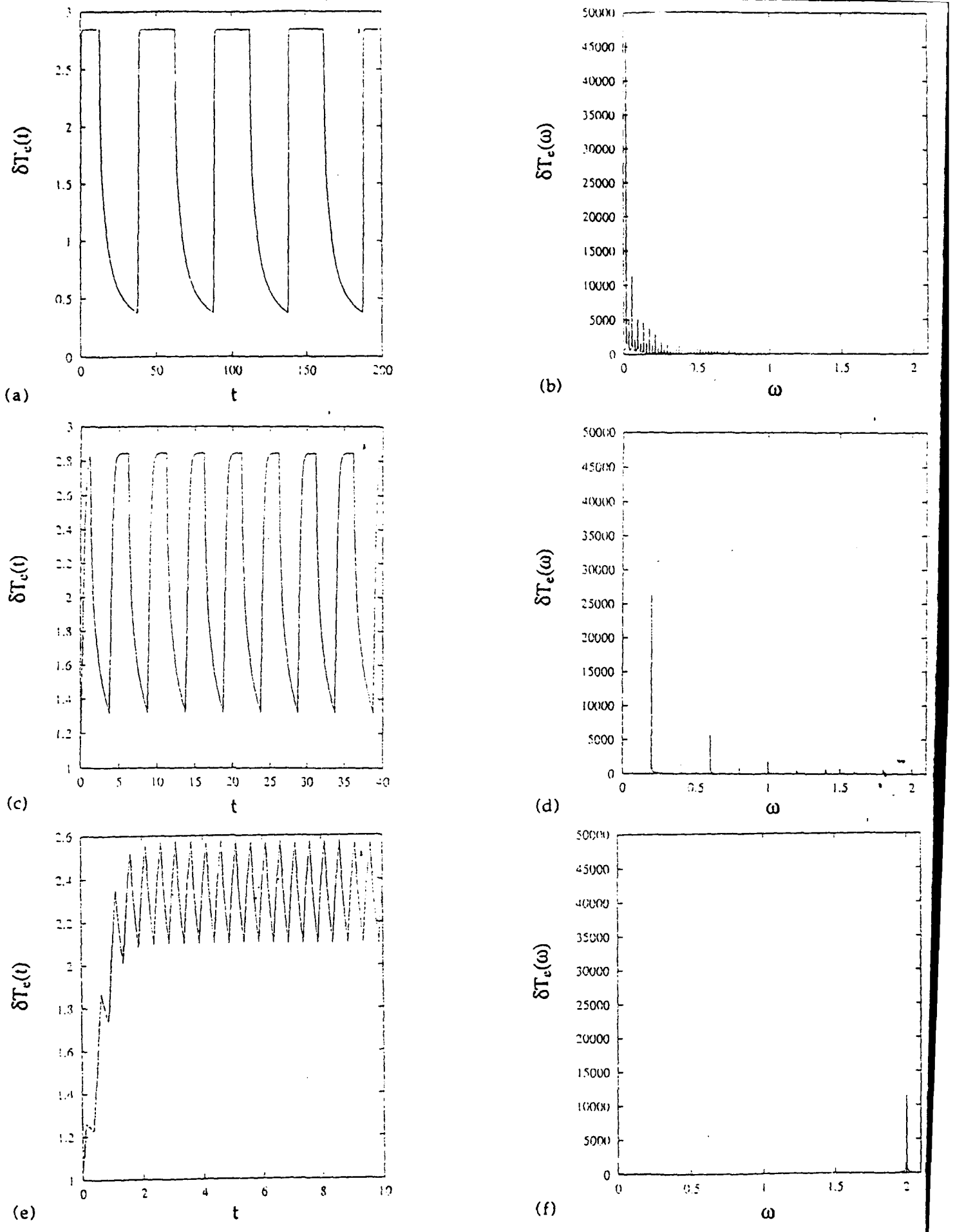


Fig. 2. Electron temperature modulation induced by a rectangular pulse modulated x-mode HF heater for three modulation frequencies (a) 10 Hz, (c) 100 Hz and (e) 1000 Hz; (b), (d) and (f) are the Fourier spectra of (a), (c) and (e) respectively.

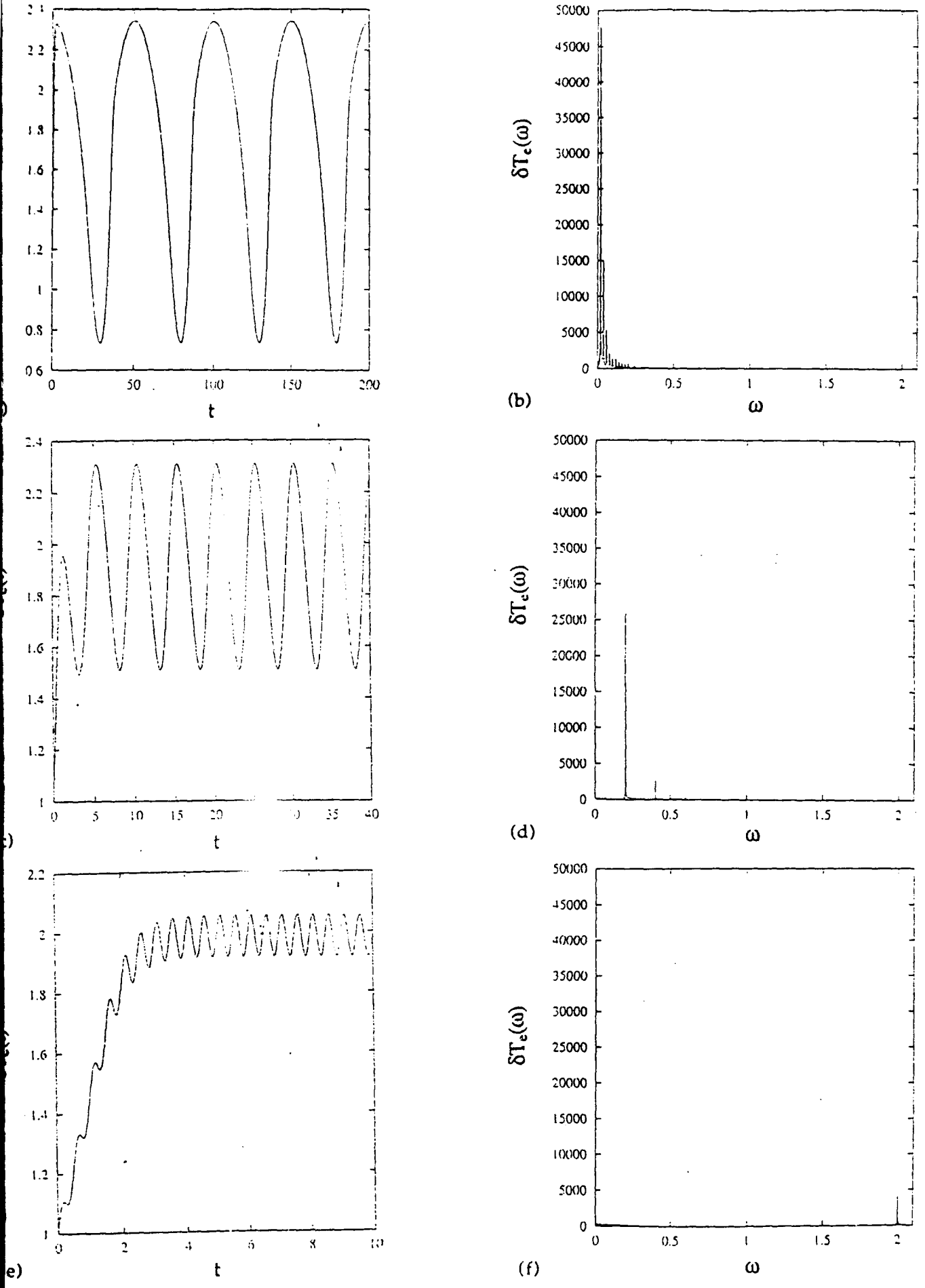


Fig. 3. Electron temperature modulation induced by two o-mode HF heaters beat for three frequency differences (a) 10 Hz, (c) 100 Hz and (e) 1000 Hz; (b), (d) and (f) are the Fourier spectra of (a), (c) and (e) respectively.

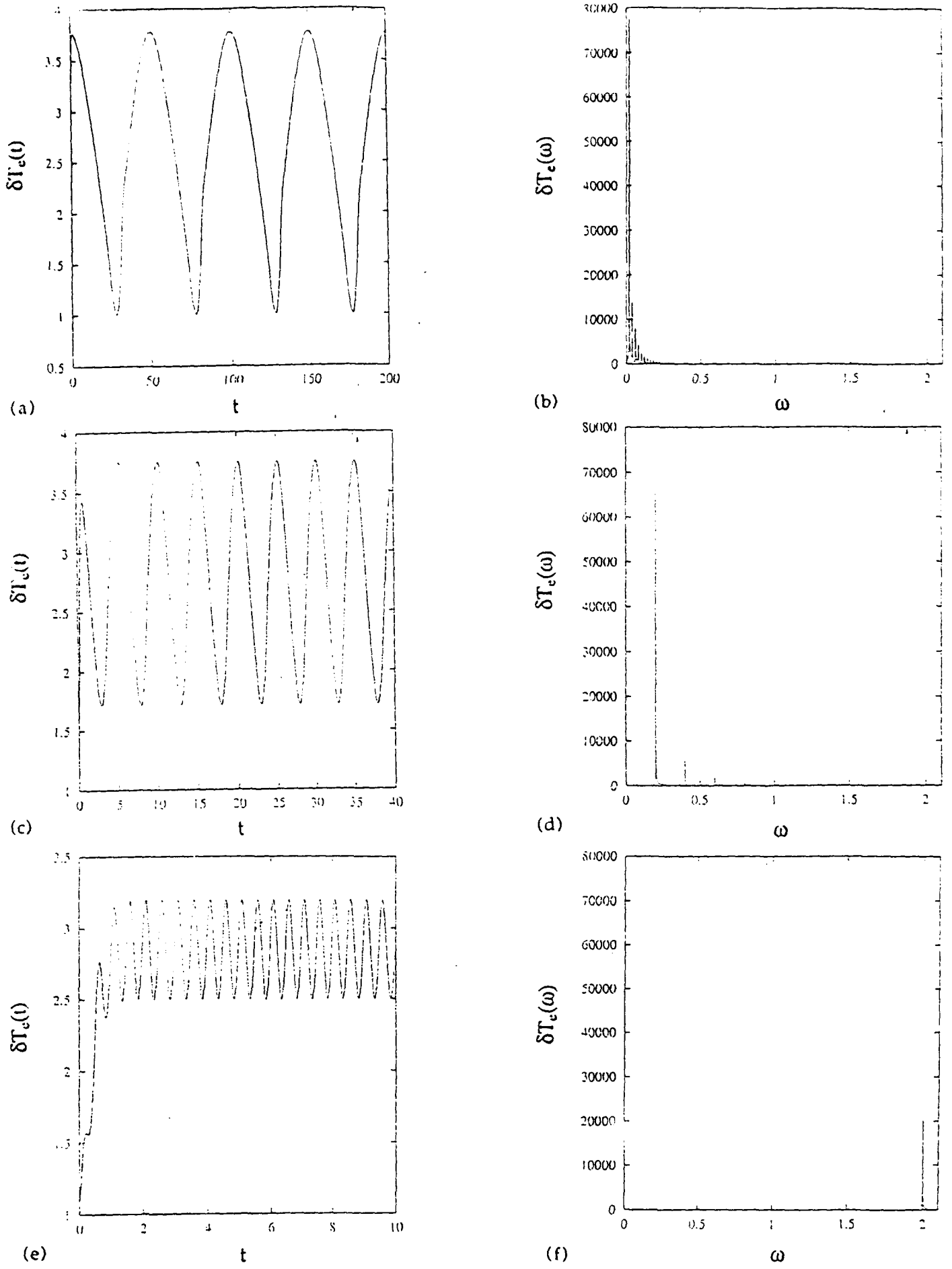


Fig. 4. Electron temperature modulation induced by two x-mode HF heaters beat for three frequency differences (a) 10 Hz, (c) 100 Hz and (e) 1000 Hz; (b), (d) and (f) are the Fourier spectra of (a), (c) and (e) respectively.

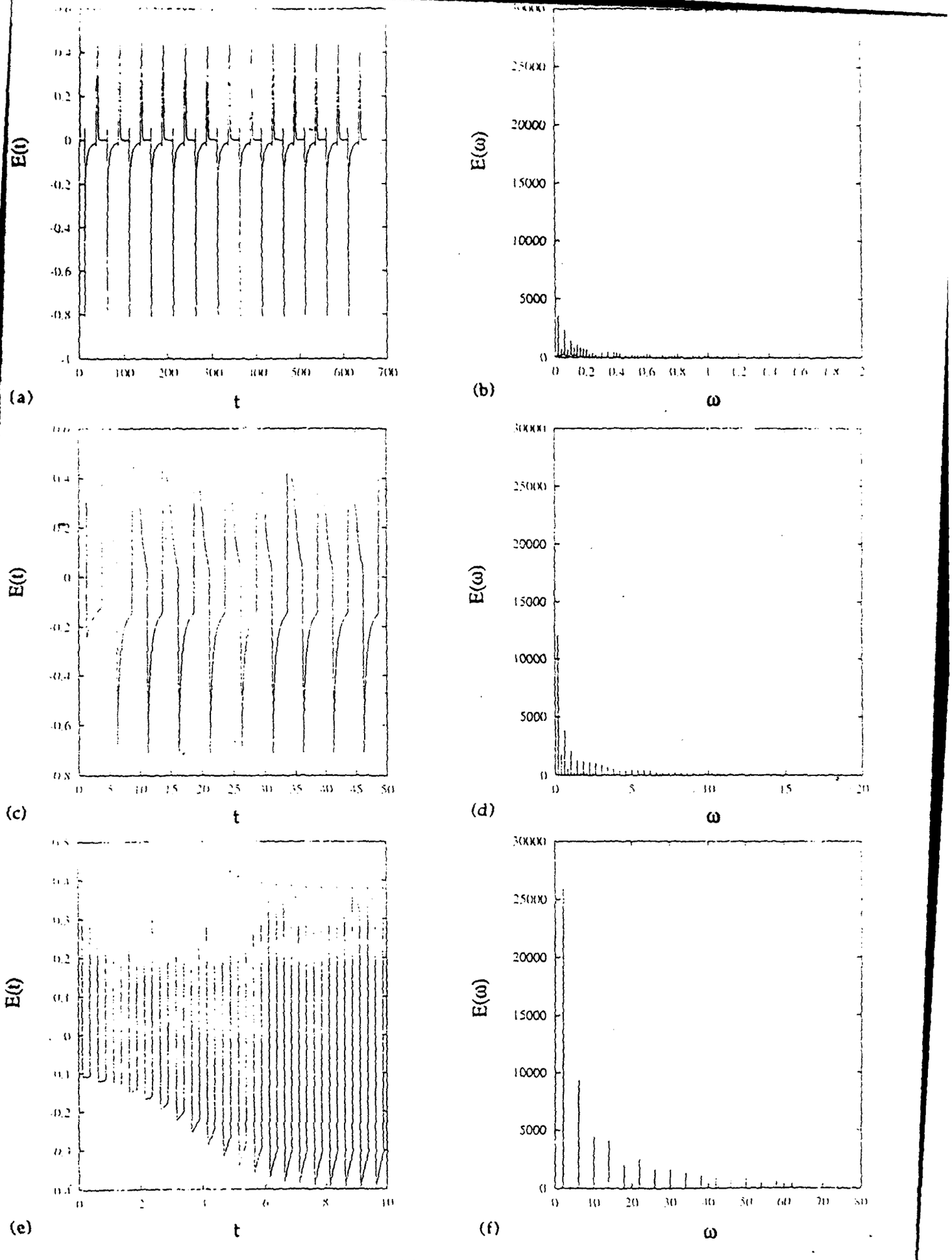


Fig. 5. Wave fields of ELF/VLF emissions (a). (c). (e) and their Fourier spectra (b), (d), (f) respond to the [illegible] modulations of Fig. 1.

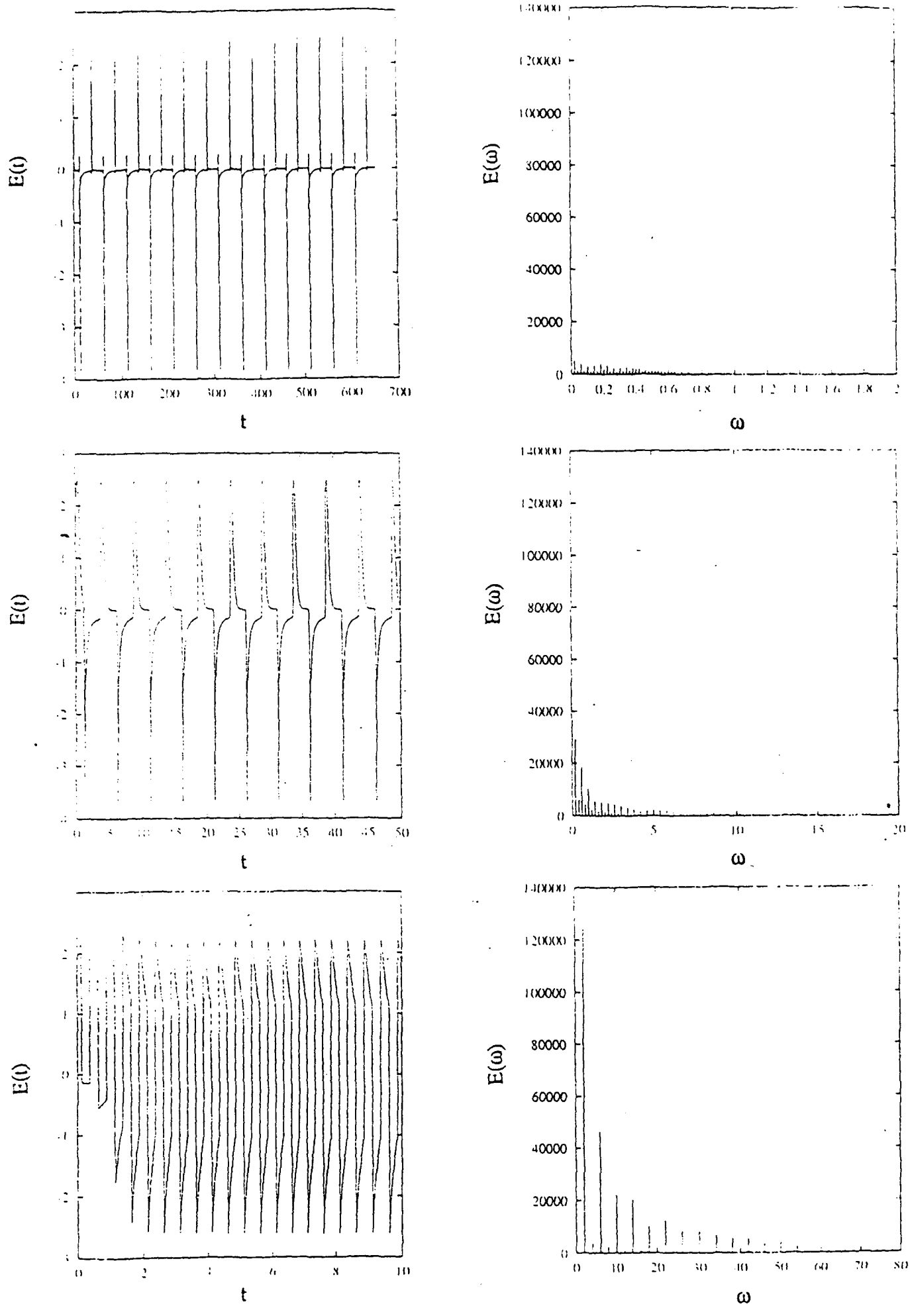


Fig. 6. Wave fields of ELF/VLF emissions (a), (c), (e) and their Fourier spectra (b), (d), (f) respond to the temperature modulations of Fig. 2.

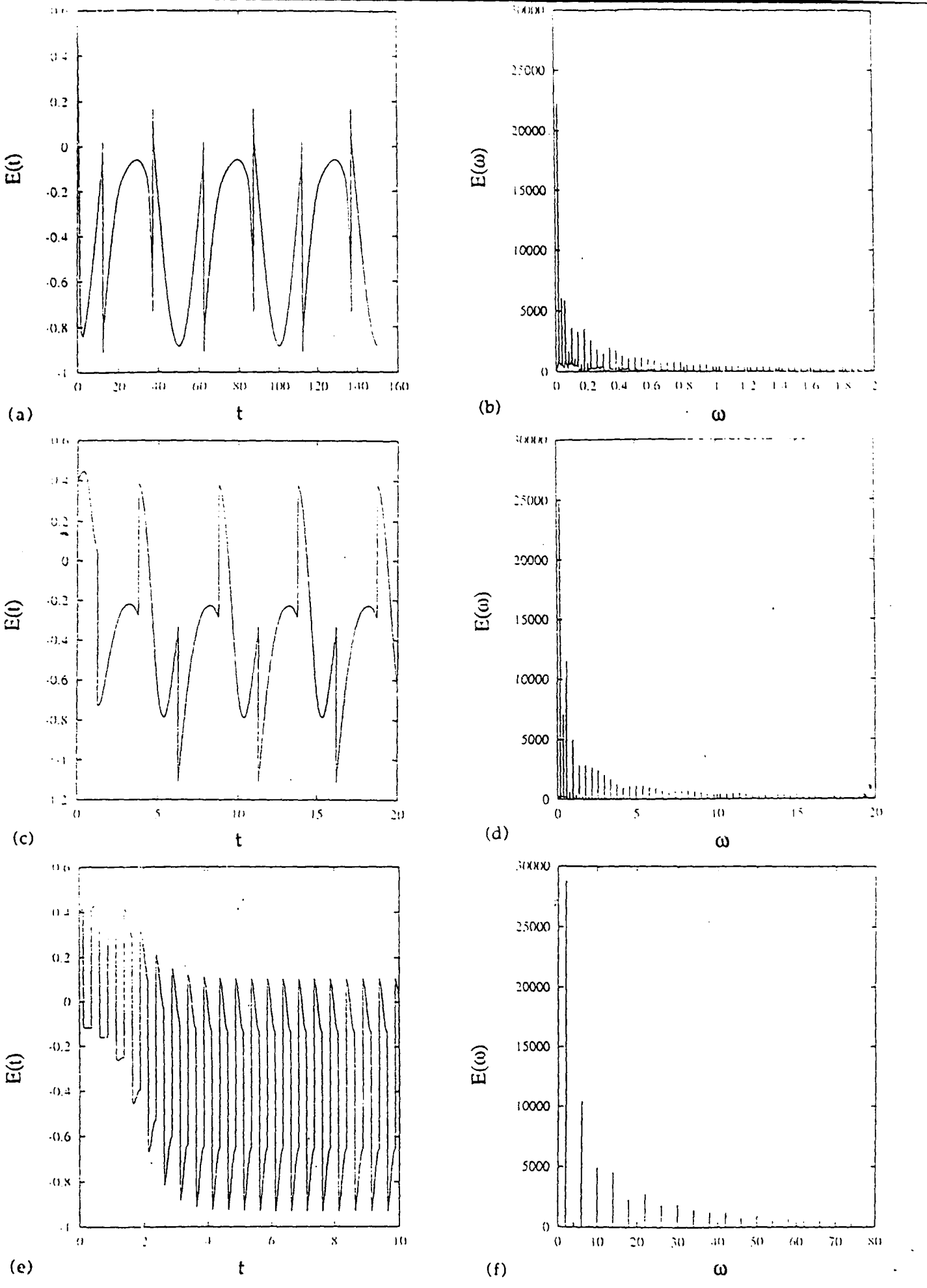


Fig. 7. Wave fields of ELF/VLF emissions (a), (c), (e) and their Fourier spectra (b), (d), (f) respond to the temperature modulations of Fig. 3.

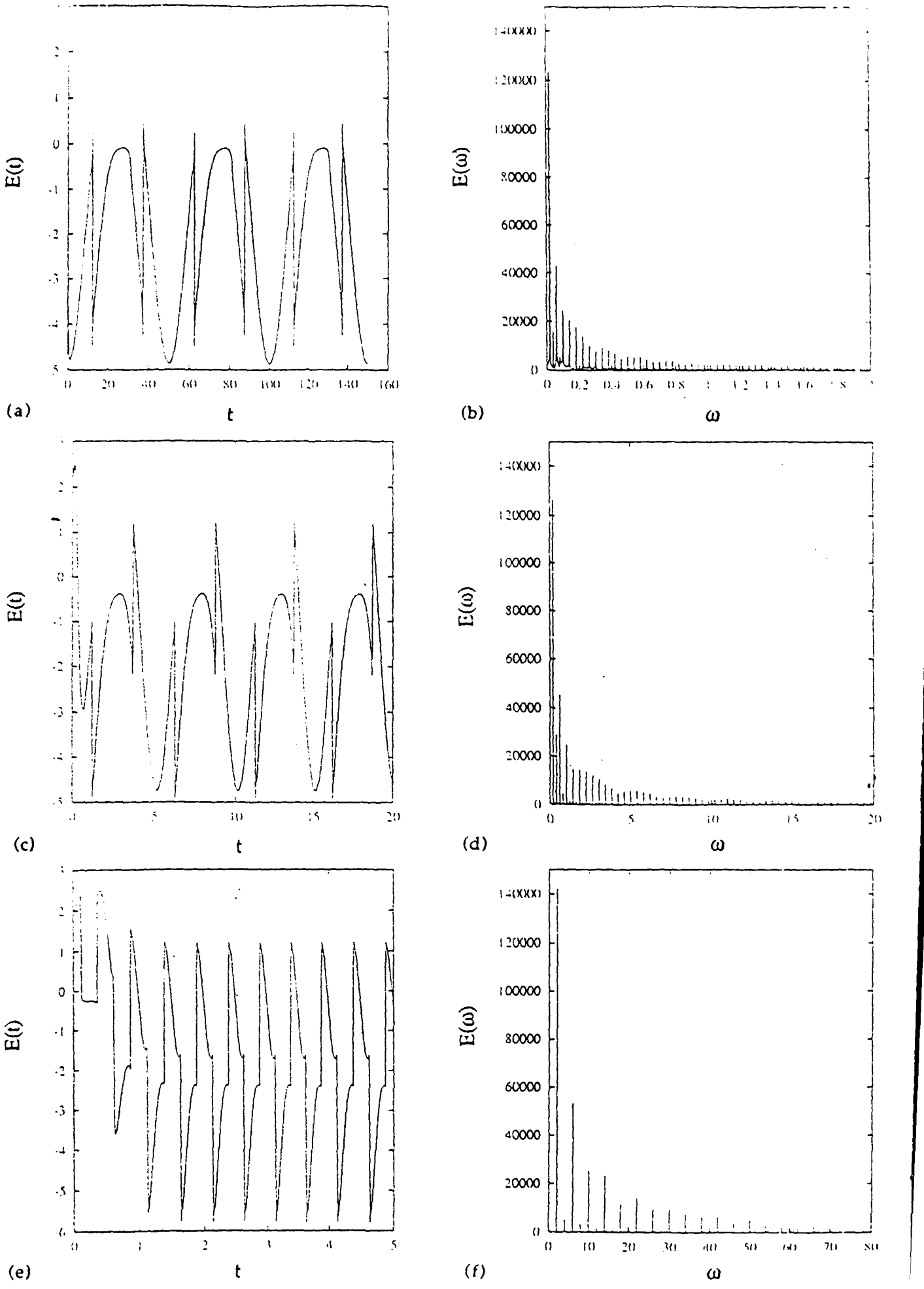


Fig. 8. Wave fields of ELF/VLF emissions (a), (c), (e) and their Fourier spectra (b), (d), (f) respond to the temperature modulations of Fig. 4.

INCIDENCE RATE ESTIMATE OF ELEVATED PEDIATRIC BLOOD LEAD AT TWO AIR FORCE BASES

Tze-San Lee

Associate Professor

Department of Mathematics

Western Illinois University

900 West Adams Street

Macomb, IL 61455

Final Report for:

Summer Faculty Research Program

Armstrong Laboratory

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D. C.

August 1993

INCIDENCE RATE ESTIMATE OF ELEVATED PEDIATRIC BLOOD LEAD AT TWO AIR FORCE BASES

Tze-San Lee
Associate Professor
Department of Mathethematics
Western Illinois University

Abstract

To account for possible large analytic errors in blood lead measurements, a bivariate lognormal model was employed to predict blood lead values of those children whose initial measurements were 10 micrograms per deciliter (μg/dl) or higher. Reductions of seventy and twenty-five per cent in the number of children with elevated blood lead levels at two Air Force Bases (AFBs) were shown by the use of the model. Assuming the mean and standard deviation of the true blood lead of the child population at two Air Force Bases are 7 μg/dl and 6.5 μg/dl, respectively, the overall incidence rate for both Air Force Bases was about the same at 0.8 percent. However, if the incidence rate is estimated by the place of residence, the incidence rates at the two Air Force Bases are very different. The off-base rate at Offutt AFB is six times the rate found at Randolph AFB; whereas, the on-base rate at Randolph AFB is three and one half times the rate found at Offutt AFB. In addition, the percentages of false negatives and false positives are calculated. Because of a larger analytic error in the blood lead measurements at Randolph AFB, the percentages of false negatives and false positives (which are higher than that at Offutt AFB) are eleven and fourteen percent. Suggestions for further study are also made.

INCIDENCE RATE ESTIMATE OF ELEVATED PEDIATRIC BLOOD LEAD AT TWO AIR FORCE BASES

Tze-San Lee

1. INTRODUCTION

Lead is a toxic metal which has no use in the physiological functioning of the human body. Yet because lead possesses some desirable properties (e.g., easily malleable, does not rust, a poor electrical conductor), lead was in great demand during the industrial revolution. The yearly consumption of lead in industry was more than one million tons in 1962 alone (Ziegfeld 1964).

One use of lead was as a pigment in paint. Lead-based paint (containing up to 50% lead) was widely used to paint the interior and exterior of houses built prior to 1940. Lead poisoning in children that is unrelated to parental exposure first occurred in Australia in 1892. It took twelve years before it was determined that eating leaded paint chips was the etiological factor (Gibson 1904). The first documented case of lead-based paint poisoning in U.S. children was reported in 1924 in the city of Baltimore (Ruddock 1924). Between 1930 and 1940, many cases of severe lead poisoning were reported in U.S. children (Lin-Fu 1980). In the 1940s, paint manufacturers began voluntarily replacing lead with titanium dioxide as a pigment in paint; but it wasn't until 1978 that the Consumer Product Safety Commission banned the manufacture of paint containing more than 0.06% lead by weight on interior and exterior residential surfaces, toys, and furniture. Nevertheless, about 3 million tons of lead in paint remain, nationwide, in an estimated 57 million occupied private housing units built before 1980 (HUD 1990).

The organic compound of lead, tetraethyl lead, was used as an antiknock additive in gasoline. It was discovered in 1922 by Thomas Midgley and co-workers at the General Motors Research Laboratory in Dayton, Ohio, that adding tetraethyl lead to gasoline raised the compression and hence, speed, by eliminating engine "knock". This allowed for the development of the "modern" automobile produced over the next 50 years (Epstein1928). Leaded gasoline was placed on sale in selected markets on February 1, 1923. This was done in

spite of the fact that industrial hygienists such as Alice Hamilton had long identified lead as an industrial toxin (Robert 1973). Although public health professionals expressed concern about the adverse effect of leaded gasoline on public health, it was to no avail. In July 1924 , the GM Director of Research, Graham Edgar, wrote to Dr. Paul Leech of the American Medical Association that "there is no danger of aquiring lead poisoning through even prolonged exposure to exhaust gases of cars using Ethyl Gas". The industry's assurances of the safety of leaded gasoline were undermined by a horrifying disaster that occurred in the Standard Oil Company's experimental laboratories in Elizabeth, New Jersey. Between October 26 and 30, 1924, five workers died and 35 others experienced severe palsies, tremors, hallucinations, and other serious neurological symptoms of organic lead poisoning. In light of the publicity over Bayway, it was also reported that other workers had died handling tetraethyl lead at both the DuPont chemical plant at Deepwater, New Jersey and the General Motors Research Division site in Dayton, Ohio. As a result of these continuing revelations, the Surgeon General of the Public Health Service called a national conference to assess the tetraethyl lead situation. The conference convened on May 20, 1925 in Washington, DC, with every major party represented. The conference opened with statements from General Motors, Dupont, Standard Oil, and the Ethyl Corporation outlining the history of the development of leaded gasoline and the reasons why they believed its continued production was essential. Three themes emerged as central arguments by the companies. First, the manufacturers maintained that leaded gasoline was essential to the industrial progress of America. Second, they maintained that any innovation involves certain risks. Third, they stated that the major reason deaths and illness occurred at their plants was that men who worked with the materials were careless and did not follow instructions. In the conference, Frank Howard of Ethyl Gasoline Corporation even dubbed the discovery of tetraethyl lead as an "apparent gift of God" (Rosner-Markowitz 1985). Since the introduction of lead as a gasoline additive, millions of tons have been used for this purpose (EPA 1986). As a result of U.S. Government regulatory action (in the 13 years between 1976 and 1989) the amount of lead used in gasoline was reduced by more than 99% (EPA 1990). Consequently, blood lead levels in the U.S. population have decreased by 37% (Annast 1982).

Lead is a toxin that affects virtually every system of the body. Results of recent studies have shown that

lead's adverse effects on the fetus and child occur at blood lead levels previously thought to be safe. Enormous studies have been made in the past 5 to 10 years that have increased our understanding of the damaging, long-term effects of lead on children's intelligence and behavior. Today in the United States, millions of children from all geographic areas and socioeconomic strata have lead levels high enough to cause adverse health effects. Childhood lead exposure costs the United States billions of dollars from medical and special education costs for poisoned children. Childhood lead poisoning continues in our society primarily because of lead exposure in the home environment, with lead-based paint being the principal high-dose source.

It is the aim of this report to study the degree of lead poisoning in children of Air Force families living in housing on and off base. In Section 2, analysis of blood lead levels in children from two Air Force Bases is presented. To account for the analytic error, a bivariate lognormal model was employed to predict the true blood lead levels of children whose initial blood lead measurements were 10 μg/dl or higher. The percentages of false negatives and false positives of pediatric blood lead measurements at two AFBs were also calculated. Section 3 is a discussion of the results presented in Section 2. Suggestions for further study are made in Section 4.

2. ANALYSIS OF PEDIATRIC BLOOD LEAD

Data

Because of the incidence of a few lead poisoning cases in children at Offutt and Randolph AFBs, a blood lead screening program for identifying children with undue lead exposure was initiated. A total of 1102 and 784 children were screened for blood lead at Offutt and Randolph AFBs, respectively. The age and place of residence of children whose blood lead values were greater than 10 μg/dl are given for each base in Tables 2.1 and 2.2. After collecting the first sample of venous blood from screened children at Randolph AFB, blood specimens were sent to Brooke Army Medical Center for the determination of blood lead levels. Twenty children were found to have elevated blood lead values. The medical staff at Randolph AFB suspected that there might be lead contamination on test tube rubber stoppers; consequently, a second sample was collected from 16 of the 20 children (4 were not available) with elevated blood lead values within two weeks after

receiving the results of the first sample. For various reasons, only sixteen children had second blood lead measurements; the results are given in the fifth column of Table 2.1. Out of 1102 children screened for blood lead measurements at Offutt AFB, twelve children were found to have elevated blood lead values. A 4 year old child having a blood lead value of 50 μg/dl measured in June 1992 received chelation therapy before a second blood lead measurement was taken on August 24, 1992; his blood lead value went down to 24 μg/dl. In October 1992, a third blood lead measurement showed further reduction to 13 μg/dl. The fifth column of Table 2.2 presents the blood lead values of those children with elevated levels taken at least a month later. A significant limitation of the data is that we do not have a complete record of blood lead values of children who did not have the elevated blood lead levels in their initial measurements. All we know is that their blood lead values were less than 10 μg/dl. The distribution of children by age and place of residence for the two Air Force Bases is given in Table 2.3(a-b). Only 258 cases out of 784 cases at Randolph AFB were categorized according to their age and place of residence. Numerical figures of proportionate generalization are given in the parentheses of Table 2.3(b). Although 1102 children were screened at Offutt AFB, only 779 cases were kept in the record. A similar projection was done on numerical figures at Offutt AFB, given in the parentheses of Table 2.3(a).

Statistical Analysis

Because of the high variability of blood lead measurements and high analytic errors, caution must be used when interpreting any single blood lead determination (Lerner 1975). To account for analytic error, which includes measurement error and sampling error, bivariate lognormal modeling was employed to predict the blood lead values of those children with elevated blood lead levels. These predictions, as a whole, are much closer to the true (unknown) blood lead levels. The bivariate lognormal model also enables us to calculate the percentage of false positives and false negatives among the initial blood lead measurements. A detailed explanation is given in the appendix. Table 2.4 was obtained by categorizing the predicted blood lead values of Tables 2.1-2 into different blood lead classes. Table 2.5 was then calculated from Tables 2.3-4. By following the method described in Appendix B of Lucas (1981), the percentages of false negatives and false positives of blood lead measurements at Offutt and Randolph AFBs are given in Table 2.6.

Results

The predicted blood lead values of children with initial elevated levels for Randolph and Offutt AFB are presented in the last column of Tables 2.1 and 2.2, respectively. As shown in Table 2.1, using a bivariate lognormal model to predict blood lead values results in a 70 percent reduction in the number of children (from 20 to 6) classified with elevated blood lead values at Randolph AFB. However, the reduction (from 12 to 9 children as noted from Table 2.2) at Offutt AFB is only 25 percent. The reason for such a difference is that the analytic error, which accounts for 50 percent of the total observed variance in the data at Randolph AFB, is larger than that at Offutt AFB where the analytic error accounts for only 14 percent (Table 2.7). After adjusting for the analytic error, the incidence rates at both AFBs are about the same (0.8%). However, if the incidence rates are broken up according to age and place of residence, a different picture emerges. As seen in Table 2.4, all children with elevated blood lead values at Randolph AFB are in class II (< 20 μg/dl), whereas two children at Offutt AFB are in Class III (20-44 μg/dl) (CDC 1991). We note from Table 2.5 that the incidence rate for on-base residents at Randolph AFB is three and a half times higher than that for Offutt AFB. On the other hand, the incidence rate for off-base residents at Offutt AFB is six times higher than that for Randolph AFB. This is probably due to differences in the age of the housing units. According to the calculations of Pope (1986) and data from the 1983 American housing survey (U.S. Census Bureau, 1986), the percentages of age-stratified housing having paint with lead > 0.7 mg/cm^2 are: pre-1940, 99%; 1940-1959, 70%; and, 1959-1974, 20%. Indeed, the housing in which children with elevated blood lead values lived at Offutt AFB were built before 1950. At Randolph AFB, the building lived in by Child No. 8 in Table 2.1 was built in 1931. These results are consistent with the premise that lead-based paint is the major source of high-dose lead exposure to children. In addition, Table 2.4 shows that although most children with elevated blood lead values are in the age category of 6 months up to 2 years old, children in other age categories can have undue lead exposure as well. From Table 2.6, it is seen that the percentages of false negatives and false positives at Randolph AFB (which are greater than ten percent) are larger than that for Offutt AFB. This verifies that a large analytic error (Table 2.7) can cause a relatively high percentage of false negatives and false positives.

3. DISCUSSION

From this study, the overall incidence rates of elevated pediatric blood lead for Offutt and Randolph AFBs are about the same (0.8 percent); i.e., per one thousand children, eight are expected to have blood lead levels of 10 micrograms per deciliter or above. However, if the incidence rate is estimated by the classification of age and place of residence, then different pictures emerge. We notice that the on-base incidence rate of Randolph AFB is three and one-half times that of Offutt AFB. However, the off-base incidence rate of Offutt AFB is six times that of Randolph AFB. This phenomenon can be readily explained by the age of the housing at each Base. Most off-base housing near Randolph AFB was built in the 1970s or later. Hence, the interior and exterior surfaces of homes probably do not have lead-based paint. On the other hand, off-base housing near Offutt AFB was built primarily in the 1950s. These houses are more likely to have a lead-based paint problem. The environmental investigation conducted by Captain C. Cogburn at Offutt AFB and MSgt Walters at Randolph AFB indicates that chipping and flaking of leaded paint and lead in water are suspected as the major sources of undue lead exposure to children. Another observation is that children in any age category (under 6 years) can have undue lead exposure, although most children with elevated blood lead levels are in the age class of six months to up two years old. Since we do not have exact numerical figures for children whose blood lead measurements are below 10 μg/dl at both AFBs, we cannot easily estimate the mean and standard deviation of the blood lead for the whole population of children. Based upon a study of more than 200 children at Boston Lying-In Hospital (Rabinowitz 1984), an educated guess about the mean and standard deviation was made for children at these two AFBs. It is assumed that the mean and the (between-person) standard deviation of children at these two AFBs are 7 and 6.5 μg/dl, respectively, in the calculation of their predicted blood lead values through the bivariate lognormal models. The accuracy of the results presented above depends on the validity of the assumed mean and standard deviation.

4. SUGGESTIONS

From this study, several suggestions are made. First, blood lead measurements for all screened children should be recorded in exact figures and preferably stored in a computer data base, as this information will be

useful in estimating the unknown mean and standard deviation of the population of children. Unfortunately, this information was not available for this study. Consequently, assumptions about the mean and standard deviation on children's true blood lead were made. Secondly, for about 10% of the total blood lead measurements, duplicate determinations are needed in order to estimate measurement error. In several large scale studies of blood lead, all blood samples have duplicate determinations (NHANES II 1984, Rabinowitz 1982). Thirdly, to take care of false negative cases, children with their measurements less than 10 μg/dl but above 9 μg/dl should be retested. The justification for choosing 9 μg/dl is that the percentage of false negatives is about ten percent, which is about one seventh of 68% (one standard deviation for the normal distribution). One seventh of the assumed standard deviation (6.5 μg/dl) is about 1 μg/dl. Therefore, children with results greater than 9 μg/dl (10 μg/dl threshold - 1 μg/dl error) should be retested to account for false negative. Fourthly, to study the stability of pediatric blood lead, a random sample of 30 children from each age strata should be tested every 6 six months for 3 years. And finally, pregnant women should be screened for blood lead levels because the fetus is very sensitive to transplacental lead exposure.

ACKNOWLEDGEMENT

The author would like to thank the staff of the Industrial Hygiene Branch of the Occupational Medicine Division, Armstrong Laboratory for their contributions to this study. Especially, Major C. Hollenbeck not only corrected some grammatical errors, but also made suggestions which greatly improved the presentation of the paper. Captain M. Mader and Lieutenant I. Hermon-Cruz provided useful discussions on the subject of this study. Captain C. Cogburn at Offutt AFB and MSgt Walters at Randolph AFB who provided the data for this study are gratefully acknowledged.

Table 2.1 Age, place of residence, and blood lead values for 20 out of 784 children measured at Randolph AFB, Texas, whose initial blood lead values were greater than 10 μg/dl

Child	Age	Place of residence	Measured blood lead (μg/dl)		Predicted blood lead (μg/dl)
			First sample	Second sample	
1	15 month	on-base	35	27	15.7
2	16 month	on-base	17	5	10.9
3	17 month	on-base	13	< 5	9.6
4	5 year	on-base	16	< 5	10.6
5	17 month	off-base	20	11	11.9
6	4 year	off-base	10	N.A.[a]	8.4
7	6 year	on-base	11	< 5	8.8
8	16 month	on-base	17	21	10.9
9	11 month	on-base	10	< 5	8.4
10	13 month	off-base	12	N.A.	9.2
11	13 month	off-base	11	N.A.	8.8
12	11 month	on-base	11	N.A.	8.8
13	13 month	on-base	10	5	8.4
14	12 year	on-base	10	10	8.4
15	3 year	on-base	10	5	8.4
16	8 year	on-base	11	6.4	8.8
17	2 year	on-base	10	5	8.4
18	5 year	on-base	16	< 5	10.6
19	1 year	off-base	14	7	9.9
20	1 year	off-base	13	9	9.6

[a] N.A. means not available.

Table 2.2 Age, place of residence, and blood lead values for 12 out of 1102 children measured at Offutt AFB, Nebraska whose initial blood lead values were greater than 10 μg/dl

Child	Age	Place of residence	Measured blood lead (μg/dl)		Predicted blood lead (μg/dl)
			First sample	Second sample	
1	12 mo	on-base	22	9	18.7
2	4 yr	off-base	50	24 (13[a])	37.7
3	12 mo	off-base	15	N.A.	13.5
4	5 yr	off-base	11	9	10.3
5	2.5 yr	on-base	10	N.A.[b]	9.5
6	27 mo	on-base	10	N.A.	9.5
7	2 yr	off-base	11	N.A.	10.3
8	8 mo	off-base	12	N.A.	11.1
9	12 mo	off-base	20	15	17.2
10	10 mo	off-base	29	28	23.7
11	5 mo	on-base	12	8	11.1
12	13 mo	off-base	10	< 5	9.5

[a] The third blood lead measurement

[b] N.A. means not available.

Table 2.3

(a) Distribution of children screened for blood lead content by age and place of residence at Offutt AFB[a], Nebraska.

Age (year)	Place of residence	
	On-base	Off-base
0.5 - 1	169 (239[b])	285 (403)
2 - 3	76 (108)	39 (55)
4 - 5	71 (100)	35 (50)
6+	57 (81)	47 (67)
Total	373 (528)	406 (574)

[a] The distribution presented here only represents 779 out of the total 1102 children screened.
[b] The numbers inside the parenthesis are the number from each cell mutiplied by 1.415 which approximately equals 1102/779.

(b) Distribution of children screened for blood lead content by age and place of residence at Randolph AFB[a], Texas.

Age (year)	Place of residence	
	On-base	Off-base
0.5 - 1	71 (216[b])	98 (298)
2 - 3	31 (94)	21 (64)
4 - 5	11 (33)	14 (43)
6+	3 (9)	9 (27)
Total	116 (352)	142 (432)

[a] The distribution presented here only represents 258 out of the total 784 children screened.
[b] The numbers inside the parenthesis are the number from each cell mutiplied by 3.0388 which approximately equals 784/258.

Table 2.4 Distribution of child blood lead levels by age and place of residence at Offutt (Randolph) AFB*

Place of residence and age (yr)	Blood lead levels (μg/dl)				
	0 - 9	10 - 14	15 - 19	20 - 24	25+
On-base					
0.5 - 1	237 (213)	1 (2)	1 (1)	0 (0)	0 (0)
2 - 3	108 (94)	0 (0)	0 (0)	0 (0)	0 (0)
4 - 5	100 (31)	0 (2)	0 (0)	0 (0)	0 (0)
6+	81 (9)	0 (0)	0 (0)	0 (0)	0 (0)
Off-base					
0.5 - 1	399 (295)	2 (2)	1 (1)	1 (0)	0 (0)
2 - 3	54 (64)	1 (0)	0 (0)	0 (0)	0 (0)
4 - 5	48 (43)	1 (0)	0 (0)	0 (0)	1 (0)
6+	67 (27)	0 (0)	0 (0)	0 (0)	0 (0)

* Total number of children screened for blood lead measurement is 1102 (784).

Table 2.5 The incidence rate of children who had elevated blood lead levels of 10 μg/dl or greater by age and place of residence at Offutt AFB, Nebraska and Randolph AFB, Texas.

Age (year)	Place of residence			
	On-base		Off-base	
	Offutt AFB	Randolph AFB	Offutt AFB	Randolph AFB
0.5 - 1	0.8%	1.5%	1%	0.3%
2 - 3	0%	0%	2%	0%
4 - 5	0%	6%	4%	0%
6+	0%	0%	0%	0%
Overall	0.4%	1.4%	1.2%	0.2%

Table 2.6 Estimation of false negatives and positives of blood lead measurements when both the true and measured blood lead threshold value of the population are 10 μg/dl

	Location of Air Force Base	
	Offutt	Randolph
Percent of values which are "false negatives"	5.7%	11.4%
Percent of values which are "false positives"	6.1%	13.8%

Table 2.7 Analysis of variance for blood lead values from two Air Force Bases

Air Force Base	Variance Components		Total
	σ^2_t	σ^2_e	σ^2_X
Offutt	3.7436 (86%)	0.6252 (14%)	4.3688
Randolph	4.1140 (50%)	4.0833 (50%)	8.1973

APPENDIX

In this appendix we shall evaluate the effect of analytic error. Because of analytic error, it is possible that a large percentage of the measurements that are less than a threshold are "false negatives" in the sense that the individual's true blood lead level is above the threshold. The occurence of "false negatives" is illustrated in Fig. 1, where the measured blood lead value is plotted on the x-axis and the true blood lead value is plotted on the y-axis. The "measured threshold limit" and "true threshold limit" divide the distribution into four quadrants - "A", "B", "C", and "D" (see Fig. 1).

"A" represents the measured values which are correctly classified as being below the threshold value since they lie below the measured and true threshold limits. The values in "D" are correctly classified as being above the threshold limits. The values in "B" and "C" are incorrectly classified. The values in "B" lie above the measured threshold but below the true threshold limit. The values in "B" will be incorrectly classified as being above the threshold; these are "false positives". In a similar manner, the values in "C" will be incorrectly classified as being below the threshold. These are "false negatives".

Assume that the relationship between measured and true blood values follows a bivariate lognormal distribution. Let

Y = Log(true blood lead)

X = Log(measured blood lead)

ϵ = Log(analytic error)

Then

$$X = Y + \epsilon$$

where Y and ϵ are independent with

$$Y \sim N(\mu_t, \sigma_t^2)$$

$$\epsilon \sim N(0, \sigma_a^2)$$

$\Rightarrow$

$$X \sim N(\mu_t, \sigma_X^2),$$

where

$$\sigma_X^2 = \sigma_t^2 + \sigma_a^2. \qquad \text{(A-1)}$$

Then it can be shown (Hogg and Craig 1 ...) that

$$E(Y|X) = \mu_t + \rho \cdot \sigma_t/\sigma_X(X \quad)$$

$$= \mu_t + (\sigma_t/\sigma_X)^2 \cdot (X - \ ) \qquad \text{(A-2)}$$

where

μ_t = the mean of Y

μ_X = the mean of X = μ_t

σ_t^2 = the variance of Y

σ_X^2 = the variance of X

ρ = the correlation between X and Y

$$= Cov(X,Y)/(\sigma_X \cdot \sigma_t) = \sigma_t/\sigma_X \qquad \text{(A-3)}$$

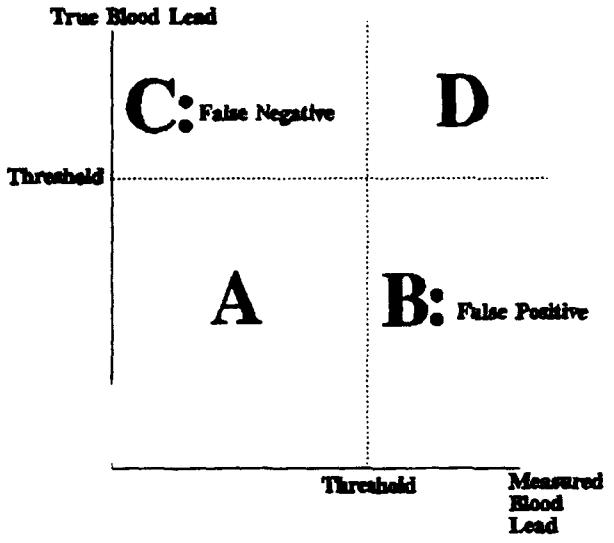


Fig. 1: True blood l ... measured blood lea...

It has been s... wn (Lucas 1981) tha...

$$\sigma_t^2 = \sigma^2_{person} + \sigma^2_{within\text{-}person} \qquad \text{(A-4)}$$

$$\sigma_a^2 = \sigma^2_{sampling} + \sigma^2_{measurement} \qquad \text{(A-5)}$$

The means of the true blood lead values are ordinarily estimated by the measured values. Since we do not have the exact numerical figures for children whose blood lead measurements are below 10 μg/dl at both AFBs, we cannot estimate the mean and standard deviation of the blood lead for the whole population of children. Based upon a study of more than 200 children at Boston Lying-In Hospital (Rabinowitz 1984), an educated guess about the mean and standard deviation was made for children at these two AFBs. It is assumed that the mean and the standard deviation of children at these two AFBs are 7 and 6.5 μg/dl, respectively. Among four variance components, the measurement error cannot be estimated separately because no duplicate blood samples were measured. For the Randolph data (Table 2.1), $\sigma^2_{\text{within-person}}$ can be estimated to be 0.3704 from the values of the first and second samples. Thus, σ_t^2 is estimated by the sum of σ^2_{person} and $\sigma^2_{\text{within-person}}$ which is $\ln(6.5)^2 + 0.3704 = 4.114$. The value of (σ_a^2) is estimated to be 4.0833 from the first and second samples. The estimate of the correlation (ρ) of (A-3) is σ_t/σ_X = the square root of $4.114/(4.114 + 4.0833) = 0.7$. For the Offutt data (Table 2.2), instead of $\sigma^2_{\text{within-person}}$, only $\sigma^2_{\text{sampling}}$ can be estimated separately as 0.5491, since the time interval between the first and second sample is more than a month apart. The value of σ_t^2 of (A-4) is then estimated to be equal to $\sigma^2_{\text{person}} = \ln(6.5)^2 = 3.7436$. Again, the value of (σ_a^2) of (A-5) is estimated to be 0.6252 from the first and second samples of the Offutt data. Similarly, the estimate of the correlation (ρ) of (A-3) is σ_t/σ_X = the square root of $3.7436/(3.7436 + 0.6252) = 0.93$. The sixth column of Tables 2.1-2 are then obtained by substituting the estimated values of σ_t/σ_X into (A-2). Table 2.7 is obtained by using (A-4) and (A-5), while Table 2.6 is obtained by using the method described in Appendix B of Lucas (1981). Incidentally, in calculating the analytic error, only those children who had blood lead values in the second sample in Tables 2.1-2 were considered. Also, for children whose second blood lead values were recorded as less than 5 μg/dl, their blood lead values are set to equal 4 μg/dl in the computation of analytic error.

REFERENCES

Annest, J. L., Pirkle, J. L., Makuc, D., Neese, J. W., Bayse, D. D., and Kovar, M. G. (1983). Chronological trend in blood lead levels between 1976 and 1980. The New England J. of Medicine, 308, 1373-1377.

Centers for Disease Control (1991). "Preventing Lead Poisoning in Young Children", U.S. Department of Health and Human Services, Public Health Service, Atlanta, GA.

Environmental Protection Agency (1986). "Air Quality Criteria for Lead", EPA Report no. EPA/600/8-83/028aF, Office of Health and Environmental Assessment, Research Triangle Park, North Carolina.

Environmental Protection Agency (1990). Lead in gasoline (Quarterly summary of lead phase down reporting data), March 8, 1990.

Epstein, R. (1928). The Automobile Industry, A. W. Shaw, New York.

Gibson, J. L. (1904). A plea for painted railings and painted walls of rooms as the source of lead poisoning among Queensland children, Austral. Med. Gazette, 23, 149-153.

Hogg, R., and Craig, A. (1989). Introduction to Mathematical Statistics, 4th edition, MacMillan Publishing Company, New York.

Lerner, S. (1975). Blood lead analysis: precision and stability, J. of Occupational Medicine, 17, 153-154.

Lin-Fu, J. S. (1980). Lead poisoning and undue lead exposure in children: history and current status, in "Low Level Lead Exposure: The Clinical Implications of Current Research" edited by H. L. Needleman, pp. 5-15, Raven Press, New York.

Lucas, J. M. (1981). Effect of analytic variability on measurements of population blood lead levels, Amer. Ind. Hyg. Assoc. J., 42, 88-96.

National Health and Nutrition Examination Survey II (1984). "Blood Lead Levels for Persons Ages 6 Months - 74 Years: United States, 1976-80", U.S. Department of Health and Human Services, National Center for Health Statistics, Hyattsville, MD.

Pope, A. (1986). "Exposure of Children to Lead-Based Paints", EPA Contract No. 68-02-4329, Strategies and Air Standards Division, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Rabinowitz, M., Leviton, A., Needleman, H. (1984). Variability of blood concentrations during infancy, Arch. of Environ. Health, 39, 74-77.

Rabinowitz, M., and Needleman, H. (1982). Temporal trends in the lead concentrations of umbilical cord blood, Science, 216, 1429-1430.

Robert, J. C. (1983). Ethyl, University of Virginia Press, Charlottesville, VA.

Rosner, D. and Markowitz, G. (1985). A 'gift of God'?: The public health controversy over leaded gasoline during the 1920s, Amer. J. Public Health, 75, 344-352.

Ruddock, J. C. (1924). Lead poisoning in children, J. Amer. Med. Assn., 82, 1682-1684.

U.S. Bureau of the Census (1986). "American Housing Survey, 1983. Part B: Indicators of Housing and Neighborhood Quality by Financial characteristics", U.S. Department of Commerce, Washington, DC.

Ziegfeld, R.L. (1964). Importance and uses of lead, Archives of Environmental Health, 8, 14-24.

Keith Priestly's report was unavailable at time of publication.

LARGE-SCALE CORONAL MAGNETIC FIELDS: NOISE STORMS, SOFT X-RAYS AND INVERSION OF RADIO POLARIZATION

Robert F. Willson
Research Associate Professor of Astronomy
Department of Physics and Astronomy

Tufts University
Medford, MA 02155

Final Report for:
Summer Faculty Research Program
Phillips Laboratory, Hanscom AFB

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1993

LARGE-SCALE CORONAL M.' ' FIELDS: NOISE STORMS, SOFT X-RA' 'D INVERSION OF RADIO POLARIZ.)N

Robert F. Willson
Research Associate Professor of Astro:.omy
Department of Physics and Astronomy
Tufts University

Abstract

Large-scale coronal magnetic fields n link wid' 'separated active regions in opposite hemispheres of the Sun. The presenc erred from Very Large Array (VLA) observations of noise storms located within them; by Yohkoh Soft X-ray Telescope (SXT) images that delineate their magnetic structure; and by RATAN 600 spectral observations of inversions of the circularly-polarized radio emission that can be used to infer the coronal magnetic field strength. Extrapolations of photospheric magnetic fields to coronal heights indicate that non-potential, or current-amplified, magnetic fields are, in some cases, required to explain the polarization inversions. These anomalously high magnetic fields are also located at the source of the noise storm; such storms require long-lasting (hours) non-thermal particle acceleration.

I. INTRODUCTION

Multi-wavelength radio observations of solar active regions provide useful information about the strength and structure of coronal magnetic fields. At centimeter wavelengths, the radio emission above sunspots is dominated by thermal gyroresonance radiation at low harmonics of the local gyrofrequency, with shorter wavelengths corresponding to higher magnetic fields and lower altitudes in the corona (Alissandrakis, Kundu and Lantos 1980; Lang and Willson 1982; Lang, Willson and Gaizauskas 1983). Two-dimensional VLA synthesis maps of total intensity, I, and Stokes parameter, V, have shown that the polarization structure above sunspots varies with frequency and position, but that the sign of circular polarization usually agrees with that of underlying sunspots or plage (right-handed circular polarization for north magnetic polarity directed out of the Sun - extraordinary mode of wave propagation) (Alissandrakis and Kundu 1982; 1984; Lang and Willson 1982). In some cases, however, the polarity of the radio emission is reversed at longer wavelengths and such reversals have been attributed to a propagation effect in which the microwave emission undergoes a polarization inversion upon traversing a "quasi-transverse" (QT) layer in the coronal field (eg. Kundu et al. 1977; Webb et al. 1983; Alissandrakis and Kundu 1984; Brosius et al. 1992).

Recent observations of active regions with the RATAN-600 have shown that the QT layers occur at relatively high coronal altitudes ($h \approx 10^{10}$ cm) where the magnetic field strengths are on the order of 10 - 20 G (Gelfreikh, Peterova and Ryabov 1987). These regions probably correspond to large-scale coronal loops or arches whose presence might be detected at X-ray or long ($\lambda \geq 20$ cm) radio wavelengths.

In this paper we discuss observations of a complex group of active regions made on two consecutive days with the new Panoramic Spectrum Analyzer (PSA) on the RATAN-600 in the wavelength range 3.5 to 6.5 cm and with the Very Large Array (VLA) at wavelengths of 20 and 91 cm. The broad spectral coverage provided by the PSA was used to study a number of sunspot-associated sources which exhibited changes in circular polarization over relatively narrow bandwidths. These data have been interpreted as evidence for wave propagation through QT layers with different magnetic field strengths and electron densities. Compar-

isons with potential field extrapolations also show that coronal currents may have amplified the magnetic field in at least one of these regions.

The VLA data were used to image coronal loops at 20 cm as well as solar noise storms at 91 cm. The presence of fluctuating noise storm emission provided evidence for long-lasting nonthermal processes in these regions as well as clues to the structure of large-scale coronal loops that may be the sites of QT regions.

The RATAN and VLA data were also compared with soft X-ray images from the Yohkoh satellite during the two days of observation. These data revealed evolving, large scale, soft X-ray structures that apparently link active regions in opposite solar hemispheres, while also providing evidence for currents in the coronal regions above the sources detected by the RATAN and VLA.

II. OBSERVATIONS WITH THE RATAN-600

The RATAN-600 was used to observe the Sun on 9 and 10 January, 1992 with the new Panoramic Spectrum Analyzer (PSA) at 27 wavelengths between 1.7 and 32 cm (for a description of the PSA see Bogod et al. 1993). On January 1992 the observations of the Sun with the RATAN-600 were made daily near 9:20 UT (local noon). In Figures 1 and 2 we show radio scans of the solar disk in total intensity, I, and circular polarization, or Stokes parameter, V, together with optical maps of the Sun as published in the bulletin *Solnechne Danneye*.

The radio scans of total intensity show enhanced emission above sunspots at all wavelengths with a general decrease in intensity at the shorter wavelengths. The polarization scans at the shorter wavelengths reflect the structure of the photospheric magnetic fields, with right-handed polarization corresponding to north polarity of the photospheric field and left-handed polarization corresponding to south polarity in the photosphere. This is expected if the radio emission corresponds to the extraordinary mode of wave propagation. Note, however, the inversion of polarization at longer wavelengths of the sunspot-associated source *a*, identified with active region AR 6996 (region number 1 on the Kislovodsk maps).

On January 9 the emission is exclusively left circularly polarized at wavelengths $\lambda \leq 4.60$ cm, reverses to right circular at $\lambda = 4.73$ - 5.26 cm, then reverses again to left circular at $\lambda = 6.5$ cm. An abrupt reversal of polarization was also detected on January 10 at a somewhat shorter wavelength of $\lambda \approx 3.45$ cm. The source *a* is connected with the large leading sunspot in AR 6996 whose negative photospheric polarity and predominantly left-handed circular polarization suggest the extraordinary mode of wave propagation at most wavelengths. In Section V we show that the changes in polarization might be attributed to a propagation effect in which the microwave emission undergoes polarization inversions upon transversing regions of quasi-transverse (QT) magnetic field high in the corona.

III. VLA OBSERVATIONS

The VLA was used to observe the Sun at 20.7 and 91.6 cm (LP mode) between 1500 and 2000 UT on January 9 and 2000 and 2400 UT on January 10. At the time of these observations the array was in the B configuration which provided synthesized beamwidths of $\approx$ 5" at 20.7 cm and $\approx$ 22" at 91.6 cm. The data were sampled using a time resolution of 1.67 seconds with individual bandwidths of 12.5 MHz at 20.7 cm and 3.125 MHz at 91.6 cm. The full solar disk was observed simultaneously at these wavelengths for 40 minute periods followed by 5 minute observations of the calibration source PKS 1819-096.

In Figure 3 we show VLA 20 cm maps of a complex of active regions for which the RATAN detected a polarization inversion at shorter wavelengths. The map on January 9 is overlaid on a Kitt Peak magnetogram while the map for January 10 is compared with an optical image of the underlying sunspots made at the Kislovodsk Observatory. Soft X-ray images are also included in Figure 3; they are described in Section IV.

The radio emission stretches across regions of opposite magnetic polarity and interconnects a complex of five active regions, AR 6996, AR 6998, AR 7003, AR 7005, and AR 7006. These sources have peak brightness temperatures of $T_b = 1$ - 2 x 10^6 K and show no detectable circular polarization above $\rho_c \approx 15\%$.

On January 9, maps at 91 cm (Figure 4 - top) show two extended sources ($\theta \approx 2'$ x $4'$)

above the east and west solar limbs. These sources persist throughout the five hour period of observation. Plots of correlated flux on different VLA baselines indicate frequent, impulsive (Δt = few seconds) bursts throughout the period of observation, suggesting that the 91 cm sources represent regions of Type I noise storm activity. On January 10 the brightest 91 cm emission (Fig. 4 - bottom) is associated with two sources, A and B, that lie between active regions in opposite hemispheres. The limb-associated sources observed on January 9 were not detected on January 10 above a brightness temperature of $T_b \approx 2 \times 10^6$ K.

IV. YOHKOH SOFT X-RAY OBSERVATIONS

On January 9 and 10, the Soft X-ray Telescope (SXT) on board the Yohkoh spacecraft observed the Sun throughtout the periods of VLA observations. The SXT is a broad-band soft X-ray grazing incidence telescope which records its images with a 1024×1024 CCD (Tsuneta *et al.* 1991). The instrument has five thin-metal filters which can be selected to provide broad-band temperature diagnostics. On 1992 January 9 and 10, full disk images ($42' \times 42'$) were obtained about three or four times during each orbit. These full-frame images were obtained at half-resolution with 5.9″ pixels using the thin Al and composite (Al/Mg/Mn) filters. The SXT Aspect Telescope also provided white-light sunspot images that were used to co-align the VLA and soft X-ray images.

In Figure 5 we show SXT images for both days of observation. The dotted boxes on each map correspond to the field of view of the VLA 20 cm maps shown in Figure 3. Bright soft X-ray emission is detected above the complex of active regions AR 6996, AR 6998, AR 7003, AR 7005 and AR 7006, as well as fainter, large-scale structures that stretch across the solar equator. Because magnetic fields constrain the hot, dense X-ray emitting plasma, these SXT images delineate the magnetic structure of the corona.

Specifically, on January 9 there are two loops, L_1, and L_2, which respectively join active region AR 6996 to AR 6994 and AR 6998 to AR 7004 in the south. By January 10, these loops appear to have faded, but there are still faint soft X-ray structures (L_3), which appear to join the widely-spaced active regions in opposite hemispheres. The bright noise storm

source A (denoted by an X) lies near the western edge of the loop system L_1 which was observed on January 9.

V. THE INVERSION OF POLARIZATION

Coronal magnetic field strengths can be inferred from the inversion of the sign of circular polarization of the centimeter-wavelength radio emission. The inversion may be observed both as a function of wavelength and time (e.g. Gelfreikh et al.,1987). This effect has been attributed to the propagation of an electromagnetic wave through the quasi-transverse (QT) magnetic field region in the corona, where the direction of the longitudinal component of the field reverses.

The observed characteristics of such inversions (Gelfreikh, Peterova and Ryabov 1987) are in agreement with model simulations based on the theory (Zheleznyakov, 1970) of QT – propagation of polarized emission of a local source. According to this theory, the degree of circular polarization of the radio waves upon transversing the QT layer is:

$$\rho_c = 1 - 2\exp(-2\delta)$$

where

$$2\delta = 1.8x10^{-25} N_e B_t^3 |d\theta/dz|^{-1} \lambda_t^4$$

Here, λ is the wavelength of observation, in cm, N_e is the electron density in the QT region, in cm^{-3}, B_t is the strength of the transverse magnetic field in Gauss, and $d\theta/dz$ is the gradient of the magnetic field direction along the line of sight, in radians cm^{-1}. The condition for the change in the sign of polarization for wavelengths $\lambda \geq \lambda_t$ is $\rho_c = 0$. From these equations the magnetic field strength in the QT region in the corona above the active region is:

$$B_t = 1.2 \times 10^8 (N_e |dz/d\theta|)^{-1/3} \lambda_t^{-4/3}$$

In addition to measuring the critical wavelength λ_t, it is necessary to know the electron density, N_e, and the scale length of the magnetic field strength in the QT region. However, because B_t depends only 1/3 power of these parameters, their uncertainty does not crucially effect the estimate of B_t and representative values may be used in model calculations.

Because of the exponential dependence on σ in equation (1), the inversion of polarization depends strongly on the wavelength. One also expects that along the ray-path there may be more than one QT region. If this is the case, then the inversion of polarization may occur at more than one wavelength, λ_t, with each one corresponding to a different magnetic field strength. Here, for the first time, we discuss observations of multiple inversions using data taken by the Panoramic Spectrum Analyzer (PSA) of the RATAN-600.

VI. MODEL COMPUTATIONS

The Kitt Peak magnetogram for January 9 (Fig. 3) shows that the cluster of active regions in the northern hemisphere has a very complicated magnetic structure, with many areas of weak photospheric fields ($B \leq 400$ G) interspersed among the more intense sunspots. In order to examine the structure of the coronal fields and to determine the location of possible QT layers above the sunspots, the photospheric fields were extrapolated to coronal heights using the potential field code developed by Sakurai (1982).

In Figure 6 we show a portion of the extrapolation which covers the northern complex of active regions. The extrapolation shows field lines that join the leading and trailing sunspots in individual active regions as well as longer loops that join sunspots in different active regions. In particular, we note the group of field lines, F_1, which connect the leading, negative-polarity, spot in AR 6996 to the trailing, positive-polarity, spot in AR 7003.

The Sakurai code was also used to calculate the magnetic field strength at a series of constant heights ranging between h = 2500 to 20000 km. Because thermal gyroresonance radiation at a given wavelength, λ, occurs at a harmonic, s, of the gyrofrequency, the magnetic field strength $B = 10714/s\lambda$ G, these potential field extrapolations can be used to ascertain

where gyroresonance radiation at different harmonics and wavelengths is likely to occur. Our results indicate maximum magnetic field strengths of B = 610 to 2100 G at respective heights of h = 16000 km and 2500 km above the main sunspots.

In Figures 8a, 8b, and 8c we show computed radio of gyroresonance radiation at λ = 3.86, 4.73, and 6.5 cm. The RATAN sources a,b,c and d lie above or close to the main sunspots and have peak brightness temperature of $T_b = 0.8 \times 10^6$ to 2.0×10^6 K and angular sizes of $\theta \approx 30$" to 80". The increase in angular size with increasing wavelength reflects the divergent magnetic field structure above the sunspots, which causes the corona to be optically thick to gyroresonance radiation over a wider area at the longer wavelengths.

Guided by the size and positional information privided by these maps, we then computed the magnetic field components B_x, B_y and B_z along lines of sight through each of the RATAN sources *a, b, c* and *d* over a range of heights from h = 2500 km to 80000 km.

Our results show that near the locations of the RATAN sources *a*, *b*, and *c*, the Sakurai code yields QT layers at respective heights of $\approx 2.6 \times 10^{10}$ cm, 1.1×10^{10} cm, and 5.0×10^9 cm. The average total magnetic field strengths obtained from the code are $\approx$ 8, 21 and 50 G with respective angular gradients along the line of sight, $d\theta/dh$, equal to 4.2×10^{-11}, 9.7×10^{-11} and 2.8×10^{-10} rad cm^{-1}. Inserting these values in equation 3 and using the wavelengths at which the polarization first changes sign we calculate electron densities of $N_e \geq 5.1 \times 10^9$, 1.8×10^8 and 3.7×10^7 cm^{-3} for the QT layers associated with the sources *a*, *b*, and *c*. Source *d*, which did not exhibit any reversal of radio polarization, as expected, showed no evidence for a QT layer at any height.

The computed densities for sources b and c are consistent with hydrostatic equilibrium models of the corona, although a density of $N_e \geq 5 \times 10^9$ cm^{-3} for source a is inconsistent with such models. $N_e \geq 5 \times 10^9$ cm^{-3} inferred for source *a* is, however, at least an order of magnitude greater than that expected from such models. To explain the density in the QT layer above source *a* to be consistent with hydrostatic models, we require a higher nonpotential magnetic field strength of B $\approx$ 15 G at the same height of h $\approx 2.6 \times 10^{10}$ cm. Higher magnetic field strengths might, for example, be generated by currents, or nonpotential fields,

in the corona. It is interesting to note that source *a* coincides with one footpoint of an evolving trans-equatorial soft X-ray loop (L_1 - Fig. 7) as well as the magnetic structure that gives rise to noise storm source A on January 10.

Non-potential field configurations which give rise to multiple QT regions are also required to explain the double inversions of polarization observed for source *a* on January 9 and 10. Neglecting the dependence on N_e and $d\theta/dh$ in equation 3, we see that for two QT layers which give rise to polarization inversions at wavelengths λ_{t_1} and λ_{t_2}, the relevant magnetic field strengths are $B_{t_1}/B_{t_2} = (\lambda_{t_2}/\lambda_{t_1})^{4/3}$. For source *a* which shows polarization inversions at $\lambda_{t_1} = 4.73$ cm and $\lambda_{t_2} = 5.26$ cm, the magnetic field strength in the second QT layer would be $B_{t_2} \approx 0.87\ B_{t_1}$. Our potential field calculations along the line of sight to source *a*, however, reveal only a single QT region, again suggesting a more complicated coronal magnetic field configuration above this radio source.

VII. DISCUSSION

We have shown that noise storms are located on large-scale coronal loops that connect opposite hemispheres on the Sun; that X-ray images delineate their magnetic structure (Yohkoh, SXT) and that magnetic field strengths inferred from polarization inversions (RATAN 600) in some cases require non-potential or current-amplified magnetic fields.

Noise storms, for example, are thought to be generated by the plasma turbulence of energetic particles that are accelerated into evolving coronal magnetic fields (eg. Melrose 1980; Wentzel et al. 1986). The accelerated particles may be produced as a consequence of magnetic reconnection of emerging magnetic flux with pre-existing loops. An interaction between low-lying regions and evolving coronal magnetic fields has, in fact, been suggested by satellite data which show a correspondence between spatial and temporal variations of coronal loops, the emergence of magnetic flux and the onset of major Type I noise storms (Brueckner 1982; Stewart, Brueckner and Dere 1986). These apparent correlations might reflect the evolution of underlying fields and loops which, via reconnection, produce enough nonthermal particles to initiate noise storm activity. It is perhaps not surprising, then, that the RATAN source *a*,

rather than sources *b* or *c* might require nonpotential fields to explain the inversion of radio polarization.

Observations from the Skylab Mission near solar minimum have revealed the presence of large-scale magnetic loops that extend to heights of 1.5 x 10^{10} cm in the corona and were apparently stable for periods of hours to days while recent observations with the SXT provide evidence for global magnetic restructuring in the corona above active regions (Acton et al. 1992; McAllister et al. 1992; Tsuneta et al. 1992,1993). The distorted shapes of some X-ray emitting loops detected by the SXT also provide evidence for coronal currents, or nonpotential fields in the corona (Acton et al. 1992).

The VLA has also delineated large-scale coronal loops that can connect widely-separated active regions (Lang and Willson 1989) as well as higher-lying loops that are the sources of Type I noise storms (Willson, Lang and Liggett 1990). Combined VLA and RATAN observations have detected long-lasting nonthermal radio sources that are associated with complex magnetic regions (Lang et al. 1993). These nonthermal sources may be attributed to electric currents in the corona, and such currents produce their own magnetic fields that distort the shape of X-ray images from relatively simple ones. Such distorted shapes have already been detected in SXT images and comparisons of VLA, RATAN and SXT data can provide new perspectives on these regions including estimates of their physical parameters.

REFERENCES

Acton, L. et al. :1992, *Science*, **258**, 618.

Alissandrakis, C.E., and Kundu, M.R.: 1982, *Ap. J. (Letters)*, **253**, L49.

Alissandrakis, C.E., and Kundu, M.R.: 1984, *Astron. Ap.*, **139**, 271.

Bogod, V.M., Gelfreikh, G.B., Willson, R.F., Lang, K.R., Shatilov, V., and Tsvetkov, S.V.: 1992, *Solar. Phys.*, **141**, 303.

Bogod, V.M., Vatrushin, S.M., Abramov-Maximov, V.E., Tsvetkov, S.V., and Dikij, V.M.

Proceedings of IAU Colloq. 141, in press 1993.

Borovik, V.N., Kurbanov, M.Sh., Livshits, M.A., Ryabov, B.I.: 1990, *Astron. Zh.*, **67**, No. 3., 1038.

Brueckner, G.E.: 1982, *Proceedings of the 4th CESRA Workshop on Solar Noise Storms*, p255.

Gelfreikh, G.B., Abramov-Maximov, V.E., Akhmedov, Sh. B., Bogod, V.M.: 1985, in *Solar Maximum Analysis*, V.E. Stepanov, V.N. Obridko and G. Ya. Smolkov, eds.

Gelfreikh, G.B., Peterova, N.G., and Ryabov, B.I.: 1987, *Solar Phys.*, **108**, 89.

Khaikin, S.E., Kaidavoskii, N.L., Parijskij, Yu.N., and Esepkina, N.A.: 1972, *Isvestia, Main Astronomical Obs. Pulkova*, **No. 188**, 3.

Kundu, M.R., Alissandrakis, C.E., Bregman, J.D., and Hin, A.C.: 1977, *Ap. J.*, **213**, 278.

Lang, K.R., and Willson, R.F.: 1982, *Ap. J. (Letters)*, **255**, L111.

Lang, K.R., Willson, R.F., and Gaizauskas, V.: 1983, *Ap. J.*, **267**, 455.

Lang, K.R., and Willson, R.F.: 1989, *Ap. J. (Letters)*, **344**, L77.

Lang, K.R., *et al.*, 1993, *submitted to Ap. J.*.

McAllister, A., Uchida, Y., Tsuneta, S., Strong, K.T., Acton, L.W., Hiei, E., Bruner, M.E., Watanabe, T., and Shibata, K.:1992, *Pub. Astron. Soc. Japan*, **44**, L205.

Melrose, D.B.: 1980, *Solar Phys.*, **67**, 357.

Parijskij, Yu. *et al.*, 1976, *Soviet Astron*, **20**, 577.

Sakurai, T.: 1982, *Solar Phys.*, **76**, 301.

Sakurai, T., and Uchida, Y.: 1977, *Solar Phys.*, **52**, 397.

Stewart, R.T., Brueckner, G.E., and Dere, K.P.: 1986, *Solar Phys.*, **106**, 107.

Tsuneta, S., et. al.: 1991, *Solar. Phys.*, **136**, 37.

Tsuneta, S., Takahashi, T., Acton, L.W., Bruner, M.E., Harvey, K.L, and Ogawara, Y.: 1992, *Pub. Astron. Soc. Japan*, **44**, L211.

Webb, D.F., Holman, G.D., Davis, J.M., Kundu, M.R., and Shevgaonkar, R.K.: 1987, *Ap. J.*, **298**, 911.

Wentzel, D.G.: 1986, *Solar Phys.* , **103**, 141.

Willson, R.F., Lang, K.R., and Liggett, M.: 1990, *Ap. J.*, **350**, 856.

Zheleznyakov, V.V.: 1970, *Radio Emission of the Sun and Planets*, Pergamon Press, New York.

Zheleznyakov, V.V., and Zlotnik, E.Ya.: 1963, *Astron Zhurn.*, **40**, 633.

FIGURE LEGENDS

Fig. 1. One-dimensional fan beam scans of total intensity, I, and Stokes parameter, V, made with the RATAN PAS on 1992 January 9. The optical map of sunspots and their magnetic fields, obtained from the bulletin *Solnechnye Dannye*, are also presented. The peak umbral fields of the sunspots are expressed in units of hundreds of Gauss, and N and S denote positive and negative magnetic polarity, respectively.

Fig. 2. One-dimensional fan beam scans of total intensity, I, and Stokes parameter, V, made with the RATAN PSA on 1992 January 10. The optical map of sunspots and their magnetic fields are also presented.

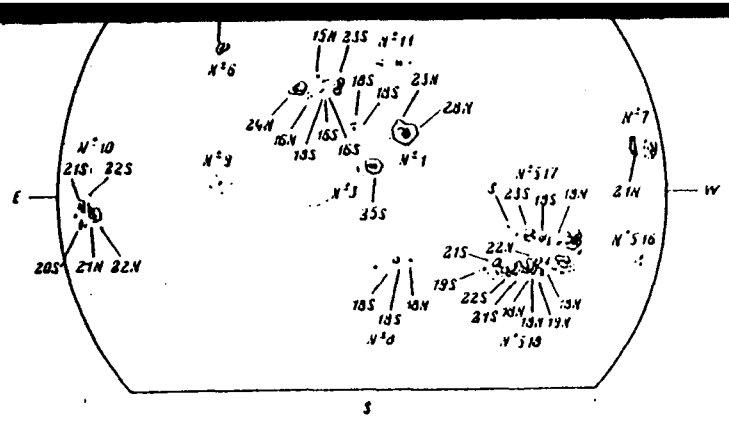
Fig. 3. VLA synthesis maps of total intensity at 20.7 cm on January 9 (left) and 10 (right) are compared with a Kitt Peak magnetogram (January 9) and an optical map showing the underlying sunspots (January 10). The bottom panels show grey-scale soft X-ray images obtained from the Soft X-ray Telescope (SXT) on board the Yohkoh satellite. The radio contours mark levels of equal brightness temperature, T_b, with an outermost contour and contour interval equal to 4.2 x 10^5 K. The light and dark areas on the Kitt Peak magnetogram correspond to positive and negative fields, respectively. The angular scale may be determined from the 100" spacing between the fiducial marks on the axes.

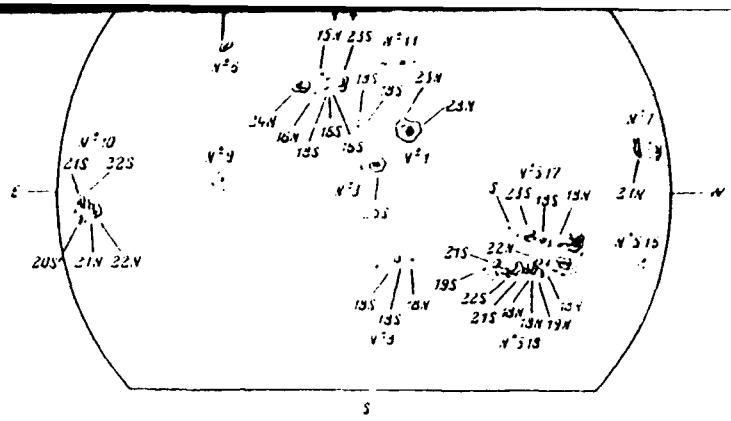
Fig. 4. VLA 2-minute snapshot maps of total intensity at 91.6 cm wavelength on January 9 (top) and January 10 (bottom) are compared with optical maps (Kislovodsk Observatory) of sunspots and their magnetic fields. The VLA snapshot maps on January 9 were made at 17:20 UT and show two noise storm sources above the east and west solar limbs. On

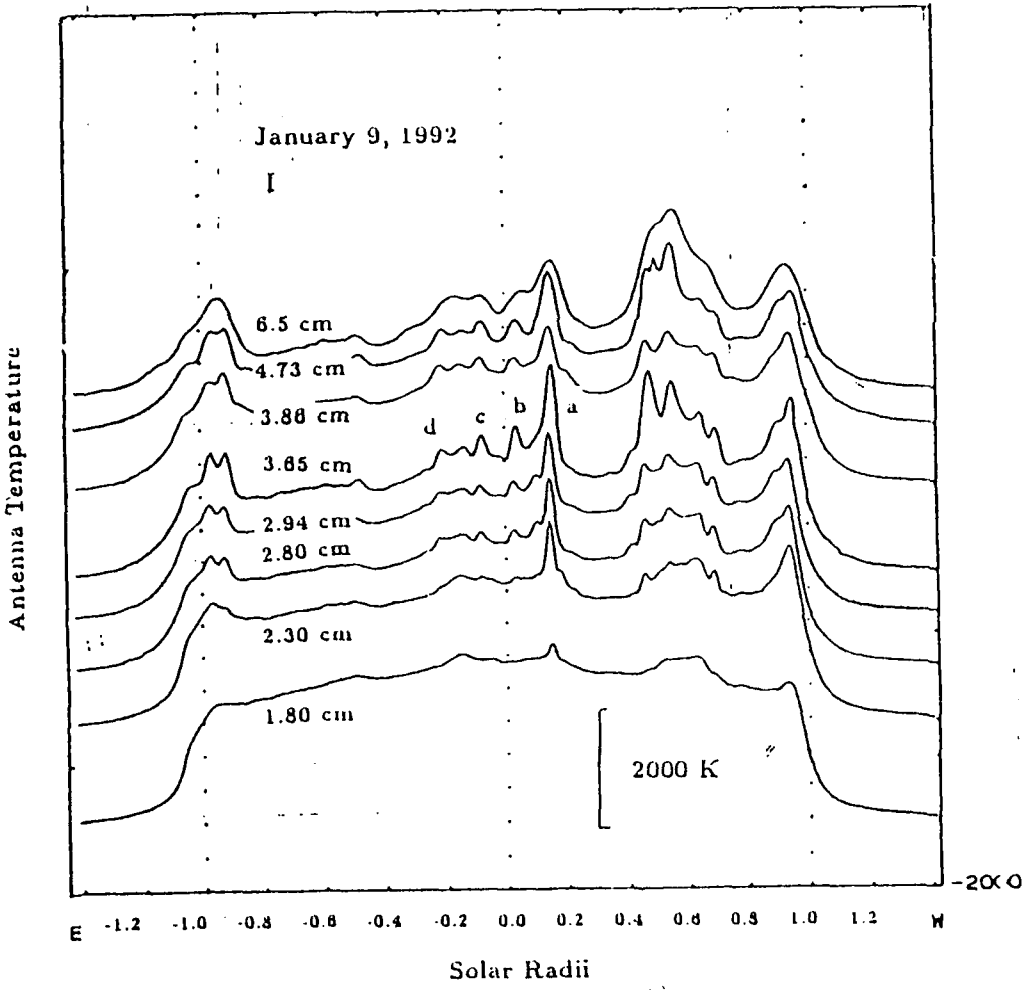
January 10, a map at 23:06 UT shows a bright noise storm source, A, between active regions AR 6996 and AR 6994 as well as a fainter noise storm source, B, to the south of AR 7004. The contours on the radio maps mark levels of equal brightness temperature, T_b, with an outermost contour equal to 3.7 x 10^6 K on January 9 and 3.1 x 10^7 K on January 10.

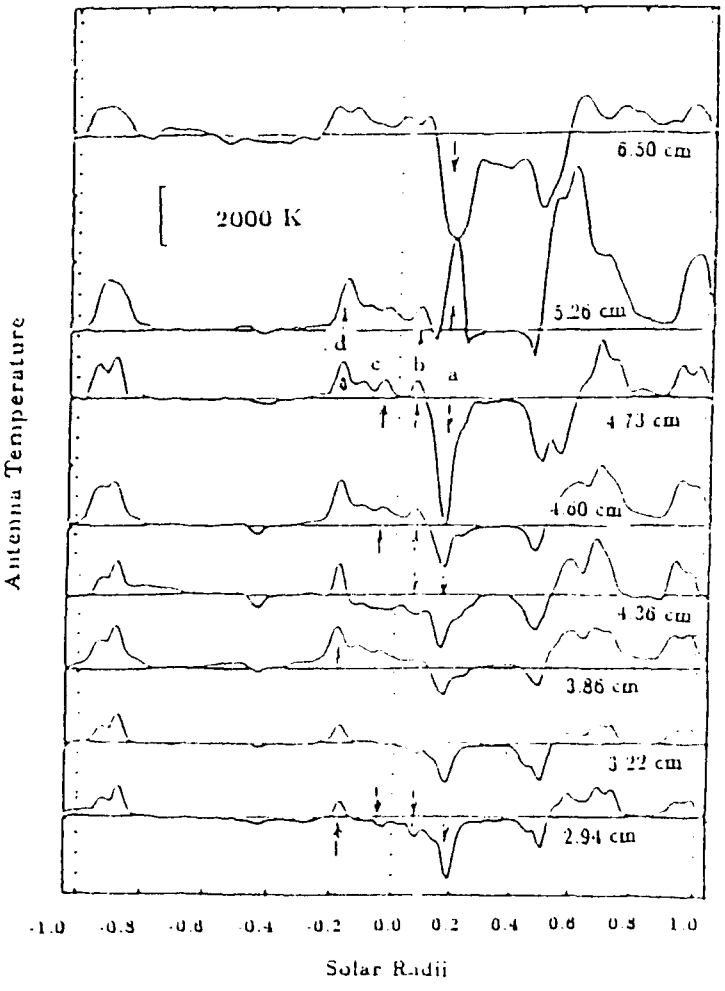
Fig. 5 Full-disk soft X-ray images taken by the SXT at the times indicated. The images were obtained at half-resolution with 5.9" pixels using the thin Al filter and an exposure time of 0.69 seconds. The boxes surrounding active regions AR 6996, AR 6998, AR 7003, AR 7005 and AR 7005 correspond to the fields of view shown in Figure 3. The map on January 9 reveals large-scale loops, L_1 and L_2, which cross the solar equator and connect active regions in opposite hemispheres. On January 10, these loops seem to have been replaced by fainter structures (L_3) which also interconnect widely-spaced active regions.

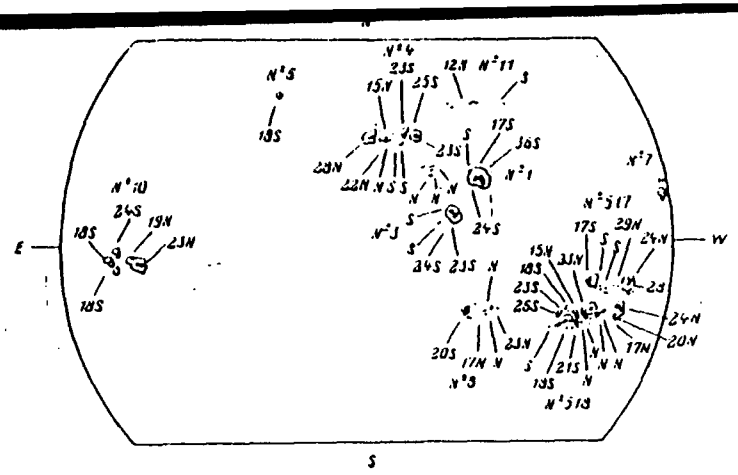
Fig. 6 Photospheric longitudinal magnetogram taken at the Kitt Peak Observatory at 15:33:32 UT on 1992 January 9 with superposed magnetic field lines calculated from the potential field code of Sakurai. The contour levels are drawn at ±400, ±800 ... G where the dashed contours represent negative, or inward-directed fields, and the solid contours represent positive, or outward-directed fields. The fiducial marks are drawn at intervals of 30". (b-d) Predicted maps of thermal gyroresonance radiation at λ = 3.86, 4.73 and 6.50 cm. These maps were generated using the potential field extrapolations and a constant conductive flux of $F_c = 2 \times 10^6$ erg cm^{-3} s^{-1}. The contours mark levels of equal brightness temperature, with an outermost contour and contour interval equal to $T_b = 2 \times 10^5$ K. The labels *a*, *b*, *c* and *d* correspond to the sunspot-associated sources observed by the RATAN (see Figures 1 and 2). Sources *a* and *b* are associated with the leading sunspots in active regions AR 6996 and AR 6998, and sources *c* and *d* are associated with the leading and trailing sunspots in AR 7003.

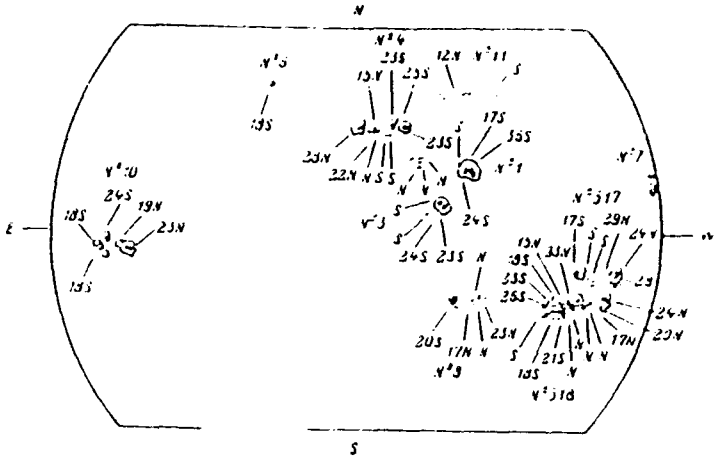


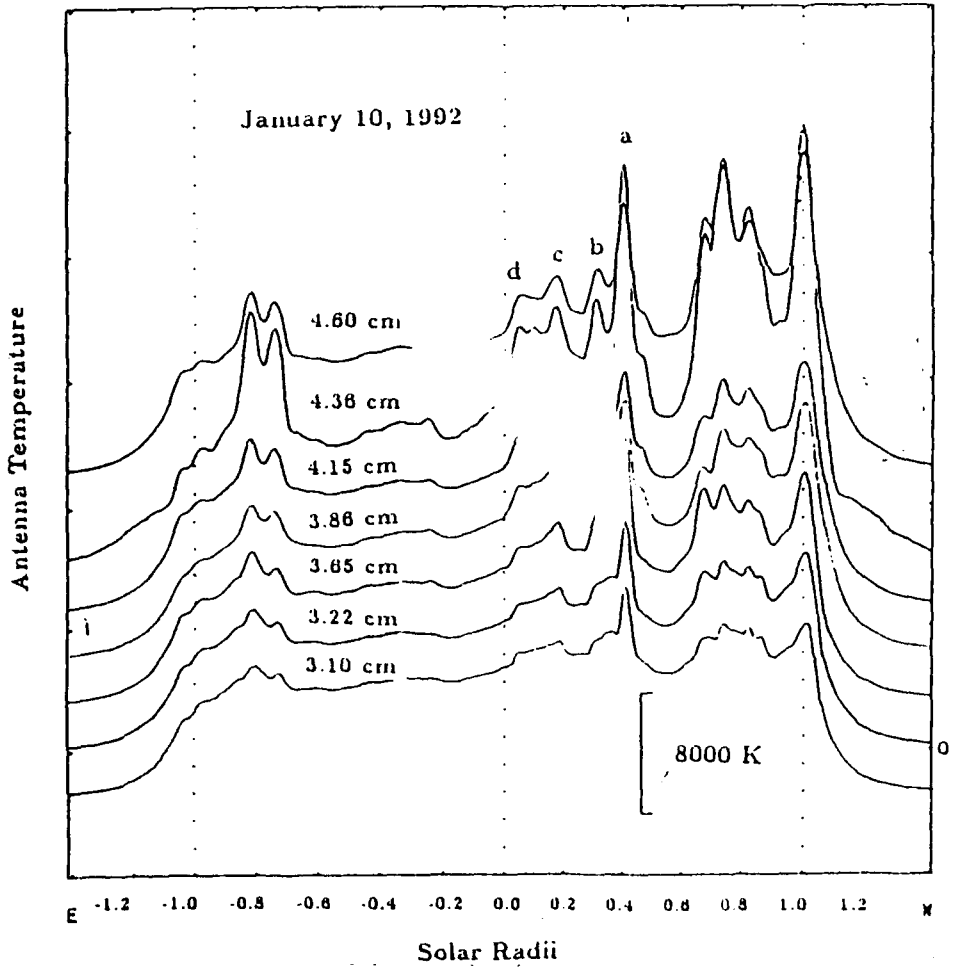


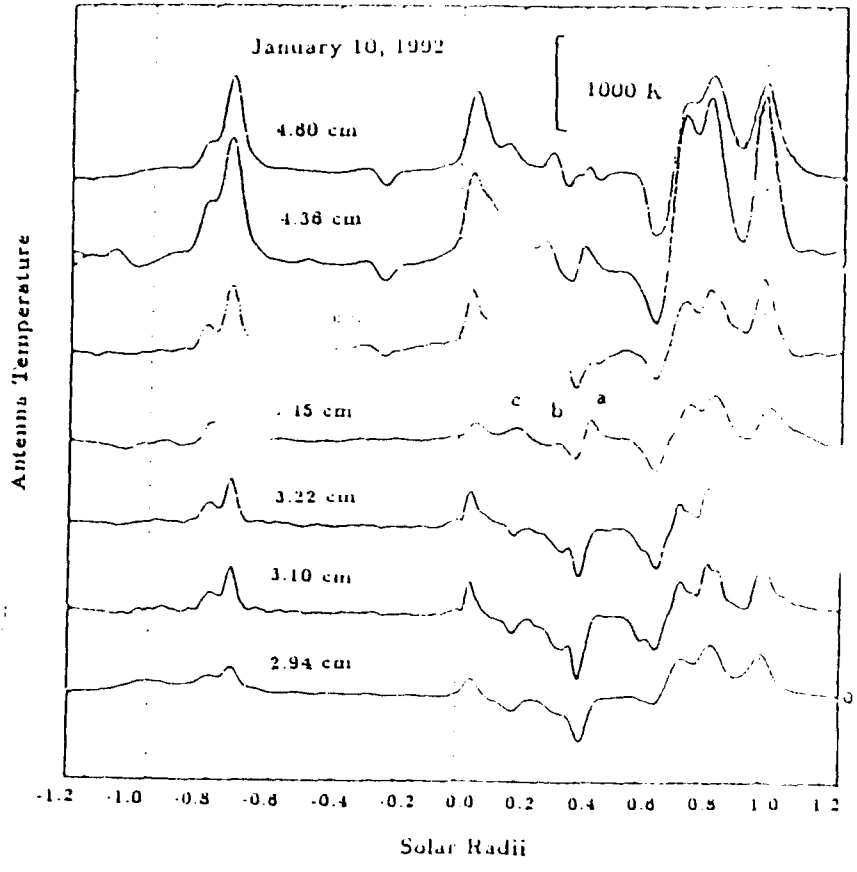

January 9, 1992
I
6.5 cm
4.73 cm
3.88 cm
3.65 cm
2.94 cm
2.80 cm
2.30 cm
1.80 cm
d
c
b
a
2000 K
-2000
E
-1.2 -1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0 1.2
W
Solar Radii
Antenna Temperature

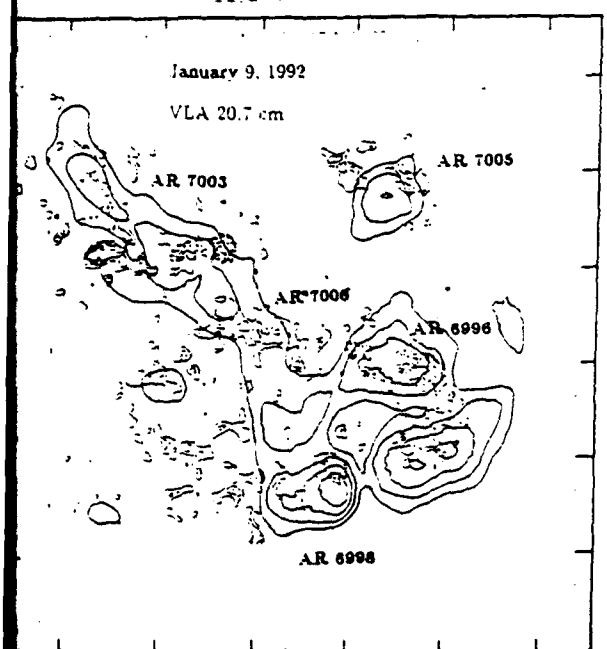

6.50 cm
2000 K
5.26 cm
d
c
b
a
4.73 cm
4.60 cm
4.36 cm
3.86 cm
3.22 cm
2.94 cm
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Solar Radii
Antenna Temperature

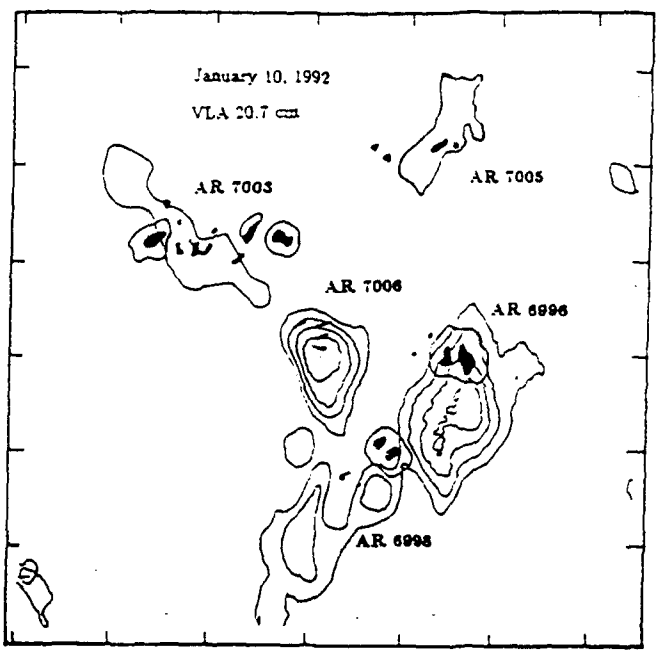


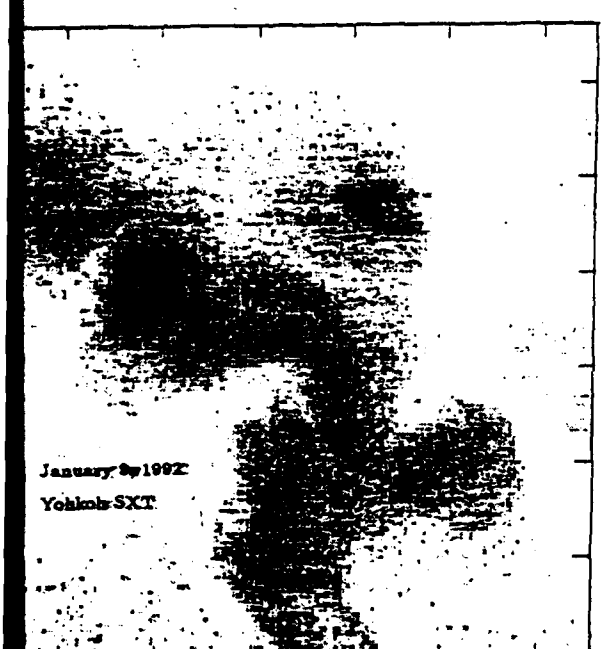


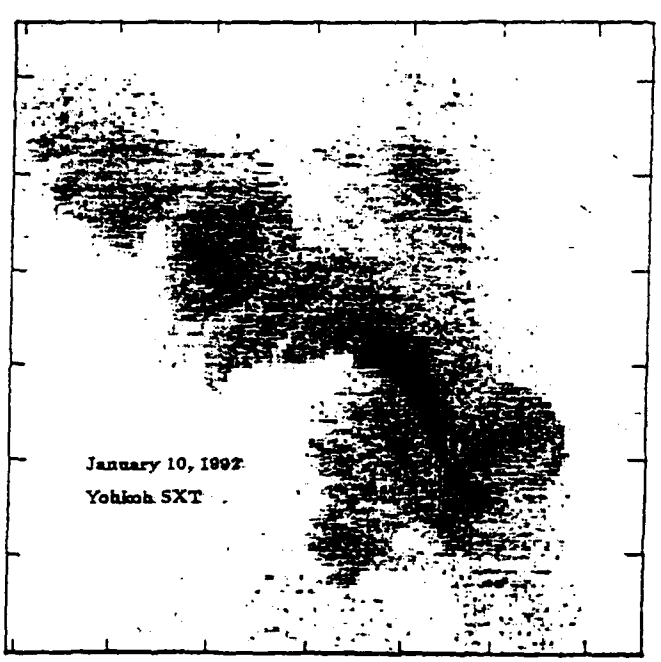

January 10, 1992
4.60 cm
4.36 cm
4.15 cm
3.86 cm
3.65 cm
3.22 cm
3.10 cm
a
b
c
d
8000 K
Antenna Temperature
E
W
-1.2 -1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0 1.2
Solar Radii

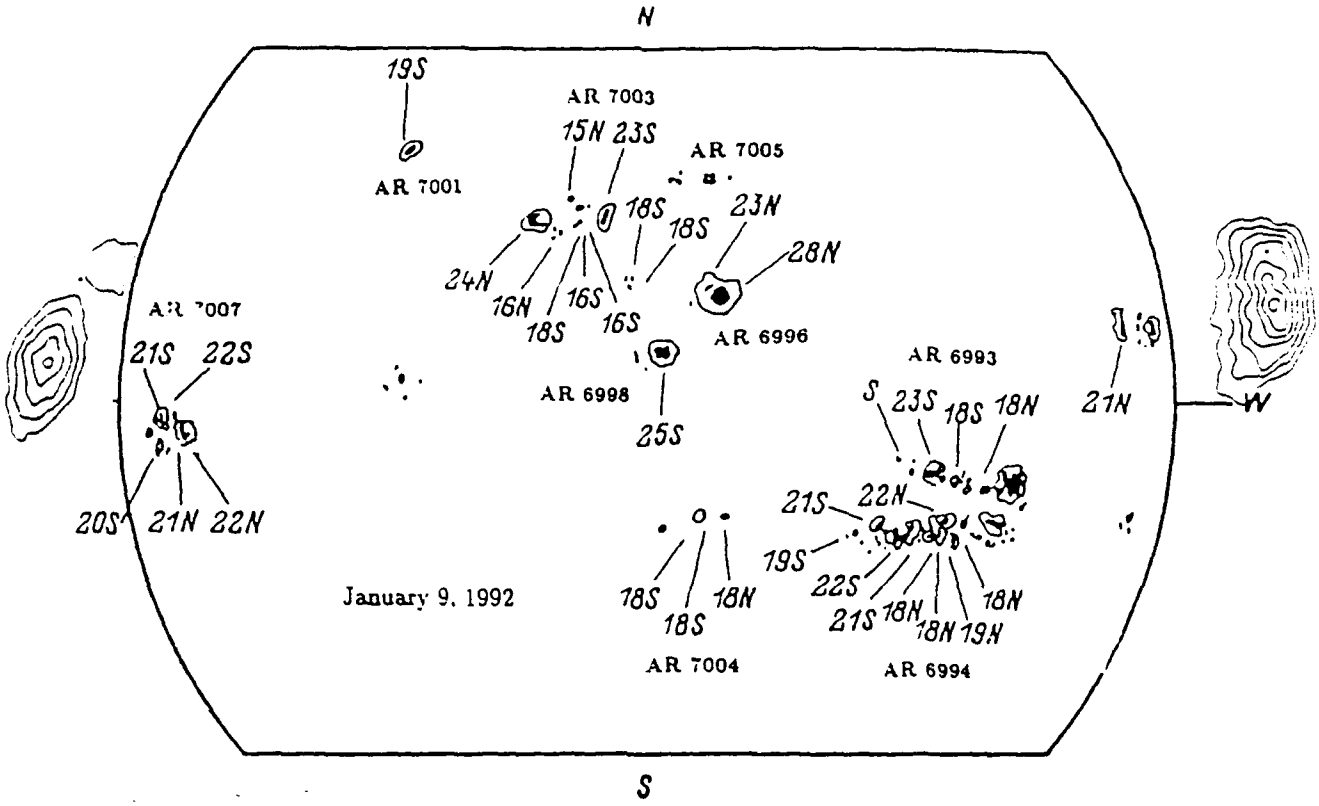

January 10, 1992
1000 K
4.80 cm
4.36 cm
4.15 cm
3.22 cm
3.10 cm
2.94 cm
a
b
c
Antenna Temperature
-1.2 -1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0 1.2
Solar Radii

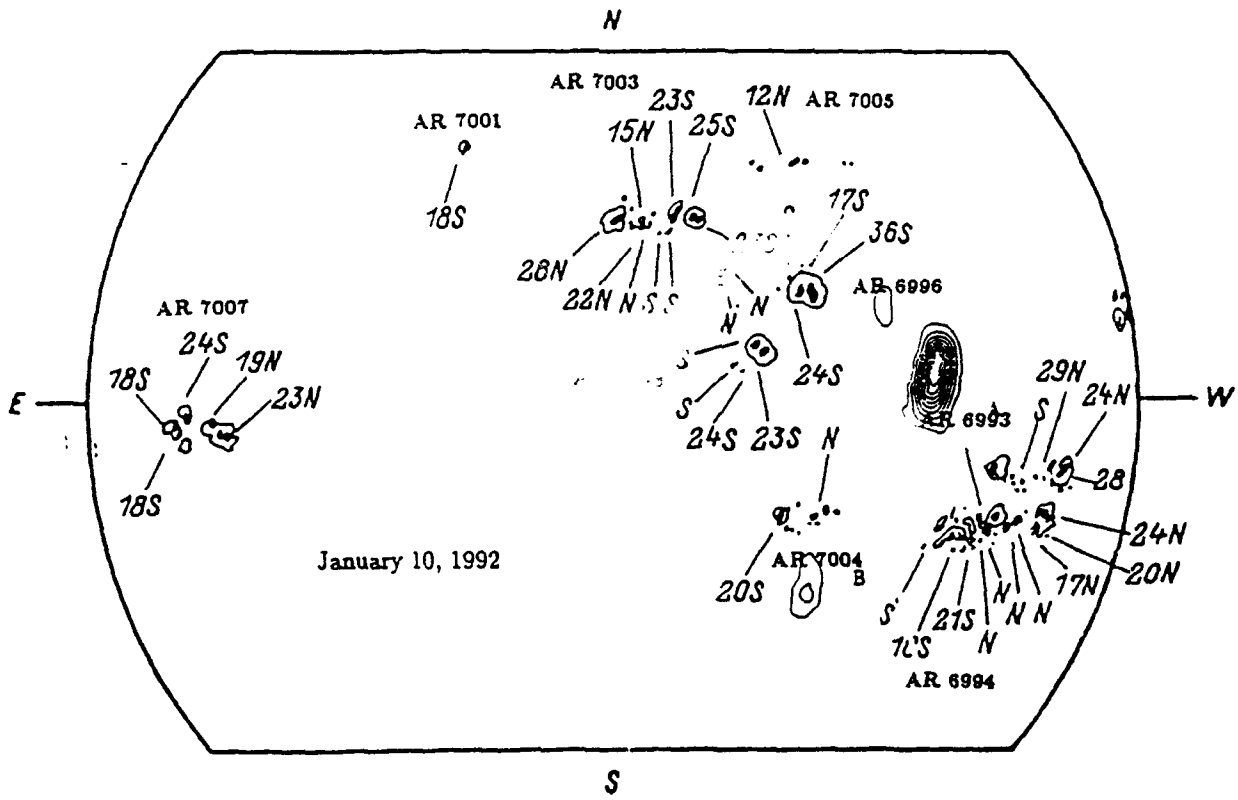

January 9, 1992
VLA 20.7 cm
AR 7003
AR 7005
AR 7006
AR 6996
AR 6998

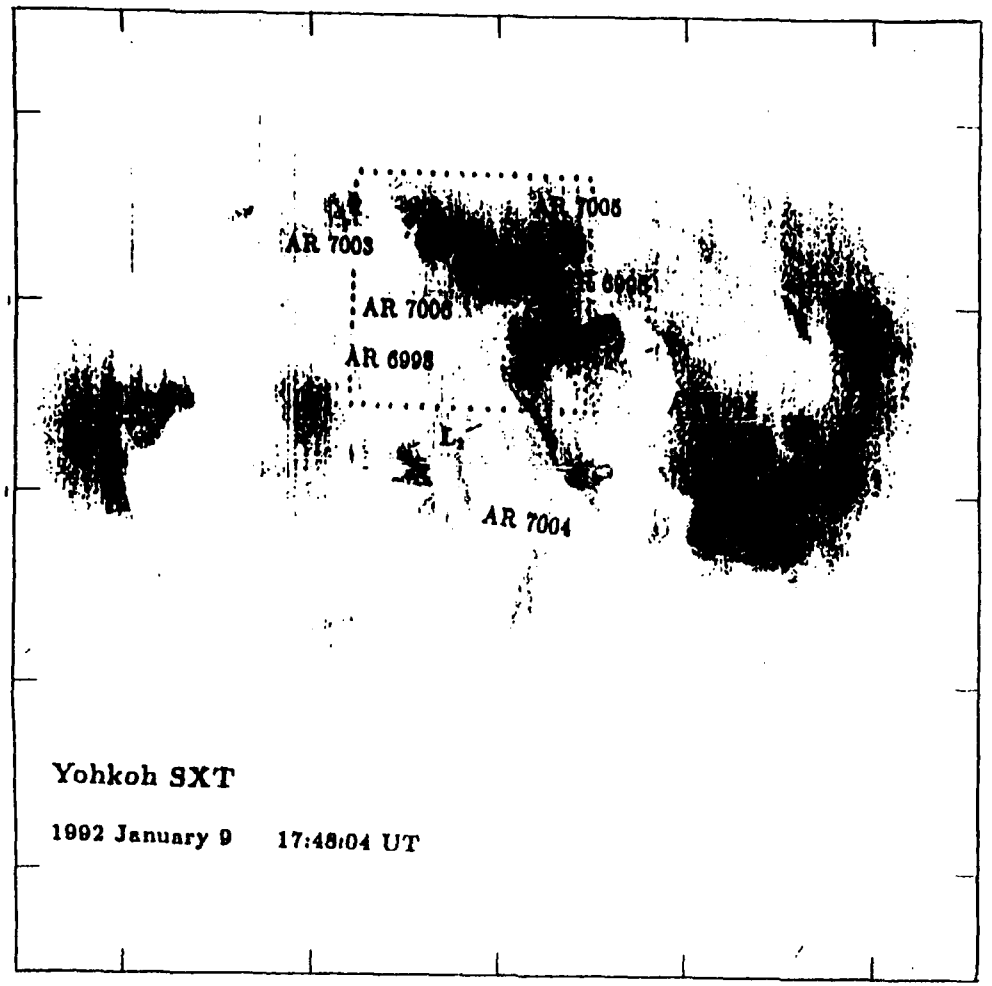

January 10, 1992
VLA 20.7 cm
AR 7003
AR 7005
AR 7006
AR 6996
AR 6998

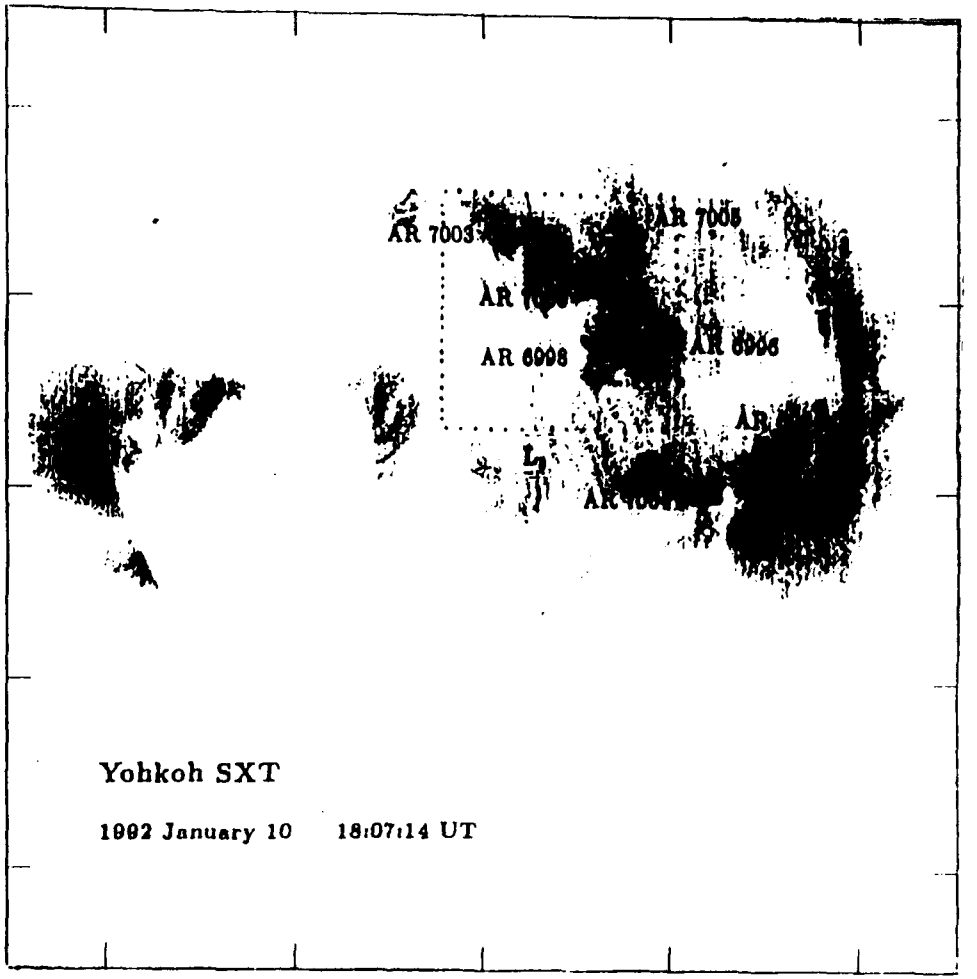

January 9, 1992
Yohkoh SXT

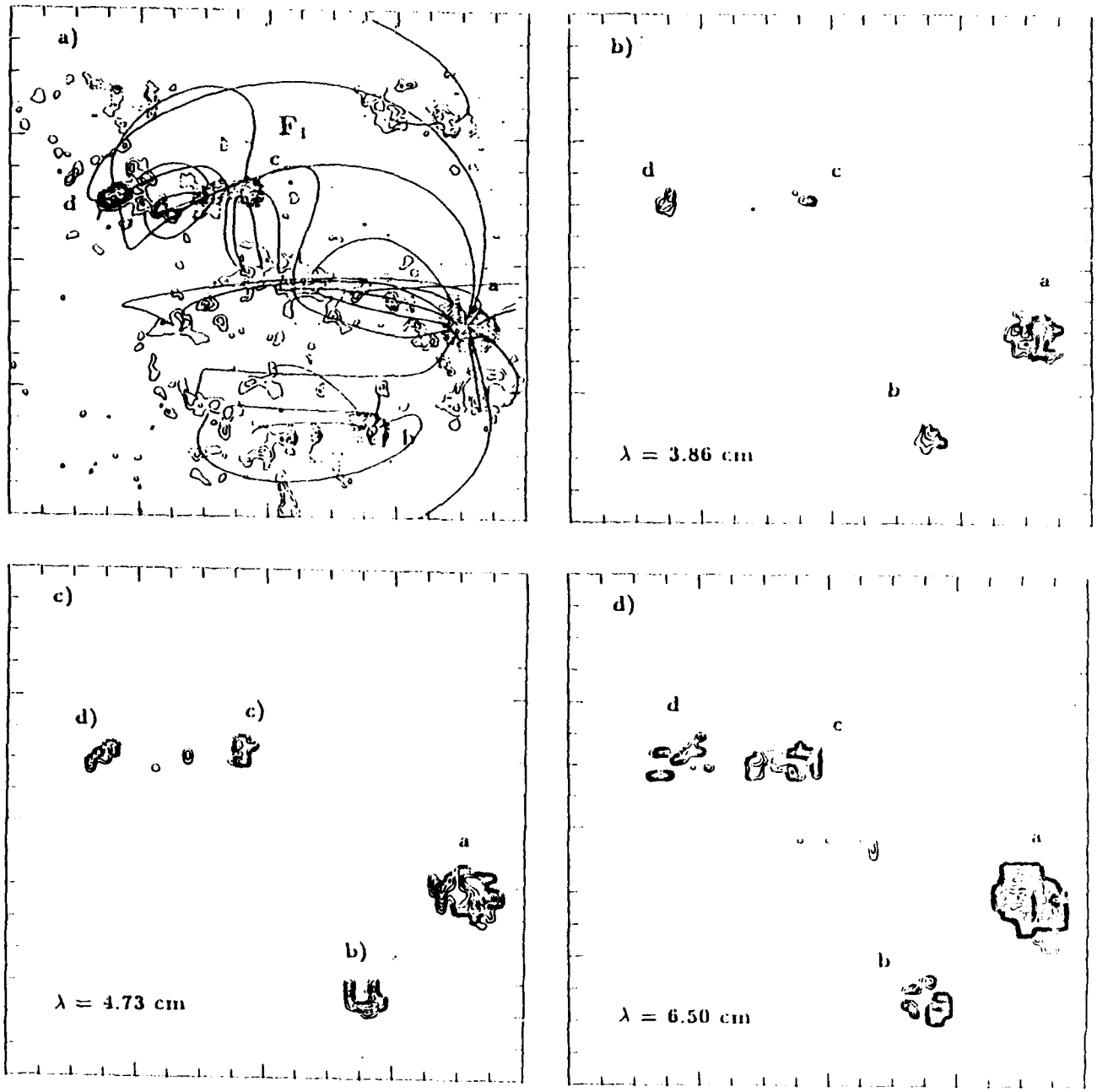

January 10, 1992
Yohkoh SXT


N
19S
AR 7003
15N 23S
AR 7005
AR 7001
18S
18S
23N
28N
24N
16N
16S
18S
16S
AR 6996
AR 7007
21S 22S
AR 6998
25S
AR 6993
S 23S 18S 18N
21N
W
20S 21N 22N
21S
22N
19S
22S
21S 18N 18N 19N
18N
January 9, 1992
18S
18S
18N
AR 7004
AR 6994
S


N
AR 7003
23S
12N
AR 7005
AR 7001
15N
25S
18S
17S
36S
28N
22N
AR 6996
AR 7007
24S
19N
18S
23N
E
24S
S
24S 23S
N
29N
24N
W
AR 6993
S
28
18S
24N
20N
January 10, 1992
AR 7004
B
20S
17N
S
21S
N
N N
N
AR 6994
S


AR 7003
AR 7005
AR 7006
AR 6998
AR 7004
Yohkoh SXT
1992 January 9 17:48:04 UT


AR 7003
AR 7005
AR 6998
AR 6996
Yohkoh SXT
1992 January 10 18:07:14 UT


a)
F_1
b
c
d
D_G
a
b)
d
c
a
b
λ = 3.86 cm
c)
d)
c)
a
b)
λ = 4.73 cm
d)
d
c
a
b
λ = 6.50 cm

THE EFFECTS OF ANISOPLANATISM ON SHEAR COHERENT INTERFEROMETRIC PHOTOGRAPHY

Richard Anderson
Professor Emeritus
Department of Physics

University of Missouri-Rolla
Rolla, MO 65401

Final Report for:
Summer Faculty Research Program
Phillips Laboratory

Sponsored by:
Air Force Office of Sceintific Research
Bolling Air Force Base, Washington, D.C.

September 1993

THE EFFECTS OF ANISOPLANATISM ON SHEAR COHERENT INTERFEROMETRIC PHOTOGRPHY

Richard Anderson
Professor Emeritus
Department of Physics
University of Missouri-Rolla

Astract

In order to image a spaced based object with a telescope and recording the heterodyne Fourier image on a detector array it is necessary that the atmospheric phase distortion is removed in the three mixed heterodyne signals corresponding to the three angularly displaced beams. Two beams make small angles θ_x and θ_y to a centered beam. The Fourier images are produced by parallel bundles of scattered waves. The phase mutual coherence function is calculated for parallel bundles and the accuracy of the calculation depends on knowing C_N^2 versus altitude for the specilic atmospheric conditions at the time of the measurement.

THE EFFECTS OF ANISOPLANATISM ON SHEAR COHERENT INTERFEROMETRIC PHOTOGRAPHY

Richard Anderson

INTRODUCTION

The technique of imaging under consideration was called AC shearing interferometry by Idell[1] and Webster[2] who theoretically analyzed the problem. The concept has been renamed shear coherent interferometric photography (SCIP). Two papers[3,4] were given at the recent SPIE meeting one discussing the technique and the other presenting simulated and laboratory experimental data and laboratory images.

In these experiments a single laser source is used of frequency ν_o'. This beam is sent through a acousto-optic modulator which shifts the frequency by a small amount to ν_o. This beam passes through another modulator and is shifted to $\nu_o + \nu_x$ and finally the undiffracted component from this modulator is shifted by a third modulator to frequency $\nu_o + \nu_y$. Three coherent frequency shifted beams are produced. The beam of frequency ν_o is called the undisplaced central beam, the beam of frequency $\nu_o + \nu_x$ is displaced a small amount Δx in the x-direction, and the beam of frequency $\nu_o + \nu_y$ is displaced a small distance distance Δy in the y-direction with respect to the central beam. These beams flood illuminate the object and the displaced beams make small angles to the central beam at the object.

Hutchins[3] indicated that the uplink beam could be considered to be diffractive spread under certain beam widths. Tavis and Yura[5] have theoretically shown the transmitted beam will be diffractively

spread for beam widths ranging from 0.1 to 2.5 times ρ_o, where ρ_o is the atmospheric lateral coherence length.[6,7] The beam widths used to flood illuminate an object would exceed the maximum value given above so the condition of diffractive beam spreading should be questioned.

Short-term exposure correponds to a picture accumulated before the turbulence profile has changed appreciably, so the turbulence may be considered stationary during the exposure (10 ms). If diffractive spread is assumed the major beam spreading mechanism and if the object is in the central diffractive lobes of the three beams, the three incident wavefront phases are initially nearly the same. The three transmitted beams will have different amplitudes at the object for the intensity of the beams from the modulators are not perfectly matched. Hutchins[3] indicated that intensity match is not as important as phase match for any phase miss match will become a part of the object produced phase change.

It is assumed the distant object is an ideal randomly diffuse scatterer, exhibiting no specular or retroreflection. The object is a Lambertian scatterer which can be considered to be a circulo-complex Gaussian function. In reality the is no such thing as circulo-complex Gaussian scatterer. In the infrared even the Labsphere diffuse gold standard is a poor standard. In this spectral region a flooded (dusted) flowers of sulfur secondary standard is a more nearly perfect randomly diffuse scatterer. This secondary standard is to fragile to make a good field target to calibrate lidars. Several reflective target materials used to calibrate coherent heterodyne aerosol backscatter lidars were studied at the CO_2 10.6 μm line by Haner and Menzies.[8]

The three coherent laser beams incident on the object are effectively at the same wavelength and in the discussion to folllow they will be considered plane wavefronts which are angularly displaced at the object by θ_x and θ_y with respect to the undisplaced beam. At a single point in the object these beams are scattered with respect to the undisplaced beam by the angles $-\theta_x$ and $-\theta_y$. The three beams flood illuminate the object so there will be a different set of points in the object for each beam where the scattered rays have the same direction of reflection and form parallel scattered bundles.

At this point the experiment must be described in more detail. The object is a space based and is a large distance (effectively infinity) from the collecting telescope. The telescope consists of a collecting aperture which focusses the collected backscatter radiation in its focal plane. A CCD detector array is placed in this plane and the far-field diffractive condition is nearly fulfilled. This image in the pupil plane of the CCD detector is the Fourier image of the object. Since the three beams were initially tilted with respect to each other and each object point scatters the three tilted beams, three displaced Fourier images are superimposed in the detector's pupil plane. One image is that from the central beam and the other images are from the displaced beams and are displaced in the x and y directions.

The rays which are focussed at a given point in the resultant Fourier image have been propagated in the same direction from the object back to the objective lens of the telescope and are parallel rays. There will be a set of object points corresponding to frequency ν_o that are scattered at a certain angle back to the telescope, another group of object points corresponding to frequency

$\nu_o + \nu_x$ scattered in the same direction, and similarly another group of object points corresponding to frequency $\nu_o + \nu_y$ scattered in the same direction. Thus, for each frequency a different set of object points will be scattered in a given direction and will produce the image at a certain point in the image plane. As a result, at a given point $\bar{u}$ in the image plane of coordinates u_x and u_y the fields of all three frequencies are superimposed and the measured intensity has a DC component and six mixed heterodyne components of frequencies $\pm\omega_x$, $\pm\omega_y$, and $\pm(\omega_x - \omega_y)$.

The short-term exposure required to recover these time dependent signals will depend on the period of time over which the atmospheric turbulence is considered stationary and object motion is very small so it may be considered at rest. For atmospheric measurements this time is approximately 10 ms. During each exposure the object must be considered stationary but since multiple exposures must be averaged, ths object must move between the exposures so each represents a different object speckle representation. The total motion of the object during the multiple exposures will still be small. In order to resolve the time-dependent mixed signals the detector must be scanned several times during each exposure. The total exposure time, object transver speed, and the camera scan speed determine the size of the of freqencies ω_x. ω_y, and $(\omega_x - \omega_y)$. In order to average out th fects of turbulence and the object speckle multiple short- exposures are taken and averaged to recover an approximate incoherent object image in the object plane.

Some experimental parameters will be considered for the experiment. A schematic of the apparatus is shown in figure 1 for a satellite observed through the atmosphere by the telescope. It is

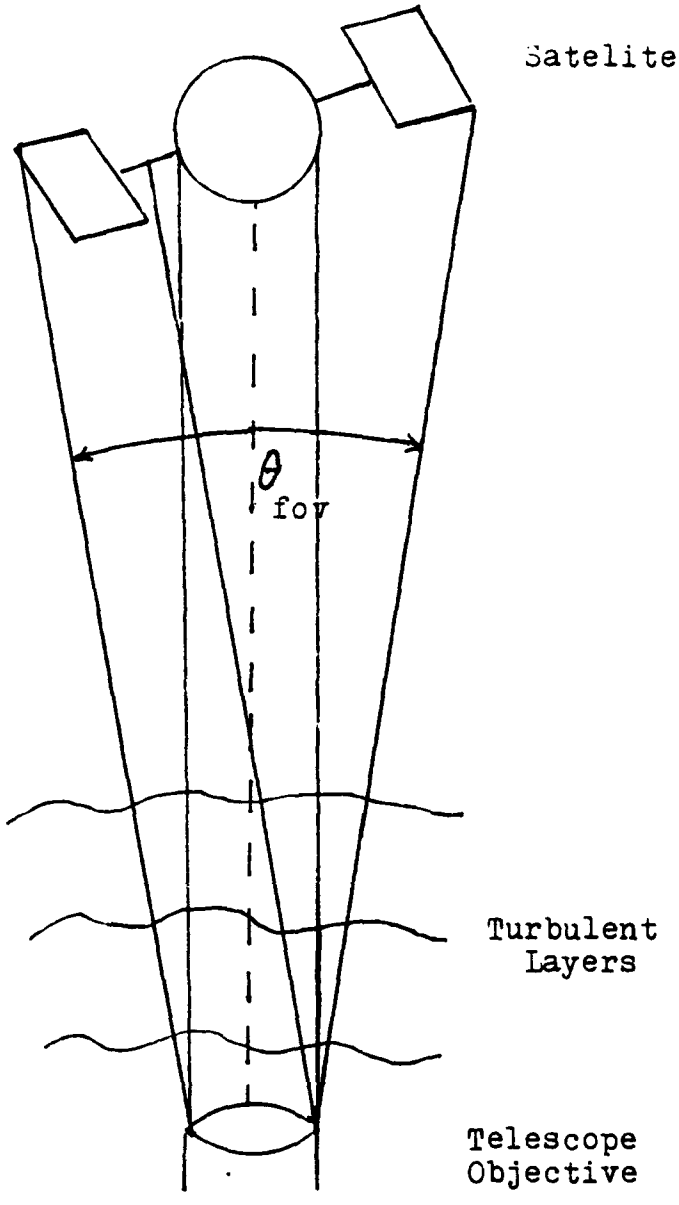


Figure 1. Parallel ray bundles from the object to telescope objective.

assumed that the satelite is tracked by the telescope. Two different orbits are considered, one is a nonsynchronous orbit of 4.8×10^2 km above the earth and the other is a synchronous orbit of 3.6×10^3 km above the earth. It will be assumed that the satelite is 10 m across and the telescope aperture is 1 m. From these values it is seen that bundles of rays parallel to the line-of-sight of the telescope are collected. In the case of the nonsynchronous orbit the field-of-view angle is $\theta_{fov} = 4.26$ arcsec and for the synchronous orbit the field-of-view angle is $\theta_{fov} = 0.57$ arcsec.

THEORY

The discusion to be presented is a variation of an analysis presented in a paper in the Physics of Optics[10] The phase variation noted in a signal traversing a turbulent atmosphere is caused by random fluctuations in the temperature of the atmosphere which produce fluctuation in the index of refraction of the atmosphere. These index fluctuations are a function of position $\bar{r}$ in the atmosphere and time

$$n(\bar{r},t) = 1 + n_1(\bar{r},t) \qquad (1)$$

where $n_i(\bar{r},t)$ is the fluctuation of the index of refraction. For clear-air turbulence $|n_1| \ll 1$ and its time ensemble average is $\langle n_1 \rangle = 0$ where the bracket indicates the ensemble average. The temporal behavior of n_1 is caused by a net transport of the inhomogeneities of n_1 by local winds so $n_1(\bar{r},t) = n_1(\bar{r} \pm \bar{v}(\bar{r})t)$. In the short-term exposures n_1 is considered constant and only a function of $\bar{r}$.

It is known that the amount of turbulence varies with elevation

in the atmosphere. There have been several field measurements of this variation.[11-17] As a result, consider the effects of a thin layer of the atmosphere at a vertical altitude h, the layer lies between h and $h+\Delta h$. The layer thickness is chosen large compared to the correlation scale of the atmospheric inhomogenoities (turbulent cell size) but small enough so diffractive effects are negligible. Consider only the effects at altitude h in the atmosphere and introduce a phase change in the backscattered field by the layer of thickness Δh. The backscattered field's amplitude is taken as unity, so phase change is

$$\psi_h(\bar{r}) = \exp[j\phi(\bar{r})] \tag{2}$$

where $\phi(\bar{r})$ is the phase shift produced at altitude h. Since all collected bundles are effectively traveling along the line-of-sight, there small inclination of 2.13 or 0.28 arcsec are neglected. The situation of a satelite overhead is consdered first and the total phase change of the field in a layer between h and $h+\Delta h$ is

$$\phi_h(\bar{r}) = k\int_p^{p+\Delta p} dp\ n(\bar{r},h) = k\int_p^{p+\Delta p} dh\ \mathrm{Cos}^{-1}\theta_f/2\ n(\bar{r},h)$$

$$\simeq k\int_h^{h+\Delta h} dh\ n(\bar{r},h) \tag{3}$$

where the viewing direction is along the slant path p inclined at an angle lying in the interval $\theta_f/2 \le \theta_{fov}/2$, h is in the vertical altitude, $\bar{r}$ is a point in the plane perpendicular to these directions, and for the case above the angle $\theta_f/2 \le \theta_{fov}/2 \simeq 2.13$ and 0.28 arcsec so $\mathrm{Cos}\theta_f/2 \simeq 1$.

The difference in the MCF over the maximum field-of-view angle of the telescope is nearly the same as along the line-of-sight and

any cosine factor introduced by taking a path tilted at this small angle to the line-of-sight direction is equal to unity. The MCF is very nearly the same in the line-of-sight direction of the telescope as at the maximum field-of-view. Since all rays from the object make samll angles to the line-of-sight direction in each short-term exposure, the covariance of the phase structure function only needs to be calculated for the normal bundle of rays incident on the telescope aperture. Effectively, this will also be the phase correlation over the other bundles gathered by the telescope from the object. In the discussion to follow the angular dependence on θ_f will be neglected.

The second order moment of the complex random phase varaition is the mutual coherence function, MCF, so

$$B_h(\bar{a}) = \langle \psi_h(\bar{r}) \psi_h^*(\bar{r}+\bar{a}) \rangle \qquad (4)$$

where the bracket is an ensemble average and $\bar{a}$ is a displacement is in the direction perpendicular to the line-of-sight. From equation (2) this MCF is

$$B_h(\bar{a}) = \langle \exp j[\phi(\bar{r}) - \phi(\bar{r}+\bar{a})] \rangle \qquad (5)$$

Since $\phi(\bar{r})$ is a sum of a great number of independent variables, the independent cell sizes found in the atmosphere, and it has Gaussian statistics. The sum in the square bracket is a Gaussian with a zero mean value. For a random function f[9] the ensemble average can be written as

$$\langle e^f \rangle = \exp\{\langle f \rangle + 1/2 \langle [f - \langle f \rangle]^2 \rangle\} \qquad (6)$$

where in this case f = $j[\phi(\bar{r}) - \phi(\bar{r}+\bar{a})]$ so f has a zero mean or $\langle f \rangle = 0$. Thus

$$\langle \exp j[\phi(\bar{r}) - \phi(\bar{r}+\bar{a})] \rangle = \exp\{-\langle [\phi(\bar{r}) - \phi(\bar{r}+\bar{a})]^2 \rangle / 2\} \tag{7}$$

and in the literature[18,19] the argument of the exponential function is called the phase strucure function so

$$B_h(\bar{a}) = \exp[-1/2\ D_\phi(\bar{a})] \tag{8}$$

where

$$D_\phi(\bar{a}) = \langle [\phi(\bar{r}) - \phi(\bar{r}+\bar{a})]^2 \rangle \tag{9}$$

D_ϕ is now related to the statistics of the index fluctuation by introducing the covariance of $\phi(\bar{r})$, $B_\phi(\bar{a})$, where

$$\begin{aligned} B_\phi(\bar{a}) &= \langle \phi(\bar{r})\phi(\bar{r}+\bar{a}) \rangle \\ &= k^2 \int_h^{h+\Delta h} dh \int_h^{h+\Delta h} dh' \ \langle n(\bar{r},h)n(\bar{r}+\bar{a},h') \rangle \end{aligned} \tag{10}$$

Now let $\zeta = (h'-h)$ so the index fluctuation in equation (10) becomes

$$B_\phi(\bar{a}) = k^2 \int_h^{h+\Delta h} dh \int_h^{h+\Delta h-h} d\zeta \ B_n(\bar{a},\zeta) \tag{11}$$

where $B_n(\bar{a},\zeta) = \langle n(\bar{r})n(\bar{r}+\bar{a}) \rangle$.

The phase structure function is related to the covariance[10] by the relationship

$$D_\phi(\bar{a}) = 2\ [B_\phi(0) - B_\phi(\bar{a})] \tag{12}$$

and one may introduce an index structure function defined by the relationship

$$D_n(\bar{a},\zeta) = 2\,[B_n(0,0) - B_n(\bar{a},\zeta)] \tag{13}$$

Then equation (11) becomes

$$D_\phi(\bar{a}) \approx k^2 \int_h^{h+\Delta h} dh \int_h^{h+\Delta h-h} d\zeta\,[B_n(0,\zeta) - B_n(\bar{a},\zeta)]$$

$$= k^2 \int_h^{h+\Delta h} dh \int_h^{h+\Delta h-h} d\zeta\,[D_n(\bar{a},\zeta) - D_n(0,\zeta)] \tag{14}$$

At this point consider the atmosphere to be layered and the integral above is for a layer at altitude h of thickness Δh. Again, it is assumed that Δh is much larger than the correlation scale of the index fluctuation so the integral over ζ is taken from $-\infty$ to ∞ and equaton (14) becomes

$$D_\phi(\bar{a}) = k^2\,\Delta h \int_{-\infty}^{\infty} d\zeta\,[D_n(\bar{a},\zeta) - D_n(0,\zeta)] \tag{15}$$

If a Kolmogorov model of atmospheric turbulence is assumed, the index structure function is related to the quantity C_n^2 by the relationship

$$D_n(\bar{a},\zeta) = C_n^2(h)\,(a^2 + \zeta^2)^{1/3} \tag{16}$$

and

$$D_n(\bar{a},\zeta) - D_n(0,\zeta) = C_n^2(h)\,[(a^2 + \zeta^2)^{1/3} - \zeta^{2/3}] \tag{17}$$

so

$$D_{\phi}(\bar{a}) = k^2 C_n^2(h)\,\Delta h \int d\zeta\,[(a^2 + \zeta^2)^{1/3} - \zeta^{2/3}]$$

$$= 2.91\,k^2 C_n^2(h)\,\Delta h\,a^{5/3} \qquad (18)$$

where a is the beam width. This means the phase MCF for each layer is

$$B_h(\bar{a}) = \exp\left\{-1/2\left[2.91\,k^2 C_n^2(h)\,\Delta h\,a^{5/3}\right]\right\} \qquad (19)$$

This means for a layer j of the atmosphere the last equation should be written in the form

$$B_{jh}(\bar{a}) = \exp\left\{-1/2\left[2.91\,k^2 C_n^2(h_j)\,\Delta h_j a^{5/3}\right]\right\} \qquad (20)$$

If the atmosphere is assumed to be composed of N layers, the total MCF is

$$B_{Total}(\bar{a}) = \prod_{j=1}^{N} \exp\left\{-1/2\left[2.91\,k^2 C_n^2(h_j)\,\Delta h_j a^{5/3}\right]\right\}$$

$$= \exp\left\{\left[-1/2[2.91\,k^2\,a^{5/3} \sum_{j=1}^{N} C_n^2(h_j)\,\Delta h_j\right]\right\} \qquad (21)$$

This equation holds only for a vertical path of height h_j and thickness Δh_j so the turbulent layers are crossed normally. For an inclined path (line-of-sight) at a tilt angle θ to the vertical in the last equation is modified to the form

$$B_h(\bar{a}) = \exp\left\{-1/2\left[2.91\,k^2\,a^{5/3} \sum_{j=1}^{N} C_n^2(h_j)\,\Delta h_j\,\mathrm{Cos}^{-1}\theta\right]\right\} \qquad (22)$$

where Δh_j is the vertic height θ is the tilt angle, and $\Delta h_j \mathrm{Cos}^{-1}\theta$ is the tilt path distance.

In order to evaluate equation (21) or (22) the values of C_n^2 with altitude must be known. The nighttime measurements of Hufnagel[11,12] have been frequently sited. In recent years there have been many field measurements of C_n^2. These measurements of C_n^2 show it is a fluctuating quantity[11-17] and it in general follows the contour of the model given by Hufnagel but at any instant of time the values can be much different than predicted from this model. Hufnagel's result give the contour of the variation of C_n^2 but not its absolute value. As a result, it is best to use measured values which are close to the condition of the day of your measurement.

CONCLUSIONS

The phase MCF has been evaluated for a parallel beams traveling along the line-of-sight of the telescope. Since the other parallel bundles which can be received by the telescope from the object make very small angles to this direction, the phase MCF across these beams will be nearly the same as the line-of-sight direction. The loss of coherence across the the bundle received along the line-of-sight will give an indication of the magnitude of the problem and will indicate if a problem exists in removing the atmospheric phase distortion between the different frequency signals in the SCIP experiment. The normalized value of $B_\phi(\bar{a})$ probalbly should be evaluated. The accuracy of any calculation based on equations (21) and (22) will only be as good as the values of C_n^2 used in the calculation.

REFERENCES

1. P.S. Idell, Analysis of detection requirements for an AC shearing interferomter, Internal Report-Document, Phased Array Branch, Weapons Laboratory, Kirtland AFB, New Mexico, 5 July 1989

2. A. Webster, Derivation leading to simulation verification, Report by Applied Research Associates Inc, Albuquerque, New Mexico, 24 August 1989

3. R.A. Hutchin, Sheared coherent interferometric photography a technique for lensless imaging, SPIE Session-"Digital Image Recovery and Syntheses II", paper 2029-17, 12-13 July 1993

4. D.G. Voelz, J.D. Gongelewski, and P.S. Idell, SCIP computer simulation and laboratory verification, SPIE Session-"Digital Image Recovery and Synthesis II", paper 2029-18, 12-13 July 1993

5. M.T.Tavis and H.T. Yura, Short-term average irradiance profile of an optical beam in a turbulent medium, Appl. Opt., 15, 2922-2931 (1976)

6. D.L. Fried, Optical heterodyne detection of an atmospherically distorted signal wave field, Proc. IEEE, 50, 57-67 (1967)

7. K.A. Winick and D.vL. Marquis, Stellar scintillation technique for the measurement of tilt anisoplanatism, J.O.S.A. A5, 1929-1936 (1988)

8. D.A. Haner and R.T. Menzies, Reflectance characteristics of reference materials used in lidar hard target calibration, Appl.

Opt., 28, 857-864 (1989)

9. H.T. Yura, An elementary derivation of phase fluctuations of an optical wave in the atmosphere, Proc SPIE-"Imaging Through the Atmosphere", 75, 9-15 (1976)

10. F. Roddier, The effect of atmospheric turbulence in optical astronomy, "Progress in Optics", XIX, ed. E. Wolf

11. R.E. Hufnagel, Varaitions in atmospheric turbulence, O.S.A. Tech Digest, Topical Meeting-"Optical Propagation through Turbulence", Washington, D.C., paper WA1 (1974)

12. M.J. Beran and A.M. Whiteman, Scintillation index calculation using an altitude dependent structure constant, Appl. Opt. 27, 2178-2182 (1988)

13. G. R. Ochs, R. S. Lawrence, and T. Wang, Stellar scintillation measurement of the vertical profile of refractive-index turbulence in the atmosphere, Proc. SPIE-"Imaging Through the Atmosphere", 75, 47-54 (1976)

14. C.E. Coulman, J. Vernin, Y. Coqueugniot, and J.L. Caccia, Outer scale of turbulence appropriate to modeling refractive-index structure profiles, Appl. Opt. 27, 155-160 (1988)

15. R.E. Good, R.R. Beland, J.H. Brown, and E.M. Dewan, Atmospheric models of optical turbulence, Proc. SPIE-"Modeling of the Atmosphere", 928, 165-186 (1988)

16. J.Krause-Polstorff and D. Walters, Refractive turbulence profiling using an orbiting light source, Appl. Opt. 29, 1877-1885

(1990)

17. M.D. Rejack, B.M. Welch, and S.K. Rogers, Sensing refractive-turbulence profiles (C_n^2) using wave front slope measurements from two reference sources, Proc. SPIE-"Atmospheric Propagation and Remote Sensing", 1688, 256-264 (1992)

18. V.I. Tatarskii, "Wave Propagation in Turbulent Medium", McGraw-Hill (1971)

19. V.I. Tatarskii, "The Effects of the Turbulent Atmosphere on Wave Propagation", National Technical Infrmation Service, U.S. Department of Commerce, Springfield, VA. (1971)

LARGE STABLE SECOND-HARMONIC COEFFICIENTS IN CORONA-POLED POLYMERIC THIN FILMS

Gene Carlisle

Professor

Department of Physics

West Texas A & M University

Canyon, Texas 79016

Final Report for:

Summer Faculty Research Program

Phillips Laboratory

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D.C.

September 1992

LARGE STABLE SECOND-HARMONIC COEFFICIENTS IN CORONA-POLED POLYMERIC THIN FILMS

Gene Carlisle
Professor
Department of Physics
West Texas A & M University

Abstract

Second-harmonic generation (SHG) and spectroscopic absorption measurements were made over a 15-month period to study the nonlinear-optical thin-film properties of an azo-dye attached polymer in which the dye chromophores were oriented by corona poling. Stable values of the nonlinear optical coefficient d_{33}, for the fundamental of 1.064 μm, were as high as 68 pm/V. The stabilized order parameter of 0.51 found for the chromophores is to our knowledge larger than any reported for guest-host or chromophore-attached polymer glasses. A series of 10 copolymers were synthesized in which the concentration of the azo-dye monomer was varied from 10% to 100%. SHG and spectroscopic data were taken to determine the effect of azo-dye number density on order parameters and d_{ij} coefficients.

SHG and spectroscopic absorption measurements were also taken on guest-host polycarbonates and guest-host poly(methyl methacrylates) poled under high pressures of carbon dioxide.

LARGE STABLE SECOND-HARMONIC COEFFICIENTS IN CORONA-POLED POLYMERIC THIN FILMS

Gene Carlisle

INTRODUCTION

Organic polymeric thin films have very attractive mechanical and electronic properties such as ultrafast response times, low dielectric constants, high damage thresholds, high polarizability, and processibility, which facilitate their use in electro-optic and second-harmonic devices [1-3]. Films having large stable second-order nonlinear optical (NLO) coefficients are obviously needed. To meet these requirements, the polymer film should contain a high number density of highly ordered NLO chromophores which have high molecular hyperpolarizabilities. Side-chain polymers, with the NLO chromophores covalently attached to the polymer chain, have an advantage over guest-host polymers because they can be synthesized with higher number densities and they usually retain their NLO properties much better than guest-host polymers [4]. Corona poling [4-6], because of the high electric fields generated within the film, has been shown to be more effective than other poling methods in producing highly ordered polymeric materials. In this report we present the results for the preparation and characterization, of an azo-dye attached polymer which possesses large stable d_{ij} coefficients and order parameters. Also, a series of 10 copolymers were synthesized in which the concentration of the azo-dye monomer was varied from 10% to 100%. SHG and spectroscopic data were taken to determine the effect of azo-dye number density on order parameters and d_{ij} coefficients.

SHG and spectroscopic absorption measurements were also taken on guest-host polycarbonates and guest-host poly(methyl methacrylates) poled under high pressures of carbon dioxide.

METHODOLOGY

The structural formulas and synthetic method are shown in Fig. 1. The azo-

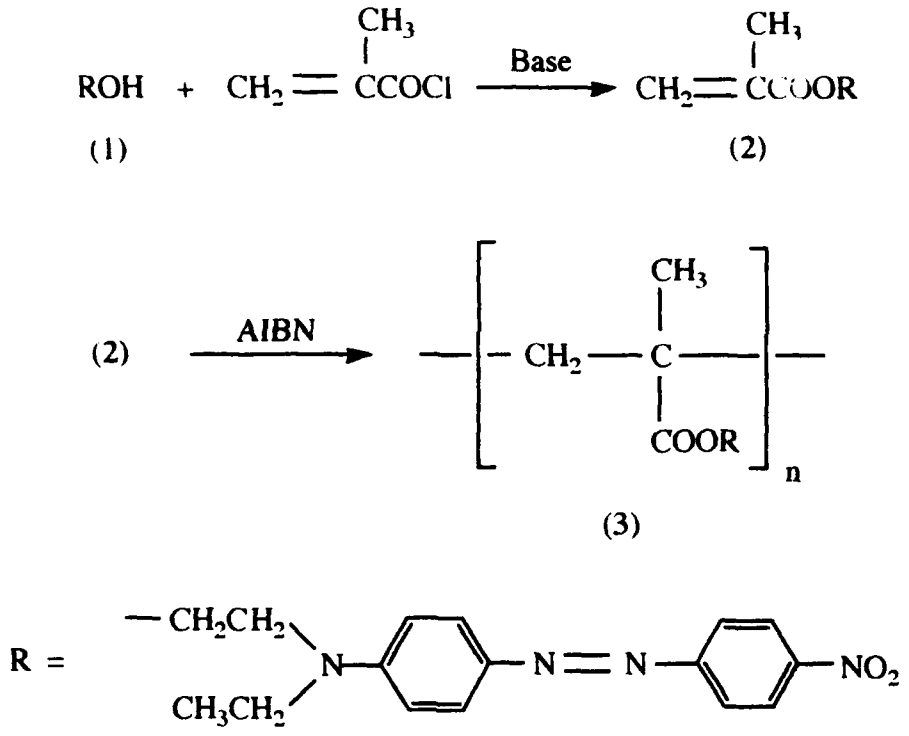


Figure 1. Structural formulas and synthetic method.

dye substituted acrylic monomer (2) was prepared by reacting equal molar quantities of disperse red 1 (DR1) (1), methacryloyl chloride and the base triethylamine in benzene solution at 75° C. After recrystallizing the monomer from an ethanol:acetone (120 mL:20 mL) mixture, the maroon crystals were dried in a vacuum oven.

Polymerization of the monomer was carried out in the presence of the initiator azoisobutyronitrile (AIBN) dissolved in a toluene:ethanol (2:1) mixture and under a 30 psi nitrogen atmosphere at 75° C. The polymer, DR1-PMMA, (3) was collected, washed with acetone, and further purified by precipitation from a cyclohexanone solution. Finally, a 40% yield of the polymer was obtained after drying in a vacuum oven at 75° C for 24 hours. Infrared and nuclear magnetic

resonance spectra and elemental chemical analyses of the monomer and polymer confirmed the structures as shown in Fig. 1. The average molecular weight and density of the polymer were found to be 22,000 and 1.19 g/cm^3, respectively.

Thin films were prepared from solutions containing 0.1047 g of polymer dissolved in 0.80 mL of cyclohexanone and 0.20 mL of nitrobenzene. The solutions were spun coated onto standard microscope slides, and the films were then baked in a vacuum oven at 90° C for 2 hrs to remove residual solvents. The resulting films were found by the use of a profilometer to have a typical thickness of 0.3 - 0.4 μm. The concentration of the DR1 chromophore was determined to be 1.87 x 10^{21}/cm^3.

The films were corona poled with a 25-μm diameter and 4.0-cm length tungsten wire held parallel to and 4.0 mm above the film. The glass-rubber transition temperature, T_g, was measured with a differential scanning calorimeter and found to be 128.5° C. Therefore, in order to ensure good mobility of the chromophore units during poling, the temperature of the films was raised to 131° C. A positive voltage to the wire electrode was slowly increased with caution until a corona breakdown voltage of usually 3 - 4 kV was reached which produced a current of about 4 μA. After poling for 30 min, the temperature was decreased and the voltage increased in order to maintain a current of 3.9 μA until the temperature reached 85° C. The voltage at this point was 5.0 kV which was then maintained until the film reached room temperature. The procedure consistently produced films of high optical quality. The poling area, which appears much more transparent than the unpoled film, is approximately 2 cm wide and extends the length of the wire electrode.

Polarized absorption spectra over the range of 300 - 700 nm were measured for the films before poling and periodically after poling for an observation period of 15 months with a spectrophotometer equipped with Glan Taylor polarizers. For a film having a thickness of 0.344 μm the absorbance at λ_{max} = 455.6 nm was measured to be 1.639 before poling; the absorbance decreased immediately after poling to 0.739 and stabilized at 0.800 after 21 days and

remained constant thereafter. The films were stored at room temperature in an unsealed slide box. Since the poled region of the film appears much more transparent than the unpoled area, there was concern that the azo dye had decomposed during the poling process; however, the possibility of a chemical change was ruled out after it was observed that the absorbance and λ_{max} would return to their unpoled values by heating the films to the T_g.

The SHG d_{ij} coefficients of the films were measured by the Maker-fringe method [7] using a neodymium-YAG laser (1064 nm, 15 ns/pulse, 160 mJ/pulse, 10 Hz). The experimental set-up, procedures, and calculations that were used have been explicitly described [8].

Similar methology was used to prepare and study a series of 10 copolymers in which the concentration of the azo-dye monomer was varied from 10% to 100%.

SHG and spectroscopic absorption measurements were also taken on guest-host polycarbonates (PC) and guest-host poly(methyl methacrylates) (PMMA). Since the T_g for PMMA polymers can be achieved under high pressures of carbon dioxide, a pressure chamber was constructed which allowed corona poling of thin films under pressures of carbon dioxide as high as 900 psi. Poling data such as temperature, pressure, current, and voltage were taken with a computer-controlled data acquisition system. SHG data were also taken on these films.

RESULTS

Fig. 2 shows the fringe patterns resulting from three processes: p-polarized incident laser light and p-polarized second harmonic (SH) detected (p-p), s-incident and p-detected SH (s-p), and the case where the incident beam is polarized 45° and s-detected SH (45-s). From these measurements, the three independent coefficients d_{31} = 11 pm/V, d_{15} = 15 pm/V, and d_{33} = 68 pm/V were calculated using a quartz crystal, d_{11} = 0.46 pm/V, as a reference. The procedure is estimated to produce a 10% uncertainty in d_{31} and d_{15} and a 20% uncertainty in d_{33}.

A thermodynamic model [9-11] has been used to describe poled materials.

From the model an orientational order parameter [12] ϕ can be estimated where

$$\phi = (3\langle\cos^2\theta\rangle - 1)/2, \quad (1)$$

$\langle\ \rangle$ is the average value and θ is the angle between the dipole moment μ of the dye chromophore and the local electric field $\mathbf{E}_p$ at the chromophore. It follows

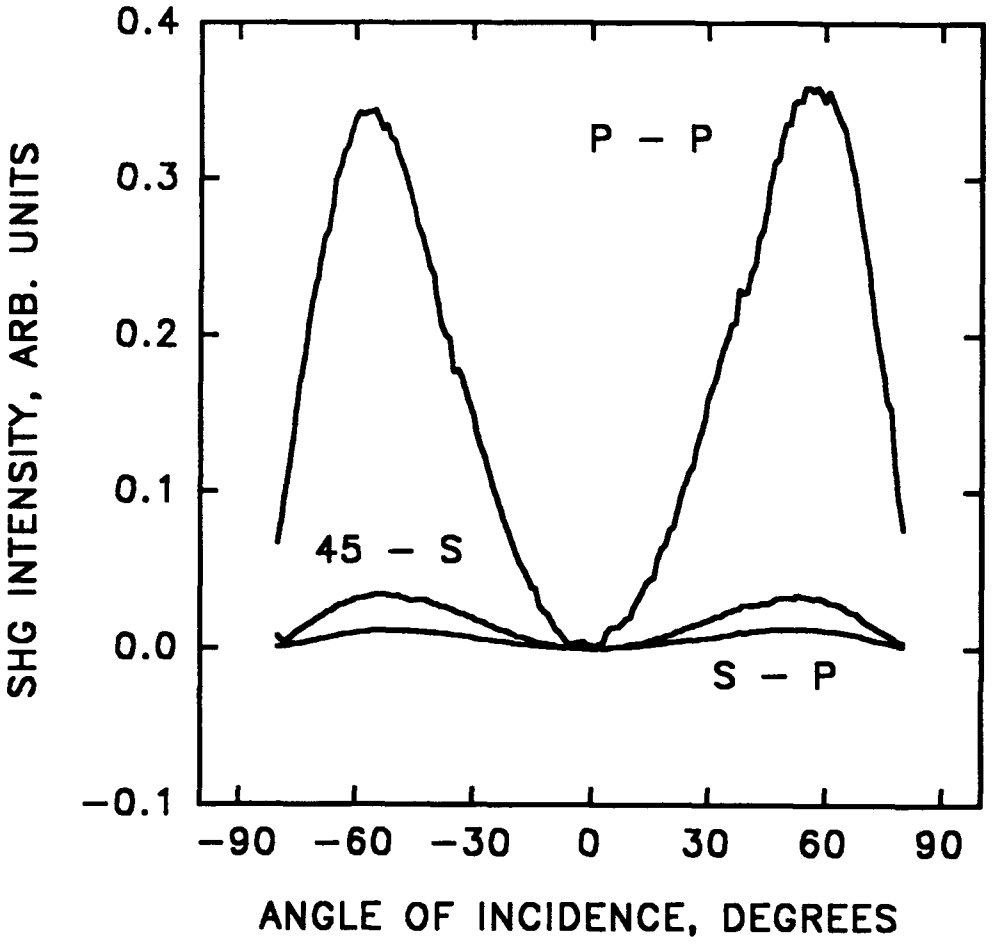


Figure 2. Second harmonic intensities *vs* the angle of incidence.

from the theory that $\phi = 1 - A_\perp/A_o$ where $A_\perp$ is the absorbance of the poled sample measured with the optical electric fields polarized perpendicular to the poling

direction and A_o is the absorbance of the unpoled film. An order parameter of 0.55 was measured immediately after poling. The parameter stabilized to a value of 0.51 after 21 days and has remained constant for 15 months.

The effectiveness of the poling procedure depends on the ratio of electrostatic dipole alignment energy to thermal energy, $p = \mu E_p/kT$ where k is Boltzmann's constant, and T is the temperature in Kelvins. Since the order parameter [6] can also be expressed as

$$\phi = 1 + 3/p^2 - (3/p)\coth p, \qquad (2)$$

p = 4.9 may be calculated from the order parameter ϕ = 0.51. For a poling temperature of 131° C and if μ is taken to be 8.7 D, the value for DR1 [13], E_p = 9.4 MV/cm may be calculated. These poling effectiveness values of p, ϕ, and E_p represent the highest reported for poled polymer glasses, and are important factors leading to the high d_{33} coefficient of 68 pm/V. Values of p = 2.5, ϕ = 0.26, and E_p = 4.8 MV/cm have been reported [6] for corona-poled DR1/PMMA guest-host polymers.

The largest second-order nonlinear coefficient, d_{33}, can be roughly estimated from the thermodynamic model to be 98 pm/V by use of Eq. (3) [10,11]

$$d_{33} = N\beta_{zzz}f^{2\omega}(f^{\omega})^2\langle\cos^3\theta\rangle/2 \qquad (3)$$

where N is the chromophore number density, and β_{zzz} is the second-order molecular hyperpolarizability of the DR1 chromophore. For this estimation the DR1 chromophore can be considered to be a one-dimensional molecule so that only β_{zzz} is nonzero; therefore, the β-value [13] of 125 x 10^{-30} esu for DR1 was taken for β_{zzz}. Appropriate local field factors for the optical field, f^{ω}, are Lorentz-Lorenz type, $(n_{\omega}^2 + 2)/3$, where, for this polymer, n_{ω} = 1.583 and $n_{2\omega}$ = 1.670. Although the d_{33} = 68 pm/V obtained from the SHG measurements is about 30% lower than that estimated from the thermodynamic model, the results are in agreement

with a previous study [11] made on DR1/PMMA guest-host polymers. In that study d_{33}-values, determined from SHG measurements for 4 different number densities less than 3 x $10^{20}/cm^3$, are about 30-36% less than those predicted by the thermodynamic model. The cause of this discrepancy is not clear at present, but with the high number density of the DR1-PMMA pure polymer, dipole-dipole interactions among the DR1 chromophores could possibly occur leading to a negative deviation from the thermodynamic model. From the SHG measurements a d_{33}/d_{31} ratio of 6.2 was found which is in excellent agreement with that predicted by the model if higher-order first and third Langevin functions are considered [14]. Under these conditions a p-value of 4.9 corresponds to a d_{33}/d_{31} ratio of about 6.

The results for the guest-host DR1/PC films yielded d_{33} = 25 pm/V, d_{15} = 6 pmV, and d_{31} = 7 pm/V while the results for the guest-host DO3/PC yielded d_{33} = 15 pm/V, d_{15} = 5 pmV, and d_{31} = 5 pm/V. The order parameters for DR1/PC and DO3/PC were 0.23 and 0.25, respectively. The higher d_{ij} coeficients for the DR1/PC films reflect the higher second-order molecular hyperpolarizability of the DR1 chromophore.

The results for the study of the effect of number density on d_{ij} coefficients and the results for the chamber-poled films are currently being completed and are not included in this report. However, these and all data are recorded in my laboratory notebook and data files at the Nonlinear Optics Center of Technology, PL/LIDN.

CONCLUSION

The stabilized order parameter of 0.51 to our knowledge is higher than any reported value for guest-host or chromophore-attached polymer glasses. Reported values are usually less than 0.3. One of the highest reported values of d_{33} for polymeric materials is that of 41 pm/V for copolymers with a chromophore similar to DR1. [8] The value of d_{33} = 68 pm/V for the pure polymer DR1-PMMA may be attributed to the high number density and order parameter.

The stabilized order parameters of 0.23 and 0.25 for the PC and PMMA films are fairly high and have remained essentially constant after an initial decay after poling.

REFERENCES

1. D. R. Ulrich, in *Organic Materials for Non-Linear Optics*, edited by R. A. Hann and E. Bloor, (The Royal Society of Chemistry, London, 1988), p 241.

2. P. N. Prasad and D. J. Williams, *Introduction to Nonlinear Optical Effects in Molecules and Polymers*, (Wiley, New York, 1991), Chap.12.

3. A. Buckley and J. B. Stamatoff, in Nonlinear Optical Effects in *Organic Polymers*, edited by J. Messier, F. Kajzar, P. Prasad and D. Ulrich (Kluwer Academic Publishers, London, 1989), p 327.

4. K. D. Singer, M. G. Kuzyk, W. R. Holland, J. E. Sohn, S. J. Lalama, R. B. Comizzoli, H. E. Katz and M. L. Schilling, *Appl. Phys. Lett.* **53**, 1800 (1988).

5. M. A. Mortazavi, A. Knowsen, S. T. Kowel, B. G. Higgins and A.Dienes, *J. Opt. Soc. Am.* **B 6**, 733 (1989).

6. R. H. Page, M. C. Jurich, B. Reck, A. Sen, R. J. Twieg, J. S. Swalen, G. C. Bjorklund and C. G. Willson, *J. Opt. Soc. Am.* **B 7**, 1239 (1990).

7. P. D. Maker, R. W. Terhune, M. Nisenoff, and C. M. Savage, *Phys. Rev. Lett.* **8**, 21 (1962).

8. L. M. Hayden, G. F. Sauter, F. R. Ore, P. L. Pasillas, J. M. Hoover, G. A. Lindsay, and R. A. Henry, *J. Appl. Phys.* **68**, 456 (1990).

9. J. Zyss and D. S. Chemla, in *Nonlinear Optical Properties of Organic Molecules and Crystals*, edited by D. S. Chemla and J. Zyss (Academic Press, New York, 1987) Vol. 1, p. 3.

10. D. J. Williams, in *Nonlinear Optical Properties of Organic Molecules and Crystals*, edited by D. S. Chemla and J. Zyss (Academic Press, New York, 1987) Vol. 1, p. 405.

11. K. D. Singer, M. Kuzyk and J. E. Sohn, *J. Opt. Soc. Am.* **B 4**, 968 (1987).

12. A. Peterlin and H. A. Stuart, *Z. Phys.* **112**, 129 (1939).

13. K. D. Singer, J. E. Sohn, L. A. King, H. M. Gordon, H. E. Katz and C. W. Dirk, *J. Opt. Soc. Am.* **B 6**, 1339 (1989).

14. P. N. Prasad and D. J. Williams, *Introduction to Nonlinear Optical Effects in Molecules and Polymers*, (Wiley, New York, 1991), p. 71.

ACKNOWLEDGMENTS

I thank Dr. Christopher M. Clayton and Mr. David R. Martinez for the opportunity of doing research at the Nonlinear Optics Center of Technology, PL/LIDN. The enjoyable working atmosphere generated by them, Dr. Stephen H. Chakmakjian, Dr. Mark T. Gruneisen, Ms. Gladys J. Kassay, Dr. Karl Koch, Dr. Gerald T. Moore, Dr. Phillip R. Peterson, Dr. Kenneth D. Shaw, Dr. P. Mohinder Sharma, and Captain Karen A. Yackiel is very much appreciated. The research accomplished this summer is greatly attributed to considerable efforts by Mr. Martinez and Dr. Koch.

The sponsorship of the Air Force Office of Scientific Research and Phillips Laboratory is gratefully acknowledged. The administrative support of Research and Development Laboratories is much appreciated.

Lidar and Atmospheric Monitoring

David B. Choate

Associate Professor

Department of Mathematics

Transylvania University

300 North Broadway

Lexington, KY 40508

Final Report for

Summer Research Extension Program

Phillips Laboratory

Sponsored by

Air Force Office of Scientific Research

Kirtland Air Force Base, Albuquerqy, NM

and

Transylvania University

August, 1993

Lidar and Atmospheric Monitoring

David Choate
Associate Professor
Department of Mathematics
Transylvania University

Abstract

Under the assumption of elastic scattering we denote by E(R) the detected radiative energy of a well aligned laser received from a target plane at a long range R.

In most cases R can be recovered given E.

If $E(R)R^2$ is very small or very large compared to the range, then there is no atmospheric turbulence near the target.

If $E(R)R^2$ is neither, then the atmosphere near the target is at worse turbid.

Lidar and Atmospheric Monitoring

David B. Choate

I. Introduction

By [4] p.277 equation 7.92 we know that the radiative energy received by the detector of a well aligned laser has the form

$$E(R) = \{(\text{pos.const})\beta(R)\exp[-\int_0^R K(R)dr]\}/R^2 \qquad (1.0)$$

where R is the range of the target plane, K(R) is the attenuation coefficient and $\beta(R)$ is the volume backscattering coefficient.

Experimental evidence [1],[3] suggests that

$$\beta(R) = (\text{pos.const})K^g(R) \qquad (1.1)$$

where $.67 \leq g \leq 1.0$ provided the backscattering is elastic; that is, the wavelength of the laser is equal to the wavelength of the returning signal.

The constant g depends on the lidar wavelength and the properties of the target.

II. Radil*

It is easy to show that the lidar equation (1.0) is usually invertible. This means that the range of the target can be determined given the detected energy. To establish this we prove

Theorem 1. Suppose that either K(R) is a positive constant or that $K(R) \geq \sqrt{g}$. Then E(R) is decreasing and hence invertible.

*$(\text{lidar})^{-1}$

Proof. We will only prove this theorem for the case $K(R) \geq \sqrt{g}$.

With $R > 0$ we have

$$(\sqrt{g}R + 1)^2 > 1 + gR^2$$

,or $\sqrt{g} > [-1 + \sqrt{(1 + gR^2)}]/R$.

So $K > [-1 + \sqrt{(1 + gR^2)}]/R = \alpha_1$

and consequently

$$K > [-1 - \sqrt{(1 + gR^2)}]/R = \alpha_2 .$$

Thus $0 < (K - \alpha_1)(K - \alpha_2) = K^2 + (2/R)K - g$

, or $-RK^2 + Rg - 2K < 0$.

Multiplying by K^{g-1} and applying (1.1) yields

$$-R\beta K + R\beta' - 2\beta < 0 \quad \text{for } R > 0 .$$

This implies

$$E'(R) = (\text{pos.const}) \exp[-\int_0^R K(R)dR] \cdot \{-R\beta K + R\beta' - 2\beta\}/R^3 < 0$$

So $E(R)$ is decreasing on $(0, \infty)$.

Comment 1. Since we will be concerned in the sequel only with long range targets, it is important to observe that a relaxed version of Theorem 1 would read:

$E(R)$ is eventually invertible provided $K(R)$ is eventually no smaller than $\sqrt{g}$.

And, of course, we assume in Theorem 1 that E and K are differentiable.

Comment 2. Theorem 1 is a holy grail result. Although the existence of an inverse is assured in most practical cases the closed form of the inverse is difficult to find even when K(R) is a positive constant.

However, given E and K, R can, of course, be approximated. We use the Newton-Raphson method [2] in the following

Example. Suppose K(R) = a, a positive constant.

Also suppose that $E = \{\exp[-\int_0^R a\,dR]\}/R^2$ is known.

Then set $f(R) = 1 - (E)R^2e^{aR}$.

So $f'(R) = -Re^{aR}\{aR + 2\}$

Thus

$$R_{k+1} = R_k + \{1 - E(R_k)^2\exp[aR_k]\}/\{R_k\exp[aR_k](aR_k + 2)\}$$

is an iterative formula that converges to the desired R.

III. Turbidity and Turbulence

With both E(R) and R in hand we are now prepared to say something about turbidity or turbulence near a long range target provide we make the modest assumption that K(R) does not oscillate; in other words, $\lim_{R\to\infty} K(R)$ exists or is infinite.

Let us first assume that $\lim_{R\to\infty} E(R)R^2 = a$, a nonzero constant.

Now suppose that $\lim_{R\to\infty} K(R) = 0$.

Then in the case of elastic scattering we know that for any well aligned laser

$\lim \beta(R) = c \neq 0$ by (1.0) (Note $c = +\infty$ will have the same effect.)

But $0 = 0^9 = c$ by (1.1).

So $\lim_{R\to\infty} K(R) \neq 0$.

In a clear atmosphere we can assume 7.105 on p. 280 of [4].

$$K_L(R) = \{\exp[-(S(R_\#) - S(R))/g]\} \div \{[1/K(R_\#)] + (2/g)\int_R^{R_\#} dR \exp[-(S(R_\#) - S(R))/g]dR\} \qquad (3.1)$$

But since a is a positive constant, the numerator approaches a finite number and the denominator approaches $+\infty$ by equation 7.22 p.243 [4]. This is a contradiction to our assumption that the atmosphere is clear.

Theorem 2. If $\lim_{R\to\infty} E(R)R^2 = a$, a nonzero constant, then in the case of elastic scattering the atmosphere near a long range target is turbid or turbulent assuming that the laser is well aligned.

Theorem 3. If $\lim_{R\to\infty} E(R)R^2 = +\infty$, then in the case of elastic scattering the atmosphere near a long range target is homogeneous assuming the laser is well aligned.

"Proof." Since $S(R) = \ln[E(R)R^2]$,

$\lim_{R\to\infty} K_L(R)$ is of type ∞/∞

since $E(R)R^2 \to +\infty$.

Applying l'Hopital's Rule we get

$$\lim_{R\to\infty} K_L(R) = \lim_{R\to\infty} [-S'(R)/2]$$

by the Second Fundamental Theorem of Calculus.

But $K_L(R) = K(R)/2$ *by equation (7.101) on p.279 of [4].*

So for large R

$K(R) \approx -d[S(R)]/dR$.

By p.278 equation (7.97) of [4]

$0 \approx [1/\beta(R)]\cdot\{d[\beta(R)]/dR\}$

, or $d[\beta(R)]/dR \approx 0$; so the atmosphere is homogeneous.

Theorem 4. If $\lim_{R\to\infty} E(R)R^2 = 0$ and $\lim_{R\to\infty} K(R) \neq 0$, then in the case of elastic scattering the atmosphere near a long range target is homogeneous assuming the laser is well aligned.

"Proof." Since $\lim_{R\to\infty} E(R)R^2 = 0$, the numerator of (3.1) approaches 0 . Since $\lim_{R\to\infty} K(R) \neq 0$, the denominator of (3.1) approaches 0 as well. Apply l'Hopital's Rule again and then proceed as in the "proof" of Theorem 3.

IV. AfterMath

Comment 3. The proofs of Theorems 3 and 4 are in quotes since technically we are begging the question of homogeneity since (3.1) holds only up to a haze.

But if we use the more sophisticated formulation of the attenuation coefficient given by Collis and Russell [1] instead of (3.1) derived by Measures, then the calculation is similar but far more technical.

Comment 4. Finally, in a practical situation, we would use this Rule of Thumb:

If $E(R)R^2$ is very large or very small with respect to the range (or, equivalently, if $E(R)R \ll 1$ or $1 \ll E(R)R$), then the atmosphere near a long range target is not turbulent.

If $E(R)R^2$ is a medium size constant, then the atmosphere is at worse turbid.

References

1. Collis and Russell (1976), "Lidar Measurement of Particles and Gases by Elastic Backscattering and Differential Absorption," Laser Monitoring of the Atmosphere (E.D. Hinkley, ED.) Springer-Verlag.

2. Hildebrand, F. B. (1956) Introduction to Numerical Analysis, McGraw-Hill.

3. Klett (1981), "Stable Analytical for Processing Lidar Returns," Appl. Optics, 20, 211-220.

4. Measures, R.M. (1992) Laser Remote Sensing, Krieger Publishing Company.

MODE BEATING IN IODINE LASERS

Wolfgang Rudolph
Associate Professor
Department of Physics and Astronomy
The University of New Mexico
Albuquerque, NM 87131

Report for:
Faculty Summer Research Program
Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Phillips Laboratory, Albuquerque

July 1993

MODE BEATING IN IODINE LASERS

Wolfgang Rudolph
Associate Professor
Department of Physics and Astronomy
The University of New Mexico
Albuquerque, NM 87131

Abstract

The potentials of iodine lasers to exhibit mode-beating in the GHz range was explored. A photolytic iodine laser pumped by an excimer laser was built. To study the temporal behavior of the laser output with picosecond resolution a streak camera was implemented. A short cavity dye laser served to test the streak camera and to find a suitable trigger mechanism. A computer simulation of the photolytic laser suggests several possibilities to achieve mode beating. Preliminary results indicate a strong dependence between laser transients and certain relaxation parameters of the active medium. Therefore, a comparison of measured and simulated data is expected to provide information on the latter.

MODE BEATING IN IODINE LASERS

Wolfgang Rudolph

1 Introduction

A great deal of interest has been devoted to the development of iodine lasers for different types of applications, see e.g., [1],[2]. Normally lasing is achieved between the hyperfine levels 4 and 3 of excited atomic iodine, I^*, which exhibits a total of six hyperfine transitions. The dominance of the 3-4 transition is because its magnetic dipole moment is about two times greater than that of the next largest transition (2-2), see Fig. 1. Recently a detailed

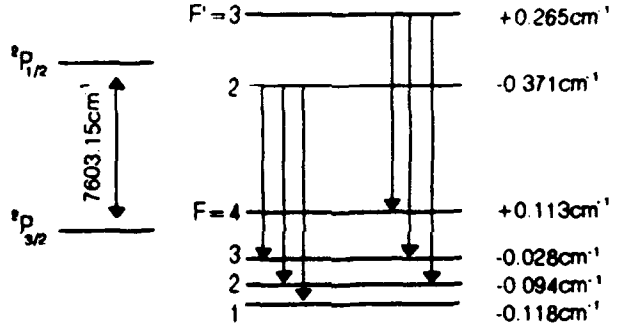


Figure 1: Hyperfine levels of atomic iodine and possible magnetic diplole transitions

study of the influence of magnetic fields and polarizing components in the cavity revealed that simultaneous lasing action on two hyperfine transitions can occur [3]. A coherent coupling of the two laser lines would lead to an output radiation modulated with their difference frequency of about 13 GHz. Given the potential of achieving very high output power from the iodine laser and the possibility of chemical pumping, this makes it an attractive light source in radar applications for example.

The aim of this project was to start a systematic study of mode beating effects in small-scale iodine lasers, to set up a suitable detection technique and to write a computer code to simulate the laser.

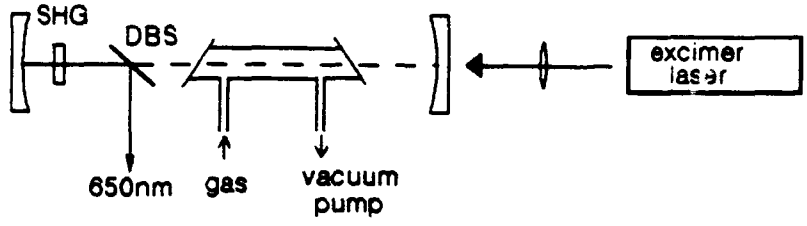


Figure 2: Experimental configuration for an excimer-laser pumped iodine laser

2 Experimental Setup

2.1 The Photolytic Iodine Laser

Photolytic pumping of an iodine laser can conveniently be achieved using an excimer laser. Depending on the chemical iodine compound to be used a KrF or XeCl laser can be employed. The criterion is a large enough absorption at the pump wavelength. For CF_3I which was available for our experiments the absorption cross section is about $3 \times 10^{-20} cm^2$ at 308nm and about $3 \times 10^{-19} cm^2$ at 248nm [1], respectively. This clearly favors the KrF pump here. A possible experimental setup is shown in Fig. 2. In the depicted longitudinal pumping arrangement the excimer beam is coupled into the iodine cell through one of the resonator mirrors. This allows a maximum overlap of the pumped region and the laser mode volume. Both end mirrors are highly reflective at 1.3μm to maximize the intracavity power. For the purpose of measuring the output by means of a streak camera (not sensitive to infrared light) the $1.3\mu m$ light has to be frequency converted. For a high conversion efficiency a nonlinear crystal ($LiIO_3$) was placed in the laser cavity. The frequency doubled signal is coupled out at a dichroic beam splitter ($>$ 98% reflectivity at 650nm and $>$ 98% transmittivity at $1.3\mu m$). The windows of the gas cell are tilted with respect to the resonator axis according to the Brewster condition in order to select a certain polarization direction. The total resonator losses including the second harmonic generation process do not exceed a few percent for the fundamental wave. Through absorption of a 20ns pump pulse the inversion builds up and lasing starts

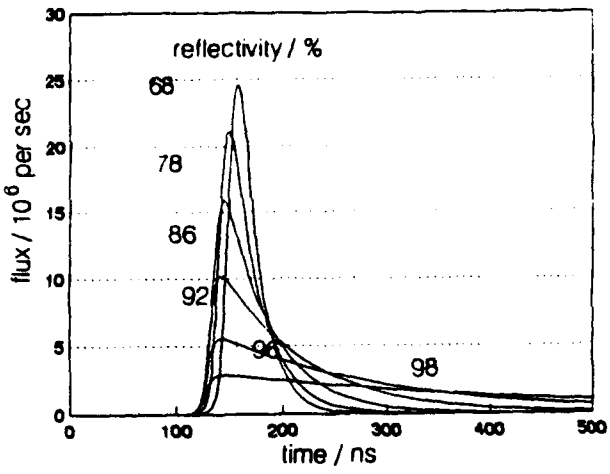


Figure 3: Calculated output power (fundamental radiation) of a gain-switched iodine laser for various reflection coefficients of the outcoupling mirror. The pump pulse duration was 20ns.

(gain switching). After pump absorption the temporal behavior of the output radiation is basically determined by the resonator losses. Using a simple model basing on rate equations the output can be simulated. A typical result is shown in Fig. 3. Obviously with linear losses in the order of a few percent laser emission over several hundred ns can be expected. This figure is important because a certain number of resonator roundtrips (usually $\gg$ 10) is required if mode locking techniques need to be applied to achieve the mode coupling.

2.2 The Streak Camera

While the temporal behavior of the iodine laser as shown in the simulation could be observed with an ordinary photodiode and a 200 MHz oscilloscope the monitoring of the GHz modulation requires a streak camera. Note, sampling techniques can not be applied since single shot events will have to be recorded.

A streak camera (Hamamatsu C2280) was available and needed to be implemented into the experiment. One of the most crucial parameters in operating such a camera is the timing, magnitude and stability of the trigger pulse for the streak unit. For ultimate high temporal resolution (5 ps) it is recommended that the trigger signal be derived from the optical signal to be measured by means of a vacuum photodiode. To compensate for the internal trigger delay in the streak unit the optical signal must then be delayed by about 30ns before being coupled into the camera. The leading edge of the iodine pulse

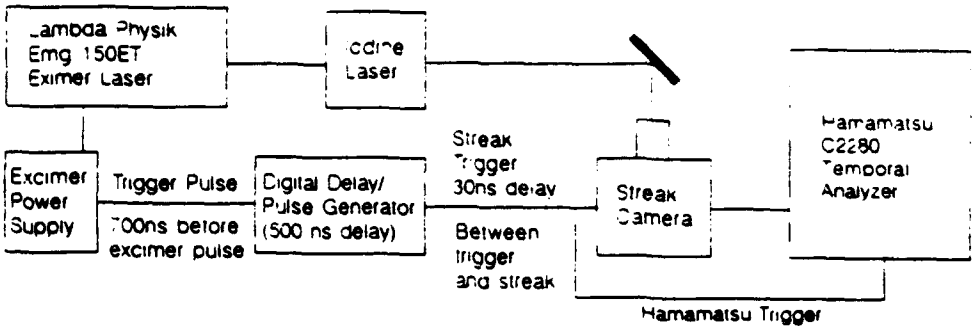


Figure 4: Sketch of the trigger diagram for the streak camera

could serve as such an optical trigger if data accumulation over several shots turns out to be necessary.

A simpler method is to use an electrical trigger signal provided by the power supply of the excimer laser. This trigger is synchronized to the trigger for the thyratron. A variable delay generator compensates for the delay between thyratron trigger, formation of the excimer pump pulse, formation of the iodine laser pulse and internal delay of the streak unit. Moreover, the delay generator allows to select a certain time window in which the iodine laser radiation is to be monitored. A part of the streak trigger serves to control the computer of the streak camera. The trigger diagram is sketched in Fig. 4. It was tested by means of a short-cavity dye laser pumped by the excimer laser. The total shot-to-shot jitter, that is the jitter seen on the display of the streak camera, was about ± 1 ns. This determines the accuracy by which the time window can be set. The time resolution is given by the streak velocity. In our test experiment with the dye laser a resolution of 50 ps was readily obtained. It was limited by the "slow" rise time of the dye laser pulse (> 300 ps) rather than by the streak camera.

2.3 Mode coupling

As detailed in Ref. [3] the gain coefficient for both hyperfine transitions starting from the upper laser level can be adjusted by an external magnetic field so that simultaneous lasing of both lines is observed. In what follows we list several techniques that seem suitable to lock the modes of the two lines.

1. Seeding of the iodine laser with light that carries the desired GHz modulation. The advantage of this technique is that a relatively small seed signal can serve to modulate a much more powerful laser. Therefore this method could serve to transfer the modulation of a small-scale, photolytic iodine laser to a high-power, chemically pumped iodine laser. It is also possible to seed a photolytic iodine laser by two, frequency-stabilized semiconductor lasers. Their mean carrier frequency and difference frequency is tuned to the corresponding iodine transitions. Other possible seed lasers are - a combination of an iodine laser lasing at the 3-4 transition and a semiconductor laser or dye laser. If necessary difference frequency mixing in a nonlinear crystal can be applied.

2. Modulation of the laser modes with the desired mode spacing frequency. This technique is widely used in combination with acousto-optic modulators operating in the 100 MHz range to synchronize the longitudinal modes and achieve modelocking for short pulse generation. Modulation in the 10 GHz range by traveling-wave modulators has become possible recently. A suitable set u. could be to modulate the output of an iodine laser extracavity and re-inject the signal back in the resonator. If the modulator is passed twice the modulating frequency needs to be one half of the desired modulation frequency of the laser output. The advantage of the external modulation would be that the modulator is only exposed to a fraction of the intracavity laser power.

3. Outcoupling at one frequency and re-injection at the other frequency can (in principle) also be achieved by Raman and Brillouin backscattering.

3 Simulation of Photolytic Iodine Lasers

3.1 Generalities and Basic Equations

The temporal behavior of the iodine laser on a time scale of a resonator round trip time or longer can favorably be studied using rate equations for the mean photon number and population inversion. This was the model used to study the general behavior of the iodine laser in the gain switched mode, cf. Fig. 3. To simulate the mode beating under various conditions one must chose a different approach. Instead of rate equations the

full set of density matrix and Maxwell equations has to be solved for the amplifier since the expected field transients are not much longer than the phase relaxation times of the transitions. Moreover averaging over one resonator roundtrip is not possible. The model employed here is similar to what is used for simulations of ultrashort pulse lasers, see for example Ref. [4], and is referred to as roundtrip model. The difference here is that we follow the evolution of individual resonator modes rather than that of a complex pulse envelope consisting of a very large number of modes.

The total electric field circulating in the resonator can be written as

$$E = \sum_{k=1}^{2} \sum_{j=1}^{M_k} \frac{1}{2} \tilde{\mathcal{E}}_{kj} e^{i(\omega_{kj} t - k_{kj} z)} + c.c. \tag{1}$$

where $k = 1(2)$ denotes the 2-2 (3-4) hyperfine transition, M_k is the number of resonator modes falling within transition k, ω_{kj} is the mode frequency, k_{kj} is the propagation constant, and $\tilde{\mathcal{E}}_{kj}$ is the complex mode amplitude. The (normalized) total field intensity is given by

$$W_p = \langle E^2 \rangle \tag{2}$$

where $\langle\rangle$ means temporal averaging over one light period.

The modification of the laser field is calculated in each resonator element separately and roundtrip after roundtrip where, for simplicity, we refer to a unidirectional ring laser. A block diagram of the program is depicted in Fig. 5. The initial field distribution for each mode can be noise simulating the spontaneous emission and/or suitable seed light. The noise is generated from random numbers independently for each mode and represents a fluctuating amplitude and phase. The temporal length of this initial field equals one resonator round trip time and was assumed to be 3 ns in the examples following.

Amplifier

The amplifier is described by a set of differential equations for the real and imaginary part of the polarization of each mode and for the occupation numbers. To start and to restrict the numbers of parameters we assumed for the gain medium four energy levels (1,2,3,4) where the 3-4 (2-2) transition corresponds to a transition between levels 4 and 1 (3 and 2). Relaxation between these two subsystems is described by corresponding relaxation times T_{43} and T_{21}. The four-level model implies a very fast relaxation time between

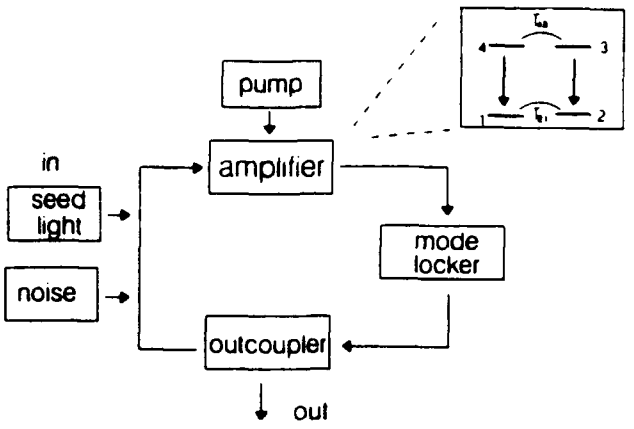


Figure 5: Block diagram of the simulation of mode beating and modelocking

Zeeman levels, an assumption that can be dropped easily if necessary. The two upper laser levels are assumed to be uniformly pumped for example by an excimer laser. The two gain transitions may have different emission cross sections which is to take into account different magnetic fields. Depending on the small signal gain for the strongest transition the amplifier is divided into a number of segments which are subsequently traversed by the field. In each slice a certain amount of noise is added to the field of each mode. The amount of noise is estimated from the light spontaneously emitted into the respective laser mode. After each slice of length Δz the field amplitude is related to the input field by

$$\tilde{\mathcal{E}}_{kj}(t, z + \Delta z) = \tilde{\mathcal{E}}_{kj}(t, z) - i\frac{\mu_0 \omega_{kj}}{2k_{kj}} \tilde{\mathcal{P}}_{kj}(t, z)\Delta z + \tilde{\mathcal{E}}_{noise} \tag{3}$$

where $\tilde{\mathcal{P}}_{kj}$ is the time dependent polarization amplitude. It can be obtained by solving a system of first order differential equations for the polarization

$$\frac{d}{dt}\tilde{\mathcal{P}}_{kj}(t) + \left[\frac{1}{\tau_k} + i(\omega_{kj} - \omega_{0k})\right] \tilde{\mathcal{P}}_{kj}(t) = -i\frac{p_k^2}{\hbar}(N_{3-k} - N_{2+k})\tilde{\mathcal{E}}_{kj}(t, z) \tag{4}$$

and for the occupation numbers N_i $(i = 1...4)$

$$\frac{d}{dt}N_3(t) = -\frac{1}{2\hbar} Im\left[\sum_{j=1}^{M_1} \tilde{\mathcal{E}}_{1j}^* \tilde{\mathcal{P}}_{1j}\right] + \frac{N_4 - N_3}{T_{43}} + \frac{1}{2}\sigma_{abs} N_x \frac{I_p}{\hbar\omega_p} \tag{5}$$

$$\frac{d}{dt}N_4(t) = -\frac{1}{2\hbar} Im\left[\sum_{j=1}^{M_2} \tilde{\mathcal{E}}_{2j}^* \tilde{\mathcal{P}}_{2j}\right] + \frac{N_3 - N_4}{T_{43}} + \frac{1}{2}\sigma_{abs} N_x \frac{I_p}{\hbar\omega_p} \tag{6}$$

$$\frac{d}{dt}N_2(t) = \frac{1}{2\hbar}Im\left[\sum_{j=1}^{M_1}\tilde{\mathcal{E}}_{1j}^{*}\tilde{\mathcal{P}}_{1j}\right] + \frac{N_1 - N_2}{T_{21}} \tag{7}$$

$$\frac{d}{dt}N_1(t) = \frac{1}{2\hbar}Im\left[\sum_{j=1}^{M_2}\tilde{\mathcal{E}}_{2j}^{*}\tilde{\mathcal{P}}_{2j}\right] + \frac{N_2 - N_1}{T_{21}} \tag{8}$$

$$\tag{9}$$

where

$$N_x(t) = N_{x0}exp\left[-\frac{\sigma_{abs}}{\hbar\omega_p}\int_{-\infty}^{t} I_p(t)dt\right]. \tag{10}$$

τ_k is the phase relaxation time of transition k, ω_{0k} is the mid frequency of transition k, p_k is the electric dipole moment, I_p is the pump intensity, N_x is the number density of parent molecules (e.g., CF_3I), and σ_{abs} is the absorption cross section for the pump radiation. For simplicity we have assumed a homogeneously pumped region within the mode volume of the resonator, that is, N_x is only time dependent.

Modelocker

After the amplifier the field passes through an optional modulator which can be a saturable absorber, an acousto-optic modelocker, etc.. For the saturable absorber we assume a homogeneously broadened two-level system. The absorber is also divided into several slices if necessary to ensure that the absorption coefficient α in each slice is small as compared to 1. The field after each slice is then related to the input through

$$\tilde{\mathcal{E}}_{kj}(t,out) = \tilde{\mathcal{E}}_{kj}(t,in)\left[1 - \frac{\alpha}{2W_p(t,in)/W_s}\right] \tag{11}$$

where W_s is the saturation intensity of the absorber. The nonlinear modecoupling originates here from the saturation term which contains the total field intensity W_p.

If an active (externally driven) modulator is used the complex field amplitudes are modified as

$$\tilde{\mathcal{E}}_{1j} = \tilde{\mathcal{E}}_{1j}(1-m) + \frac{1}{2}m\tilde{\mathcal{E}}_{2j} \tag{12}$$

$$\tilde{\mathcal{E}}_{2j} = \tilde{\mathcal{E}}_{2j}(1-m) + \frac{1}{2}m\tilde{\mathcal{E}}_{1j} \tag{13}$$

$$\tag{14}$$

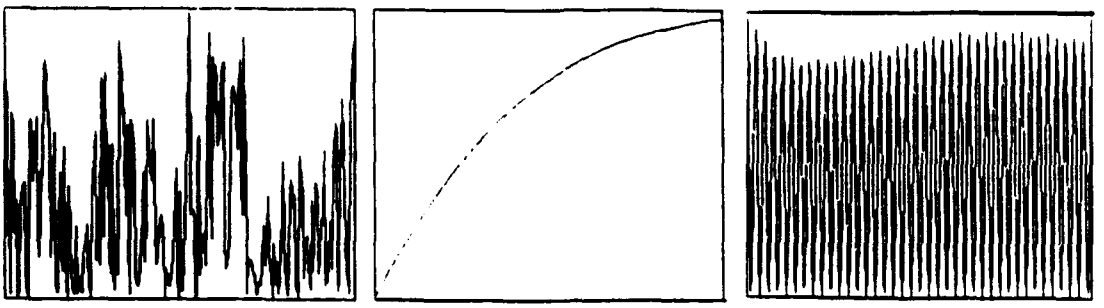


Figure 6: Laser power at different round trip times after switch-on of the 20 ns pump pulse. The time windows shown have a duration of 3 ns and correspond to one resonator round trip. left - initial noise, middle - after 20 roundtrips, right - after 40 roundtrips (gain saturated). Parameters: $T_{43} = 50ns, T_{21} = 1ns, \tau_1 = \tau_2 = 1ns, p_2^2 = 2p_1^2$

where m is the modulation depth.

Before coupled again into the amplifier the field passes the outcoupler and is attenuated by a certain amount. The roundtrip model has the advantage that more accurate models of the amplifier (e.g., actual Zeeman splitting in a magnetic field) or other pumping mechanisms (e.g., chemical pumping) can easily be implemented. Note, the integral of the field power over one cavity round trip shown as a function of the roundtrip number would yield a graph similar to that shown in Fig. 3.

3.2 Results

Even though a thorough exploration of the potentials of this computer code and a systematic study of different modelocking mechanism and the influence of various material and laser parameters will be undertaken in the near future, preliminary simulations show promising and interesting results. We will briefly describe here two examples. Firstly, let us assume that the external magnetic field is zero and the emission cross sections of the two transitions differ by a factor of two. Let us further assume that the gas pressure is low ($<$ 20 mbar) so that the relaxation times are in the ns range and that only one resonator mode is within the 2-2 and 3-4 transition. The laser output after different roundtrips is shown in Fig. 6. From the initial noise pattern the mode beating develops after a few roundtrips (not shown here). However, due to the different gain the two

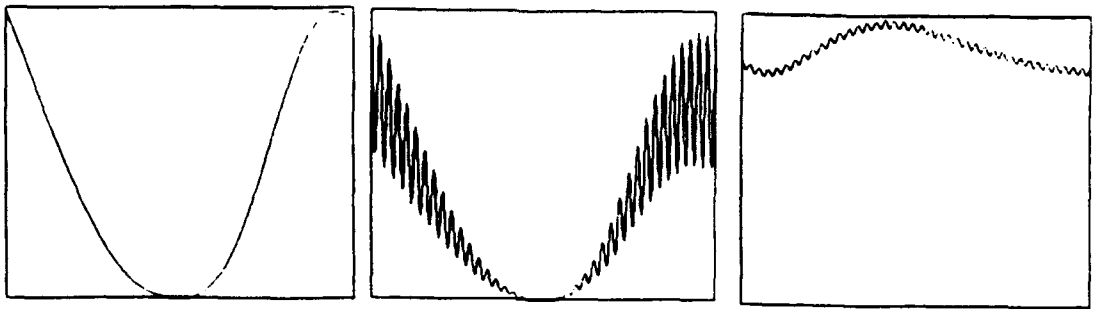


Figure 7: Laser power after gain saturation is reached for free-running (left), an amplitude-modulated (middle), and seeded (right) laser. Parameters: $T_{43} = 10ns, T_{21} = 100ps, \tau_1 = \tau_2 = 100ps, p_2^2 = 2p_1^2$

modes experience, the contrast of the modulation decreases rapidly so that after about 20 roundtrips the mode beating disappears in the laser output. It re-appears when the gain in the strongest transition is saturated while the amplitude of the second mode still growths until both have about the same magnitude. The final modulation is about one hundred percent. Since the gain is saturated the temporal change of the laser power is only determined by the resonator losses which do not affect the mode beating. It should be noted that this mode beating occured without additional mode-coupling mechanism and external magnetic fields. Of course, there is no modulation if the T_{21} and T_{43} relaxation times are much shorter than a resonator roundtrip time. Since these relaxation times can be controlled by the iodine and buffer gas concentration the results could be checked experimentally in a straightforward way.

In the second example we want to demonstrate the effect of seed light and active modecoupling. Corresponding results are shown in Fig. 7. We assume a zero magnetic field and a total of four modes, two under each transition. The modulation depth is 0.2 for the electric field. The seed light is assumed to be 100% modulated and has a power of about 10^{-8} times the intracavity power reached by the seeded laser. If we consider incoupling through a 99.99% reflective mirror the intracavity seed power is damped by another factor of 100. We estimated this seed signal to be ten times larger than the noise generated by the spontaneous emission emitted in the corresponding laser mode. Obviously this seed level is large enough to induce the desired mode beating despite of the

fact that the emission cross sections differ by a factor two (zero magnetic field). It is also obvious that the seeding substantially suppresses the beating between modes belonging to one hyperfine transition.

Acknowledgement

The author would like to thank G.Hager and J.McIver for stimulating and clarifying discussions and R.Tate and J.Nicholson for support in performing the experiments.

References

[1] G.Brederlow, E.Fill, and K.J.Witte. *The high-power iodine laser.* Springer, 1983.

[2] L.A.Schlie and R.D.Rathge. *Opt.Commun.*, 66:289, 1988.

[3] M.A.Kelly, J.K.McIver, R.F.Shea, and G.D.Hager. *IEEE J. Quantum Electron.*, QE-27:263, 1991.

[4] V.Petrov, W.Rudolph, U.Stamm, and B.Wilhelmi. *Phys.Rev.*, A40:1474, 1989.

ESTIMATION OF THE SQUARED MODULUS OF THE COMPLEX COHERENCE FACTOR FROM AMPLITUDE OR INTENSITY MEASUREMENTS AT HIGH LIGHT-LEVELS

Timothy J. Schulz
Assistant Professor
Department of Electrical Engineering

Michigan Technological University
1400 Townsend Drive
Houghton, MI 49931-1295

Final Report for:
Summer Faculty Research Program
Phillips Laboratory

:nsored by:
Air Force ... e of Scientific Research
Bolling Air Force Base
Washington, D. C.

September 1993

Estimation of the Squared Modulus of the Complex Coherence Factor from Amplitude or Intensity Measurements at High Light-Levels

Timothy J. Schulz
Assistant Professor
Department of Electrical Engineering
Michigan Technological University

Abstract

Lower bounds are presented for the variance of any unbiased estimator of the squared modulus of the complex coherence factor from amplitude or intensity measurements when the fluctuations of the complex-valued amplitude are governed by circular Gaussian statistics, and when the light level is high enough that all non-classical fluctuations of the measurements can be ignored. For both amplitude and intensity measurements, this bound is compared with the variance of some common estimators along with the variance of the maximum-likelihood estimators. Additionally, the lower bounds for amplitude and intensity measurements are compared, and the signal-to-noise ratio gain that an unbiased estimator that optimally processes amplitude measurements can obtain over an unbiased estimator that optimally processes intensity measurements is examined.

Estimation of the Squared Modulus of the Complex Coherence Factor from Amplitude or Intensity Measurements at High Light-Levels

Timothy J. Schulz

1 Introduction

The squared modulus of the complex coherence factor (or mutual intensity) of a light field at two points in space can be estimated by interfering or correlating the field's amplitude or intensity at the two points. Although the phase of the complex coherence factor can also be estimated from amplitude measurements, it cannot, in general, be estimated from intensity measurements. When the complex amplitudes of the light field are interfered optically, as in Young's interference experiment, the technique is referred to as amplitude interferometry. When the intensities of the light field are correlated after detection, as in the pioneering work of Hanbury Brown and Twiss [1, 2], the technique is referred to as intensity interferometry.

When direct measurements of the field amplitude or intensity are available at the two points, one is not restricted to the processing schemes commonly associated with amplitude or intensity interferometry. Instead, the data can be stored on a digital computer and processed in a variety of ways. When few realizations of the light field are available, as can occur with pseudothermal or laser light, the processing scheme can greatly influence the estimation accuracy.

In this report, we consider fundamental limits on the estimation accuracy that can be obtained by any estimator that processes direct measurements of a light field's amplitude or intensity. For this study, we only consider the effects of classical fluctuations of the optical field and do not consider detection noise, nor do we consider the manner in which the magnitude and phase of the complex-valued amplitude might be measured.

The report is organized as follows. In Section 2 a statistical model is presented for the measured data, and the problem of estimating the squared modulus of the complex coherence factor μ is formally stated. In Section 3 the estimation of $|\mu|^2$ from amplitude measurements is considered, and a fundamental limit on the estimation accuracy is derived. Additionally, a simple estimator is examined along with the maximum-likelihood estimator, and their performances are compared with the limit. In Section 4 the estimation of $|\mu|^2$ from intensity measurements is considered, and a fundamental limit on the estimation accuracy is presented. Estimators from a class of polynomial estimators, discussed in detail by S. Ebstein [3], are examined along with the maximum-likelihood estimator and their performances are compared with the limit. Finally, in Section 5 the results of this study are summarized, and the fundamental limits on the estimation accuracy from amplitude and intensity measurements are compared.

2 Data Model and Problem Statement

Consider a pair of jointly circular, complex-valued Gaussian random-variables, U_1 and U_2, that have zero mean, unit variance, and whose complex correlation coefficient is μ. Mathematically, these random variables satisfy

$$E[U_1] = E[U_2] = 0, \tag{1}$$

$$E[|U_1|^2] = E[|U_2|^2] = 1, \tag{2}$$

$$E[U_1 U_2^*] = \mu,$$

and their joint density is

$$p_U(U_1, U_2) = \pi^{-2}(1 - |\mu|^2)^{-1} \exp\left(\frac{-|U_1|^2 - |U_2|^2 + \mu U_1^* U_2 + \mu^* U_1 U_2^*}{1 - |\mu|^2}\right). \tag{3}$$

If we write the complex correlation coefficient as

$$\mu = \sqrt{\alpha} e^{j\beta}, \tag{4}$$

where $\alpha = |\mu|^2$ and $\beta = \arg(\mu)$, then the joint density can be written as

$$p_U(U_1, U_2; \alpha, \beta) = \pi^{-2}(1 - \alpha)^{-1} \exp\left[\frac{-|U_1|^2 - |U_2|^2 + \sqrt{\alpha}(e^{j\beta} U_1^* U_2 + e^{-j\beta} U_1 U_2^*)}{1 - \alpha}\right], \tag{5}$$

where the notation for the density function explicitly shows the dependence on the parameters α and β.

Next, consider the pair of real-valued random variables $I_1 = |U_1|^2$ and $I_2 = |U_2|^2$. According to Goodman [4, p. 43], these random variables have the joint density function

$$p_I(I_1, I_2; \alpha) = (1 - \alpha)^{-1} \exp\left(-\frac{I_1 + I_2}{1 - \alpha}\right) I_0\left(\frac{2\sqrt{I_1 I_2 \alpha}}{1 - \alpha}\right), \tag{6}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind, zero order, and the notation for the density function explicitly shows the dependence on the parameter α. Notice that the density does not depend on the phase β of the complex correlation coefficient.

Data such as these can arise in a variety of situations. One such situation is described by Goodman [4, 5] in which a rough object is illuminated by a coherent light source and amplitude (U_1 and U_2) or intensity (I_1 and I_2) measurements are made at two far-field locations. For this situation, the complex correlation coefficient μ is commonly referred to as the complex coherence factor. The assumption that the variances, $E[|U_1|^2]$ and $E[|U_2|^2]$, are known is based on the fact that these quantities can be controlled or easily measured. That they are assumed to be unity is

the result of an arbitrary choice and does not affect the analysis.

When measurements of U_1 and U_2 are made, one can attempt to estimate both the squared modulus α and phase β of μ, whereas when measurements of I_1 and I_2 are made, one can only attempt to estimate α. For purposes of comparison, we assume throughout the remainder of this paper that we are only interested in estimating α, regardless of the nature of our measurements. Therefore, our problem is simply stated as one of estimating the squared modulus of the complex coherence factor from measurements of $\{U_{1k}, U_{2k}\}_{k=1}^{K}$ or $\{I_{1k}, I_{2k}\}_{k=1}^{K}$, where k is an index over independent realizations (measurements) of the amplitudes or intensities.

3 Amplitude Measurements

As mentioned previously, when amplitude measurements are available one can estimate both the squared modulus α and the phase β of the complex coherence factor μ. Here, we only consider the estimation of α and discuss fundamental limits on the estimation accuracy, as defined by the variance, of any unbiased estimator. Additionally, we consider the estimator

$$\hat{\alpha}_{\mathrm{a}} = \left| \frac{1}{K} \sum_{k=1}^{K} U_{1k} U_{2k}^{*} \right|^2 , \tag{7}$$

which is asymptotically unbiased, and compare its variance to the fundamental limit. Finally, the maximum-likelihood estimator is discussed and its variance is also compared with the lower bound.

3.1 Cramer-Rao Lower Bound

Detailed discussions of the Cramer-Rao lower bound (CRLB) can be found in a variety of texts on statistical estimation theory [6, 7, 8]. Through the use of the CRLB, one can compute lower bounds on the variance, a measure of estimation accuracy, of any unbiased estimator of a scalar or vector parameter. Any unbiased estimator whose variance attains the bound is called an *efficient* estimator. Here, we use the CRLB to derive a fundamental limit on the estimation accuracy of any unbiased estimator of the squared modulus of the complex coherence factor when amplitude measurements are made.

Recall that, when amplitude measurements are available, the unknown parameters of the measurement density are the squared modulus α and phase β of the complex coherence factor μ:

$$p_{U}(U_1, U_2; \alpha, \beta) = \pi^{-2}(1-\alpha)^{-1} \exp\left[\frac{-|U_1|^2 - |U_2|^2 + \sqrt{\alpha}(e^{j\beta} U_1^{*} U_2 + e^{-j\beta} U_1 U_2^{*})}{1-\alpha} \right]. \tag{8}$$

The CRLB states that the variance of any unbiased estimator of α, say $\hat{\alpha}$, must satisfy

$$\mathrm{var}(\hat{\alpha}) \geq \left[J^{-1}(\alpha, \beta) \right]_{11} , \tag{9}$$

where $J(\alpha,\beta)$ is the Fisher information matrix and is defined by

$$J(\alpha,\beta) = -E\left[\begin{array}{cc} \dfrac{\partial^2 \ln p_U(U_1,U_2;\alpha,\beta)}{\partial\alpha^2} & \dfrac{\partial^2 \ln p_U(U_1,U_2;\alpha,\beta)}{\partial\alpha\partial\beta} \\ \dfrac{\partial^2 \ln p_U(U_1,U_2;\alpha,\beta)}{\partial\beta\partial\alpha} & \dfrac{\partial^2 \ln p_U(U_1,U_2;\alpha,\beta)}{\partial\beta^2} \end{array}\right]. \tag{10}$$

Here $E[\cdot]$ denotes an expectation over the random variables U_1 and U_2. Although we are not considering the estimation of the phase, the CRLB for the variance of any unbiased estimator of β is

$$\mathrm{var}(\hat{\beta}) \geq \left[J^{-1}(\alpha,\beta)\right]_{22}. \tag{11}$$

As derived in Appendix A, the Fisher information matrix for this problem is

$$J(\alpha,\beta) = \left[\begin{array}{cc} \dfrac{1+\alpha}{2\alpha(1-\alpha)^2} & 0 \\ 0 & \dfrac{2\alpha}{1-\alpha} \end{array}\right], \tag{12}$$

so that

$$\mathrm{var}(\hat{\alpha}) \geq \frac{2\alpha(1-\alpha)^2}{1+\alpha}. \tag{13}$$

This is the bound on the estimator variance when one measurement of U_1 and U_2 is made. For K independent measurements, $\{U_{1k},U_{2k}\}_{k=1}^{K}$, the bound is simply scaled by K^{-1}. Therefore, Eq. (13) represents a bound on the normalized variance, where for K independent measurements the normalization factor is K. This bound on the normalized variance of any unbiased estimator of α is plotted in Fig. 1 as a function of the magnitude of the complex coherence factor $|\mu|$ (or $\sqrt{\alpha}$).

3.2 Analysis of a Simple Estimator

Consider the simple estimator

$$\hat{\alpha}_{\mathrm{a}} = \left|\frac{1}{K}\sum_{k=1}^{K} U_{1k}U_{2k}^{*}\right|^2. \tag{14}$$

As is shown in Appendix B, this estimator is asymptotically unbiased and its variance, ignoring terms that are scaled by K^{-n} for n greater than one, is

$$\mathrm{var}(\hat{\alpha}_{\mathrm{a}}) = K^{-1}2\alpha(1+\alpha). \tag{15}$$

Its normalized variance is plotted along with the CRLB in Fig. 2. This estimator is nearly optimal when $|\mu|$ (or $\sqrt{\alpha}$) is near zero, but as $|\mu|$ becomes larger the estimator becomes suboptimal.

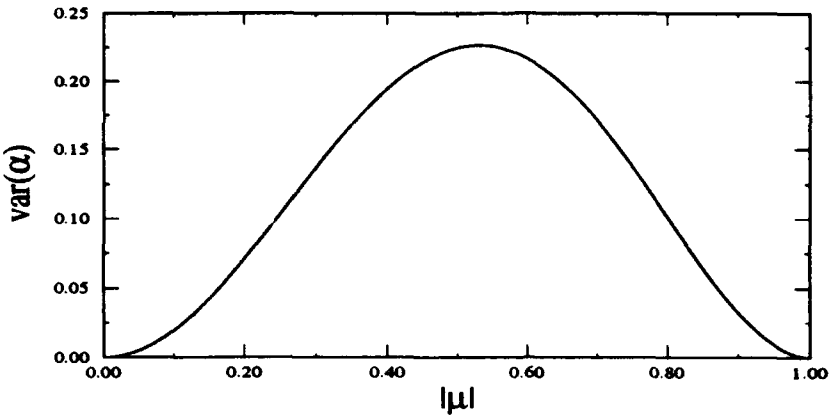


Figure 1: Cramer-Rao lower bound for the normalized variance of any unbiased estimator of α when amplitude measurements are made.

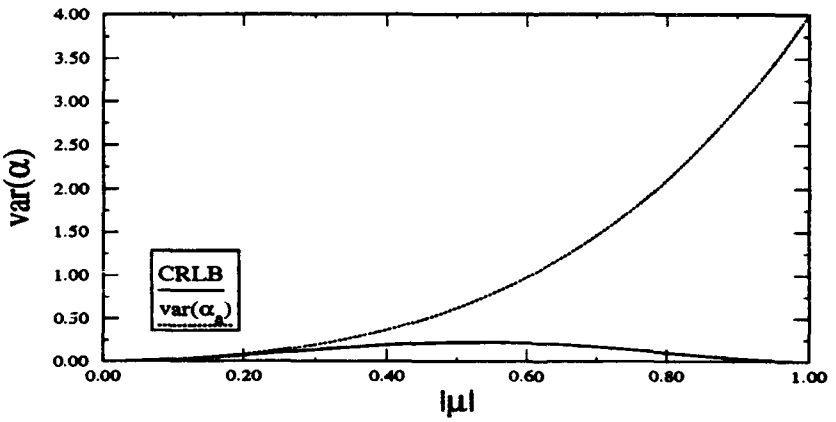


Figure 2: Normalized variance for the estimator defined by Eq. (14) along with the CRLB for amplitude measurements.

3.3 Analysis of the Maximum-Likelihood Estimator

The maximum-likelihood estimator (MLE) of α from amplitude measurements is defined by the joint maximization

$$(\hat{\alpha}_{\text{mle}}, \hat{\beta}_{\text{mle}}) = \underset{\alpha \in [0,1],\ \beta}{\arg\max} \prod_{k=1}^{K} p_U(U_{1k}, U_{2k}; \alpha, \beta). \tag{16}$$

Since the logarithm is a monotonically increasing function of a nonnegative argument, the MLE can also be defined as

$$(\hat{\alpha}_{\text{mle}}, \hat{\beta}_{\text{mle}}) = \underset{\alpha \in [0,1],\ \beta}{\arg\max}\ L(\alpha, \beta), \tag{17}$$

where

$$\begin{aligned} L(\alpha,\beta) &= \ln\left[\prod_{k=1}^{K} p_U(U_{1k}, U_{2k}; \alpha, \beta)\right] \\ &= \sum_{k=1}^{K} \ln p_U(U_{1k}, U_{2k}; \alpha, \beta) \\ &= -2K \ln \pi - K \ln(1-\alpha) - \frac{\sum_{k=1}^{K} \left(|U_{1k}|^2 + |U_{2k}|^2\right)}{1-\alpha} \\ &\quad + \frac{\sqrt{\alpha}}{1-\alpha}\left[e^{j\beta}\left(\sum_{k=1}^{K} U_{1k}^* U_{2k}\right) + e^{-j\beta}\left(\sum_{k=1}^{K} U_{1k} U_{2k}^*\right)\right]. \end{aligned} \tag{18}$$

It is straightforward to show that, for any α, the β that maximizes $L(\alpha,\beta)$ is the phase of $\sum_{k=1}^{K} U_{1k} U_{2k}^*$:

$$\hat{\beta}_{\text{mle}} = \arg\left(\sum_{k=1}^{K} U_{1k} U_{2k}^*\right).$$

Therefore, $\hat{\alpha}_{\text{mle}}$ must maximize

$$L(\alpha) = -K \ln(1-\alpha) - \frac{S}{1-\alpha} + 2|R|\frac{\sqrt{\alpha}}{1-\alpha}, \tag{19}$$

where $\hat{\beta}_{\text{mle}}$ has been substituted for β in $L(\alpha,\beta)$, terms that do not depend on α have been dropped, and where the substitutions $S = \sum_{k=1}^{K} (|U_{1k}|^2 + |U_{2k}|^2)$ and $R = \sum_{k=1}^{K} U_{1k} U_{2k}^*$ have been made. A closed form expression for a maximizer of $L(\alpha)$ is not obvious; however, the maximizer can be found numerically by a variety of techniques.

The variance and mean of the MLE was evaluated by a Monte-Carlo calculation. With K chosen to be 64, the sample variance and sample mean for 10,000 independent experiments was computed for each of 20 equally spaced values of $|\mu|$. The results, with the variance normalized for $K = 64$, are shown in Fig. 3. An examination of this figure shows that for $K = 64$ the MLE is

slightly biased for small $|\mu|$, and that the normalized estimator variance is close to the CRLB, with a slight deviation near $|\mu| = 0$.

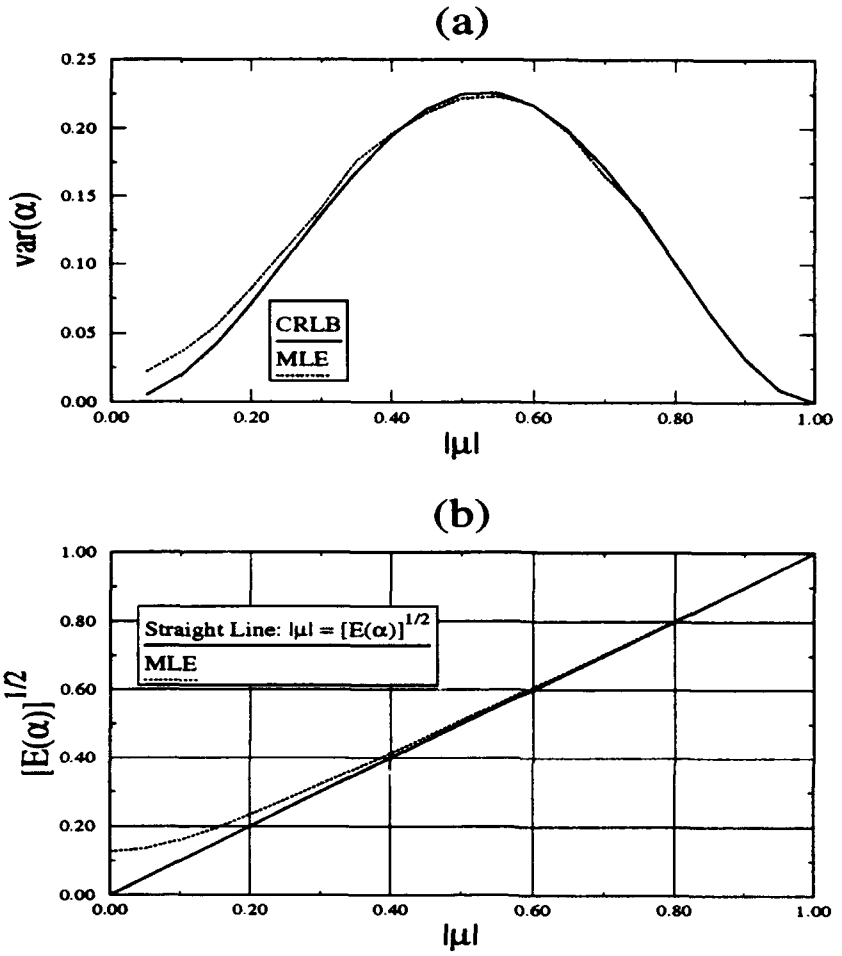


Figure 3: Normalized variance (a) and mean (b) of the MLE from amplitude measurements evaluated by a Monte-Carlo calculation from 10,000 independent experiments with $K = 64$ samples used to form each estimate. The CRLB is also shown for comparison.

4 Intensity Measurements

When intensity measurements are available, one can only estimate the squared modulus α of the complex coherence factor μ. The phase β cannot be estimated. In this section we discuss fundamental limits on the estimation accuracy, again as defined by the estimator variance, of any unbiased estimator of α. Additionally, we consider a class of unbiased estimators recently discussed by S. Ebstein [3] and compare their variances with the fundamental limit. Finally, the maximum-likelihood estimator is discussed and its variance is also compared with the lower bound.

4.1 Cramer-Rao Lower Bound

When intensity measurements are made, the unknown parameter of the measurement density is α:

$$p_I(I_1, I_2; \alpha) = (1-\alpha)^{-1} \exp\left(-\frac{I_1+I_2}{1-\alpha}\right) I_0\left(2\sqrt{I_1 I_2}\frac{\sqrt{\alpha}}{1-\alpha}\right). \tag{20}$$

Due to the presence of the Bessel function, an analytic, closed form expression for the CRLB is difficult to derive. However, the first derivative of $\ln p_I(I_1, I_2; \alpha)$ with respect to α is easily derived as

$$\frac{\partial \ln p_I(I_1, I_2; \alpha)}{\partial \alpha} = \frac{1}{1-\alpha} - \frac{I_1+I_2}{(1-\alpha)^2} + \sqrt{I_1 I_2}\left[\frac{\alpha^{1/2}+\alpha^{-1/2}}{(1-\alpha)^2}\right] \frac{I_1\left(2\sqrt{I_1 I_2}\frac{\sqrt{\alpha}}{1-\alpha}\right)}{I_0\left(2\sqrt{I_1 I_2}\frac{\sqrt{\alpha}}{1-\alpha}\right)}, \tag{21}$$

where $I_1(\cdot)$ is the modified Bessel function of the first kind, order one. To overcome the difficulty of deriving the CRLB an alternative form for the Fisher information matrix, which for a scalar parameter is the scalar

$$J(\alpha) = E\left[\left(\frac{\partial \ln p_I(I_1, I_2; \alpha)}{\partial \alpha}\right)^2\right], \tag{22}$$

was used and the expectation was evaluated through a Monte-Carlo calculation. The results are shown in Fig. 4. Here, the expectation in Eq. (22) was evaluated by generating 5 million indepen-

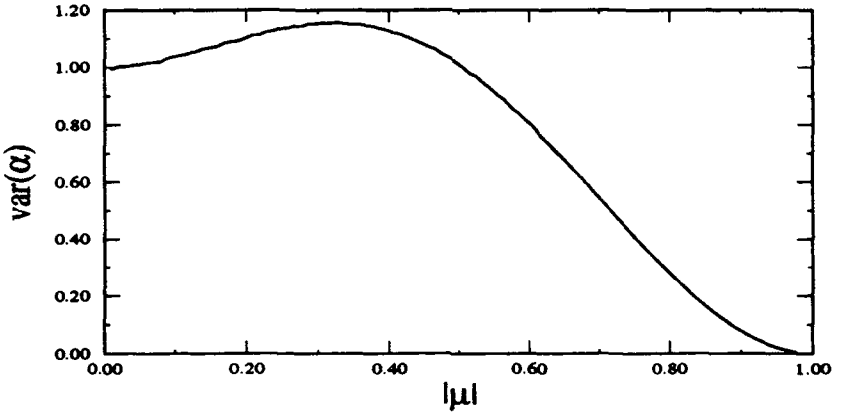


Figure 4: Cramer-Rao lower bound for the normalized variance of any unbiased estimator of α when intensity measurements are made.

dent realizations of $(\frac{\partial \ln p_I(I_1, I_2; \alpha)}{\partial \alpha})^2$ for each of 100 equally spaced values of $|\mu|$ (or $\sqrt{\alpha}$). Again, for K independent measurements, $\{I_{1k}, I_{2k}\}_{k=1}^{K}$, this represents the bound on the normalized variance where the normalization factor is K.

4.2 Polynomial Estimators

In Ref. [3], S. Ebstein considers the class of single-measurement estimators

$$\hat{\alpha}(a,b,d,f) = a(I_1^2 + I_2^2) + bI_1I_2 + d(I_1 + I_2) + f, \tag{23}$$

and shows that for $\hat{\alpha}(a,b,d,f)$ to be an unbiased estimator, the following conditions on the coefficients (a,b,c,d) must be satisfied:

$$b = 1 \quad \text{and} \quad 4a + 2d + f = -1. \tag{24}$$

Additionally, he shows that the variance of the unbiased estimator, as a function of the free parameters a and d, is

$$\begin{aligned} \text{var}[\hat{\alpha}(a,b,d,f)] &= \alpha^2(8a^2+3) + \alpha(2d^2 + 16ad + 12d + 32a^2 + 64a + 14) \\ &\quad +(2d^2 + 16ad + 4d + 40a^2 + 16a + 3), \end{aligned} \tag{25}$$

and that the choice for the parameters a and d that minimizes the variance for a given α is

$$a = -\frac{\alpha}{1+\alpha^2}, \tag{26}$$

and

$$d = -\frac{(1-\alpha)(1-3\alpha^2)}{(1+\alpha)(1+\alpha^2)}, \tag{27}$$

with the resulting variance:

$$\text{var}[\hat{\alpha}_{\text{opt}}(a,b,d,f);\alpha] = \frac{1 + 5\alpha - 8\alpha^2 - \alpha^4 + 3\alpha^5}{(1+\alpha)(1+\alpha^2)}. \tag{28}$$

Notice that this is not the variance of a particular estimator as a function of α, but is rather the lower envelope of the variance of the estimators for all values of a and d. When Eq. (28) is plotted with the CRLB, as is shown in Fig. 5, we see that the bound is satisfied for all α.

When the expected values of I_1 and I_2 are known, as we are assuming for this work, the single measurement, polynomial estimators discussed by Ebstein can be extended to the multiple measurement estimators

$$\hat{\alpha}(a,b,d,f) = \frac{1}{K}\sum_{k=1}^{K}\left[a(I_{1k}^2 + I_{2k}^2) + bI_{1k}I_{2k} + d(I_{1k} + I_{2k}) + f\right]. \tag{29}$$

Since the data represent K independent measurements, the normalized variance of the estimators

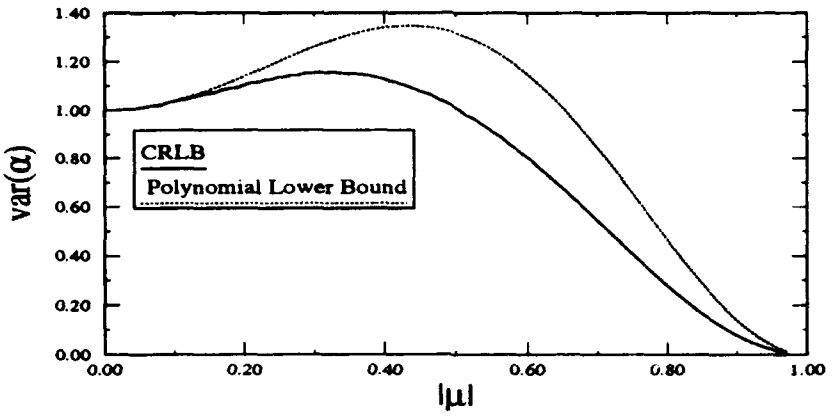


Figure 5: Lower bound for the variance of the unbiased polynomial estimator (Eqs. 23 and 24) along with the Cramer-Rao lower bound for the variance of any unbiased estimator of α when intensity measurements are made.

will remain the same. Three particular estimators discussed by Ebstein are

$$\hat{\alpha}_1 = \frac{1}{K}\sum_{k=1}^{K}(I_{1k}I_{2k} - 1), \qquad (a = 0, b = 1, d = 0, f = -1), \tag{30}$$

$$\hat{\alpha}_2 = \frac{1}{K}\sum_{k=1}^{K}(I_{1k} - 1)(I_{2k} - 1), \qquad (a = 0, b = 1, d = -1, f = 1), \tag{31}$$

and

$$\hat{\alpha}_3 = \frac{1}{K}\sum_{k=1}^{K}\left(I_{1k}I_{2k} - \frac{I_{1k}^2 + I_{2k}^2}{4}\right), \qquad (a = -1/4, b = 1, d = 0, f = 0), \tag{32}$$

and their corresponding variances are

$$\mathrm{var}(\hat{\alpha}_1) = \frac{1}{K}(3\alpha^2 + 14\alpha + 3), \tag{33}$$

$$\mathrm{var}(\hat{\alpha}_2) = \frac{1}{K}(3\alpha^2 + 4\alpha + 1), \tag{34}$$

and

$$\mathrm{var}(\hat{\alpha}_3) = \frac{1}{K}\left(\frac{7}{2}\alpha^2 + \frac{3}{2}\right). \tag{35}$$

As shown by Ebstein, the estimator $\hat{\alpha}_2$ is the optimal polynomial estimator for $|\mu| = 0$. Also,

within the class of polynomial estimators that are optimal for a single value of $|\mu|$, the estimator

$$\hat{\alpha}_4 = \frac{1}{K}\sum_{k=1}^{K}\left(I_{1k}I_{2k} - \frac{I_{1k}^2 + I_{2k}^2}{2} + 1\right), \qquad (a = -1/2, b = 1, d = 0, f = 1), \tag{36}$$

is optimal for $|\mu| = 1$, and its variance is

$$\mathrm{var}(\hat{\alpha}_4) = \frac{1}{K}(5\alpha^2 - 10\alpha + 5). \tag{37}$$

Whereas Ebstein discussed unbiased polynomial estimators whose variance is optimal for a particular $|\mu|$, other criteria can be considered for selecting the coefficients a and d. For instance, one could seek an estimator whose maximum variance is smaller that the maximum variance of any other unbiased polynomial estimator over the interval $|\mu| \in [0, 1]$. Such an estimator would be considered the *minimax* estimator within the class of unbiased polynomial estimators. Mathematically, the minimax estimator is defined by

$$\max_{|\mu|\in[0,1]} \mathrm{var}(\hat{\alpha}_{\mathrm{mm}}) \leq \max_{|\mu|\in[0,1]} \mathrm{var}[\hat{\alpha}(a, b, d, f)], \tag{38}$$

for all a, b, d, f such that $\hat{\alpha}(a, b, d, f)$ is unbiased. The minimax estimator for this problem is

$$\hat{\alpha}_{\mathrm{mm}} = \frac{1}{K}\sum_{k=1}^{K}\left(I_{1k}I_{2k} - \frac{I_{1k}^2 + I_{2k}^2}{4} - \frac{I_{1k} + I_{2k}}{2} + 1\right), \qquad (a = -1/4, b = 1, d = -1/2, f = 1), \tag{39}$$

where the coefficients that satisfy the minimax condition were found by a numerical search. The variance of this estimator is

$$\mathrm{var}(\hat{\alpha}_{\mathrm{mm}}) = \frac{1}{K}\left(\frac{7}{2}\alpha^2 - \frac{7}{2}\alpha + 2\right). \tag{40}$$

Fig. 6 contains a plot of the normalized estimator variance for the estimators $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4$, and $\hat{\alpha}_{\mathrm{mm}}$, along with the CRLB and the polynomial lower bound.

4.3 Analysis of the Maximum-Likelihood Estimator

The MLE of α from intensity measurements is defined by the maximization

$$\hat{\alpha}_{\mathrm{mle}} = \underset{\alpha\in[0,1]}{\arg\max} \prod_{k=1}^{K} p_I(I_{1k}, I_{2k}; \alpha), \tag{41}$$

which can be redefined as

$$\hat{\alpha}_{\mathrm{mle}} = \underset{\alpha\in[0,1]}{\arg\max}\ L(\alpha), \tag{42}$$

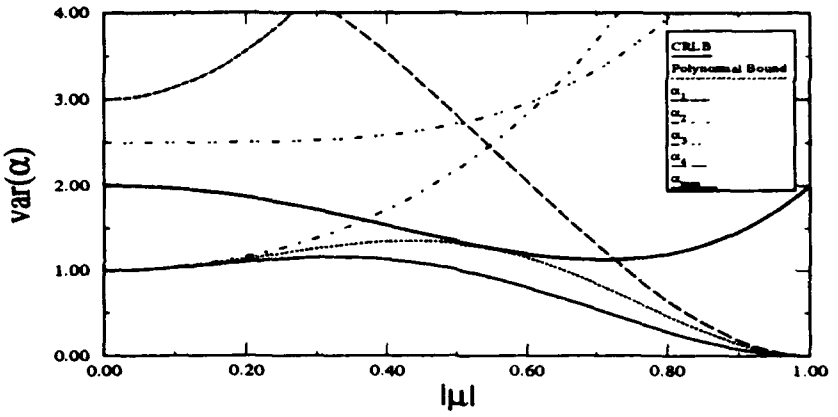


Figure 6: Normalized variance for selected polynomial estimators along with the CRLB and the pointwise polynomial lower bound.

where

$$\begin{aligned} L(\alpha) &= \ln\left[\prod_{k=1}^{K} p_I(I_{1k}, I_{2k}; \alpha)\right] \\ &= \sum_{k=1}^{K} \ln p_I(I_{1k}, I_2 k; \alpha) \\ &= -K\ln(1-\alpha) - \frac{S}{1-\alpha} + \sum_{k=1}^{K} \ln I_0\left(2\sqrt{I_{1k}I_{2k}}\frac{\sqrt{\alpha}}{1-\alpha}\right), \end{aligned} \tag{43}$$

and where $S = \sum_{k=1}^{K}(I_{1k} + I_{2k})$. A closed form expression for a maximizer of $L(\alpha)$ is not obvious; however, the maximizer can be found numerically by any one of a variety of techniques.

The variance and mean of the MLE were evaluated by a Monte-Carlo calculation. With $K = 64$ samples used to form each estimate, 10,000 independent experiments were performed and the sample variance and sample mean were computed for each of 50 equally spaced values of $|\mu|$. The results, with the variance normalized for $K = 64$, are shown in Fig. 7. Notice that the MLE appears to violate the CRLB for $|\mu| < 0.5$, but upon examination of the mean we see that the MLE is biased for $|\mu| < 0.5$ (or $\alpha < 0.25$), and, therefore, its variance is not required to satisfy the bound.

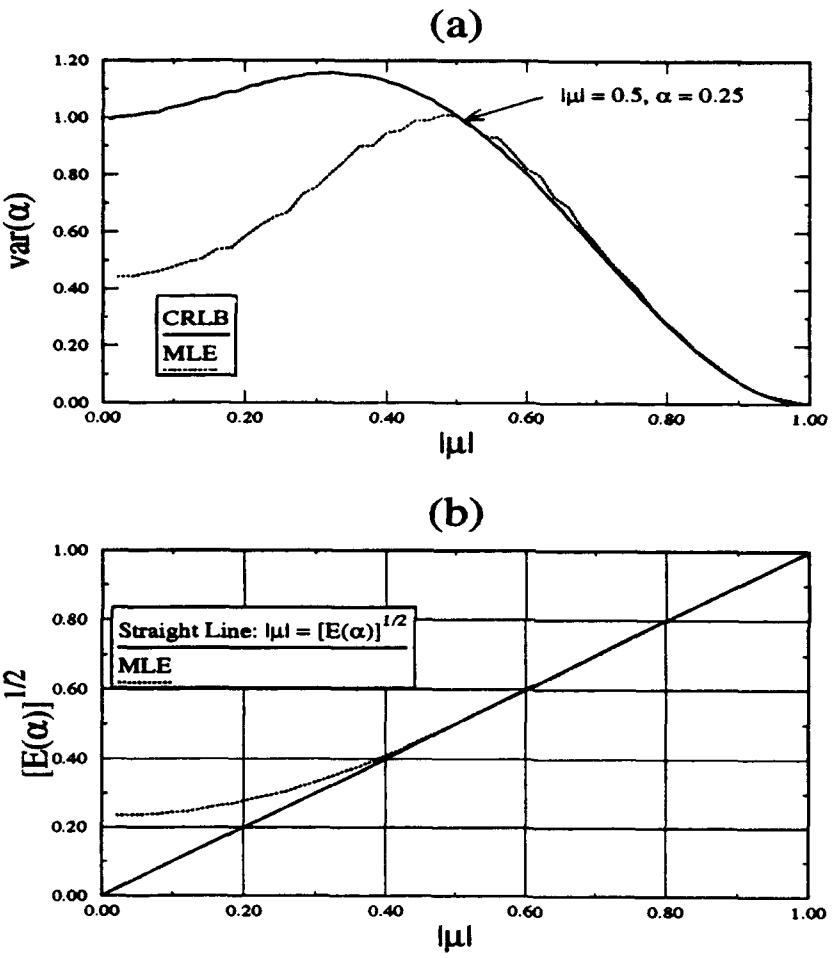


Figure 7: Normalized variance (a) and mean (b) of the MLE from intensity measurements evaluated by a Monte-Carlo calculation from 10,000 independent experiments with $K = 64$ samples used to form each estimate. The bias in the MLE explains the apparent violation of the CRLB for $|\mu| < 0.5$ (or $\alpha < 0.25$).

5 Summary

Lower bounds on the normalized estimator variance of any unbiased estimator of the squared modulus of the complex coherence factor have been derived and are shown in Fig. 8 for both amplitude and intensity measurements. These results are not derived for a particular estimator and, therefore, represent fundamental limits on the variance of any unbiased estimator. We have

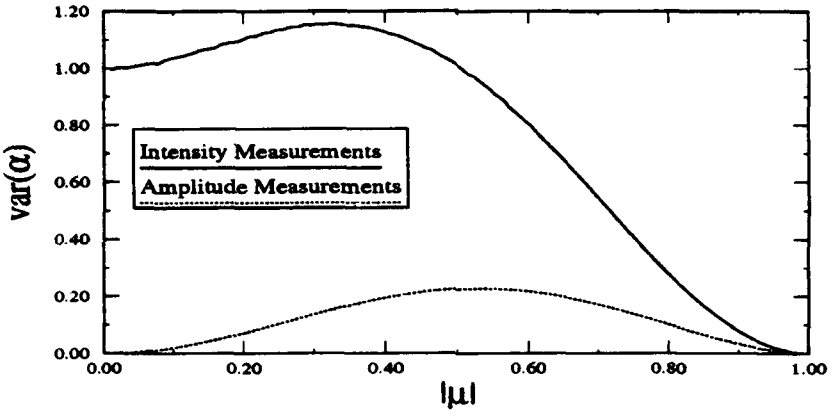


Figure 8: Cramer-Rao lower bounds for the normalized variance of any unbiased estimator of α when amplitude or intensity measurements are made.

also demonstrated that many common estimators, while optimal or nearly optimal over a limited range of $|\mu|$, are suboptimal for most $|\mu|$.

For many problems a signal-to-noise ratio, defined as

$$\text{SNR} = \frac{\alpha}{[\text{var}(\hat{\alpha})]^{1/2}}, \tag{44}$$

is used as a measure of the performance of an unbiased estimator of α. In Fig. 9 we show the SNR gain, defined as

$$\text{SNR Gain} = \frac{\text{SNR}_{\text{amplitude}}}{\text{SNR}_{\text{intensity}}},$$

that can be obtained by an unbiased estimator that optimally processes amplitude measurements over an unbiased estimator that optimally processes intensity measurements. Whereas the advantages and disadvantages of amplitude interferometry, in comparison with intensity interferometry, have been discussed by others [5], most analyses have been directed at particular processing schemes. In this report, we have presented fundamental limits on performance and have not restricted the analysis to particular estimators. Important issues that have not been considered in this analysis include the effects of detection noise, the manner by which the magnitude and phase of the complex

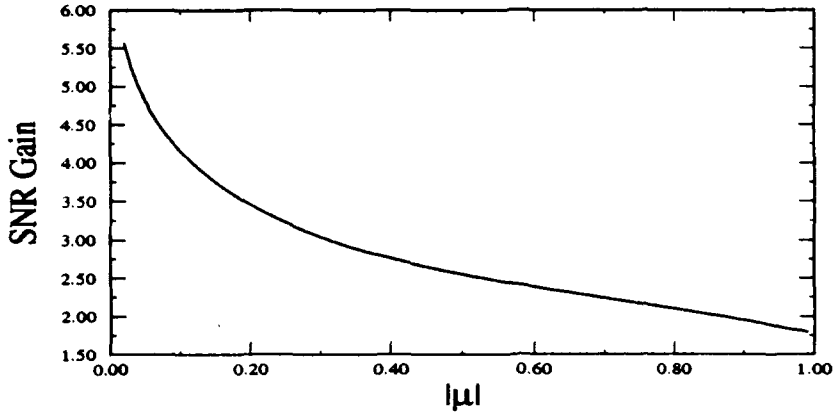


Figure 9: Signal-to-noise ratio gain that can be obtained by an unbiased estimator that optimally processes amplitude measurements over an unbiased estimator that optimally processes intensity measurements.

amplitude is measured, and the importance of the phase information that can also be recovered from amplitude measurements.

Appendix A: Fisher Information Matrix for Amplitude Measurements

Recall that

$$p_U(U_1, U_2; \alpha, \beta) = \pi^{-2}(1-\alpha)^{-1} \exp\left[\frac{-S + \sqrt{\alpha}\left(e^{j\beta}R^* + e^{-j\beta}R\right)}{1-\alpha}\right], \tag{A.1}$$

where $S = |U_1|^2 + |U_2|^2$ and $R = U_1 U_2^*$. Therefore,

$$\begin{aligned}\frac{\partial^2 L(\alpha,\beta)}{\partial \alpha^2} &= (1-\alpha)^{-2} - 2(1-\alpha)^{-3}S \\ &+ \left[\frac{1}{4}\alpha^{-3/2}(1-\alpha)^{-1} - \alpha^{-1/2}(1-\alpha)^{-2} - 2\alpha^{1/2}(1-\alpha)^{-3}\right]\left(e^{j\beta}R^* + e^{-j\beta}R\right),\end{aligned} \tag{A.2}$$

$$\frac{\partial^2 L(\alpha,\beta)}{\partial \beta^2} = -\frac{\alpha^{1/2}}{1-\alpha}\left(e^{j\beta}R^* + e^{-j\beta}R\right), \tag{A.3}$$

and

$$\frac{\partial^2 L(\alpha,\beta)}{\partial \alpha \partial \beta} = \frac{\partial^2 L(\alpha,\beta)}{\partial \beta \partial \alpha} = j\left[\frac{1}{2}\alpha^{-1/2}(1-\alpha)^{-1} + \alpha^{1/2}(1-\alpha)^{-2}\right]\left(e^{j\beta}R^* - e^{-j\beta}R\right), \tag{A.4}$$

where $L(\alpha,\beta) = \ln p_U(U_1, U_2; \alpha, \beta)$. Since

$$\begin{aligned} E[S] &= E\left[|U_1|^2\right] + E\left[|U_2|^2\right] \\ &= 2, \end{aligned} \tag{A.5}$$

and

$$\begin{aligned} E[R] &= E\left[U_1 U_2^*\right] \\ &= \alpha^{1/2} e^{j\beta}, \end{aligned} \tag{A.6}$$

it follows that

$$\begin{aligned} -E\left[\frac{\partial^2 L(\alpha,\beta)}{\partial \alpha^2}\right] &= \frac{-1}{(1-\alpha)^2} + \frac{4}{(1-\alpha)^3} + \frac{1}{2\alpha(1-\alpha)} - \frac{2}{(1-\alpha)^2} - \frac{4\alpha}{(1-\alpha)^3} \\ &= \frac{1+\alpha}{2\alpha(1-\alpha)^2}, \end{aligned} \tag{A.7}$$

$$-E\left[\frac{\partial^2 L(\alpha,\beta)}{\partial \beta^2}\right] = \frac{2\alpha}{1-\alpha}, \tag{A.8}$$

and

$$-E\left[\frac{\partial^2 L(\alpha,\beta)}{\partial \alpha \partial \beta}\right] = -E\left[\frac{\partial^2 L(\alpha,\beta)}{\partial \beta \partial \alpha}\right] = 0. \tag{A.9}$$

Appendix B: Mean and Variance of a Simple Estimator Based on the Correlation of Amplitude Measurements

Recall that

$$\begin{aligned} \hat{\alpha}_{\mathrm{a}} &= \left|\frac{1}{K}\sum_{k=1}^{K} U_{1k} U_{2k}^*\right|^2 \\ &= \frac{1}{K^2}\sum_{k=1}^{K}\sum_{k'=1}^{K} U_{1k} U_{2k}^* U_{2k'} U_{1k'}^*. \end{aligned} \tag{B.1}$$

Therefore,

$$E[\hat{\alpha}_{\mathrm{a}}] = \frac{1}{K^2}\sum_{k=1}^{K}\sum_{k'=1}^{K} E\left[U_{1k} U_{2k}^* U_{2k'} U_{1k'}^*\right], \tag{B.2}$$

which can be evaluated, with the aid of the complex Gaussian moment theorem (see [9] or [5, p. 44]), as

$$E[\hat{\alpha}_{\mathrm{a}}] = \frac{1}{K^2}\sum_{k=1}^{K}\sum_{k'=1}^{K} \left[\alpha + \delta(k-k')\right]$$

$$= \alpha + \frac{1}{K}, \tag{B.3}$$

where $\delta(k-k')$ is the Kronecker delta function:

$$\delta(k-k') = \begin{cases} 1, & k = k' \\ 0, & k \neq k' \end{cases}.$$

Since $\frac{1}{K}$ can be subtracted from $\hat{\alpha}_a$, this estimator can be made to be unbiased for any K, and its variance must obey the CRLB. Also, for large K the mean approaches the true, underlying parameter. To evaluate the variance of this estimator we must first evaluate the expectation

$$E\left[\hat{\alpha}_a^2\right] = \frac{1}{K^4} \sum_{k=1}^{K} \sum_{k'=1}^{K} \sum_{m=1}^{K} \sum_{m'=1}^{K} E\left[U_{1k} U_{2k}^* U_{2k'} U_{1k'}^* U_{1m} U_{2m}^* U_{2m'} U_{1m'}^*\right]. \tag{B.4}$$

Notice that this expectation consists of the summation of K^4 terms. $K(K-1)(K-2)(K-3)$ of these terms will have none of their summation indices equal, $6K(K-1)(K-2)$ terms will have exactly two indices equal, $3K(K-1)$ terms will have two pairs of indices equal, but not equal to each other, $4K(K-1)$ terms will have three indices equal, but not equal to the other, and K terms will have all four indices equal. Since each term in the summation is scaled by K^{-4}, if we wish to ignore terms that depend on K^{-n} for $n > 1$ we only need to consider the terms that occur on the order of K^3 times. For these terms, the complex Gaussian moment theorem can again be used to evaluate the expectations. The specific examples and results are shown in Table 1. Therefore, to order K^{-1}, the expected value of $\hat{\alpha}_a^2$ is

$$\begin{aligned} E[\hat{\alpha}_a^2] &= \alpha^2 + \frac{4}{K}(\alpha^2 + \alpha) - \frac{4}{K}\alpha^2 \\ &= \alpha^2 + \frac{1}{K}(2\alpha^2 + 4\alpha), \end{aligned} \tag{B.5}$$

and the variance of $\hat{\alpha}_a$ is

$$\begin{aligned} \text{var}[\hat{\alpha}_a] &= E\left[\hat{\alpha}_a^2\right] - E^2[\hat{\alpha}_a] \\ &= \alpha^2 + \frac{1}{K}(2\alpha^2 + 4\alpha) - \alpha^2 - \frac{2\alpha}{K} - \frac{1}{K^2} \\ &\simeq \frac{1}{K} 2\alpha(1+\alpha). \end{aligned} \tag{B.6}$$

References

[1] R. Hanbury Brown and R. Q. Twiss. A new type of interferometer for use in radio-astronomy. *Philos. Mag.*, 45:663–682, 1954.

Table 1: Terms used to compute $E[\hat{\alpha}_a^2]$.

Equal Indices	Number of Terms	Expected Value
None	$K^4 - 6K^3 + 11K^2 - 6K$	α^2
$k = k'$	$K^3 - 3K^2 + 2K$	$\alpha^2 + \alpha$
$k = m$	$K^3 - 3K^2 + 2K$	$2\alpha^2$
$k = m'$	$K^3 - 3K^2 + 2K$	$\alpha^2 + \alpha$
$k' = m$	$K^3 - 3K^2 + 2K$	$\alpha^2 + \alpha$
$k' = m'$	$K^3 - 3K^2 + 2K$	$2\alpha^2$
$m = m'$	$K^3 - 3K^2 + 2K$	$\alpha^2 + \alpha$
$k = k',\ m = m'$	$K^2 - K$	don't care
$k = m,\ k' = m'$	$K^2 - K$	don't care
$k = m',\ k' = m$	$K^2 - K$	don't care
$k = k' = m$	$K^2 - K$	don't care
$k = k' = m'$	$K^2 - K$	don't care
$k = m = m'$	$K^2 - K$	don't care
$k' = m = m'$	$K^2 - K$	don't care
$k = k' = m = m'$	K	don't care

[2] R. Hanbury Brown. *The Intensity Interferometer.* Taylor and Francis, London, 1974.

[3] S. M. Ebstein. High-light-level variance of estimators for intensity interferometry and fourth-order correlation interferometry. *J. Opt. Soc. Am. A*, 8(9):1450–1456, 1991.

[4] J. W. Goodman. Statistical properties of laser speckle patterns. In J. C. Dainty, editor, *Laser Speckle and Related Phenomena, 2nd Edition*, pages 9–75. Springer-Verlag, Heidelberg, 1984.

[5] J. W. Goodman. *Statistical Optics.* John Wiley & Sons, New York, 1985.

[6] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I.* John Wiley, New York, 1968.

[7] L. L. Scharf. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis.* Addison-Wesley, Reading, Massachusetts, 1991.

[8] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice Hall, Englewood Cliffs, New Jersey, 1993.

[9] I. S. Reed. On a moment theorem for complex Gaussian processes. *IRE Trans. Inf. Theory*, IT-8:194–195, 1965.

PHASE RETRIEVAL VIA SENSOR FUSION IN INTERFEROMETRIC APERTURE SYNTHESIS

Mehrdad Soumekh
Associate Professor
Department of Electrical & Computer Engineering
State Unversity of New York at Buffalo
Amherst, NY 14260

Final Report for:
Summer Faculty Research Program
Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1993

PHASE RETRIEVAL VIA SENSOR FUSION IN INTERFEROMETRIC APERTURE SYNTHESIS

Mehrdad Soumekh
Associate Professor
Department of Electrical & Computer Engineering
State Unversity of New York at Buffalo
Amherst, NY 14260

Abstract

The objective of this work was to develop signal processing algorithms to remove phase ambiguity in interferomteric aperture synthesis for celestial imaging. A phase retrieval algorithm was developed and implemented that utilized the target information on the image plane from a smaller aperture, e.g., a telescope. The magnitude information in the target's spatial frequency domain obtained from the larger synthetic aperture were fused with the image plane data from the telescope to form an image that had the resolution of the synthesized aperture.

1. Introduction

Observing celestial targets using waves being emitted from them has been an active area of research since the invention of the telescope. In the 1920s, Michelson introduced a procedure for imaging celestial targets via an interferometric processing of the waves being radiated from them. Our work is concerened with phase retrieval in a synthetic aperture-based celestial imaging system that utilizes Michelson interferometry.

2. System Model and Inversion

We formulate the problem in the two-dimensional spatial domain with aperture synthesis in a one-dimensional linear array. We assume the target region is a planar structure on the line $x = X_1$ in the (x, y) domain. The receivers are located at $(0, u_i)$, $i = 1, \ldots, N$. The analog circuitry of th receivers is tuned such that they are capable of measuring a bandpass portion in the electromagnetic spectrum that is commonly in the visible band. For simplicity, we assume that the radiation pattern of each receiver is *omni-directional*. A general receiver radiation pattern can be incorporated in the following analysis; this will be shown later.

The i-th receiver records the following time-dependent signal from the radiating target:

$$s(u_i, t) \equiv \int_y p_y\Big[\, t - \frac{\sqrt{X_1^2 + (y - u_i)^2}}{c} \,\Big]\, dy, \tag{1}$$

where $p_y(t)$ is the bandpass portion of the wave that is transmitted by the target located at (X_1, y). This signal can be expressed in the following analytical form:

$$p_y(t) \equiv a_y(t)\ \exp(j\omega t),$$

where ω is the center frequency of the receiver's band.

Provided that $X_1 \gg y, u_i$ and $a_y(t)$ is a slowly fluctuating signal, we have the following Fresnel approximation for the spherical wave on the right side of (1):

$$\begin{aligned} p_y\Big[\, t - \frac{\sqrt{X_1^2 + (y - u_i)^2}}{c} \,\Big] =& a_y\Big[\, t - \frac{\sqrt{X_1^2 + (y - u_i)^2}}{c} \,\Big] \\ & \times\ \exp\Big[\, j\omega[t - \frac{\sqrt{X_1^2 + (y - u_i)^2}}{c}] \,\Big] \\ \approx& a_y(t - \frac{X_1}{c}) \\ & \times\ \exp(j\omega t - jkX_1)\ \exp[-j\frac{k(y - u_i)^2}{2X_1}]. \end{aligned} \tag{2}$$

The signal that is recorded by the imaging system is the magnitude of the optical correlation of the signals measured by the two receivers over an integration time $t \in [0, T_0]$ that is chosen by the user; that is,

$$\Gamma_{ij} \equiv |\int_0^{T_0} s(u_i, t)\ s^*(u_j, t)\ dt|, \tag{3}$$

where the coherent processing time, over which the receivers and the target are assumed to be motionless, is a parameter chosen by the user. This is referred to as ***interferometric*** processing of the aperture data.

Suppose the signals being emitted from the various points in the target region are uncorrelated over the time interval of observations; that is,

$$\int_0^{T_0} a_{y_1}(t)\ a^*_{y_2}(t)\ dt = \begin{cases} f(y_1) & \text{for } y_1 = y_2; \\ 0 & \text{otherwise.} \end{cases}$$

Using this assumption and the approximation in (2) in the expression for the interferogram of (3), one obtains

$$\begin{aligned} \Gamma_{ij} &\approx |\int_y f(y)\ \exp[-j\frac{k}{2X_1}[(y-u_i)^2 - (y-u_j)^2]]\ dy| \\ &= |\underbrace{\int_y f(y)\ \exp[-j\frac{k(u_i-u_j)}{X_1}y]\ dy}_{Fourier\ integral}| \\ &= |F[\frac{k(u_i-u_j)}{X_1}]|. \end{aligned} \tag{4}$$

Based on (4), the interferometric data provide the samples of $|F(k_y)|$ at

$$k_{yij} \equiv \frac{k(u_i - u_j)}{X_1}.$$

An important impasse for generating the spatial image, that is, $f(y)$, is the lack of knowledge of the phase of $F(k_y)$ in this imaging scheme. Recovery of phase of a signal from its magnitude has been an active area of research in celestial imaging. An energy reduction algorithm has been suggested by Gerchberg and Saxton that is based on a priori information on the properties of the target function (e.g., it is ***positive*** and ***space-limited*** in the spatial domain) to solve the phase ambiguity problem.

3. Phase Retrieval via Sensor Fusion

Next, we examine an iterative method to address the problem of phase ambiguity in interferometric aperture synthesis celestial imaging. This is achieved by fusing the image data from the physical aperture of a telescope with the interferometric synthetic aperture data. As we will show, this procedure is based on iterative corrections in the spatial and spatial frequency domain that is also the basis of the Gerchberg-Saxton algorithm. The additional data provided by the physical aperture of a telescope helps the convergence of the algorithm to the actual image function.

Suppose we are capable of obtaining a low resolution image of the target region via a single telescope with diameter $2L_r$, where L_r is by an order of magnitude smaller than L. We call this image the reference target image and denote it with $f_r(y)$.

- *The reference image is a real signal that is obtained by associating alternative positive and negative signs to the fringe patterns of the telescope's intensity image. For the processing that follows, the important information in the reference image is its phase function in the spatial frequency domain, that is, the phase of $F_r(k_y)$. It turns out that this phase function is approximately equal to the phase function of the spatial Fourier transform of the recorded amplitude function $|f_r(y)|$. Thus, the recovery of $f_r(y)$ from the fringe patterns of $|f_r(y)|$ is not essential.*

Based on the Rayleigh resolution limit of the telescope, we have the following for the Fourier transform of the reference target function:

$$F_r(k_y) = \begin{cases} F(k_y) & \text{for } |k_y| \leq \frac{kL_r}{X_1}; \\ 0 & \text{otherwise.} \end{cases}$$

The phase of the above function in the Rayleigh band of the telescope is the critical information that is used in the following procedure.

We combine the magnitude data from the larger interferometer with the phase data from the telescope to generate the following signal:

$$F_0(k_y) \equiv \begin{cases} F_r(k_y) = F(k_y) & \text{for } |k_y| \leq \frac{kL_r}{X_1}; \\ |F(k_y)| & \text{for } \frac{kL_r}{X_1} < |k_y| \leq \frac{2kL}{X_1}; \\ 0 & \text{otherwise.} \end{cases}$$

We call $F_0(k_y)$ or its inverse Fourier transform $f_0(y)$ the zero-th solution of the iteration.

The next step of the iterative solution is to mask $f_0(y)$ with the telescope's intensity image; that is,

$$f_1(y) \equiv f_0(y)\ |f_r(y)|.$$

This operation is similar to the imposition of the space-limited constraint on the target's extent that is used in the Gerchberg-Saxton iterative solution. Other constraints on $f(y)$, for example, its dynamic range, may be imposed in this step.

Next, we take this solution in the spatial frequency domain. Suppose we have

$$F_1(k_y) \equiv |F_1(k_y)| \ \exp[j\phi_1(k_y)],$$

where $\phi_1(k_y)$ is the phase of $F_1(k_y)$. The solution at the next step of the iteration is

$$F_2(k_y) \equiv \begin{cases} F_r(k_y) = F(k_y) & \text{for } |k_y| \leq \frac{kL_r}{X_1}; \\ |F(k_y)| \ \exp[j\phi_1(k_y)] & \text{for } \frac{kL_r}{X_1} < |k_y| \leq \frac{2kL}{X_1}; \\ 0 & \text{otherwise.} \end{cases}$$

In general, we have the following operation at the odd steps of the algorithm:

$$f_{2n+1}(y) \equiv f_{2n}(y) \ |f_r(y)|.$$

The solution at the even steps of the iteration has the following form:

$$F_{2n}(k_y) \equiv \begin{cases} F_r(k_y) = F(k_y) & \text{for } |k_y| \leq \frac{kL_r}{X_1}; \\ |F(k_y)| \ \exp[j\phi_{2n-1}(k_y)] & \text{for } \frac{kL_r}{X_1} < |k_y| \leq \frac{2kL}{X_1}; \\ 0 & \text{otherwise,} \end{cases}$$

where $\phi_{2n-1}(k_y)$ is the phase of $F_{2n-1}(k_y)$.

Similar to the Gerchberg-Saxton method, the convergence of this algorithm cannot be shown. The main difference between this method and the Gerchberg-Saxton algorithm is the use of the phase of the reference signal in the spatial frequency domain to provide a ***fair*** initial guess for the algorithm. The use of an appropriate zero-th solution, that is, the initial guess, is a critical issue for the convergence of iterative reconstruction methods.

THE IMPACT OF CONDENSING WATER VAPOR IN CHEMICAL OXYGEN IODINE LASERS.

Philip D. Whitefield
Research Associate Professor
Departments of Physics and Chemistry
and
W. Mark Barnett
Graduate Student
Department of Chemistry

Cloud and Aerosol Sciences Laboratory
University of Missouri - Rolla
Norwood Hall G11
Rolla, MO 65401

Final Report for:
Summer Faculty Research Program
Phillips Laboratory (PL/LIDB)

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

and

University of Missouri - Rolla

September 1993

THE IMPACT OF CONDENSING WATER VAPOR IN CHEMICAL OXYGEN IODINE LASERS.

Philip D. Whitefield
Research Associate Professor
Departments of Physics and Chemistry
and
W. Mark Barnett
(Graduate Student - Chemistry)
Cloud and Aerosol Sciences Laboratory
University of Missouri - Rolla

Abstract

This final report describes the results and interpretation of a joint research project sponsored by (i) AFOSR through its SRE, SFR and GSR programs and (ii) the University of Missouri - Rolla, to investigate the impact of condensing water vapor in the reacting flow regimes of chemical oxygen iodine lasers (COIL's). The "condensation shock phenomenon" first reported in 1991 has been successfully simulated and recreated under non-reacting flow conditions in two independent COIL devices. The two-phase singlet oxygen generator has been identified as the major source of particulates (condensation nuclei), especially those upon which heterogeneous nucleation can take place. These particulates have been characterized in terms of size distribution, total concentration, hydration and dependence on typical generator operating parameters such a chlorine flow rate, disc rotation rate, basic hydrogen peroxide concentration and temperature. Based on the heterogeneous nucleation characterization and known heat released into the laser supersonic flow during the "condensation shock" it is reasonable to conclude that both homogeneous and heterogeneous nucleation are responsible for water vapor condensation in the COIL devices, however, homogeneous nucleation is the dominant step in "condensation shock". Furthermore, the onset of the shock is readily observed at very low BHP temperatures.

THE IMPACT OF CONDENSING WATER VAPOR IN CHEMICAL OXYGEN IODINE LASERS.

Philip D. Whitefield and Wm. Mark Barnett

INTRODUCTION

Water vapor, when supersaturated, will spontaneously condense both heterogeneously and homogeneously, growing water droplets or aerosols provided the appropriate supersaturation is sustained. Heterogeneous nucleation requires the presence of condensation nuclei (CN) in the flow. The soluble mass fraction of the CN will determine its critical supersaturation spectrum which in turn determines the specific water vapor supersaturation for a given CN to grow without bound. Heterogeneous nucleation occurs at low supersaturations of water vapor typically < 5%. At very high water vapor supersaturations homogeneous nucleation will occur. In this case it is thermodynamically favorable for molecules of water vapor to spontaneously cluster to form aerosols. In both cases the concomitant heat of condensation will be released into the gas flow. The criteria for both types of nucleation exist in supersonic COIL's and a massive heat release has been reported immediately following the supersonic expansion in several devices, especially where the water vapor pressure in the gas flow was thought to be high (typically 5 mole%).[1] Any increase in temperature in the cavity will manifest itself as a ramp in the laser cavity pressure profile, and will diminish the gain of the laser and reduce the optimum laser performance.[2] The relative importance of either type of nucleation channel must be determined for COILs if the impact of condensing water on the devices is to be characterized.

This report describes how the UMR Mobile Aerosol Sampling System (MASS) is being used in conjunction with a small scale supersonic COIL device (SSSC) to continue a study, initiated in 1992,[3,4] investigating the impact of condensing water vapor on the performance of COIL devices. The conclusions and recommendations from the 1992 study were as follows[3,4]:

- Heterogeneously nucleated aerosols are present in COILs.
- They are mechanically produced in the generator and their production rate and

size are governed by generator operating parameters.

- The relationship between these parameters and aerosol production should be explored in detail.
- The limits of detectability of aerosols using the UMR-MASS in its current state of the art configuration may have led to an underestimate of the total concentration of aerosols produced in the generator. As a result of the studies to date, design up grades should be made to the sampling probe system and counting system to customize the UMR-MASS to this COIL investigation

- The inability to simulate the pressure ramp in the supersonic nozzle flow system indicates that the pressure ramp is not a simple gas dynamic effect arising from the enthalpy of condensation associated with a homogeneous nucleation event.

- The calculations upon which the homogeneous nucleation rational are based should be re-examined.

- The principles of operation and performance of the spray generator concepts, as singlet oxygen generators for COIL's, should be examined to determine whether submicron aerosol formation will occur in these devices and whether it will affect their predicted performance.

The investigative approach in the preliminary study relied upon the UMR MASS to characterize any heterogeneous nucleation. In this very limited study although heterogeneous nucleation was observed it was not possible to correlate nucleus formation to generator operating parameters, or to quantify total nucleus production concentrations. As a result, it was impossible to determine what proportion of the heat release driving the pressure ramp in the cavity was due to heat of heterogeneous condensation. Using the experience gained from the preliminary studies a methodology was rationalized, to complete the characterization of the heterogeneous processes. The methodology and results of these experiments are described below.

Attempts were also made, in the preliminary study, to simulate or recreate the cavity pressure profile ramp downstream of a particle free supersonic expansion of a water laden gas flow, the rationale being that the cavity pressure profile ramp was a physical indication of a

purely gas dynamic "condensation shock"[5]. These experiments were not successful. Once again, based on the experience gained in the preliminary study a methodology was rationalized and a series of experiments were carried out to test the negative results of the preliminary study. This approach has led, in 1993, to several successful demonstrations of the condensation shock phenomena. The methodology and results of these experiments are, also described below.

METHODOLOGY

The basic methodology behind the condensation shock phenomenon experiments has been described previously [3,4] and will not be repeated here. A novel approach in the 1993 studies involved the use of a pitot tube pressure sensor to define the boundary layers in the various COIL's examined in this study. The pitot tube methodology is described below. The heterogenous nucleation experiment methodologies are those of the UMR MASS and have been described in detail elsewhere [6-8] and therefore only a brief summary will be given here. The setup and operation of the laser device and supersonic nozzle facility is not described in this report. Detailed descriptions of this activity can be found elsewhere [9,10].

The Pitot Tube - design and principles of operation: In order to better understand the nature of the boundary layers in the nozzle and cavity flow regimes of the various COIL devices at our disposal during this study, a pitot tube was designed and constructed to measure the total pressure profile as a function of the vertical displacement of the pitot tube at various positions along the flow axis of the cavity, but in particular close to the expansion nozzles. Flow velocity can be interpreted in terms of "static" and "dynamic" pressure at any point in a flow regime. "static" pressure is measured at the cavity walls perpendicular to the supersonic flow direction. It is assumed to be fairly constant throughout a given cross-section. The "dynamic" pressure is extracted from pressure measurements made with the pitot tube. The pitot tube measures the total pressure of the supersonic flow by sampling parallel with and into the

supersonic flow direction. The thrust or "dynamic" pressure attributed to the supersonic flow is the difference between the total and "static" pressures, measured at any point in the flow. A schematic diagram of the "static" and total pressure measurement approach is given in figure (1). The orifice of the pitot tube has a diameter is 8.1 x 10^{-5}m and all pressure measurements were made using MKS Baratron pressure transducers calibrated against an NIST standard.

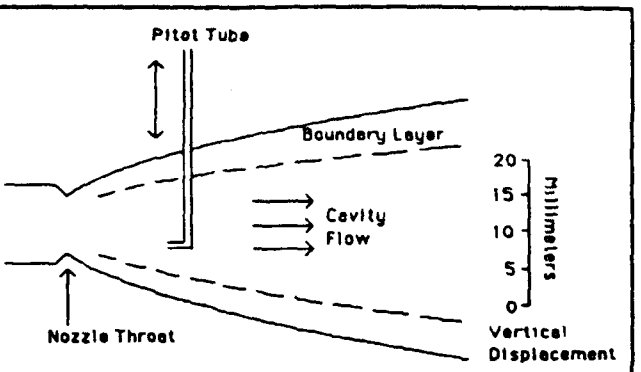


Figure (1). Pitot Tube Experimental Approach.

For the purposes of this experiment the boundary layer is defined in terms of the pitot tube measurements as that region of the vertical cross section "dynamic" pressure profile where the difference between the "dynamic" pressure and the "static" pressures approaches zero.

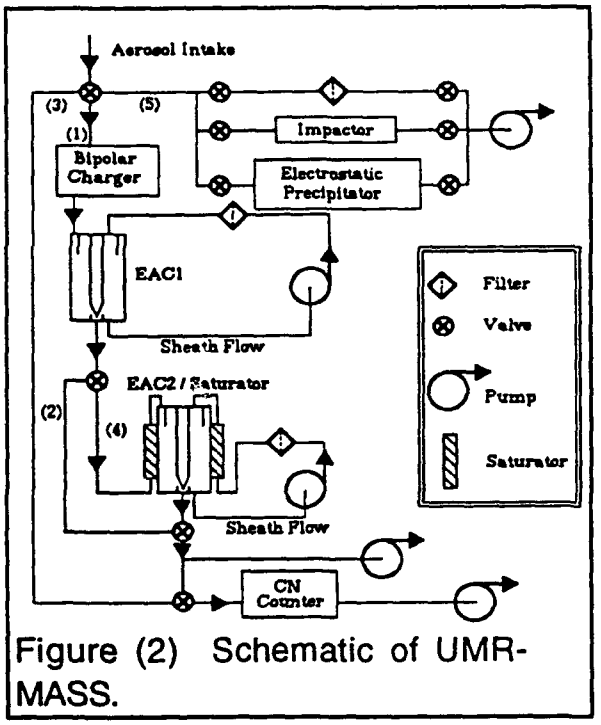


Figure (2) Schematic of UMR-MASS.

Heterogeneous Nucleation Characterization Using the UMR/MASS: The Mobile Aerosol Sampling System (UMR-MASS), figure (2) was employed to characterize aerosols in the gas flow of a COIL in terms of their total number density, size distribution and/or hydration properties. The MASS was operated in its rapid sample acquisition/storage mode, with off line sample analysis. This mode is employed under test conditions where it is not possible to characterize aerosol in real-time because the aerosol generation system (in this case the COIL) cannot be run continuously with fixed operating parameters for the time periods required to perform the characterization tests in real time. Typical minimum continuous operation times of 3 minutes are required. The experimental sequence of events is as follows:

(a) Particulate Concentration Size Distributions - The particulate stream emanating from the sampling probe and dilution facility is sent through, initially, an alcohol counter to examine the total size range of aerosols sampled and determine the total concentration. For a size distribution the aerosol is charged with a bipolar charger and is then sized via its electric mobility using an electric aerosol classifier (EAC). The EAC passes only particles contained in a narrow size range (typically 10% of mean size). A specific particle size is set by the EAC voltage. This voltage and hence the selected particle size is variable. The resulting monodisperse particle stream is then passed through the alcohol counter where it is exposed to a fixed supersaturation of alcohol of sufficient magnitude to condense alcohol on the particle, thus forming aerosols large enough to be detected by the optical particle counter, OPC, sub system of the counter. The enlarged aerosols are passed through the OPC where individual aerosols are counted and the aerosol concentration is determined.

(b) Hydration Properties - For particulate hydration property measurements, specifically the critical supersaturation spectrum (i.e. the minimum supersaturation of water that will cause condensation of water onto a particle to produce a continuously growing droplet presented as a function of particle diameter) is measured. The particulate stream is first passed through an EAC to select a specific size element of the sample distribution. The resulting monodisperse element is then passed through a saturation chamber holding water vapor at 100% relative humidity. As the aerosols pass through the chamber they deliquesce and increase in diameter. This growth is monitored by passing the droplets, after they exit the saturator, through a second EAC, the output of which can be monitored by the alcohol counter. Should a particulate prove difficult to hydrate (i.e. require significant supersaturation (>1%) of water a continuous flow diffusion chamber (CFD) is employed. (For a more detailed description of the overall UMR approach to particulate measurements, see references 11-14).[11-14]

RESULTS

Simulation of "Condensation Shock" in COILs operated under cold flow conditions.

The Small Scale Supersonic Device (Preliminary Results): In the preliminary study, [3,4] of the condensation shock phenomenon an SSSC was used to attempt the simulation. A schematic diagram of this device indicating the various components, flow regimes and sampling ports is given in figure (3). During the 1992 experimental campaign this device was considered to be identical to the COIL device in which the pressure ramp had originally been detected[1] and similar results were expected from the SSSC. As reported initially [3,4] no pressure ramp could be detected. This was later explained by an over-simplification of the gas mixtures employed in those initial simulation studies where just nitrogen and water vapor were employed. The actual laser mixture constituents are helium, oxygen and water vapor, with helium as the bath gas constituting the major part of the flow (75%)[10]. In retrospect, it was reasonable to substitute oxygen with nitrogen but not helium with nitrogen. The significant difference in their heat capacities forfeits such a substitution ($C_{(Nitrogen)}$ = 29.124 J/[mole•K] , $C_{(Helium)}$ = 20.786 J/[mole•K]).[15]

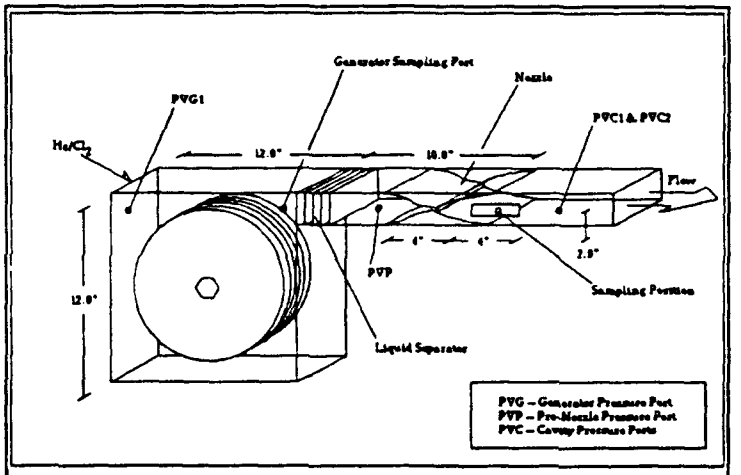


Figure (3) Schematic of Small Scale Supersonic COIL (SSSC).

Further simulation experiments were performed after the submission date of the 1992 report [3] using helium, nitrogen, water vapor mixtures. They were successful in detecting pressure ramps in the cavity, however, the increases in pressure were small compared to the original experimental results. These data have subsequently been published elsewhere[16]. A typical maximum pressure increase was found to be on the order of 10-20% significantly smaller than the originally reported 40% [1]. The assumption that the devices could be considered identical was re-examined. There was some indication that the supersonic expansion nozzle configurations were different. If this were so it would be reasonable to assume that the boundary layers, created in the nozzle region during the supersonic expansion, in each of these devices could be different. Dissimilar boundary layers would explain the difference in the

pressure ramps since these boundary layers define the supersonic flow cross-section and the effective gas volume into which any heat of condensation could be released. For example a thick boundary layer will reduce the effective supersonic flow volume which in turn will result in a diminished pressure ramp.

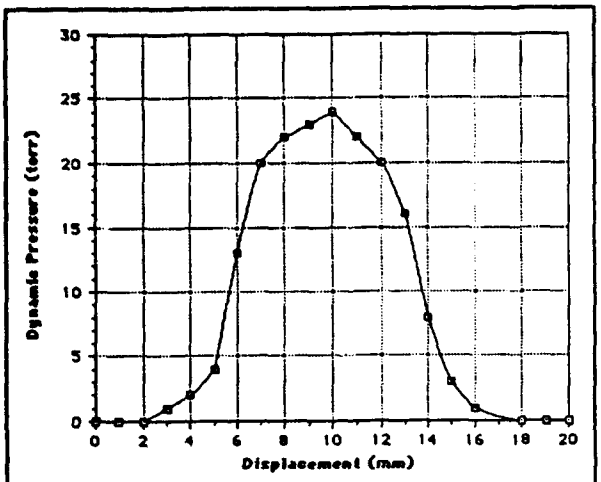


Figure (4) Boundary Layer of SSSC Cavity Downstream of Throat.

Boundary Layer Definition: Before proceeding further with the simulation experiments it was decided that the boundary layer issue should be resolved. The pitot tube pressure sensor, described above, was used to measure the "dynamic" pressure, as a function of the vertical displacement in the cavity close to the supersonic nozzle exit plane in the small scale device. These data are presented in figure (4), where the flow parameters were, for the primary flow: He - 110 x 10^{-3} mole.s^{-1}, N_2 - 35 x 10^{-3} mole.s^{-1} and the secondary flow: He - 60 x 10^{-3} mole.s^{-1} with the pitot tube port located 2.5 inches (0.06 meters) downstream of the nozzle throat. These data clearly reveal a massive boundary layer occupying as much as 50% of the cavity flow volume in the region of the vertical displacement of the pitot tube. The rounded appearance of the pressure profile indicates that the boundary layers converge. The profile was only measured at one position in the cavity however successive measurements demonstrated excellent reproducibility.

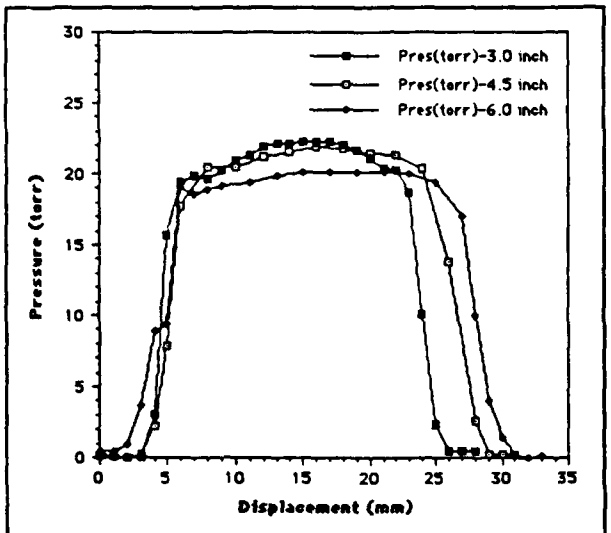


Figure (5) Boundary Layer for RADICL Downstream of Throat.

It was not possible to define the actual boundary layer conditions of the device in which the original shock observations were made.[1] However, another COIL device (RADICL) [9, 10], was available, although, much larger than the SSSC device that had been used to date. The boundary layer for this device was characterized using the pitot tube method

described above. The data from these experiments is given in figure (5). In the case of the RADICL measurements, profiles were obtained at four discrete locations along the flow axis in the cavity. Even though the geometry of the RADICL cavity differs slightly from that of the SSSC device, it is clear that the boundary layer in RADICL is much smaller. Also, in the region of high total pressure (typically 85% of the flow volume) the profiles are flat indicating the supersonic core is broad and uncompromised.

Condensation Shock Simulation with RADICL: The boundary layer results for RADICL indicated that a measurable shock would be more likely to be observed in RADICL than in the SSSC device, therefore, shock simulation studies were transferred to RADICL. Unlike the small scale device RADICL cannot be run by one or two individuals with a few minutes notice. Instead all experiments have to be scheduled in advance or have to be "piggybacked" on previously scheduled tests. Nevertheless, data were obtained. The shock indicator, as described previously, is a change in magnitude of the ratio of the "static" pressure in the supersonic cavity to that in the subsonic generator region ($P_{cav} : P_{gen}$). In the absence of a condensation shock, a condition that can be achieved by eliminating all water vapor from the flow, this ratio remains constant for a fixed set of run conditions. A typical plot of the constant pressure ratio profile as a function of time is given by the solid line in figure (6a). When water vapor is added to the flow, if the condensation shock model is correct, at an appropriate water vapor concentration, to be determined, the threshold supersaturation of water vapor in the cavity will exceed that required for homogeneous nucleation. Nucleation occurs instantaneously under these conditions and is accompanied by a rapid release of heat (heat of condensation), this is the so called shock. This heat is released into the supersonic gas flow and manifests itself as a localized pressure increase in the cavity.

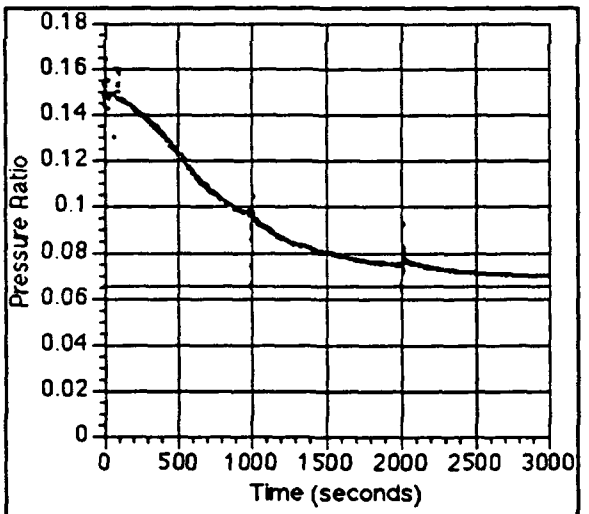


Figure (6a) RADICL Pressure Ratio Profile with Respect to Time.

A thermally controlled 30% aqueous solution of potassium hydroxide, initially cooled to 243K, was used as a variable water vapor concentration source. The cold solution was placed in

the RADICL generator, a helium nitrogen (3:1) mixture was passed at subsonic speeds through the generator and duct, where it entrained water vapor, the concentration of which was controlled by the bulk solution temperature. In the absence of any active cooling of the KOH solution, the solution temperature steadily increased with time as it sought to come to thermal equilibrium with its surroundings. During this period the KOH solution temperature, and the pressure ratio of the cavity to the duct were monitored. The pressure ratio profile is plotted in figure (6a) and the temperature profile is plotted in figure (6b).

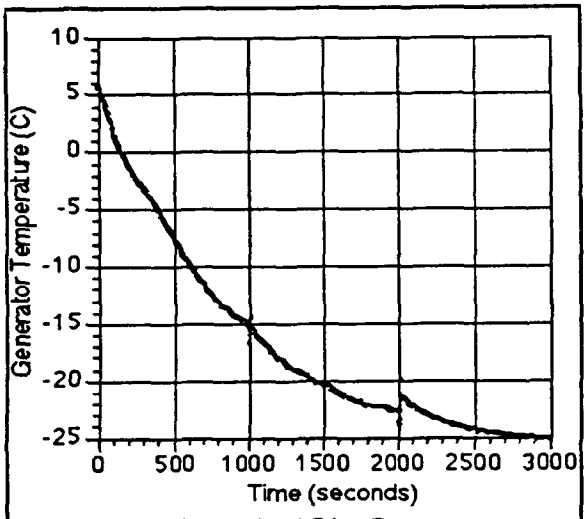


Figure (6b) RADICL Generator Temperature Profile with Respect to Time.

Characterization of Heterogeneous Nucleation: Based on the findings of the preliminary study of heterogeneous nucleation [3, 4] the measurements in the current study were confined to samples taken in the subsonic, post-generator, pre-cavity regions of the SSSC device (see figure (3)). Since it was clear from the preliminary study, that the particulate production rate was not coupled to either, the bulk temperature in the generator, or the presence of iodine in the flow these parameters were not re-investigated in the current study. As previously reported the largest particulate production rates occurred in the immediate vicinity of the rotating disc generator during the passage of chlorine gas over the BHP wetted disc walls. In the current study, therefore, the parameters: disc rotation rate, BHP concentration as a function of the age of BHP mixture, neutralization salt concentration, also as a function of the age of the BHP mixture and chlorine flow rate were examined for any correlation to particulate production rate.

Disc Rotation Rate: Disk rotation rates were varied within the range 0-45 rpm (where 45rpm is considered the optimum disc rotation rate for singlet oxygen production). The discs were rotated in a BHP solution but in the absence of a chlorine flow. The particle number densities increased with increasing rotation rate, however, at 45rpm the maximum particle number densities recorded did not exceed 5 cm^{-3}

BHP Concentration: During the course of a days testing on a single batch of BHP, periodic samples of BHP were drawn from the generator and titrated for the hydroperoxide anion [17]. These titrations repeatedly showed the concentration of the hydroperoxide ion to be relatively constant during the lifetime of the mixture during which time the BHP was being exhausted. For example a typical batch had a 9.5M hydroperoxide ion concentration, initially, and a 9.1M concentration at the end of the test, with a 9.3M mid-test value.

Neutralization Salt Concentration: As a batch of BHP solution is exhausted, during the course of a days testing, large concentrations of potassium chloride (KCl) build up in solution and eventually start to precipitate as the BHP solution approaches saturation with respect to KCl (~2% mass solubility in BHP). Observations of the particulate number densities present in the generator measured, between chlorine flow periods, but with a disc rotation rate of 45rpm in an aging BHP mixture demonstrated a reproducible in crease in particulate number density with KCl build-up. For example a fresh BHP solution would yield typically 5 cm^{-3} at 45rpm and after 5 chlorine runs this number would increase to 60 cm^{-3}.

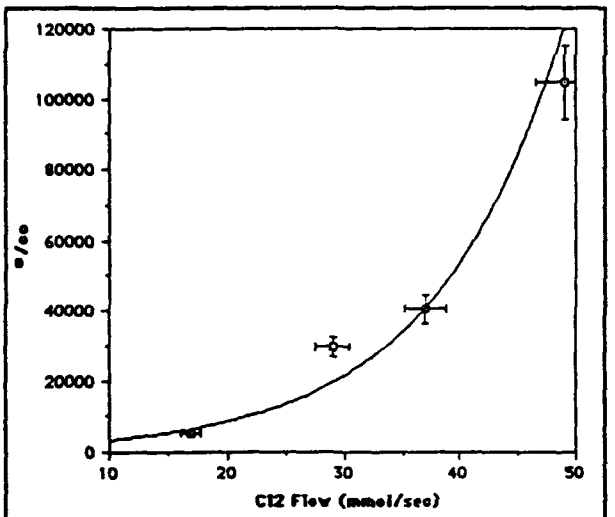


Figure (7) Particulate Concentration as a Function of Chlorine Flow Rate.

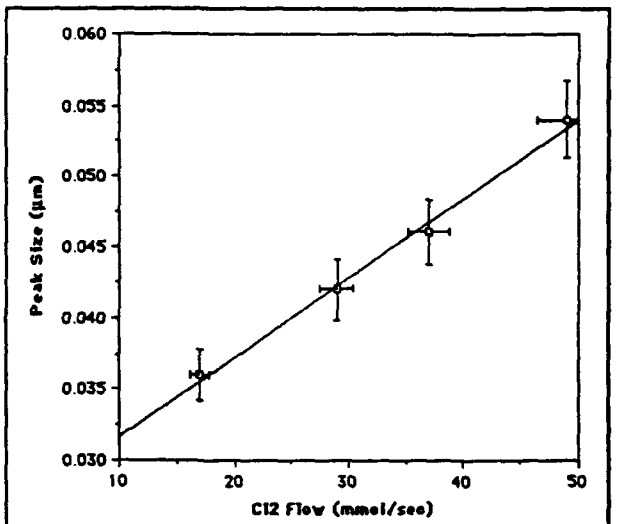


Figure (8) Peak Particle Diameter as a Function of Chlorine Flow Rate.

Chlorine Flow rate: At a typical chlorine flowrate of 37 x 10^{-3}moles.s^{-1} and a total generator pressure of 56 torr the mean particulate concentration is 1.45 x 10^4 cm^{-3}. This is three orders of magnitude greater than the concentrations generated by any other process. The total concentration of particles is determined either by integrating the size distribution function

over a specific size range or by monitoring the total concentration of particulates with a CNC counter. Both methods were employed in this study and good agreement was obtained between them. Figure (7) is a plot of the total concentration as a function of chlorine flow rate. The particulate yield correlates exponentially with the chlorine flowrate over the flow range studied (i.e. 0 - 50 x 10^{-3}moles.s^{-1}, yielding a maximum particulate concentration in excess of 1 x $10^{6}cm^{-3}$. These data are a clear indication that chlorine flow rate is the primary generator operating parameter controlling particulate production. The peak size of the particulates demonstrates a linear correlation with the chlorine flow rate, indicating a shift in the size distributions to larger mean particulates at higher chlorine flows. Typical data are presente in figure (8).

DISCUSSION

Condensation Shock Simulation

These data represent the first detailed investigation of the potential for, and characteristics of "condensation shock" in COIL devices. The simulation experiments described here, when combined with the findings of the heterogeneous nucleation studies, also described here, lead to the following inferences.

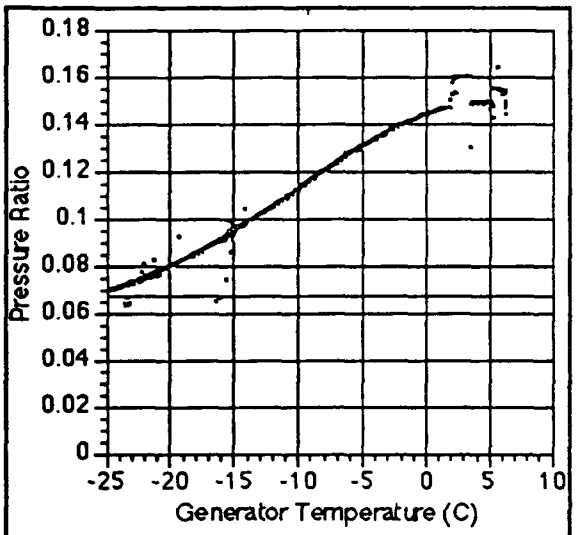


Figure (9) RADICL Pressure Ratio as a Function of Generator Temperature.

A gas dynamic condensation shock can be induced in COIL devices in the absence of any condensation nuclei. Figure (9) is a plot of pressure ratio versus temperature for RADICL derived from the data in figures (6a) and (6b). The dry system pressure ratio line is included for comparison. The P(ratio) versus T plot can be seen to diverge from the "dry line" as T increases. The rate of divergence increases from 253K to 268K. It, then, slows and approaches zero at T = 278K. This divergence is direct evidence of the onset of condensation shock and can be entirely attributed to the heat released during the "shock". Calculations using the 3-D MINT code by Plummer and coworkers at RDA-Logicon [18] successfully model these results and explain the asymptotic behavior as the onset of choked flow in the cavity due to the heat of condensation released into the flow during the shock.

The concentrations of aerosols emanating from the generator,($\sim 10^6$ cm^{-3}), are not large enough to contribute significantly to the total heat released into the flow through condensation. It is reasonable to assume, however, that heterogeneous and homogeneous condensations are taking place simultaneously in COILS. Furthermore, in the case of RADICL, evidence of the shock (in the P ratio vs T plot) at very low temperatures (i.e. low partial pressures of water) is clearly observed. It is not possible to determine the total water concentration with the current experimental configuration. An upper limit for this concentration could be taken as the equilibrium vapor pressure of water above the KOH solution. A sensitive absolute water vapor concentration diagnostic is being constructed currently. It will be used to profile the massive change in the water vapor concentration in the laser cavity at the onset of the shock. The latter will be monitored by the pressure versus temperature profile as in figure (9). Such data will unequivocally validate the theory of condensation shock as it affects COIL devices. These experiments are planned for this fall when the state of the art water vapor diagnostic is to be delivered. It is hoped that these results will then be incorporated into the AIAA paper planned for publication of this work and the SREP contract report due in December 1993.

Heterogeneous Nucleation

The results of this study and the preceding preliminary study clearly demonstrate that heterogenous nucleation is not necessary to create a condensation shock in COIL devices. They further demonstrate, that the maximum concentration of particulates generated under typical COIL operating conditions, although large, is orders of magnitude too small to account for the heat release detected via the pressure ratio divergence data as seen in figure (9). These nucleation data, however, are important. A production mechanism was suggested in the preliminary study: " These results indicate that heterogeneously nucleated aerosols are mechanically generated during the explosion of gas bubbles on the basic hydrogen peroxide (BHP)/chlorine interface and that certain generator operating parameters such as disc velocity and degree of neutralization of the BHP can strongly influence the aerosol production rate." [3,4]. The data from the current study support this mechanism but take our understanding of the process further. The particulate production rate and size distributions depend primarily on the chlorine flow rate. The data acquired to data indicates the dependence to be exponential for total concentration and linear with respect to the shift in the peak size of the size distribution. The sampling techniques and dilutions employed in this study eliminate the potential for agglomeration of particulates, that might skew the data in such a manner [19].

CONCLUSIONS

The following conclusions can be drawn from this study:

- Condensation shock occurs in COIL devices even at extremely low water vapor concentrations.
- The onset of the shock will directly impact the gain of the device in which it occurs.
- Heterogeneous nucleation is controlled primarily by the chlorine flow rate into the generator.
- Nuclei production rates grow exponentially with increasing chlorine flow rate.
- Nuclei diameters increase with increasing chlorine flow rate (0.033-0.055 μm for the chlorine flow range investigated in this work).
- The onset of foaming in the generator is indicative of massive sub-micron nuclei production rates.
- The next generation of COILs may employ spray generators. These generators should be examined to define their characteristics with respect to condensation shock and heterogeneous nucleation if their performance is to be accurately compared to current wetted-wall type generators as are found in the SSSC and RADICL.

REFERENCES

(1) Private Communication, D. Plummer RDA, 19 July 1991.

(2) RDA Memo 87-A/K-3-02-752 Particulates/Aerosols in Oxygen/iodine Lasers, P.G. Crowell, January 1989.

(3) Whitefield, P.D., AFOSR-SFRP Final Report September 1992.

(4) Whitefield, P.D., AFOSR-SREP Final Report December 1992.

(5) Fundamentals of Gas Dynamics, Editor H.W. Emmons, Princeton University Press, 1958.

(6) Hagen, D.E., M.B. Trueblood, and J. Podzimek, "Combustion Aerosol Scavenging", Atmos. Env. 25A, 2581 (1991).

(7) Hagen, D.E., Whitefield P.D., and M.B. Trueblood, "A field Sampling of Jet Exhaust Aerosols." Particulate Science and Technology, 10, 53, 1992.

(8) Whitefield, P.D., M.B. Trueblood and D.E. Hagen, "Size and Hydration Characteristics of Laboratory Simulated Jet Engine Combustion Aerosols," Pariculate Science and Technology, 11, 25, 1993.

(9) Review of Modern Physics vol. 59, No.3 paper II, July 1987 pp 541-546

(10) Oxygen Iodine Supersonic Technology Program, Part III - Management and Technical

Proposal, MDRL/MDAC Report No. D9299-953003, June 1982.

(11) Alofs, D.J., "Performance of a dual-range cloud nucleus counter", J. Appl. Meteor. 17, 1286-1297 (1978).

(12) Alofs, D.J., M.B. Trueblood, D.R. White, and V.L. Behr, "Nucleation experiments with monodisperse NaCl aerosols", J. Appl. Meteor. 18, 1106-1117 (1979).

(13) Hagen, D.E., and D.J. Alofs, "A Linear Inversion Method to Obtain Aerosol Size Distributions from Measurements with a Differential Mobility Analyzer", Aerosol Sci. and Tech. 2, 465-475 (1983).

(14) Alofs,D.J. and M.B. Trueblood, "UMR Dual Mode CCN Counter (Modes:CFD plus Haze)," J. Rech. Atmos. 15, n' 3-4, 219-223 (1981)

(15) CRC Handbook of Physics and Chemistry 73rd Edition, 1992.

(16) Whitefield, P.D., D.E. Hagen, M.B. Trueblood, W.M. Barnett and C. Helms, SPIE vol 1871 Intense Laser Beams and Applications (1993) pp 277-288.

(17) Richardson, R.J., Wiswall, C.E., Carr, P.A.G., Hovis, F.E., and Lilenfeld, H.V., "An Efficient Singlet Oxygen Generator For Chemically Pumped Iodine Lasers", J. Appl. Phys. 52, 4962-4969, (1981).

(18) RADICL Data presented at a briefing on the condensation shock phenomena by D. Plummer RDA, August 1992.

(19) Whitefield, P.D., D. E. Hagen and H.V. Lilenfeld Geophysica Acta (submitted September 1993)

SPECIMEN SIZE EFFECTS ON FRACTURE PARAMETERS FOR SOLID PROPELLANT MATERIALS

Somnath Chattopadhyay
Department of Mechanical Engineering
201 Votey Building
The University of Vermont
Burlington, VT 05405-0156

Final Report for:
Summer Faculty Research Program
Phillips Laboratory
Edwards Air Force Base, CA.

Sponsored by:
Air Force Office of Scientific Research
Bollings Air Force Base, Washington, DC.

August 1993

SPECIMEN SIZE EFFECTS ON FRACTURE PARAMETERS FOR SOLID PROPELLANT MATERIALS

Somnath Chattopadhyay
Department of Mechanical Engineering
The University of Vermont
Burlington, Vermont

Abstract

Two dimensional finite element elastic and inelastic analyses were performed on solid propellant tension specimens to compute the J-integrals. Single-edge notched specimens of constant width and thickness were studied. The lengths of the specimens were varied to study the size effect. A wide range of crack depths was used for a given specimen length. The mode I stress intensity factors for various specimen geometries were calculated from the elastically computed J-integral. Experimental stress-strain curve was used to perform the inelastic analyses. For a nonlinear material with linear strain hardening, two dimensional plane stress inelastic analyses were performed to compute the J-integrals and the load-deflection curves. The specimens were assumed to be thin so that a plane stress representation could be justified. From the initial slopes of the load-displacement curves, the elastic compliance of specimens of different lengths as a function of the crack depth was determined. The J-integral and the load-displacement curves were used to calculate the "eta" factors for various specimen sizes as a function of crack depth. These factors can be readily employed to establish an approximate single specimen procedure for J-integral calculations. Of particular significance is the assumed incompressibility of the solid propellant material which was modeled by assuming a Poisson's ratio approaching 0.5 for both elastic and inelasic analyses. Comparisons have been made wherever possible with analyses with Poisson's ratio of 0.25, which corresponds to common structural materials.

SPECIMEN SIZE EFFECTS ON FRACTURE PARAMETERS FOR SOLID PROPELLANT MATERIALS

Somnath Chattopadhyay

Introduction

Standardized procedures for fracture toughness testing require sufficient specimen thickness to ensure predominantly plane strain conditions at the crack tip. However no requirement currently exists to address the specimen length to be used in fracture toughness testing. In this study, a range of specimen lengths has been employed to study the variation of fracture parameters as a function of the crack depth. For work-hardening materials, irrespective of the applied loads and geometric effects, J integral provides scaling factors for crack tip stresses and strains and also ensures very severe conditions for fracture as described by the Hutchinson-Rice-Rosengren (HRR) crack tip fields [1,2]. These conditions make the applied driving force needed to initiate fracture in a laboratory specimen lower than the value needed to initiate fracture in the real structure where such severe stress singularities are generally not expected to be present. Therefore such structures can be expected to carry higher loads without failure than what is predicted from fracture toughness information using standardized procedures. Such procedures, however do not specifically address size requirements in terms of the length of the specimen. To this end the study is directed. In this work, a range of specimen lengths and crack sizes have been employed to obtain the fracture parameters. The evaluation of elastic-plastic material parameter, Jic and their application for an assessment of defects in a component require the knowledge of local stresses and strains in terms of J as functions of the prevailing geometry and loading. Generally, only forces and displacements can be measured at points far away from the crack tip. Therefore calibration experiments are needed and computational concepts have to be developed and checked to correlate the measured quantities with the crack tip field.

Only single-edge notched specimens subjected to uniform displacement at the boundaries have been studied. The specimens are assumed to be thin so that a plane stress representation is justified. Mode I stress intensity factors for various specimen geometries have been obtained from the two dimensional elastic plane stress J-integral values. The J-integrals have been computed using a finite element procedure. The finite element program computes the loads associated with the displacement

loadings. From this information, the non-dimensional geometric factor associated with the mode I stress intensity factors for various specimen lengths have been obtained as a function of crack depth. This information is useful for situations involving small-scale yielding. In order to consider situations where significant yielding takes place, the experimental data on nonlinear stress strain characteristics of the material have been used. For a material with linear strain-hardening, two dimensional plane strain inelastic analyses were performed to compute the J-integral and the load-deflection characteristics for various specimen sizes. From this information, the "eta" factors relating the J-integral with the strain energy has been calculated for various specimen sizes as a function of the crack depth. These factors establish an approximate single specimen procedure for J-integral evaluations and subsequent structural integrity assessments.

Specimens and Material Properties

Single edge-notched tension specimens of thickness 0.1" and width (W) of 2" were used. The lengths of the specimen were taken to be 1", 2", 3", 4" and 5". Each of the specimens were cracked to a relative crack depth, a/W = 0.1, 0.2, 0.3, 0.4 and 0.5 (See Figure 1). Solid propellant material was used for which the stress-strain curve is shown in Figure 2. An equivalent bilinear stress-strain characteristic was approximated from this curve and was employed for inelastic analyses. The elastic modulus, E was taken to be 1250 psi (corresponding to an assumed yield strain of 4%), the tangent modulus, Et was taken to be 400 psi. Since these materials are almost incompressible a Poisson's ratio of 0.4999 was used. It is to be mentioned here that the finite element code does not admit a Poisson's ratio exactly equal to 0.5. However for elastic calculations, for comparison purposes a Poissons ratio of 0.25 was also used.

Computational Models

A total of 25 specimen models were generated (five lengths with five crack depths) for performing finite element computations. Due to the symmetry of the loading, only the top half of the specimen was modeled. A typical finite element discretization is shown in Figure 3 of mostly four node quadrilateral elements. Triangular elements are used in the vicinity of the crack tip. Each of the specimen model was divided into four triangular regions radially extending from the crack-tip. These regions were then subsequently divided into specified number of sectors.

An inner circle of radius typically 10% of the crack length was employed over which elements were mapped progressively decreasing in size (in geometric progression) as they approached the crack tip. For each of the 25 finite element models, a total of 266 elements and 281 nodal points was used.

The bottom nodes of the model along the uncracked ligament was constrained to have zero vertical displacements from the symmetry of geometry and loading. Across the top nodes, different boundary conditions were used to model a possible variation of the angular stiffness of the test machine and fixtures. To model the conditions of parallel displacement of the end faces (infinite angular stiffness), these nodes were constrained from moving along the horizontal direction. In an attempt to better quantify the influence of the angular stiffness a simple model of the specimen together with the aluminum end tab was developed and the case of point load through the center of the edge of the specimen was also considered.

Two dimensional plane strain and plane stress finite element analyses of all the specimens were performed using conventional small strain theory. The analyses were carried out using a modified version of the finite element program FEAP [3]. Uniaxial stress strain behavior was described by the bilinear model as described earlier. J2 deformation plasticity (nonlinear elasticity) model was used to describe the multiaxial material model.

Evaluation of J-Integral

For two dimensional models, the J-integral [4] is defined as:

$$J = \int_{\Gamma} (W\, dy - \underline{T} \cdot \partial \underline{u} / \partial x\, ds) \qquad (1)$$

where W is the strain energy density, Γ is any contour in the x-y plane which encircles the crack tip and connects the lower and the upper crack surfaces; $\underline{T}$ is the traction vector acting on Γ ; $\underline{u}$ is the displacement vector along the path of integration.

The J-integral was computed at each load step using a domain integral form. The line integral (1) was recast as an area integral which is advantageous in numerical computations. The domain integral formulation developed by Moran and Shih [5] was employed for this purpose. J values were calculated over domains adjacent to and remote from the crack tip to demonstrate path independence of the J-integral.

Evaluation of the Mode I Stress Intensity Factor, K_I

From both plane stress and plane strain elastic analyses, the mode I stress intensity factors can be calculated by the following equations:

$$K_I = (JE)^{1/2} \quad \text{for plane stress} \qquad (2)$$

and

$$K_I = (JE/ (1 - \nu^2))^{1/2} \quad \text{for plane strain} \qquad (3)$$

Elastic calculations were performed for all specimen sizes and for the range of crack lengths. Different values of Poisson's ratio were used to study their effects on the mode I stress intesity factors using the plane stress formulation. The stress intensity factors were nondimensionalized with respect to $\sigma\sqrt{\pi a}$ to get the geometric factors associated with the specimen configuration.

Evaluation of Force-Displacement Diagrams

For each of the specimens, a far-field strain of 4% was applied in 40 load-steps. Plane stress inelastic analyses were performed using the bilinear approximation of the stress-strain curve for the material. For each load step, the program computes the reaction load based on inelastic analysis. The area under the load-displacement diagram gives the strain energy. All the models were loaded with parallel displacement of the end faces thus simulating infinite angular stiffness of the loading device.

Evaluation of the Elastic Compliance

From the initial slopes of the force-displacement diagrams the elastic compliance C of the specimens is given by:

$$C = V / F \qquad (4)$$

Where F is the force and V is the applied displacement. The value of C can be non-dimensionalized using the parameter EBC, where E is the elastic modulus, and B is the specimen thickness. For different specimen lengths, the elastic compliance can be calculated from the load-displacement diagram and evaluated for various crack depths. For an uncracked specimen, the quantity EBC should equal to L/W, where L is the length of the specimen (gauge length). In fact, it is possible to relate EBC in terms of the remaining ligament, (W - a), or more appropriately in terms of the parameter, 1 / (1 - a/W).

Evaluation of "eta" Factors

The multispecimen compliance method for experimentally evaluation of J-integral is based on energy interpretation and is expressed as follows:

$$J = - (1/B) (dU/da) \qquad (5)$$

Where U is the work done on the specimen as evaluated from the area under the force-displacement diagram. The differentiation uf U with respect to a eliminates the dependence of U with respect to specimen length if elastic behavior is assumed.

An approximate single-specimen procedure may be possible following Turner [6] and Paris et al [7] by introducing "eta" as:

$$J = \eta U / [B (W-a)] = - (1/B) (dU/da) \qquad (6)$$

For limited amount of yielding taking place in the specimens it would be interesting to see the variation of "eta" for various lengths of the specimens as a function of crack size.

Results and Discussion

Figure 4 presents the plane stress elastic analyses performed to compute the J-integrals for various specimen lengths with a range of crack depths. The normalized J-integrals represent the mode I stess intensity factors non-dimensionalized with respect to $\sigma\sqrt{\pi a}$ where σ is the far-field stress. For comparison purposes the corresponding computations for a Poisson's ratio of 0.25 are also shown. Both the materials exhibit similar trends with respect to increasing crack lengths. The Poisson's ratio effects seem to be more prominent for short cracks and for shorter specimens. A marked dependence on specimen size is apparent. Plane strain analyses and subsequent normalization using equation (3) could be performed to obtain similar results.

Figures 5 through 9 provide the normalized J integrals for inelastic analyses and load-displacement curves for various specimens. In all these figures, the J integrals start off with their elastic values as shown in Figure 4. The J values steadily increase as the applied displacements are increased. However the initial differences in the values tend to get smaller with increasing displacements and crossovers occur in some cases.

It is worthwhile mentioning here that in the study reported by Hollstein et al [8] for thin specimens, the load displacement curves obtained by a three-dimensional and plane stress analyses were not too different for constant parallel displacements of the end faces. Therefore the plane stress analysis in this study should closely represent the three-dimensional situation.

Figure 10 shows the elastic based compliances for various lengths of the specimen as a function of the crack depth. The corresponding solutions for a Poisson's ratio equal to 0.25 are also indicated. For the incompressibe case, the compliance values are seen to decrease slightly and then increase with increasing crack lengths. The cases for $\nu = 0.25$ are similar to the solutions of Hollstein et al. The case of L/W = 2.5 have been compared with an analytical expression in [8] developed for longer specimens (L/W > 5.0). It is possible to develop similar analytical expressions for various specimen lengths.

Figure 11 shows the "eta" factors as a function of the crack depth for various specimen length The curves display similar trends. The factors seem to asymptotically converge as the specimen lengths increase. These "eta" values are of course in between the linear elastic and fully plastic material behaviors. The elastic η values can be estimated from the elastic based compliance curves of Figure 10 using the analytical procedure outlined in [8]. Estimates of fully plastic η values are possible from limit load experiments where applicable. Recently Kirk and Dodds [9] report a procedure for η factors based on dividing the J into small scale and large scale yieldings. Such a procedure, however, is not applicable to propellant materials since these do not exhibit large work hardening region typical for structural steels.

ACKNOWLEDGMENTS

The author would like to express his sincere appreciation to Dr. C.T. Liu of the Phillips Laboratory, Edwards Air Force Base for his support and encouragement. He would also like to greatfully acknowledge the assistance provided by Professor G. Ravichandran of the California Institute of Technology on finite element studies.

References

[1] Hutchinson, J. W., "Singular Behavior at the End of a Tensile Crack in a Hardening Material," Journal of the Mechanics and Physics of Solids, Vol. 16, 1968, pp. 13-31.

[2] Rice, J. R., and Rosengren, G. F., "Plane Strain Deformation Near a Crack Tip in a Power-Law Hardening Material," Journal of the Mechanics and Physics of Solids, Vol. 18, 1968, pp. 1-12.

[3] Taylor, R. L., "Computer Procedures for Finite Element Analysis," in The Finite Element Method by O.C. Zienkiewicz, Third Edition, McGraw Hill, London, 1977, pp. 677-757.

[4] Rice, J. R., "A Path Independent Integral and the Approximate Analysis of Strain Concentrations by Notches and Cracks," Journal of Applied Mechanics, Vol. 35, 1968, pp. 379-386.

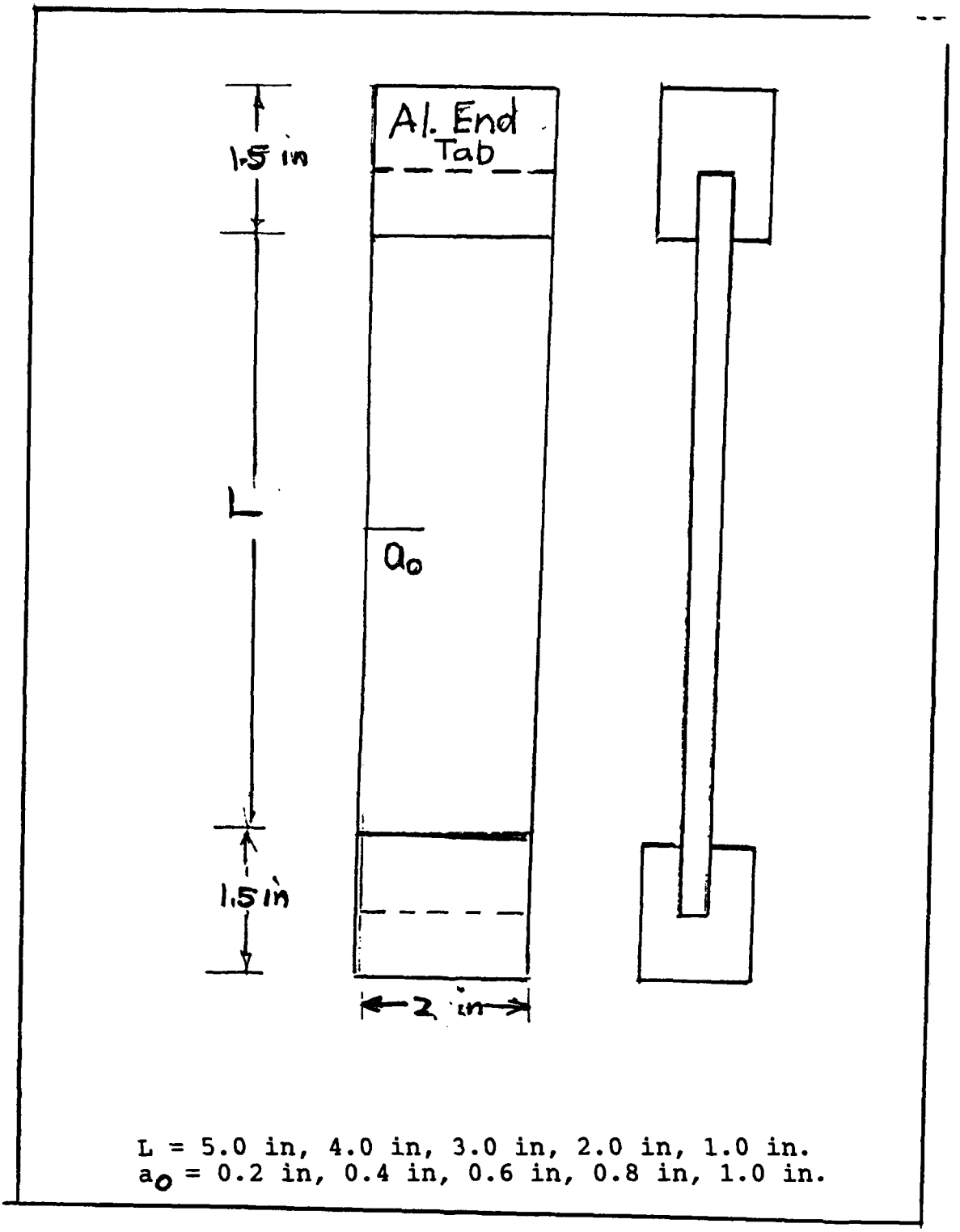
[5] Moran, B., and Shih C. F., "A General Treatment of Crack Tip Contour Integrals," International Journal of Fracture, Vol. 35, 1987, pp. 295-310.

[6] Turner, C. E., "The Ubiquitous η Factor," in Fracture Mechanics: Proceedings of the Twelfth National Symposium on Fracture Mechanics, STP 700, American Society for Testing and Materials, Philadelphia, 1980, pp. 314-337.

[7] Paris, P. C., Ernst, H., and Turner, C. E., "A J-integral Approach to the Development of η Factors," in Fracture Mechanics: Proceedings of the Twelfth National Symposium on Fracture Mechanics, STP 700, American Society for Testing and Materials, Philadelphia, 1980, pp. 338-351.

[8] Hollstein, T., Schmitt, W., and Blauel, J. G., "Numerical Analysis of Ductile Fracture Experiments Using Single-Edge Notched Tension Specimens, Journal of Testing and Evaluation, Vol. 11, 1983, pp. 174-181.

[9] Kirk, M. T., and Dodds, Jr., R. H., "J and CTOD Estimation Equations for Shallow Cracks in Single-Edge Notched Bend Specimens," Journal of Testing and Evaluation, Volume 21, 1993, pp. 228-238.



L = 5.0 in, 4.0 in, 3.0 in, 2.0 in, 1.0 in.
a_o = 0.2 in, 0.4 in, 0.6 in, 0.8 in, 1.0 in.

FIGURE 1 SPECIMEN GEOMETRIES

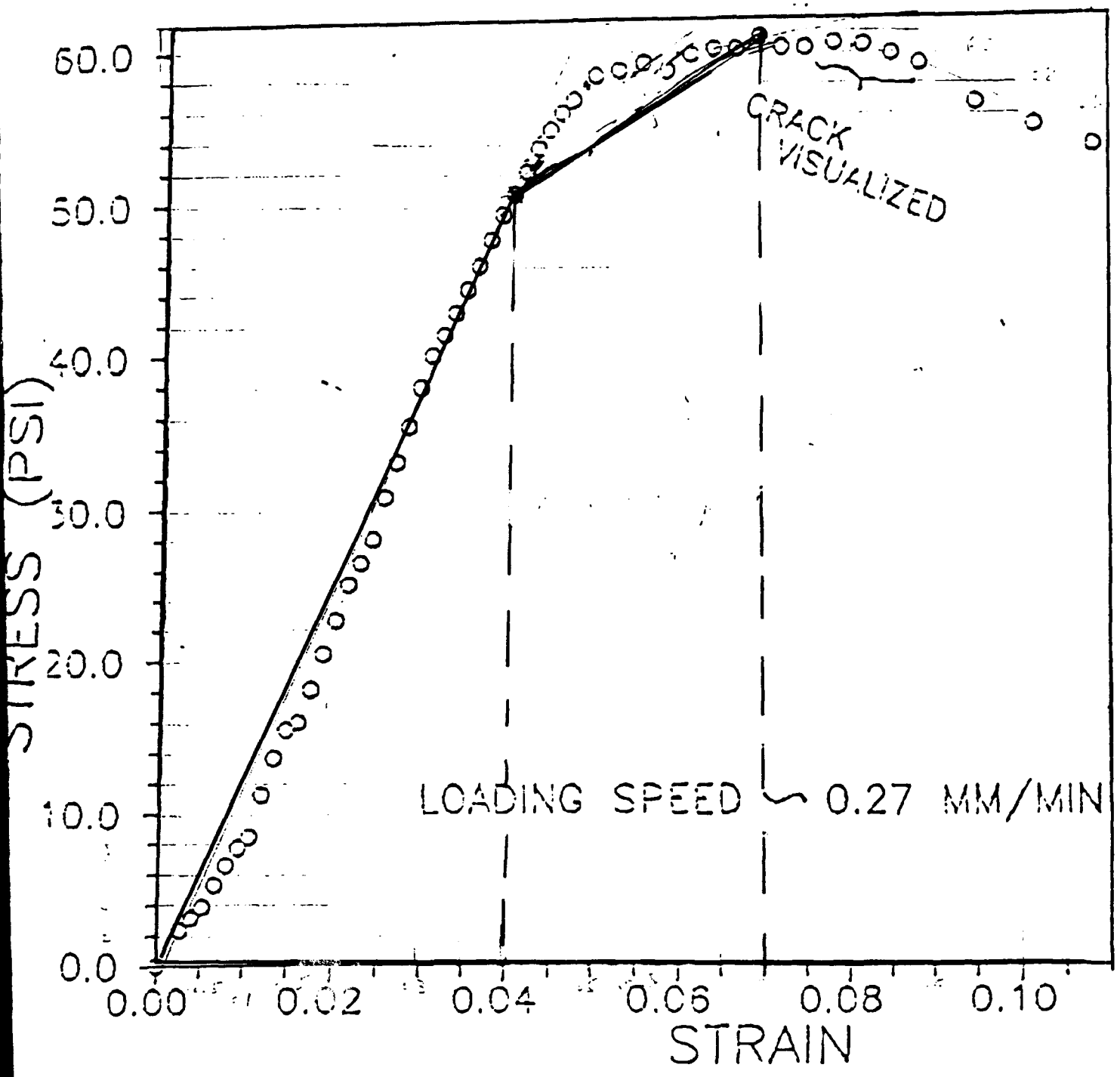


FIGURE 2 STRESS-STRAIN CURVE FOR SOLID PROPELLANT MATERIAL

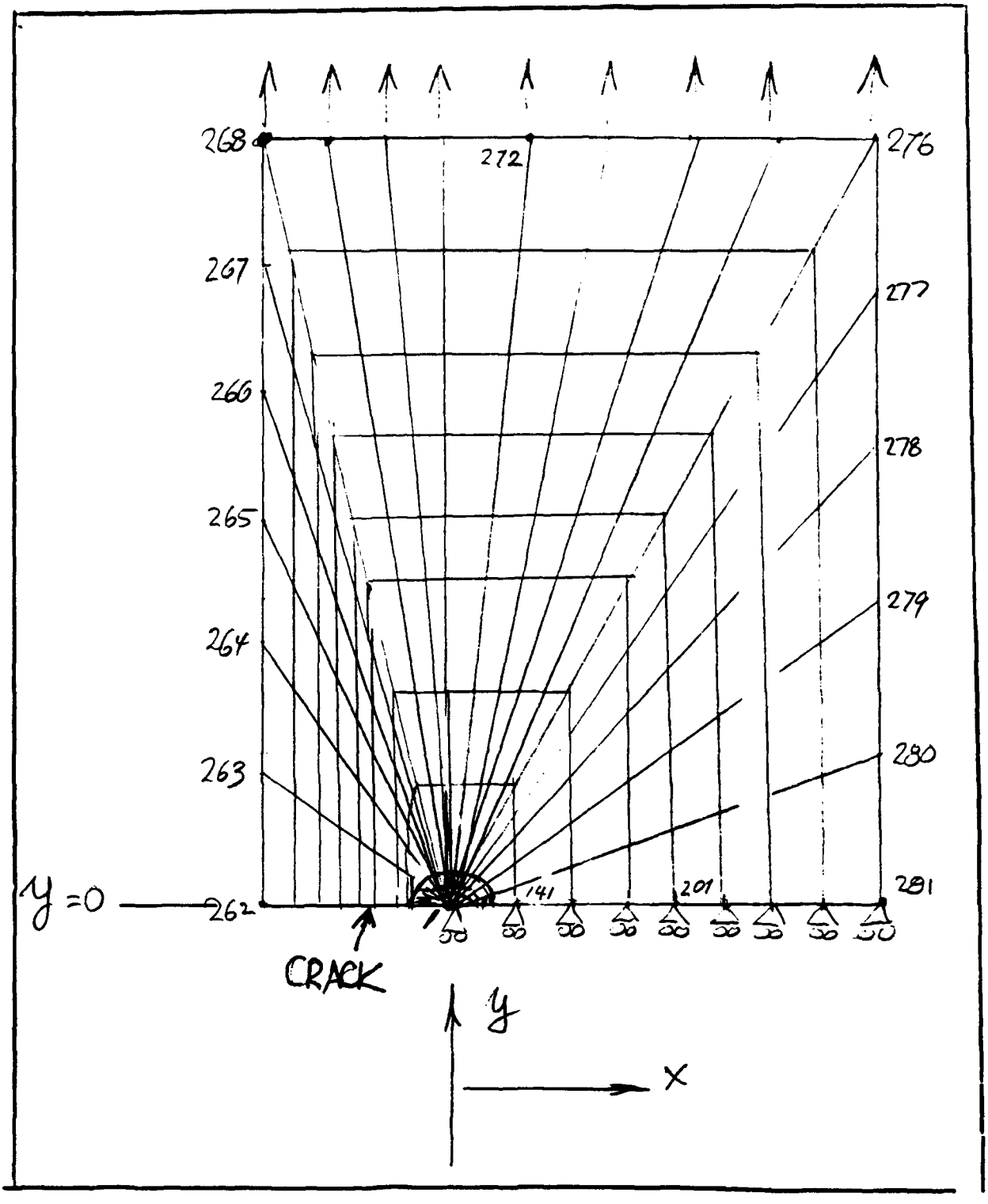


FIGURE 3 TYPICAL FINITE ELEMENT MODEL

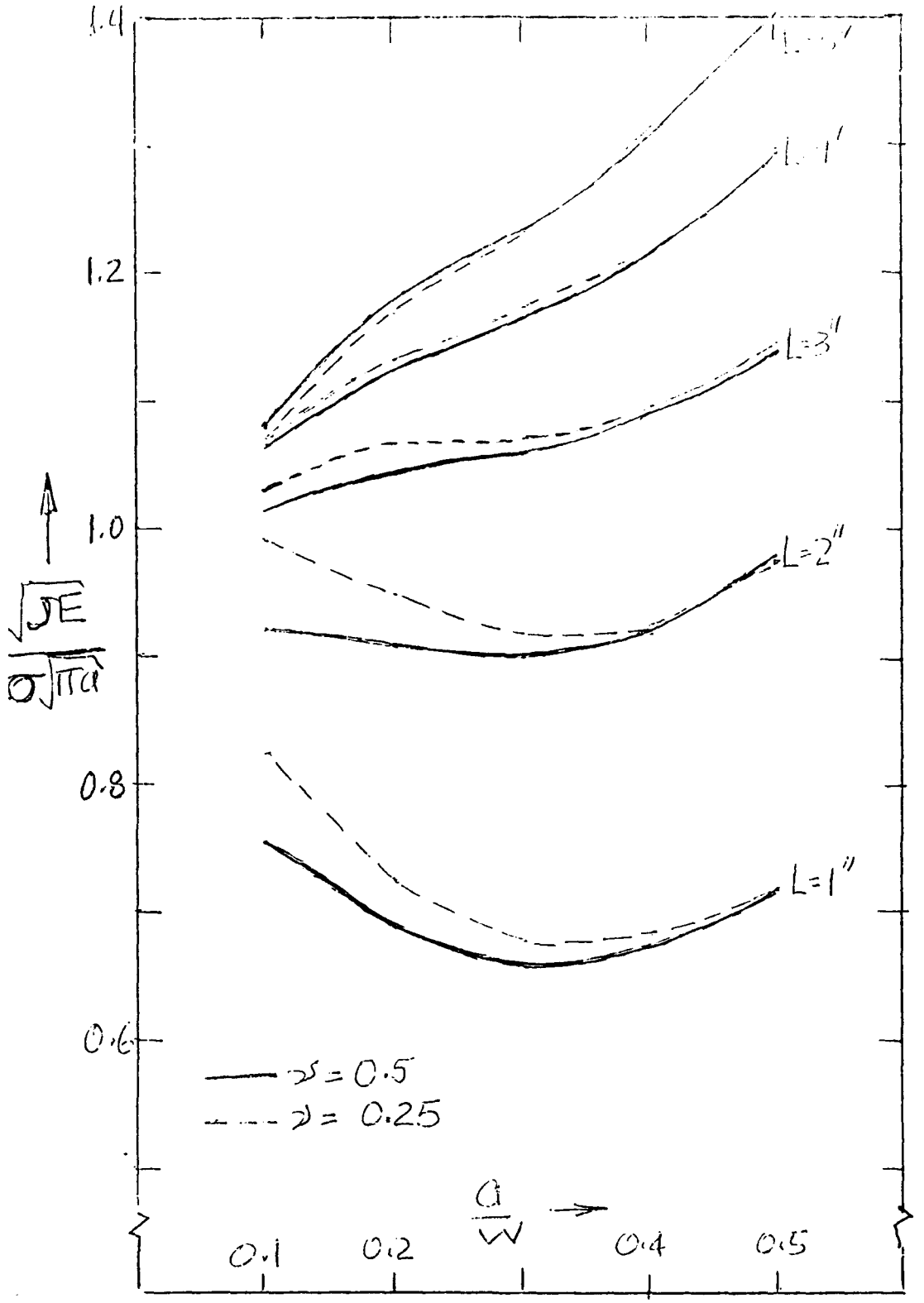


FIGURE 4 NORMALIZED J-INTEGRALS USING ELASTIC ANALYSES

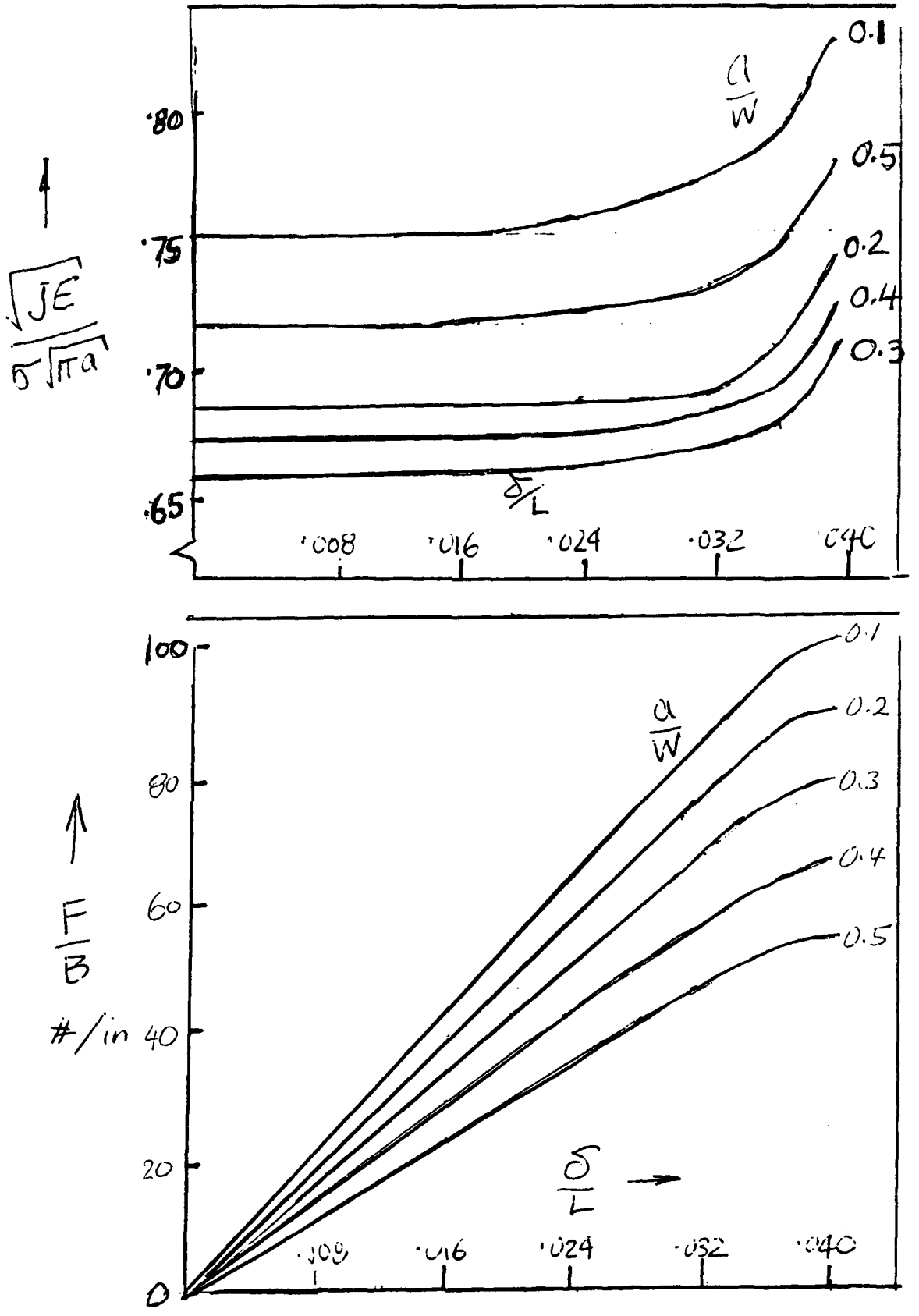


FIGURE 5 J-INTEGRAL AND LOAD VS. DISPLACEMENT FOR 1" SPECIMEN

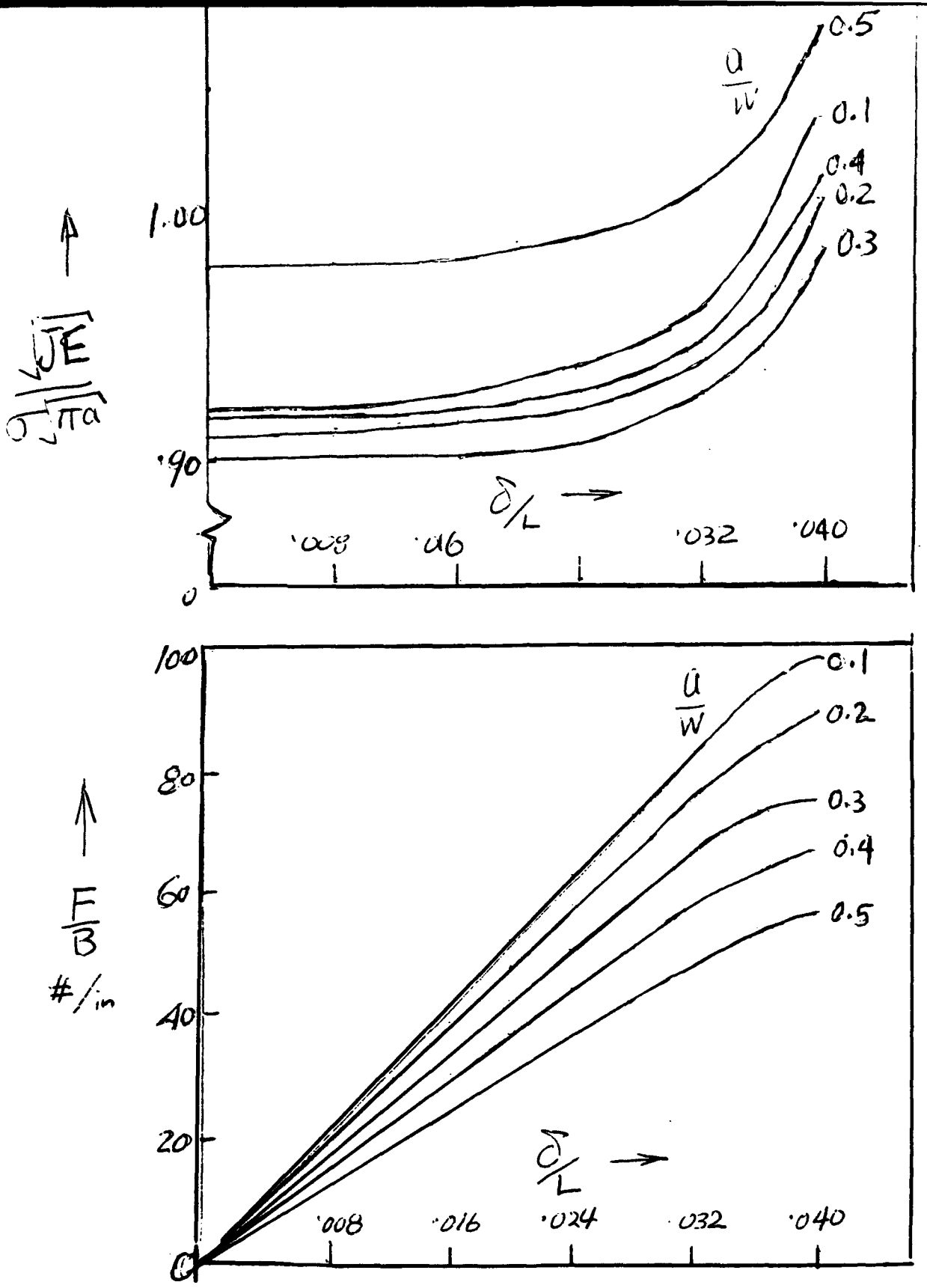


FIGURE 6 J-INTEGRAL AND LOAD VS. DISPLACEMENT FOR 2" SPECIMEN

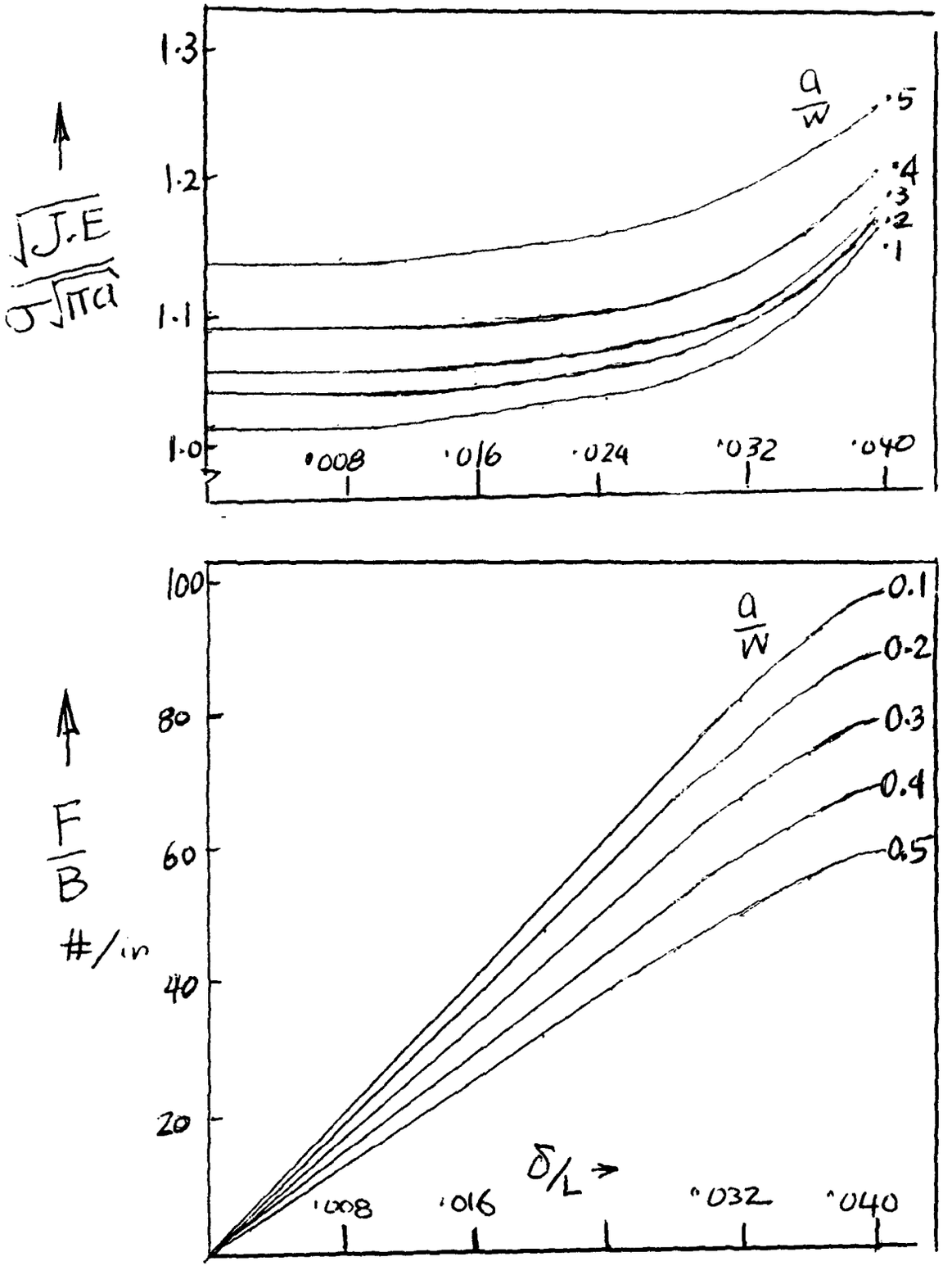


FIGURE 7 J-INTEGRAL AND LOAD VS. DISPLACEMENT FOR 3" SPECIMEN

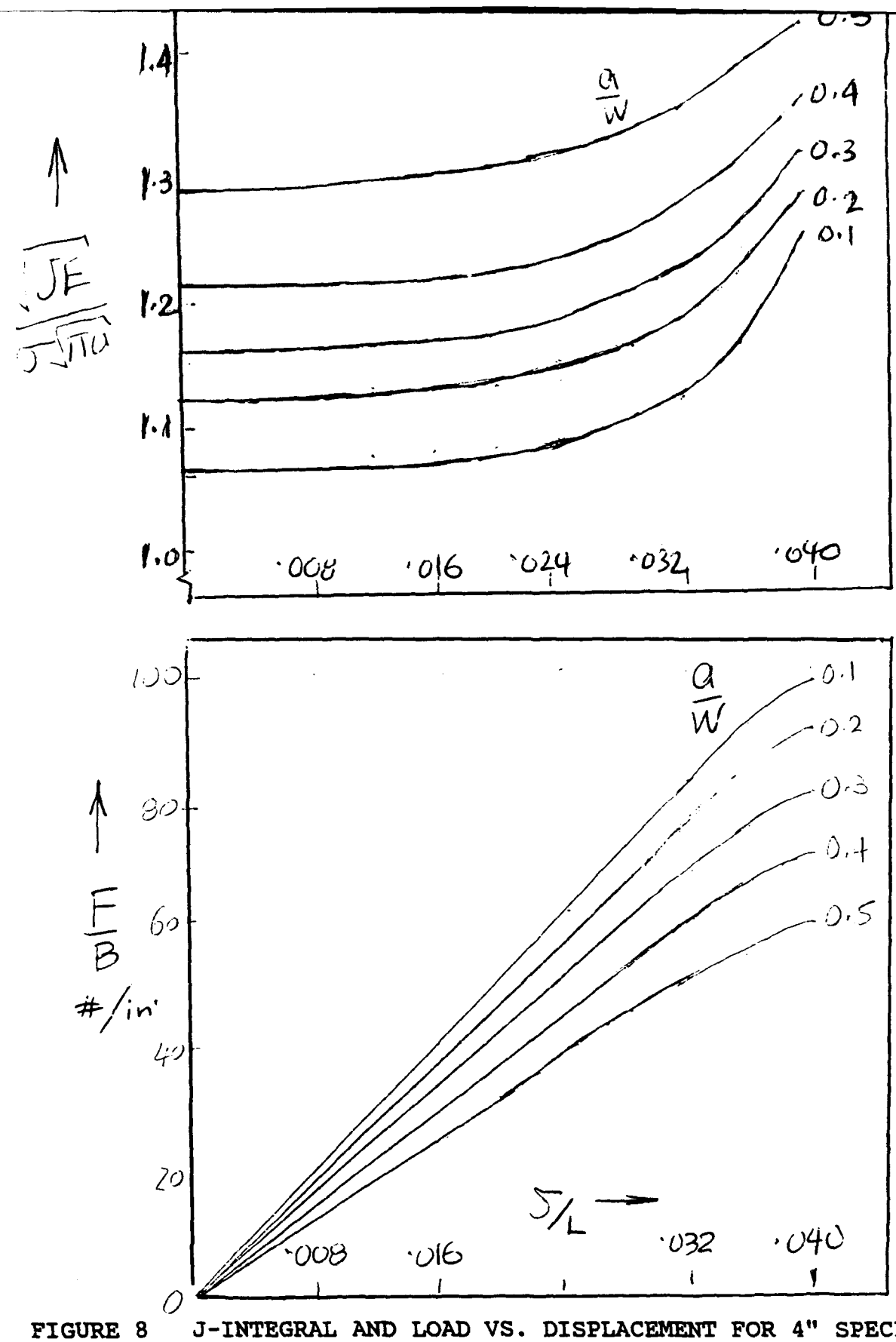


FIGURE 8 J-INTEGRAL AND LOAD VS. DISPLACEMENT FOR 4" SPECIMEN

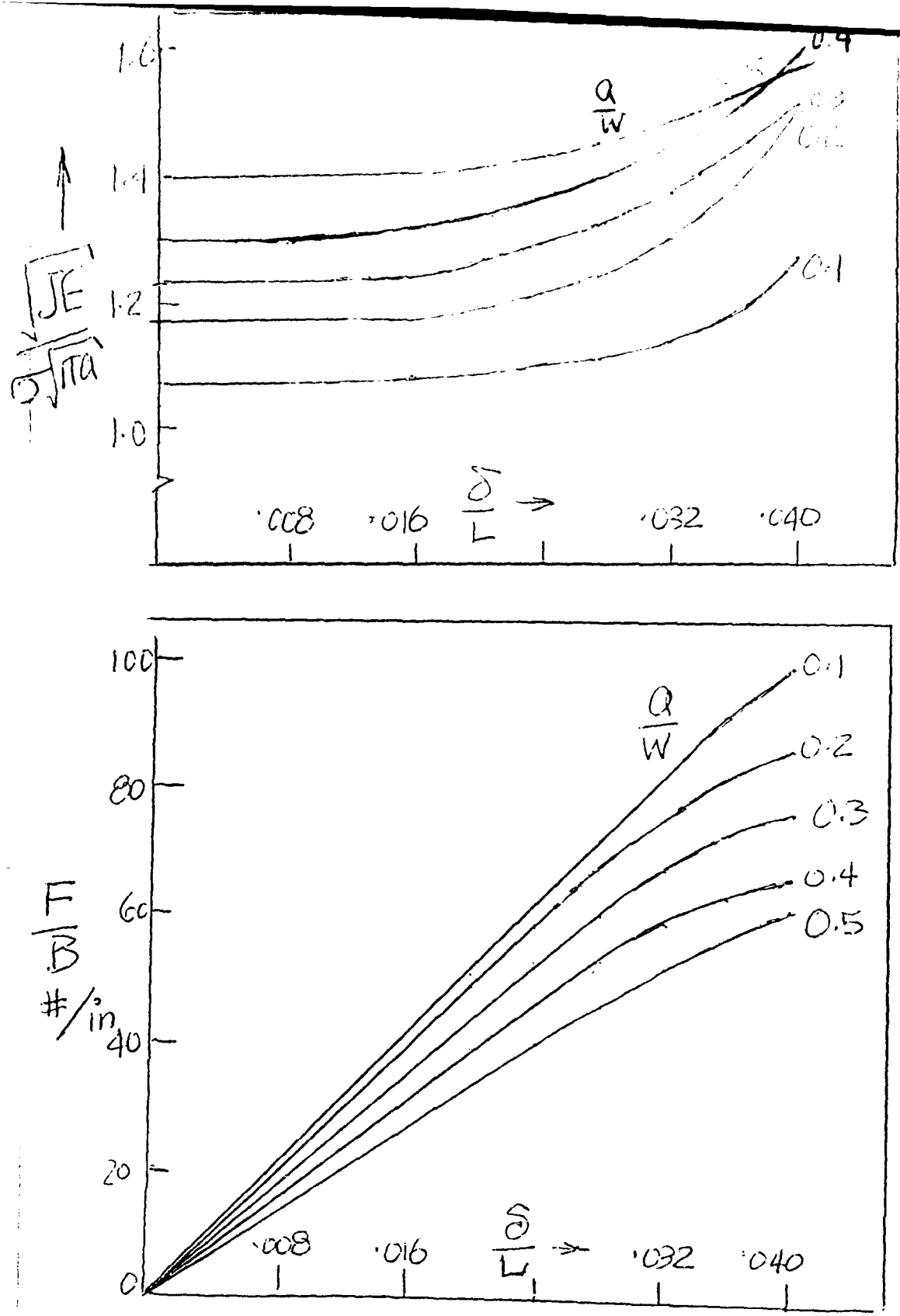
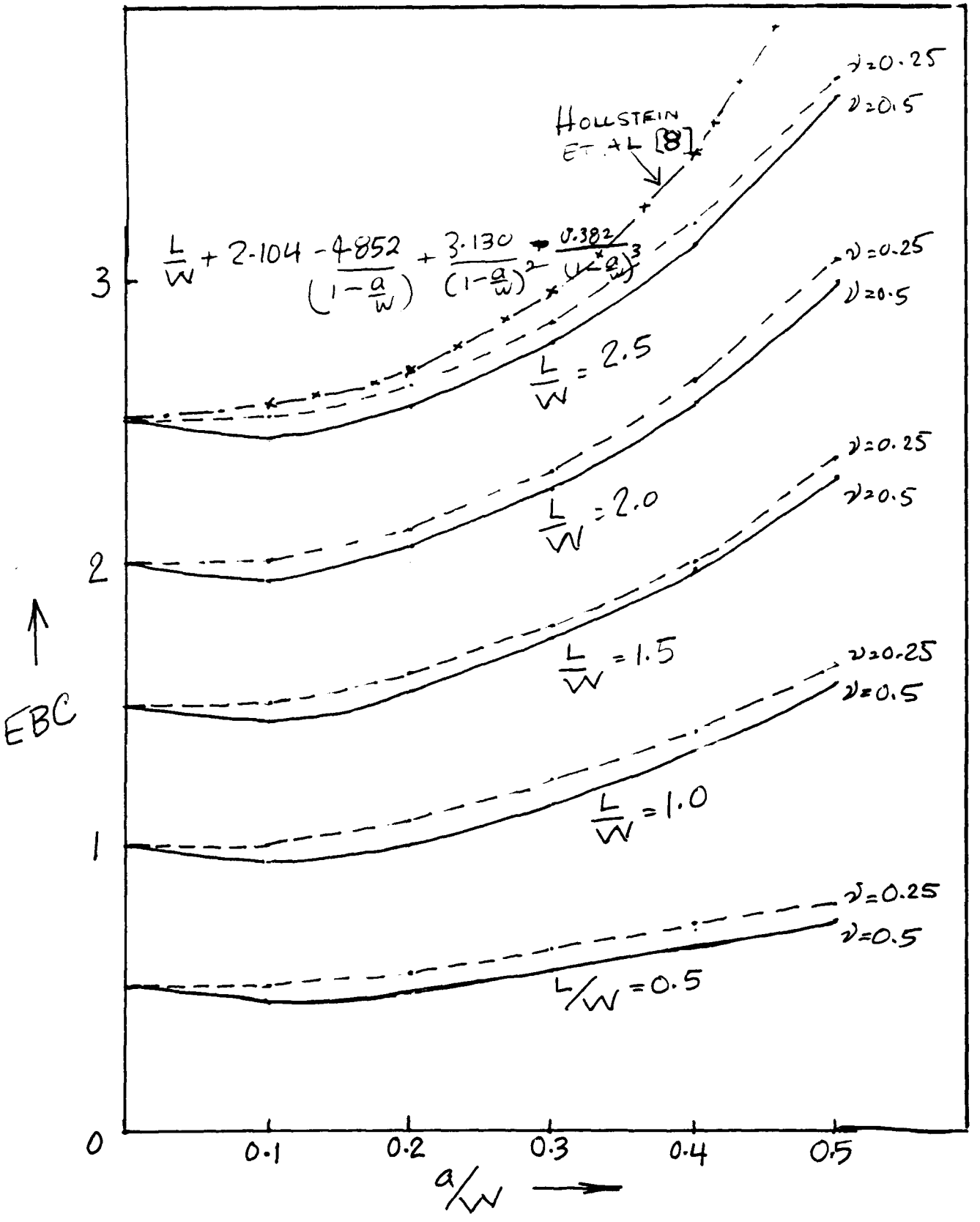


FIGURE 9 J-INTEGRAL AND LOAD VS. DISPLACEMENT FOR 5" SPECIMEN



FIGUPE 10 COMPLIANCE CALCULATED FROM ELASTIC ANALYSES

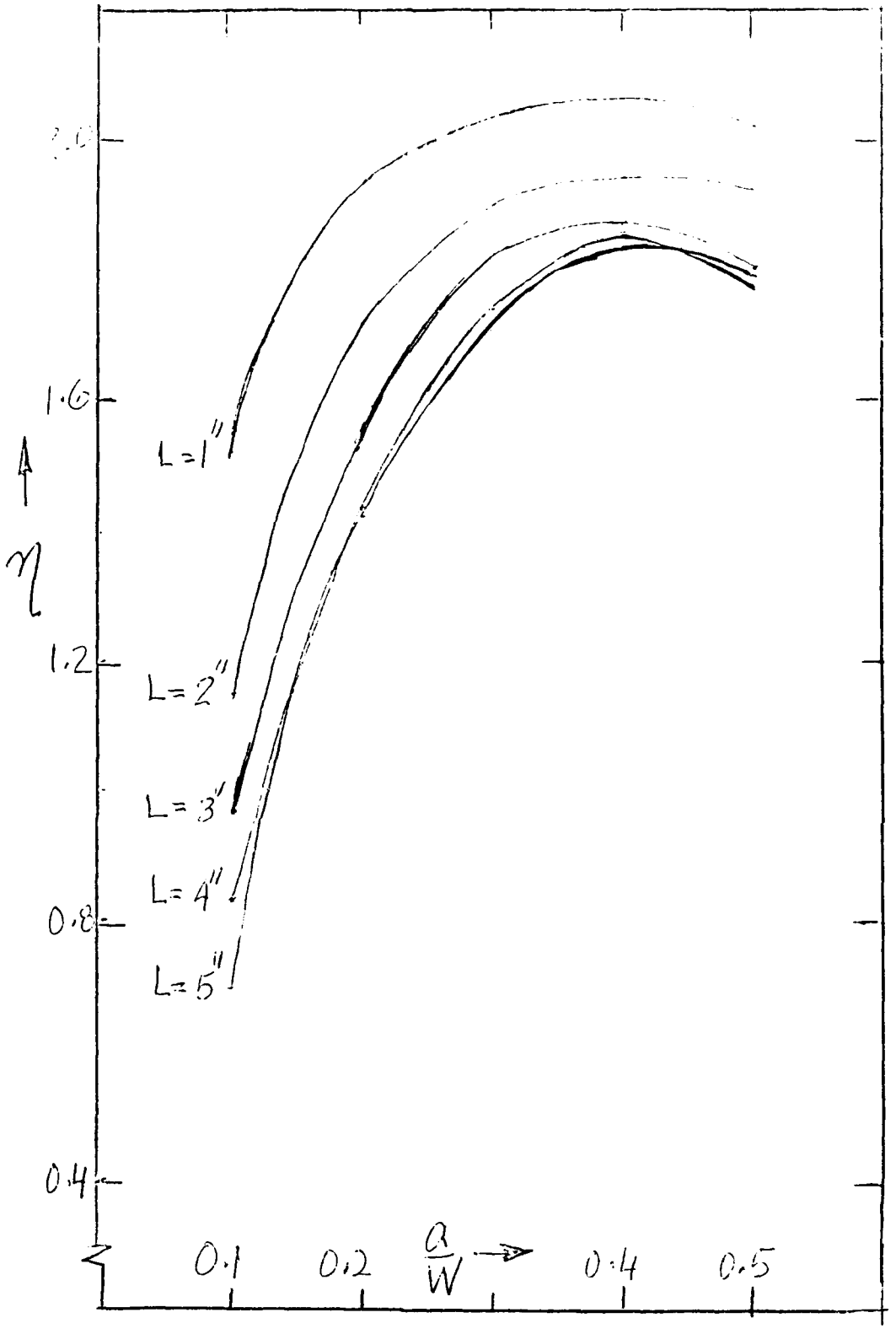


FIGURE 11 "ETA" FACTORS FOR VARIOUS SPECIMEN LENGTHS

COMPUTER-AIDED POLYMER AND HIGH ENERGY DENSITY MATERIAL DESIGN

Thomas R. Cundari

Department of Chemistry
Memphis State University
Memphis, TN 38152
phone: 901-678-2629
fax:901-678-3447
e-mail:cundarit@memstvx1.memst.edu

Final Report for:
AFOSR Summer Faculty Research Program

Phillips Laboratory, Edwards AFB

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D. C.

June - July 1993

Computer-Aided Polymer and High-Energy Density Material Design

Thomas R. Cundari
Department of Chemistry
Memphis State University
Memphis, TN 38152
e-mail:cundarit@memstvx1.memst.edu

Abstract

This research focuses on the use of computational chemistry techniques to predict properties for polymers and high energy density materials (HEDM) currently under experimental investigation by Air Force scientists. Molecular mechanics (MM) calculations were successful in predicting structural parameters for poly-oligomeric silsesquioxane (POSS) polymers. The MM calculations also suggested experiments for further probing the structure-property relationships of POSS polymers. The PM3 semi-empirical, quantum chemical method provided a quick and accurate estimate of the heats of formation (and hence I_{sp}) for HEDM targets, particularly when comparing related families of molecules. Based on the results, several recommendations are made for profitable exploitation and extension of this research.

Introduction

Computer-aided design of materials holds profound benefits for the future of the chemical industry. The use of computational techniques to estimate properties for useful materials (before going to the time, effort and expense, of making them in the laboratory) can reduce research and development time, and therefore costs. Additionally, computational chemistry techniques (particularly when coupled with current experimental research) can be used to yield a more complete understanding of currently employed processes. This can lead to processes which are cheaper and more efficient (reducing side-products and the attendant environmental problems associated with their disposal).

The area of chemistry in which computer-aided design has made the most important impact to date is computer-aided drug design.[1] Two main limitations, and thus opportunities for the future, exist in computer-aided design as applied to chemistry. First, it has been primarily limited to materials containing elements from the extreme upper right in the Periodic Table. The importance of lanthanide and transition metals in areas such as catalysis, advanced materials, and biomedical technology clearly indicates that effort is warranted in this direction.[2] The chemistry of the heavier main-group elements has also undergone a renaissance in recent years, particularly with respect to advanced materials applications.[3] A second limitation of computer-aided design is that existing tools have not been applied to a wide variety of materials other than drugs. Previous to the summer faculty associateship all of our Memphis State research has been directed towards computational studies of transition metal and lanthanide materials.[4] It is the purpose of this report to describe work carried out during the summer of 1993 at Phillips Laboratory, Edwards Air Force Base work in the area of computer-aided design of polymers and propellants for the High-Energy Density Materials (HEDM) program.

Goals

Our goals for this Summer Faculty Research appointment are two-fold. The first goal is to evaluate the use of commercially available, semi-empirical, computational chemistry

programs to provide a quick (i.e., several minutes of CPU time on a standard workstation) determination of the geometry and heat of formation of potential HEDM targets. The second goal is to use computational techniques to investigate the properties of polymers currently under experimental investigation in Air Force laboratories. It is envisaged that our experience[4] can be used to advantage for the achievement of these goals, and that this research will extend our expertise into new areas.

Background

1. Polyhedral Oligomeric Silsesquioxane (POSS) Polymers

A family of polyhedral oligomeric silsesquioxane (POSS) polymers have recently been studied by Lichtenhan et al.[5] with the formula shown in Figure 1. The POSS units can be built into two component co-polymers, with the silsequioxane groups linked by a variety of "spacers."

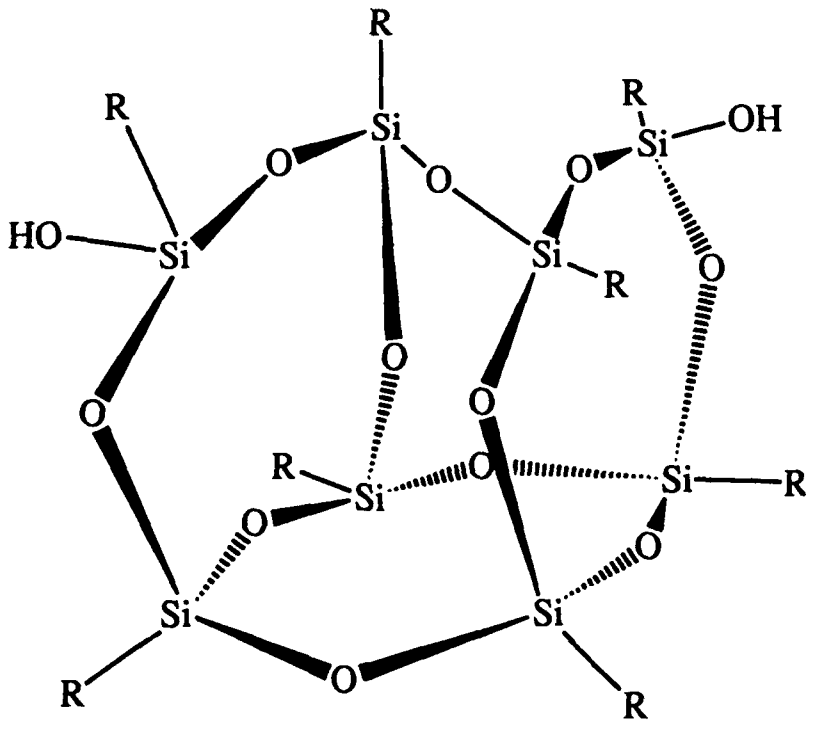
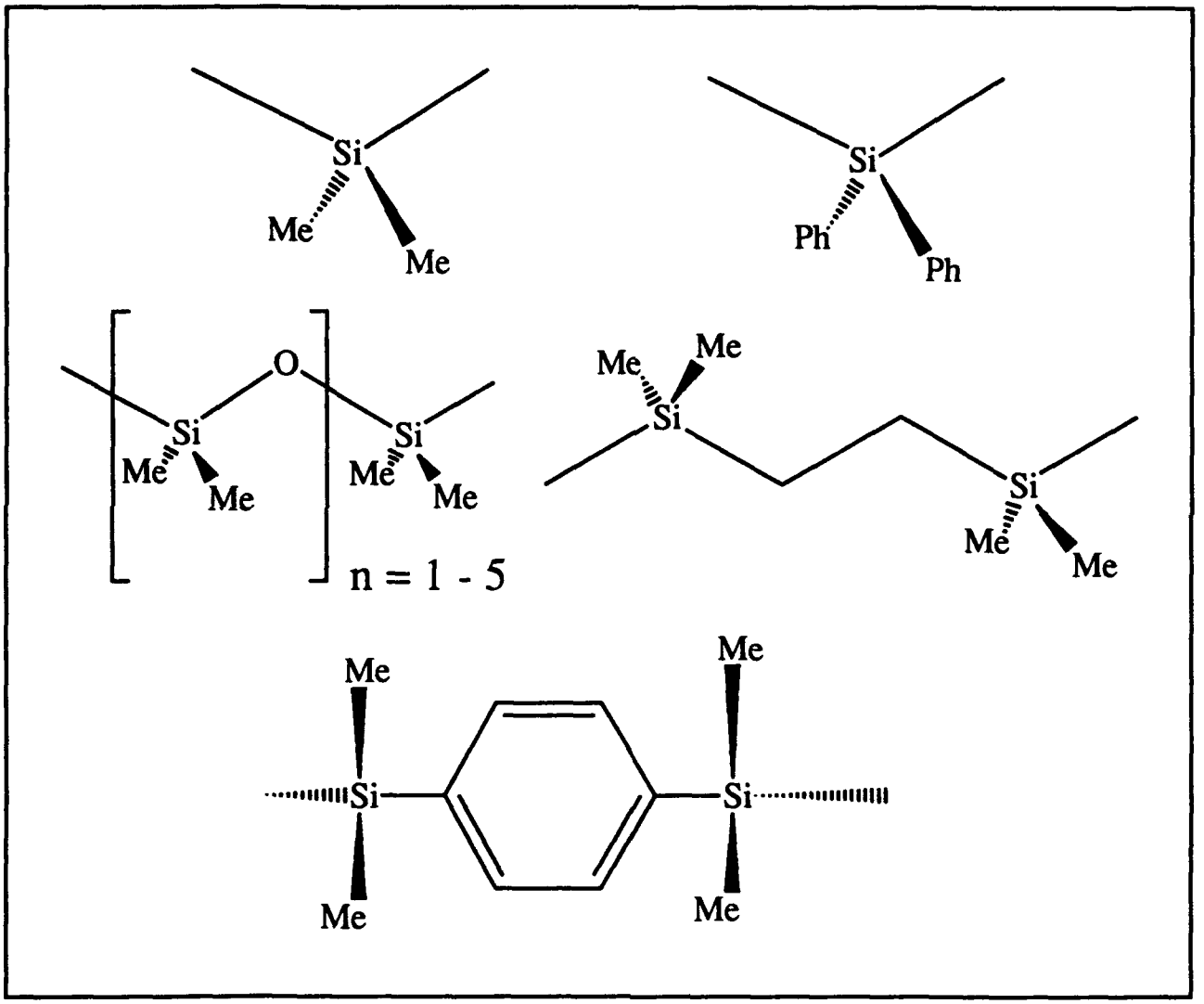


Figure 1: Polyhedral oligomeric silsesquioxane (POSS) building blocks.

These materials represent an exciting new family of polymers since they display a great range of properties, which can be tuned by the nature of the spacer, Figure 2. For example, the POSS polymer with the oligomeric dimethylsiloxane spacer is an elastomer, while the dimethylsiloxane spaced material is more rigid.[5] The POSS polymers also have the potential to act as preceramics for solid-state materials such as SiC.[5]



Two main problems have inhibited the use of POSS-based polymers. First, synthetic routes are far from general. Second, it has not been possible to build desired properties into these materials in a rational manner. The relationship between polymer structure and observed

properties is unclear. The first problem has come under investigation at Phillips Laboratory, Edwards Air Force Base.[5] Our immediate goals are to evaluate the graphical user interface (GUI) Chem-X[6] for the purposes of correlating the properties of polymers, and then to use Chem-X to investigate the tertiary structure of POSS polymers currently under investigation.

2. High-Energy Density Materials

It is of interest to search for mono- and bi-propellants (fuel/oxidizer) with higher specific impulse (I_{sp}). Additionally, a high I_{sp} compound can be used as an additive to a standard rocketry mixture, e.g. liquid oxygen (LOX) and rocket propellant # 1 (RP1). Higher I_{sp} translates into greater payload for a given amount of propellant. Extracting even small (a few percent) increases in I_{sp} has enormous economic benefits by increasing the ability of smaller rockets such as the Delta (costing $55M) to carry payloads marginally above its current operating limit in order to preclude the use of a larger launch vehicle (e.g., the Atlas II costing $80M). The specific impulse as defined by Forbes[7] is "...related to the energy content of the combustion gases being exhausted..." The formula for specific impulse is given in Equation 1. The

$$I_{sp} = (2\ J\ (H_c - H_e) / g)^{1/2} \qquad (1)$$

of H_c and H_e are the enthalpies of combustion of combustion products before and after, respectively, expansion; J is the mechanical equivalent of heat, and g is the gravitational constant. To calculate enthalpies of combustion it is necessary to know the enthalpies of formation of the mono- and bi-propellants and potential additives.

The heat of formation of a compound can be calculated with quantum chemical methods. Thus, it can complement the experimental determination of this property, which is subject to experimental uncertainty. Calculation of the heat of formation is computationally intensive, particularly for ab-initio methods.[8a] One potential avenue of approach is the use of semi-empirical methods,[8b] which are quick and applicable to larger, more realistic HEDM targets

A systematic investigation of HEDM targets is being undertaken by Wucherer and co-workers (OLAC PL/RKFE).[9] To assess the accuracy of semi-empirical, quantum chemical

methods, a series of test organic molecules was studied. In the test cases the experimental heats of formation are known with a good degree of accuracy (± 1 kcal mol^{-1}), a situation which is not always true for HEDM targets. Forty-nine targets were chosen to contain a variety of functional groups (strained hydrocarbons, cyano and nitro containing compounds, and N-N bonded systems) of interest in HEDM. Out of the 49 HEDM targets, several "worst case scenarios" in which calculated heats of formation ranged over a wide variety of values were studied to test the sensitivity of the calculated I_{sp} to the heat of formation. HEDM calculations were carried out by High School Research Associate Alexandra Kitty (Associate # 461); a more complete account of this research can be found in pages 10-1 through 10-14 in this series of reports.

Computational Methods

1. Molecular Mechanics

Molecular mechanics (MM) employs the laws of classical mechanics to predict the three-dimensional structure of compounds.[8b] Molecular mechanics can be thought of as a "ball-and-spring" description of chemical bonding, where "balls" represent atoms and the "springs" the bonds which connect them. The MM algorithm attempts to find the lowest energy geometry of a molecule in terms of bond lengths, bond angles and torsional angles. More sophisticated MM algorithms can take into account electrostatic, hydrogen-bonding and van der Waals interactions. Molecular mechanics can treat very large polymeric systems, a task which is greatly assisted by the use of a graphical user interface (GUI). The software evaluated for this summer's research was Chem-X, distributed by Chemical Design Limited.[6] Chem-X is a GUI front-end for a MM back-end. *Calculations were carried out on a Silicon Graphics at OLAC PL/RKFE.*

2. Semi-empirical methods

Three semi-empirical parameterizations were evaluated for their ability to predict enthalpies of formation. Semi-empirical methods employ assumptions in construction of the electronic Hamiltonian to reduce computational time. As is often the case, assumptions

introduced to save time can also degrade the accuracy of the method. Parameters can be introduced to remedy loss of accuracy, making semi-empirical methods in some cases more accurate than ab-initio techniques. The parameter schemes employed are AM1 (Austin Method 1), PM3 (Parameterization Method 3) and MNDO (Modified Neglect of Differential Overlap).[8b] These methods are contained in the larger GAMESS[10] quantum chemistry code. Previously, Guidry and co-workers[11] have evaluated semi-empirical schemes for calculating enthalpies of formation for nitro compounds. The present research is a more extensive investigation, in that a greater variety of compounds are being investigated. An additional parameterization scheme (i.e., PM3), not available at the time of the study by Guidry et al.,[11] is being studied.

Results and Discussion

1. Polyhedral Oligomeric Silsesquioxane (POSS) Monomer

To begin this research it is necessary that the standard parameters in Chem-X can successfully predict the structure of oligomeric siloxanes. To avoid the "multiple minima" problem inherent in computational analyses of large molecules, the following strategy was employed. The crystal structure of $Si_8O_{11}Cy_8(OSnPh_3)_2$, Cy = cyclohexyl, **1**, has been determined by Feher et al.,[12] Figure 3. The conformation of this solid-state structure was used as the starting point for the POSS units in forming the POSS-spacer copolymers. Spacers (see Figure 2) were built from scratch.

The standard molecular mechanics approach in Chem-X performed very well in predicting the structure of **1**. Calculated Si-O and Si-C distances in **1** are 1.61 ± 0.01 Å and 1.87 ± 0.01 Å, respectively. Experimental values in **1** are in excellent agreement: Si-O = 1.62 ± 0.01; Si-C = 1.85 ± 0.01 Å.[12] Bond angles about the Si and O in the cage are close to tetrahedral (≈ 109.5°), as found experimentally.[12] Comparison of optimized structures for spacers, Figure 2, also shows good agreement with standard bond lengths and bond angles.[13] One limitation of Chem-X is apparent upon analysis of bond lengths involving the tin. Bond lengths involving tin were much shorter (by as much as 0.36 Å) than experimental values. For any future research

incorporating tin atoms, extensive reparameterization would be necessary. Since POSS polymers currently under investigation contain no tin, Chem-X is suitable for further study this family of polymers.

2. POSS Homopolymer and POSS-Spacer Copolymers

Molecular mechanics optimized **1** was used (after replacement of the $SnPh_3$ groups with H) to build all subsequent homo- and co-polymers. The standard polymer-building routine in Chem-X was employed. In actuality, "polymers" were decamers in order to make computations tractable. Studies on dimers were also carried out to assess the sensitivity to the torsional angle formed about the fusion point. In no case, including homopolymer, was the energetic dependence of the torsional angle great.

The first studies of polymers focused on POSS homopolymer. Hydrogen atoms between cyclohexyls on neighboring POSS units are well within the sum of their van der Waal radii. The van der Waals interactions are the quantum equivalent of steric repulsions, and it is clear that a homopolymer would possess severe steric strain. Inspection of the homopolymer geometry also makes it clear why spacers are needed to provide a sterically accessible polymers by moving the offending H atoms out of van der Waals range Experimentally, synthetic attempts to construct a POSS homopolymer (i.e., no spacer) have thus far been unsuccessful.[5] However, when even a small spacer such as $SiMe_2$ is used, POSS-spacer polymers have been synthesized.[5] The molecular mechanics calculations provide a rationalization for these observation based on repulsions between H atoms on neighboring POSS units.

Interesting variations in tertiary structure of POSS-spacer copolymers are noticed as a function of spacer. For small spacers ($SiMe_2$, $SiPh_2$, $SiMe_2OSiMe_2$) and no spacer (homopolymer) the optimized structure is a rigid rod. Introduction of longer spacers (e.g., oligomeric dimethylsiloxanes and $SiMe_2C_6H_4SiMe_2$) leads to a helical structure. Recall that the dimethylsiloxane spaced copolymer is a more rigid polymer, while the oligomeric

dimethylsiloxanes are elastomeric.[5] Thus, the difference in the three-dimensional structure of the POSS-spacer copolymers is interesting in connection with their observed physical properties.

The above calculations results suggest potential experimental probes into the structure of POSS-spacer copolymers which have been synthesized to date. A technique such a WAXS (wide-angle X-ray scattering) can be used to probe the tertiary structure of polymers. For example, Liu and Ruland have used WAXS to study polyacrylonitrile polymers.[14] Analysis of the diffraction pattern should yield valuable information about the tertiary structure of POSS polymers and the reliabbility of the MM approach.

2. High-Energy Density Materials

As mentioned above, the HEDM portion of out research was carried out in the main by Summer High School Research Associate Alexandra Kitty, and more details can be found in the appropriate report. In summary, the main question to be answered is - Can semi-empirical quantum chemical methods be used to quickly estimate the heat of formation of a potential HEDM target? Of the three semi-empirical parameterization schemes employed (PM3, AM1, and MNDO), PM3 was clearly superior. Analysis of heat of formation data for 49 HEDM molecules for the PM3 method, showed an average total error of 9 kcal mol^{-1}. The average absolute error was +3 kcal mol^{-1}, showing that PM3 tended to be slightly high. Omission of several worst cases (e.g., cubane, malononitrile, and the dimethylhydrazido analogue of norbornadiene) which had errors greater than 40 kcal mol^{-1} improves the quality of the results drastically. Note that experimental errors are roughly ± 1 - 2 kcal mol^{-1} at best.[15] The best fit line for the PM3 versus experimental data (in kcal mol^{-1}) is given in Equation 2. The slope is

$$\Delta H_f(\text{expt}) = \Delta H_f(\text{PM3}) * 0.93478 - 0.17402 \; (R = 0.95468). \qquad (2)$$

close to one, and the intercept is very close to unity, in line with ideal values. The correlation coefficient is also quite good. Perhaps most importantly, analysis of the sensitivity of calculated I_{sp} to deviations in ΔH_f for even these worst case scenarios showed the PM3 ΔH_f to provide an

effective gauge of I_{sp}, and thus a potentially valuable tool in directing synthetic efforts for the HEDM group.

Summary, Conclusions and Recommendations

To summarize, this summer's research has focused on the use of computational chemistry techniques to predict properties for polymers and HEDM materials currently under experimental investigation by Air Force scientists. Agreement between computational methods and available experimental data is good in most cases, for example in predicting structural parameters for POSS polymers and heats of formation (using PM3) of HEDM targets. Several recommendations for profitable exploitation of this research seem valid.

1) The PM3 method can provide experimentalists with a good indication of heats of formation (and hence I_{sp}) for synthetic targets, particularly when comparing related families of target molecules. Of course, theoretical predictions (which will focus on increasing ΔH_f and thus I_{sp}) must be tempered by experimental realities (e.g., ease of synthesis and scale-up), so that close coordination between computational and experimental scientists is needed.

2) The Chem-X software represent an easy to use, relatively intuitive GUI for studying the structure of POSS polymers. The MM calculations suggest that WAXS studies of different POSS-spacer copolymers may be informative with respect to the tertiary structure of these materials, and hence may provide important clues about structure-property relationships.

3) POSS polymers with siloxanes other than **1** have been investigated at Phillips, and show very different properties from those focused on in this research.[5] Modifications can include changes to the siloxane cage and/or the R group, see Figure 1. Given the good results obtained in predicting the structural properties, molecular mechanics studies be extended to polymers with different siloxane units in order to further probe structure-property relationships.

4) The main shortcoming of the Chem-X software lies in its inflexibility to non-standard atoms. Given the potential importance of polymeric materials containing heavier main group elements (e.g., tin) or transition metals,[16] a program which is more adaptable in allowing the

introduction of new parameter sets is clearly needed. Examination of additional software, development of parameters sets to include a larger selection of elements, and coupling of quantum and molecular mechanical methods constitute profitable areas for extension of this research.

Acknowledgments

This research has profited greatly from interaction with the Air Force scientists at Phillips Lab (Edwards AFB), in particular Joe Lichtenhan (polymers) and E. J. Wucherer (HEDM). The expert assistance of Alexandra Kitty in the HEDM portion of research is gratefully acknowledged . Our computational chemistry research at Memphis State have supported by the American Chemical Society-Petroleum Research Fund, Air Force Office of Scientific Research, IBM, Digital Equipment Corporation, Memphis State University Faculty Research Grants, National Institutes of Science and Technology, the Department of Energy, and the National Science Foundation.

References

1. Cohen, N. C.; Blaney, J. M.; Humblet, C.; Gund, P.; Barry, D. C. J. Med. Chem. **1990**, 33, 883.

2. The growing importance of the lanthanide and transition metals in these areas is discussed in any standard inorganic text, e.g. Cotton, F. A.; Wilkinson, G. "Advanced Inorganic Chemistry;" 5th Ed. Wiley: New York, **1991**.

3. Typical applications of main-group chemistry to the manufacture of materials of interest in areas such as electronics can be found in "Mechanisms of Reactions of Organometallic Compounds;" Cole-Hamilton, D. J.; Williams, J. O., Eds.; NATO ASI Series B; Plenum: New York, **1988**, Vol. 198.

4. a) Cundari, T. R. J. Am. Chem. Soc. **1992**, 114, 10557.
b) Cundari, T. R. Intern. J. Quantum Chem., Proc. 1992 Sanibel Symp. **1992**, 26, 793.
c) Stevens, W. J.; Cundari, T. R. J. Chem. Phys. **1993**, 98, 5555.
d) Cundari, T. R. J. Am. Chem. Soc. **1992**, 114, 7879.
e) Cundari, T. R. Organometallics **1993**, 12, 1998.
f) Cundari, T. R.; Gordon, M. S. J. Am. Chem. Soc. **1993**, 115, 4210.

5. Lichtenhan, J. D.; Vu, N. Q.; Carter, J. A.; Gilman, J. W.; Feher, F. J. Macromolecules **1993**, 26, 2141 and personal communication.

6. "Chem-X;" Chemical Design Ltd., Oxford, **1993**.

7. Forbes, F. S. in "Encyclopedia of Physical Science and Technology;" Vol. 7, Academic press: New York, **1987**, p. 391.

8. a) Pople, J. A.; Hehre, W. J.; Radom, L.; Schleyer, P. v. R.;" Ab-Initio Molecular Orbital Theory;" Wiley: New York, **1986**.
b) Clark, T. "Handbook of Computational Chemistry;" Wiley: New York, **1985**.

9. Behdadnia, R.; Lang, D.; Wucherer, E. J. "Chemistry and Propulsion;" **1993** HEDM Contractors Meeting, Woods Hole, MA, **1993**.

10. GAMESS (General Atomic and Molecular Electronic Structure System): Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Jensen, J. H.; Koseki, S.; Gordon, M. S.; Nguyen, K. A.; Windus, T. L.; Elbert, S. T. QCPE Bulletin, **1990**, 10, 52.

11. Guidry, R. M.; Storch, D.; Davis, L. P. J. Energetic Materials **1987**, 5, 89.

12. Feher, F. J.; Newman, D. A.; Walzer, J. F. J. Am. Chem. Soc. **1987**, 111, 1741.

13. "CRC Handbook ;" Weast, R. C., Ed. 64th Ed. CRC Press: Boca Raton, **1983**, p. F-187.

14. Liu, X. D.; Ruland, W. Macromolecules **1993**, 26, 3030.

15. Cox, J. D.; Pilcher, G. "Thermochemistry of Organic and Organometallic Compounds;" Academic Press: London, **1970**.

16. POSS-spacer copolymers containing transition metal containing spacers (e.g., $TiCp_2$ and $ZrCp_2$) have been synthesized[5] and display good thermal and hydrolytic stability.

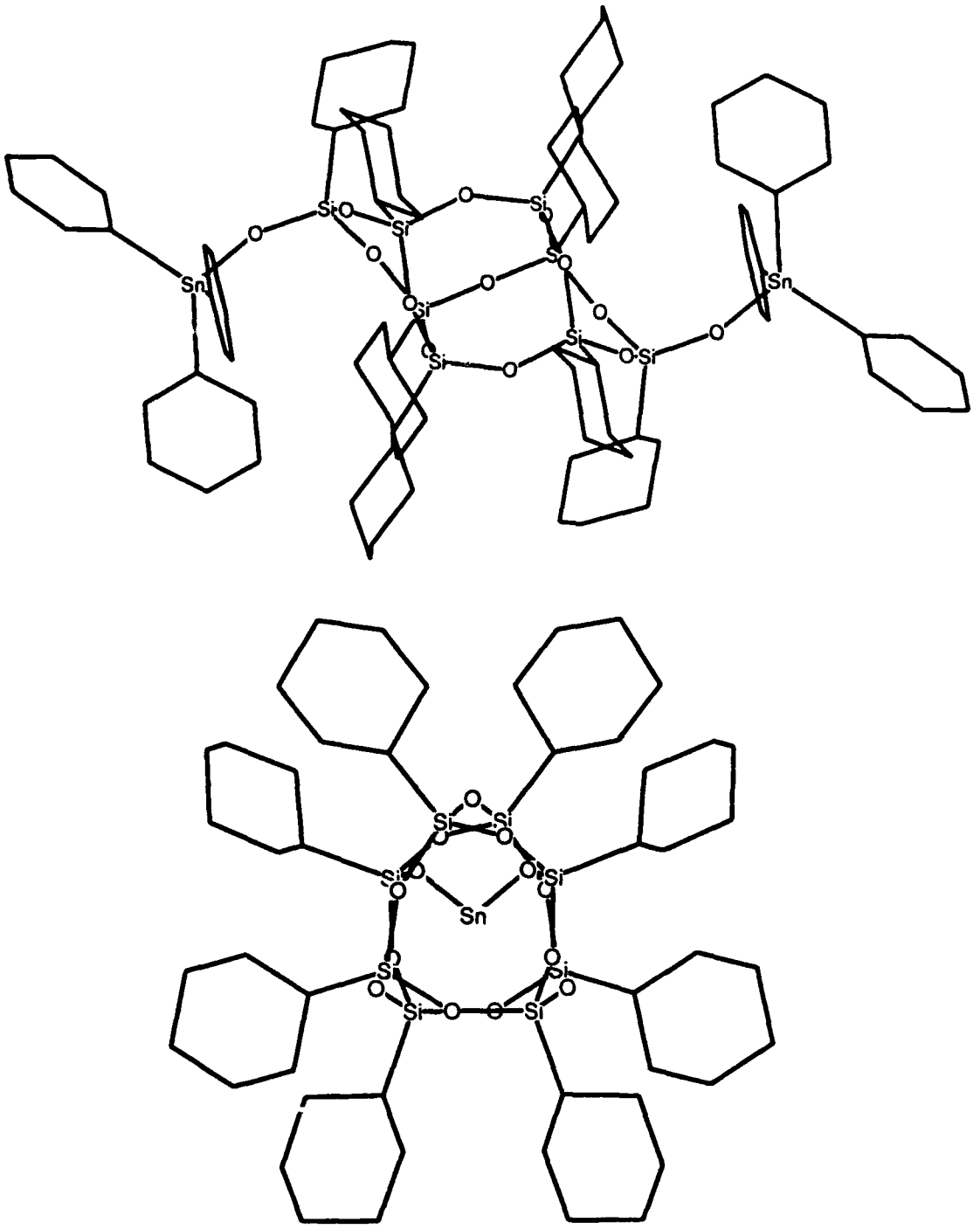


Figure 3: Plot of $[Si_8O_{11}Cy_8(OSnPh_3)_2]$, **1**, viewed down the C_2 axis (top) and perpendicular to the C_2 axis. In the top view, the H atoms have been omitted for clarity; in the bottom view the H atoms and Ph groups have been omitted for clarity.

Environmentally Safe Propellants:

Synthesis of Polymers for Gelation of Liquid Salt Systems

Daniel Lee Fuller

Professor of Chemistry

Department of Natural Sciences

College of DuPage

Glen Ellyn, IL

Final Report for:

Summer Research Program

Phillips Laboratory: Edwards Air Force Base

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington D. C.

September 1993

Environmentally Safe Propellants:

Synthesis of Polymers for Gelation of Liquid Salt Systems

Daniel Lee Fuller

Professor of Chemistry

Department of Natural Sciences

College of DuPage

Abstract

Poly(acrylhydroxamic acid) and poly(acrylamideoxime) were prepared were prepared by the folloing methods. Poly(acrylhydroxamic acid) was prepared from the reaction of poly(acrylamide) and hydroxylamine in basic aqueous solution (pH ≈ 14) at room temperature. Poly(acrlyamideoxime) was prepared from the reaction of poly(acrylonitrile) and hyroxylamine hydrogen chloride and anhydrous sodium carbonate in N,N-dimethylformamide at 75°C.

Binary mixtures of S-HAN-5 with the two polymers were prepared. Poly(acrylhydroxamic acid) is both miscible and compatible, and leads to gelation of the HAN. Poly(acrylamideoxime) appears to be compatible but is not miscible nor did gelation occur in HAN.

Characterization of the polymers have begun. The results of the infrared analysis and the DSC are included. Carbon-13 nmr analysis is in progress. Elemental analysis of the polymers, as well as, impact and friction tests of the binary mixture of S-HAN-5 and poly(acrylhydroxamic acid) are in various stages of progress.

Environmentally Safe Propellants:

Synthesis of Polymers for Gelation of Liquid Salt Systems

Daniel Lee Fuller

INTRODUCTION

Formulation of environmentally safe energetic solid rocket propellants is the goal of rocket scientists. These new environmentally safe propellants must retain or exceed the specific impulse of the present HTPB/Isocyanate cured ammonium 2perchlorate solid propellants. The conventional propellants, utilizing the technologies of the 1950's, had an acid plume of 16-20% HCl. In the 1980's, the scavenged propellants had reduced the acid plume to 1-15% HCl. The present generation of HTPB solid rocket propellants still uses ammonium perchlorate hence can not be classified environmentally safe with respect to HCl and they have limited operational safety. The ultimate goal of the formulator is to produce an environmentally safe rocket propellant producing 0% HCl. The goal is to obtain a propellant possessing low explosive hazard for improved operational safety but retaining the high specific impulse critical for rocket propellants.

Although HTPB base propellants have acceptable mechanical properties, the are obtained with processing difficulty. The processing variables are numerous and, as a result, control of product quality is difficult. The formulators's goal is to simplify the processing required in the preparation of solid propellants. There is a real need to examine new and/or

relatively unknown systems for their use as new solid propellants.

Aerojet Solid Propulsion Company, Sacramento, CA has developed a unique propellant system. The oxidizer is a liquid salt. Adding the fuel, an appropriate polymer, to the liquid oxidizer produces a solid gel. Curing of the propellant results not from a chemical reaction but from swelling of the polymer by the liquid oxidizer. Clean-up does not require ozone depleting chlorinated hydrocarbons as is necessary with HTPB base propellants. Rather, HAN solid gels require only deionized water for clean-up.

Screening tests have indicated that both neutral, i.e. polyacrylamide[1], and ionomeric polymers will bring about solidification of the liquid oxidizers; however, polyacrylamide does not appear to have long term stability. It is necessary to identify compounds that are not only miscible and compatible with liquid oxidizer salts but also have long term stability. Synthesis of PHA reported in the literature include mild hydrolysis of polymeric amidoxime[2], treatment of polyacrylate esters[3] or poly(acrylamide)[4] with hydroxylamine, and the preparation of the monomer with subsequent polymerization[5]. The current study includes the synthesis of poly(acrylhydroxamic acid), PHA, and poly(acrylamideoxime), PAO. We have begun the characterization of these polymers, determining their miscibility and compatibility in HAN. The screening of these polymer/HAN solutions for long term stability, impact, and friction sensitivity are in progress.

EXPERIMENTAL

I. Synthesis of Poly(acrylamide)

A. Procedure

A 0.5-liter jacketed reaction kettle is fitted with a reaction flask-top containing a thermometer, mechanical stirrer, and an inlet for admission of nitrogen below the surface of the solvent. A brisk stream of nitrogen is passed through the system. The flow of nitrogen is reduced to a steady stream as 327 grams of deionized distilled water (DDI water) is added to the reaction flask. The water was deaerated by bubbling a slow stream of nitrogen gas into the water for approximately four hours (Note 1). In succession, the following are added to the water: 0.098 g. of EDTA (Note 2), dilute NaOH until the solution had a pH of 6.99, and 18.5 grams of a 5.5% by wt solution of potassium peroxydisulfate (Note 3). An hour after the addition of the peroxydisulfate, acrylamide (18.6 grams) is added to the solution (Note 4). The reaction mixture is held at 25°C by using a constant temperature bath. After two and a half hours, 29.7 grams of a 0.250 % by wt of MEHQ is added (Note 5). The reaction solution is poured into a two liter beaker containing 1-liter of isopropyl alcohol. The precipitated polymer is washed three times with 150 ml. portions of isopropyl alcohol. The polymer is dried using a vacuum oven at 32° for twenty-four hours. The material is ground in a Micro-Mill and weighed. A yield of 7.62 grams (41%) of polymer was obtained.

B. Notes

1. Deaeration is necessary to remove the dissolved O_2. Oxygen inhibits the polymerization of the monomer by scavenging the free radicals produced by the initiator.

2. The EDTA is added to chelateany heavy metal cations that might be present in the reaction solution.

3. The potassium persulfate , Lot # 907052, was purchased from Fisher Scientific. The solution was prepared by dissolving 1.0144 grams of $K_2S_2O_8$ in 17.545grams of DDI water and deaerated by bubbling N_2 through the solution.

4. Acrylamide, Reagent grade, Lot # 901834, was purchased from Fisher Scientific and was not purified before use.

5. 4-Methoxyphenol, MEHQ, Lot # 9708EX, was purchased from Aldrich Chemical Company and was not purified before use.

II. Synthesis of Poly(Acrylhydroxamic Acid)

A. Procedure

A 0.5-L. jacketed reaction kettle is equipped with a reaction flask-top containing a thermometer, mechanical stirrer, and an inlet for admission of nitrogen to blanket the reaction mixture. Three hundred ml. of DDI water is placed into the flask. As the water is stirred, 20.54 grams (0.29 moles) of poly(acrylamide) is added over a period of fifteen minutes (Note 1). The mixture was stirred over night (Note 2); the resulting mixture was a transparent viscous solution.

The viscous solution is charged with a solution composed of 33.92 grams (0.49 moles) $NH_2OH \cdot HCl$ (Note 3) and 75 ml. of DDI

water; the resulting heterogeneous mixture is stirred for one hour. To the viscous solution, 17.41 (0.43 moles) NaOH dissolved in 75 ml. of DDI water is added dropwise over the course of 15 minutes. The solution is vigorously stirred during the addition of the base. The pH is adjusted to 14 by adding an additional 46.35 (1.16 moles) of NaOH dissolved in 90 ml. of DDI water over the course of two hours (Note 4). The reaction mixture is stirred throughout (Note 5). After the pH is adjusted to approximately 14, the characteristic odor of ammonia is easily detectable. To bring about the termination of the reaction, the pH of the reaction mixture is decreased to pH $\approx$ 1 by the addition of 3 M HCl. The resulting solution is poured into a two-liter beaker containing methanol (Note 6). The white tacky polymer is washed three times with 150 ml. portions of methanol. The methanol saturated polymer is dried in a vacuum oven at 45°C and the solid is ground in the Micro-Mill.

B. Notes

1. We used poly(acrylamide) synthesized as previously described as well as that purchased from Polysciences, Inc. Cat. # 2806, MW 5-6,000,000 ave. The poly(acrylamide) purchased from Polysciences is preferred since the polymer synthesized in our laboratory is observed to destabilized HAN.

2. Addition of the polymer at one time leads to the formation of a water swollen polymer which requires numerous days to form a homogenous viscous solution.

3. The reaction time varied from between 15 hours to 100 hours. Even at the end of 100 hours there was a distinct ammonia odor

over the reaction solution.

4. Both methanol and isopropyl alcohol was used to precipitate the polymer. Methanol is preferred for two reasons. Salts are more soluble in methanol and the isopropyl alcohol used had been stored in metal containers.

III. Synthesis of Polyacrylonitrile

A. Procedure

The hydroquinone inhibitor present in acrylonitile is removed by adsorption of a column of silica gel (Note 1). The commercial acrylonitrile is slowly trickled through the column at the rate of approximately one drop per second. The treated liquid is collected in a flask flushed with a slow stream of nitrogen. Twenty ml. of acrylonitrile is collected.

Five hundred ml. of DDI water are deaerated by boiling for 10 minutes. A 0.5 liter jacketed reaction kettle is equipped with a thermometer, mechanical stirrer, reflux condenser, and an inlet for admission of nitrogen to deaerate the reaction mixture. Three-hundred ml. of the boiled water is poured into the reaction flask. The temperature is brought to 75°C using a constant temperature bath (Note 2). The 20 ml. of acrylonitrile is placed in the flask and stirred for 10 minutes. To this solution $Na_2S_2O_8$ (0.3087 grams) dissolved in 10 ml. of non-deaerated water (Note 3) is added. The resulting solution is stirred for about ten minutes followed by the addition of an aqueous solution of sodium bisulfite, 0.152g of $NaHSO_3$ in 10 ml. of non-deaerated water. After the addition of the sodium bisulfite solution, the reaction mixture is stirred at 75°C for two hours

(Note 4). The colorless, aqueous solution becomes opaque and the white polymer begins to precipitate. The solid product is filtered using a Buchner funnel, washed with 50-ml. portions of DDI water, and dried in a vacuum oven at 36°C overnight. The solid is ground in a Micro-Mill.

B. Notes

1. A prepackaged glass column was purchased from Aldrich, Stock NO. 30,632-0. The silica gel filled column is specifically designed to remove hydroquinone and hydroquinone monomethyl ethers.

2. Lower temperatures did not lead to polymerization. It is reported in the literature that polymerization occurs readily at 40°C. It was assumed that we might have had hydroperoxides present that inhibited the reaction at the lower temperatures; however, at 75° the hydroperoxides decomposed which allows the polymerization reaction.

3. The use of water that was not deaerated did not noticeably impact the polymerization reaction.

4. When the reaction is carried out at 40°C the reaction rate is unacceptably slow.

IV. Synthesis of Poly(Acrylamideoxime)

A. Procedure

A 0.5 liter jacketed reaction kettle is equipped with a thermometer, mechanical stirrer, reflux condenser, and an inlet for admission of nitrogen to supply a blanket over the reaction mixture. Two-hundred ml. of N,N-dimethylformamide (Note 1) is

added in the reaction flask. The liquid is heated to 75°C using a constant temperature bath, and 6.3 grams of polyacrylonitile (0.312 moles of repeat units) is added. The mixture is stirred until the polyacrylonitrile dissolves (approximately 2 hours). Solid $NH_2OH \cdot HCl$ (16.6 grams, 0.377 moles) is added and the mixture stirred for thirty minutes. Anhydrous Na_2CO_3 (12.68 grams, 0.1844 moles) is added (Note 2 and 3). The mixture stirred for 25 hours. The stirring is stopped and the solids allowed to settle. The liquid layer was removed in portions using a 60 cc. syringe. During the removal of the liquid, care is taken not to agitate the solid layer. The liquid layer is slowly added to two liters of methanol. The white insoluble polymer is removed by filtration through a Buchner funnel. The filter cake is washed with three 150-ml. portions of methanol. The white solid is dried in a vacuum oven at 45°C and then pulverized in a Micro-Mill. The yield of the polymer is 9.4 grams.

B. Notes

1. N,N-Dimethylformamide, anhydrous, 99+%, Lot # 14220EY was purchased from Aldrich Chemical Company and was not purified before use.

2. Hydroxylamine Hydrochloride was "Baker Analyzed" reagent grade, Lot # 37738.

3. Production of a gas, CO_2, occurs as the carbonate is added to the acid acidic solution. The degree of foaming can be controlled by the rate of addition of the Na_2CO_3.

RESULTS

Poly(acrylhydroxamic acid)

The synthesis of PHA was based upon a method which had been reported by *Domb, Cravalho, and Larger*[4] . In this synthesis the polyacrylamide polymer undergoes base catalyzed acyl substitution. The literature procedure was modified by increasing the reaction time by a factor of four. The PHA while still in solution showed a strong tendency to chelate with iron producing a brownish colored substance.

Both infrared spectrophotmetry and thermal analysis was used to characterize the polymer. An infrared spectrum was obtained on an FT-IR with a photoacoustic cell (PAS-FTIR). The spectrum could not be used to confirm the product as PHA. Although the spectrum was consistent with PHA, the absorption of both the starting material, an amide, and a secondary product, carboxylic acid, led to overlapping absorption peaks. The DSC thermograms of the polymer varied appearing to be effected by the specific manner of work-up of the product and the length of time the substitution reaction was allowed to occur.

Poly(acrylamideoxime)

The PAN was converted to PAO following the procedure of *Sorenson and Campbell*[6]. Infrared spectra of the product shows an intense absorption band at 2250cm^{-1}. Absorption in this range is indicative of cyanide. Changing the reaction time from heating at 75°C for three hours to 24 hours led to a product which did not show the nitrile band.

Both infrared spectrophotmetry and thermal analysis was used to characterize the polymer. An infrared spectrum was obtained on an FT-IR with a photoacoustic cell (PAS-FTIR). The spectrum could not be used to confirm the product as PAO. Although the spectrum was consistent with oximes, the absorption of the starting material, an amide, led to overlapping absorption peaks. The DSC thermograms of PAO showed a large endotherm at around 160°C.

Compatibility and Miscibility of HAN Mixtures

Preliminary screening tests were performed for compatibility and miscibility of Liquid oxidizer, HAN, and PAO and PHA. The concentration of the additive polymers to HAN was approximately 10% by weight of the total mixture. Assessment of the compatibility and gelation was qualitative. The evaluation was based principally on the appearance of the resulting mixture after one hour, twenty-four hours, and one week. PHA showed compatibility and gelation over the entire test period. PAO showed a tendency to gel in HAN after three weeks exposure to HAN.

REFERENCES

1. D. L. Fuller, *Environmentally Safe Propellants: Ionophilic Polymers In Liquid Salt Systems*, Final Report: Summer Research Program, Summer 1992, Sponsored by the Air Force Office of Scientific Research.

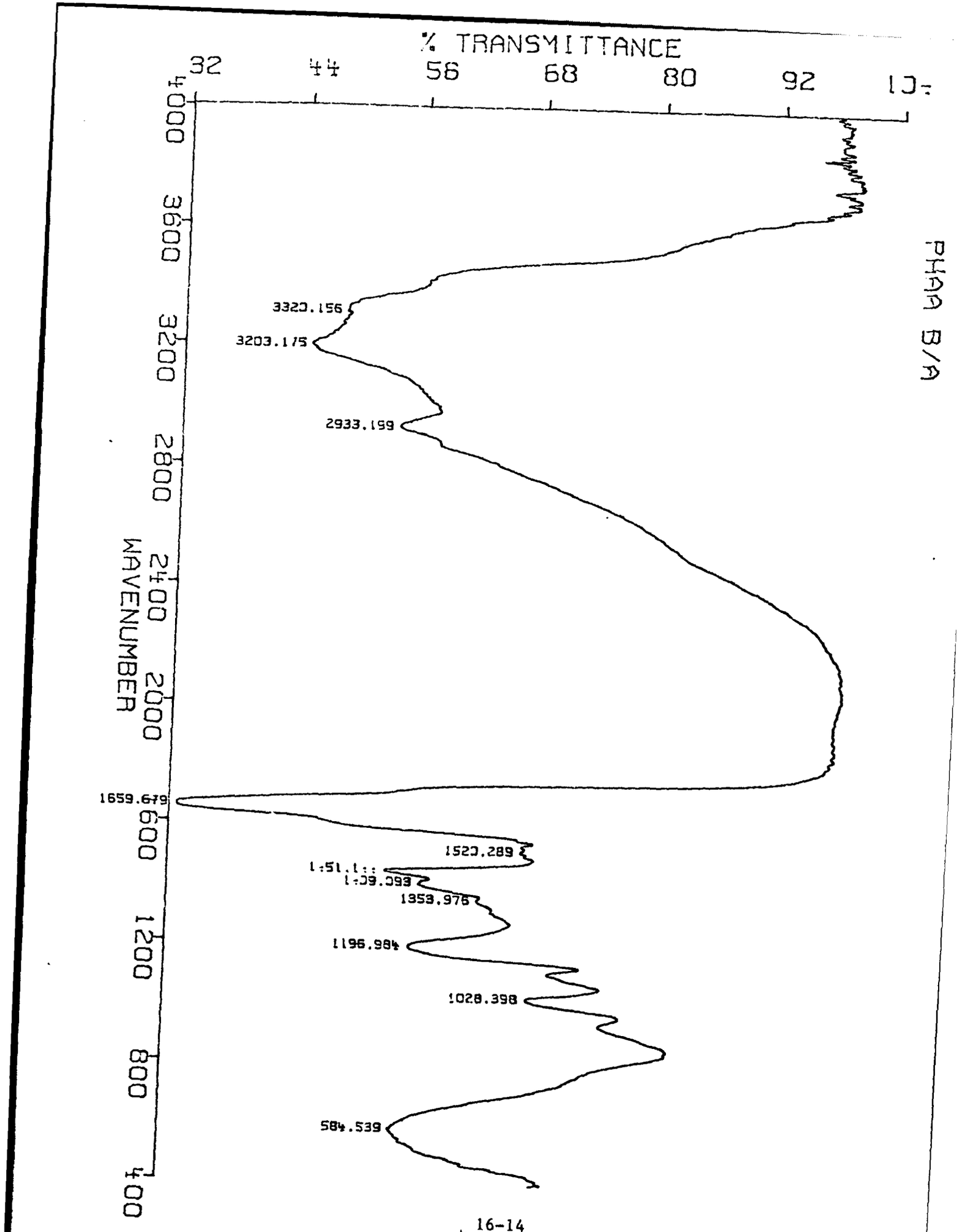
2. C.A. Fetscher and S.A. Lipowski, U.S. Pat. 3,345,344 (1967); *Chem. Abstr.*, **67**, 10927a (1967).

3. M. Hatano, Y. Nose, T. Nozawa, and S. Kambara, *Kogyo Kagakeu Zasshi*, **69**, 571 (1966); *Chem. Abstr.*, **65**,15532g (1966).

4. A.J. Domb, E.G. Cravalho, and R. Langer, *J. Polym. Sci.; Part A: Poly. Chem.*, **26**, 2623 (1988).

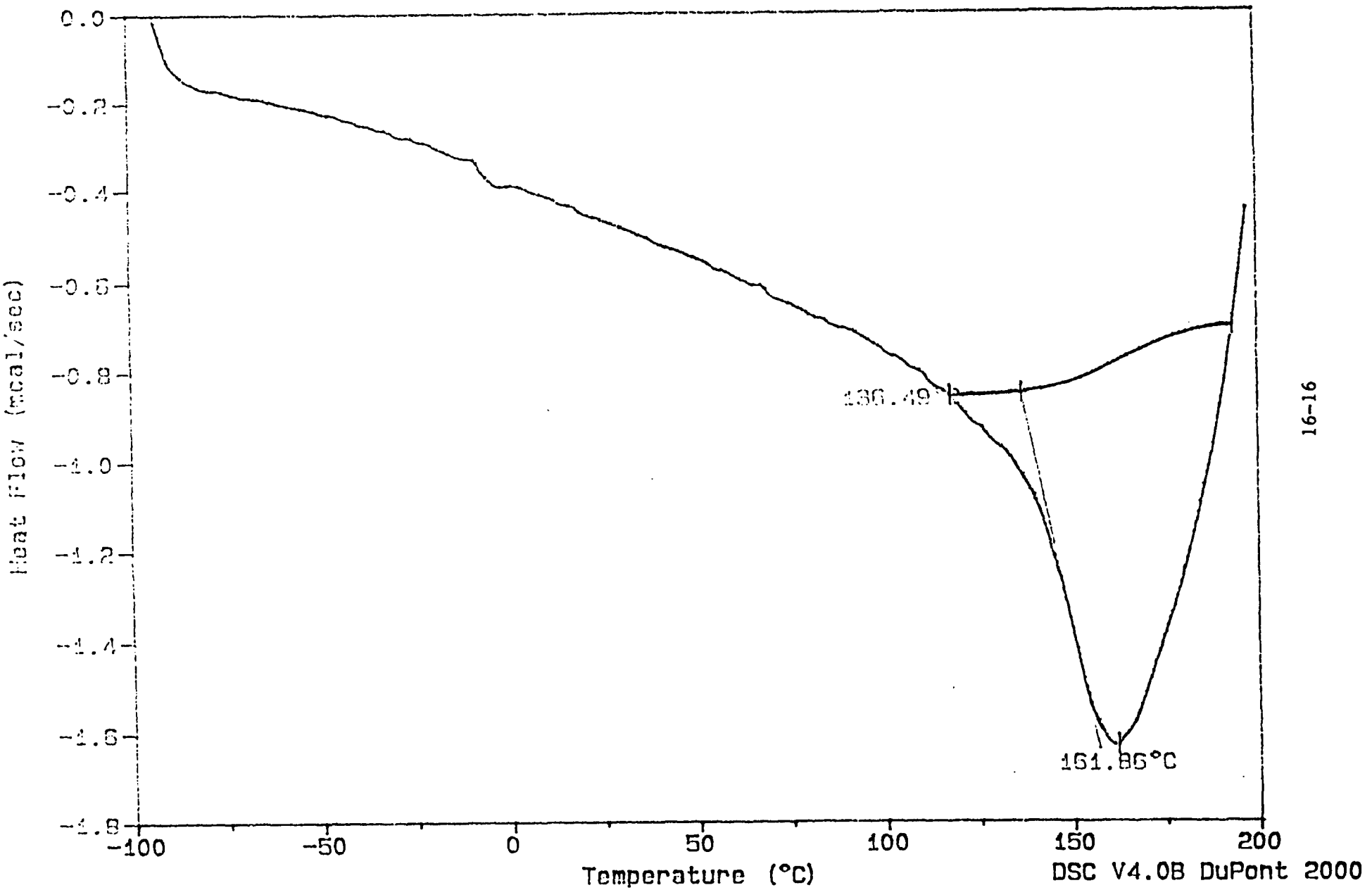
5. F. Becke and G. Mutz, *Chem. Ber.*, **98**, 1322 (1965).

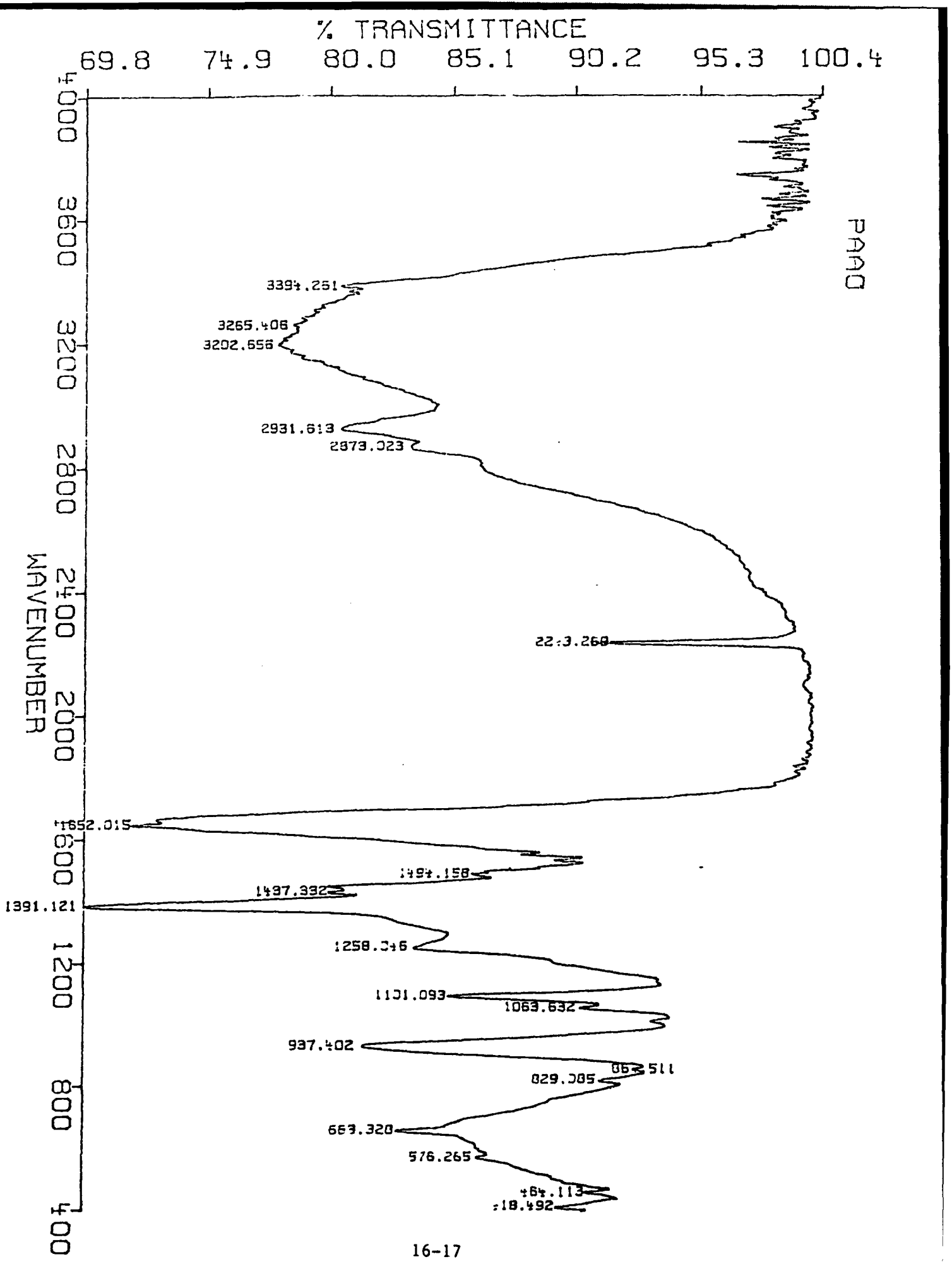
6. W. Sorenson and T. Campbell, *Preparative Methods of Polymer Chemistry*, 2nd. Ed., Interscience Publ., New York, 1961, p.237.

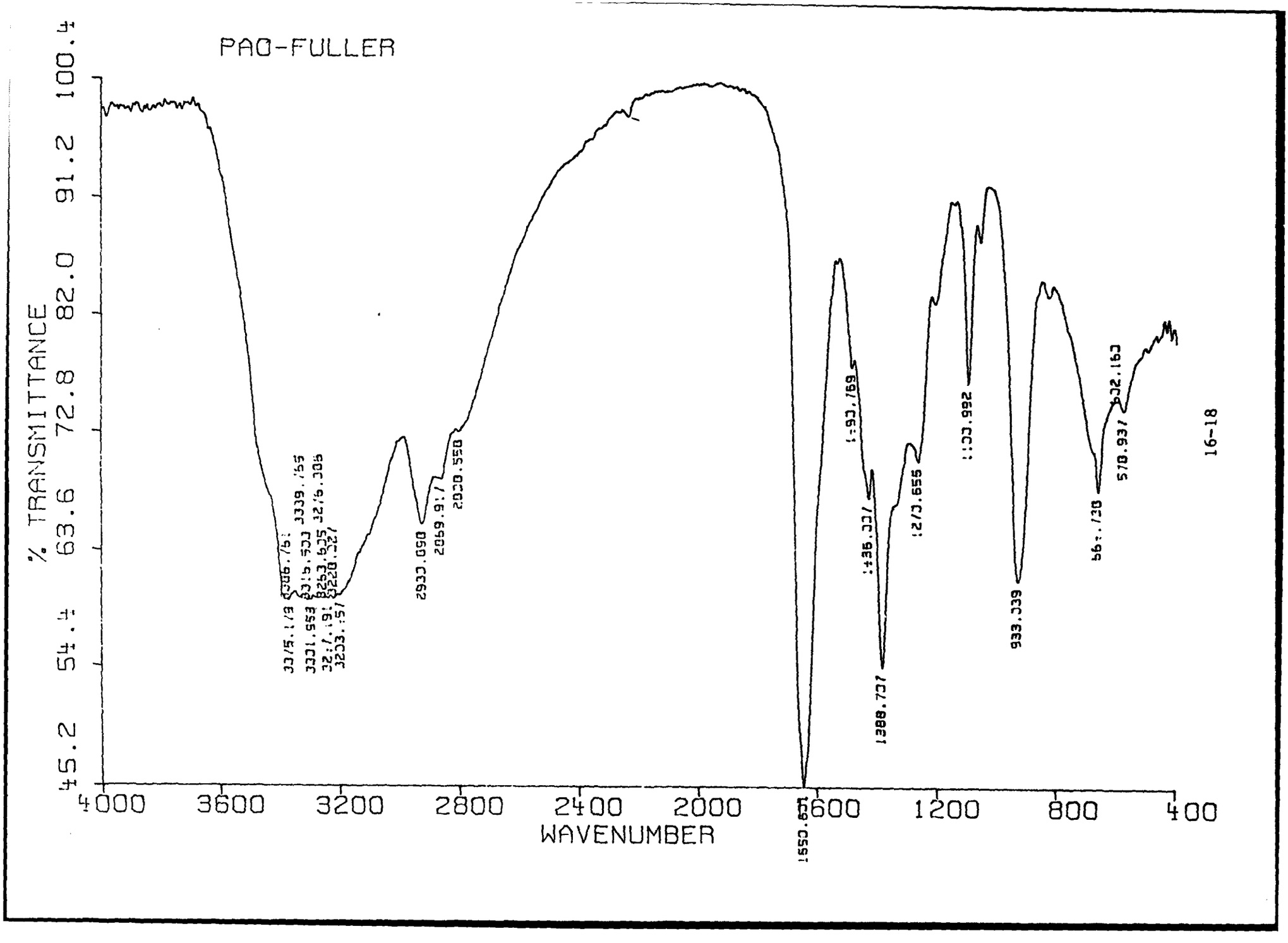

% TRANSMITTANCE
32
44
56
68
80
92
4000
3600
3200
2800
2400
2000
1600
1200
800
400
WAVENUMBER
3320.156
3203.175
2933.159
1659.679
1520.289
1451.111
1409.093
1358.976
1196.984
1028.398
584.539
PHAA B/A

Sample: PHA DAN'S
Size: 7.9000 mg
Method: GREG
Comment: RATE 10°C/MIN, GN2 ATM 50 ML/MIN, HERMETIC AL PANS
DSC
File: DSCSCOTT.37
Operator: JONES / LOSEY
Run Date: 13-Jul-93 13:19
173.55°C
130.75°C
Heat Flow (mcal/sec)
Temperature (°C)
DSC V4.0B DuPont 2000

Size: 13.8000 mg
Method: GREG
Run Date: 13-Jul-93 10:16
Comment: RATE 10°C/MIN, GN2 ATM 50 ML/MIN, HERMETIC AL PANS




% TRANSMITTANCE
69.8
74.9
80.0
85.1
90.2
95.3
100.4
4000
3600
3200
2800
2400
2000
1600
1200
800
400
WAVENUMBER
PAAD
3394.251
3265.406
3202.656
2931.613
2873.023
2243.268
1652.015
1494.158
1437.392
1391.121
1258.346
1131.093
1063.632
937.402
829.305
663.320
576.265
464.113
418.492


PAO-FULLER
% TRANSMITTANCE
100.4
91.2
82.0
72.8
63.6
54.4
45.2
4000
3600
3200
2800
2400
2000
1600
1200
800
400
WAVENUMBER
1388.737
933.339
570.937
16-18

SPRAY MEASUREMENTS USING MICROWAVE BEAM

San-Mou Jeng
Associate Professor
Department of Aerospace Engineering and Engineering Mechanics

University of Cincinnati
Mail Location #70
309 Rhodes Hall
Cincinnati, OH 45221

Final Report for:
Summer Faculty Research Program
Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1993

SPRAY MEASUREMENTS USING MICROWAVE BEAM

San-Mou Jeng
Associate Professor
Department of Aerospace Engineering and Engineering Mechanics
University of Cincinnati

ABSTRACT

The use of a microwave beam for spray diagnostics has been investigated. Microwave attenuation through the spray was estimated. It was found that microwave and spray interaction can be modeled using Rayleigh theory, and microwave can penetrate optically thick sprays. Also, microwave transmission efficiency is strongly dependent on the selected wavelength. Based on these microwave properties, it was concluded that non-intrusive probes based on microwave beams (frequency higher than 10 GHz) may offer new diagnostic capabilities for optically thick sprays. A probe to measure spray patternation based on the microwave absorption can be designed.

SPRAY MEASUREMENTS USING MICROWAVE BEAM

San-Mou Jeng

INTRODUCTION

The liquid rocket spray combustion process has been subjected to many experimental investigations. Rapid advances in optical technology have produced many laser-based probes for spray combustion measurements. Examples are Laser Doppler Velocimetry (LDV)[1], Phase Doppler Particle Analysis[2] (PDPA) and Fraunhoffer Diffraction methods[3]. These probes allowed access to previously unavailable valuable information regarding spray combustion characteristics. However, they still face a number of challenges when applied to realistic rocket combustion environments. For example, they cannot be used in optically thick sprays, delicate optical alignments can be sensitive to mechanical and acoustic vibrations, and they require the installation of fragile optical windows. Alternatively, other probes such as the use of X-rays or Nuclear Magnetic Resonances (NMR) are being investigated by a number of researchers. However, little attention has been given to the use of microwave techno in this report, the possibility of using a microwave beam for spray diagnostic is described.

The electromagnetic (EM) wave that passes through a two-phase medium (spray combustion) will be attenuated by both the liquid and gaseous phases. The extinction of the EM wave can be due to the Mie scattering and absorbing cross section of tiny liquid droplets as well as to the gaseous absorption. For a typical LO_x/H_2 fueled rocket, the combustion chamber contains two-phase chemically reacting flows having liquid LO_x droplets and gaseous H_2,

O_2 and H_2O combustion products. Very little work has been conducted to study microwave and rocket spray combustion spray interactions. However, microwave propagation through atmospheric gases, rain, and fog is well documented, as shown in Figure 1.[4] Both O_2 and H_2O gases absorb microwave radiation at several discrete wavelengths. Fog, which has a similar drop size distribution to a rocket combustion spray, attenuates microwaves significantly less than visible light. These results imply that microwaves can penetrate through sprays that might otherwise be optically thick, and thus allow measurements in dense sprays that would not be possible using optical techniques. Motivated by these findings, the interactions between spray droplet field and microwave were studied.

TECHNICAL DISCUSSION

Attenuation of an electromagnetic ray through a two-phase medium is described in equation (1) where the multiple scattering effect from liquid droplets is ignored.

$$\frac{E_t}{E_0} = e^{-\int \alpha_{ext} dl} \qquad (1)$$

E_o is the initial EM ray energy before entering the medium and E_t is the EM ray energy leaving the medium. l is the distance along the EM ray, and α is the local extinction coefficient of medium per unit length. Both liquid droplets and gases can contribute to ray attenuation. Therefore, the local extinction coefficient is written as

$$\alpha_{ext} = \alpha_{gas} + \alpha_{drop} \qquad (2)$$

where subscripts of *gas* and *drop* separate the portions of the extinction coefficient contributed by droplets and gases, respectively.

Extinction due to liquid droplets

The indexes of refraction of liquid oxygen in the microwave frequencies are not well known. Those of liquid water were used[5,6], as listed in Table 1, to conduct the attenuation cross section calculations. The droplet diameter distribution of rocket spray injectors ranges from 5 to 100 micron. The EM beam that propagates through the spray will be attenuated by both scattering and absorption from each individual droplet. Since microwave wavelengths (>3 mm) are much longer than the maximum expected droplet diameters in rocket sprays (<200 μm), the interaction between droplets and microwaves can be properly modeled using Rayleigh theory. The extinction cross section of each droplet can be written as:

$$C_{ext} = C_{sca} + C_{abs} \tag{3}$$

where

$$C_{sca} = \frac{2\lambda^2}{3\pi} x^6 Re\left[\frac{m^2-1}{m^2+2}\right]^2 \tag{4}$$

and

$$C_{abs} = -\frac{\lambda^2}{\pi} x^3 Im\left(\frac{m^2-1}{m^2+2}\right) - \frac{1}{15}\frac{\lambda^2}{\pi} x^5 Im\left[\left(\frac{m^2-1}{m^2+2}\right)^2 \frac{m^4+27m^2+38}{2m^2+3}\right] \tag{5}$$

x is the droplet size parameter, defined as $\pi d/\lambda$. m is the index of refraction of the liquid and λ is the microwave wavelength. This equation has fourth order accuracy in x. The extinction of microwaves by water droplets is typically dominated by absorption while the scattering effect is negligible. The smaller the size parameter, the more accurate are the above equations. For a very small droplet size parameter, the second term of the absorption coeffi-

cient in equation 5 can be neglected with little error. In this case, the extinction coefficient of each individual droplet is proportional to the size parameter to the third power, and at each selected microwave frequency, the extinction coefficient of each droplet is proportional to the droplet mass. The accuracy of this approximation is examined. Table 2 shows the droplet diameters when the above approximation will reach ten percent inaccuracy at different microwave frequencies. For typical rocket sprays, which have maximum droplet size less than 200 μm, the approximation of the extinction coefficient, which is directly proportional to the liquid droplet mass, will induce little error. The attenuation coefficient per unit length of microwave through a collective droplet field is the summation of each droplet extinction within the ray path and is written as:

$$\alpha_{drop} = \sum C_{drop} N \qquad (6)$$

where N is the droplet density per unit volume. The extinction coefficients at a microwave frequency of 10 to 100 GHz are shown in Table 3. The extinction coefficient is a strong function of the microwave wavelength, varying by four orders of magnitude over the considered frequency. This indicates that a spray field can either be totally transparent to microwaves or a strong absorber depending on the selected frequency.

Potential rocket applications

A microwave probe to measure beam attenuation through a spray, and thereby total mass along the beam path, can probably be designed. Tomographic reconstruction of the mass distribution in a combustion chamber could eventually become possible by instrumenting the chamber with multiple probes and receivers. Since point source transmitters and receivers are

essentially just antennae, the probes could be made reasonably durable. Cooling would be required in hot-fired applications. This would not seem to be an insurmountable problem, especially in cryogenic engines.

The resulting measurements would have several potential uses. Condensed phase mass distributions cannot currently be easily measured in hot fired applications. Especially appealing is the possibility of making time resolved measurements, thus making possible the detection and study of acoustic instabilities. Acoustic instabilities are currently measured using wall mounted pressure transducers. Microwave probes, if successful, would allow the measurement of acoustic response in the bulk flow. Comparison between the relative magnitudes and phases of the wall mounted and bulk measurements could lead to important insights into the acoustic instability mechanisms. Finally, flightweight systems could probably be designed due to the small size of the probes, with potential health monitoring applications.

Conclusions

In this study, the following fundamental properties of microwave and spray interactions were found: (1) Microwave can penetrate optically thick spray where its transmission efficiency is highly dependent on its wavelength. (2) Microwave and spray droplet interaction can be modeled by Rayleigh theory where the extinction coefficient is proportional to the liquid mass along the ray path. These results imply that microwaves can penetrate through sprays that might otherwise be optically thick, and thus allow measurements in dense sprays that would not be possible using optical techniques.

Suggested study:

It is suggested that an exploratory study, based on the attenuation of microwaves through sprays as discussed above be conducted. Both experimental and theoretical work will be accomplished. Major issues that should be considered are the following:

1. Microwave probe system analysis. The accuracy of Rayleigh theory for microwave/droplet interactions should first be addressed, along with an assessment of the current state of microwave technology as potentially applied to sprays. A probe system should be designed, and the accuracy and capability of the system should be analyzed, including but not limited to assessing the accuracy of the absolute and relative liquid mass distribution measurements, and the temporal and spatial resolution of the probe.

2. Demonstration of concept (steady state). The designed probe should be used to measure a small laboratory scale steady spray (cold flow). Computer software should be written to convert from the line of sight attenuation measurement to the local liquid mass fraction. Reasonableness of the measurements should be judged by comparison to experimental data available in the literature.

3. Unsteady tests. An attempt should be made to conduct unsteady measurements of an acoustically excited spray. The qualitative and quantitative ability of the probe to track the acoustic excitations will be assessed for potential application to the study of acoustic instabilities in rockets.

References:

1. Yeh, Y., and Cummins, H.Z., *Applied Physics Letter*, 4. 1964, p176.
2. Bachalo, W. D., *Proc. Third Intl. Workshop on Laser Velocimetry*, Purdue Univ. , 1978.
3. Hodkinson, J.R., and Greenleaves, I., *J. Opt. Soc. Am.*, 5, 1963, p.577.
4. Preissner, *J. AGARD Conference Reprint No. 245*, Millimeter and Submillimeter Wave Propagation and Circuits, 1978.
5. Lowan, A.N., Natl. Bur. Standards (U.S.) Appl. Math. Series 4, 1949.
6. Aden, A.L., *J. Appl. Phys.*, 22, 1951, p.601.

λ (mm)	h (GHz)	m
3	100	3.41 - 1.94 i
5	60	4.21 - 2.51 i
9	33	5.55 - 2.85 i
12.5	24	6.41 - 2.86 i
17	17.5	7.20 - 2.65 i
30	10	8.18 - 1.96 i
100	3	8.90 - 0.69 i

Table 1. Index of Refraction (Water Droplets at 18°C)

λ (mm)	h (GHz)	d (μm)
2	150	120
5	60	200
7	41	220
10	30	280
12.5	20	320
20	15	380
30	10	480
50	6	490
100	3	640

Table 2. Water Droplet Diameters When 10 % Error Occurs
(Volumetric Absorption Approximation)

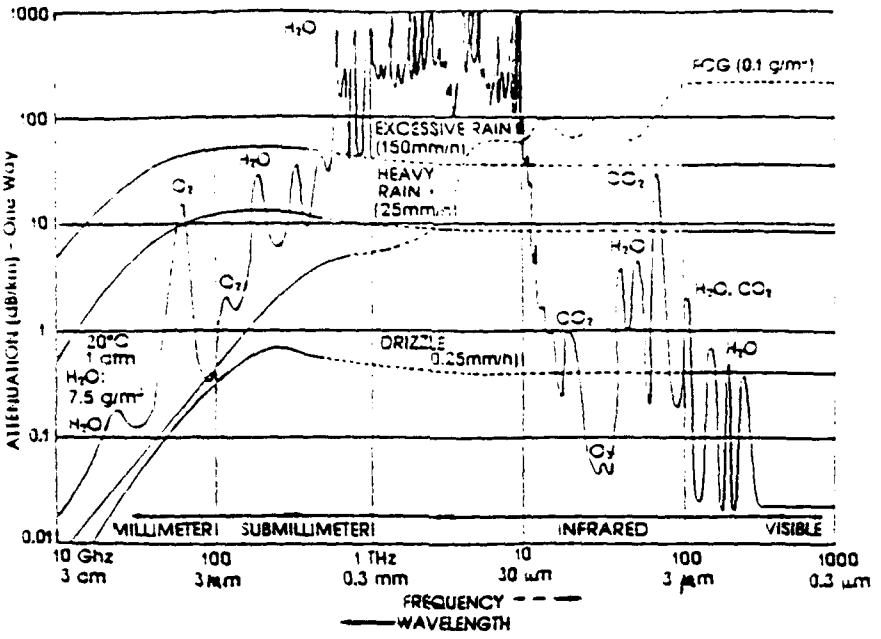


Fig. 1 Electromagnetic wave attenuation by atmospheric gases, rain and fog (from Preissner)

λ (mm.)	Attenuation coefficient α (mm^{-1})
3	7.3254
5	3.0504
9	0.9877
12.5	0.5177
17	0.2784
30	0.0919
100	0.0083

$$I_{transmit} = I_{initial} * e^{-\int \alpha V_d \, dl}$$

α : attenuation coefficient
V_d : liquid volume fraction
l : pathlength

Table 3: Attenuation coefficient (water droplet at 18° C)

RESEARCH IN EMERGING TECHNOLOGIES AT PHILLIPS LABORATORY

Peter W. Langhoff

Professor of Chemistry

Department of Chemistry
Indiana University
Bloomington, IN 47405

Final Report for:

Summer Faculty Research Program

Phillips Laboratory

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, DC

and

Indiana University

Bloomington, IN 47405

September 1993

RESEARCH IN EMERGING TECHNOLOGIES AT PHILLIPS LABORATORY

Peter W. Langhoff

Professor of Chemistry

Department of Chemistry
Indiana University
Bloomington, IN 47405

Abstract

Progress is reported in theoretical and computational studies related to the research interests of the Emerging Technology Branch at Phillips Laboratory, Edwards AFB. Topical areas under investigation include (i) alkali metal vapor spectroscopy & solar rocket propulsion (ii) Al_2 & AlLi vapor and high-energy density materials (iii) high energy density materials evaluation, (iv) spectra of trapped atomic radicals, (v) instrumentation for the Larson/Edwards plasma spectroscopy cell, (vi) an ARPA/Edwards solar receiver project, and (vii) aspects of cavity QED experiments. In each case progress is reported, the current status of the research effort is provided, and work in progress or continuing is indicated.

RESEARCH IN EMERGING TECHNOLOGIES AT PHILLIPS LABORATORY

Peter W. Langhoff

Professor of Chemistry

Introduction

Progress is reported on research performed in the Emerging Technologies Branch of the Fundamental Technology Division, Phillips Laboratory, Edwards AFB, under the auspices of the RDL-AFOSR Ressearch Associates Program during the summer of 1993. Reference is made to the attached Appendix listing three principal and four secondary research topics identified for study during the course of the research period (6 June to 28 August 93). Status reports are provided in each topical area and some recommendations are made for continuing studies.

Alkali Metal Vapor Spectroscopy & Solar Propulsion

A summary report of experimental and theoretical studies of alkali metal and other vapors relevant to solar propulsion and high-energy-density materials considerations was prepared and presented by PWL at the Phillips Laboratory Technical Interchange Symposium held in Albuquereque, NM, on 14 July 93. An abstract and letter of appreciation from James M. Romero, Deputy Director, Phillips Laboratory, are attached for reference. A manuscript and report describing attributes of volumetric alkali-metal absorbers in hydrogen gas for solar propulsion is in preparation for presentation at the JANNAF Conference on Propulsion to be held in Monterey, CA, 15-18 November 93. The spectral absorbance and emission code required in modelling volumetric solar absorbers has been completed and is ready for additional applications to specific experimental metal vapor studies.

Al_2 & AlLi Vapor and High-Energy Density Materials

Thermodynamic and spectroscopic calculations were performed in support of experimental efforts to form and detect Al_2 and AlLi vapor in the Larson Plasma Spectroscopy Cell (PSC). Concentrations of Li, Al, Li_2, Al_2, and AlLi vapors were predicted at thermodynamic conditions appropriate to the PSC on basis of available liquid metal vapor pressure curves and calculated gas-phase partition functions. The later were obtained employing potential energy curves for the three relevant diatomic species in the absence of JANNAF or other reliable data. Detailed absorbance spectra were constructed for Al_2 employing theoretically derived electronic transition moments and the computer code indicated above in vibrational and rotational resolution. Related spectral absorbance calculations are in progress for AlLi, in which case further refinements are required of the relevant electronic transition moments (Rosenkrantz). Concomitant experimental studies (Larson, Erdman) are in progress at this date. A separate more detailed report on this topic is provided

in this volume by J.D. Mills, a collaborating graduate student from Indiana University, Bloomington.

High-Energy Denity Materials Evaluation

The thruster designed by Bornhorst, et al., and the experimental set-up employed at the Phillips/Edwards test facility under the supervision of M. Nichols and S. Harris was inspected at the start of the summer research period. Available data sets for each candidate HEDM propellant material are apparently comprised of mass flow rates of fuel and oxidizer, overall chamber pressure, nozzle temperature, and, in selected cases, optical spectra. So-called c^* values (mass-weighted exit velocities) can be determined from the known or measured flow rates and chamber pressures, whereas the optical spectra give qualitative indication of completeness of combustion from observed particulate luminosity and identified emission bands. These fuel performance data relate in unknown ways to a variety of ΔH values and to kinetic events of various type, including liquid fuel droplet formation rates and size distributions, fuel droplet evaporation rates in F_2 and/or O_2 atmospheres, gas-liquid and gas-gas mixing rates, and gas-liquid droplet and elementary gas-phase chemical reaction rates. Data analysis reported by other workers (Surratt, Grism) indicate relatively modest improvements in HEDM materials c^* values ($\approx 4\%$ at optimal mixing ratio) relative to RP1 to date. In view of this circumstance, a detailed model of the Bornhorst thruster, or a carefully planned experimental kinetics program, is recommended to help in the evaluation of HEDM materials. In view of the great extent of incomplete combustion evident in some date sets inferred from the high particulate concentrations revealed by broad-band (grey-body) spectral luminosity, it can be anticipated kinetic analysis and remedial design of combustion circumstances may lead to considerably higher c^* from existing HEDM materials than have been measured to date in the thruster apparatus.

Spectra of Trapped Radicals

An optical potential approach was developed in attempts to improve upon the Balling & Wright technique employed by M. Fajardo and coworkers in the analysis of the spectra of trapped metal atoms trapped in low-temperature inert gas and hydrogen matrices. Although the present development provides an *in principle* solution for construction of more accurate many-body potentials from available perturber-metal atom pair potentials, its implimentation in present form apparently requires use of theoretically derivable information. Additionally, it appears the correct many-body perturber potential in pair-wise approximation must be inferred from knowledge of the interaction potentials of a single metal atom with at least *two* perturbers, in contrast to the Balling & Wright approach, which requires only the potential due to a *single* perturber. It is hoped to continue work on this problem in future and to avoid use of calculated data in application of the approach.

Instrumentation for Larson PSC

The Larson PSC has proved to be a reliable source of controlled amounts of equillibrium metal vapors (Li, Li_2, Na, Na_2) at considerably higher densities than have been achieved employing other devices, such as commonly used heat pipe ovens. The facility would be

of significantly greater benefit to the HEDM and Solar Rocket Propulsion Programs with additional instrumentation for the detection and characterization of metal vapors relevant to these programs. Specifically, an additional detector to perform optical spectrometry, and a higher resolution optical spectrometer, would improve the quality and value of absorbance measurements. Additionally, use of optical holographic interferometery would provide the basis for performing dipersion spectroscopy measurements complementary to absorbance experiments now performed. In view of the relatively low concentrations of Al_2 and AlLi vapors anticipated in the current experiments, consideration should also be given to assembly of a laser induced fluorescence apparatus to increase detection sensitivity. Separate white papers and a NASA-SBIR proposal describe possible dispersion and LIF detection experiments involving the Larson PSC.

ARPA/Edwards Solar Receiver Project

The Edwards/Phillips Solar Furnace Facility provides a useful test facility for development of solar energy receivers for use in Stirling and Brayton power conversion devices, for use in solar cooling units employing desiccant and absorption chiller technologies, and in solar rocket propulsion trusters. Relevant proposal have been written in response to the ARPA "Dual-Use" Technology Reinvestment Project in an attempt to gain support for a collaborative program involving Phillips/Edwards and public (Indiana University, San Diego State University) and private-sector (Cummins Engine, Bechtel National) team members. A separate white paper describes the aims and objectives of this program in further detail. In brief, the receiver unit proposed for development at Phillips/Edwards involves a fluid absorption approach explored earlier by PWL and coworkers under the auspices of NASA-Lewis and Wright-Patterson AFB support. The technology is relevant to commercial applications (Cummins, Bechtel) and also to the requirements of the Edwards Solar Rocket Propulsion Program. Accordingly, there is a dual use mission in this case possibly worthy of support under the above indicated ARPA TRP solicitation.

Cavity QED Experiments

Preliminary background studies were performed relevant to cavity QED theory and experiments for extracting energy from the zero-point modes of the quantized electromagnetic field. It is hoped to devise a go/no-go theorem on theoretical grounds for this phenomenon, and in the positive case to devise a cavity experiment for producing a beam of electronically excited metal vapor atoms employing an appropriately shaped micro-cavity. Very small cavities are required in order to increase the intensities of the zero-point modes, which vary inversley with the square root of the cavity volume. Moreover, the detailed cavity shape controls the spectral distribution of intensity over the allowable modes, which can be chosen to maximized a given electronic transition. Current literature indicates technological issues associated with fabrication of such cavities have been overcome, and experiments have been performed demonstrating the de-excitation of optically excited metal atoms by simple shaped cavities. Zero-point mode excitation of ground-state or optically excited metal vapor beams has yet to be demonstrated experimentally, but as of this date a "no-go" theorem precluding such a possibility is apparently lacking.

Appendix - Research in Emerging Technologies

Topic List - Principal Items

1. Complete Li/Li_2 & Na/Na_2 spectral computations, including construction of synthetic absorbance and emission spectra over temperatures and partial vapor pressures appropriate to the Larson, et al., exprimental data, employ integrated absorbance/emisssion profiles in model studies of solar rocket flow configurations, and prepare manuscripts for the PL 2nd Annual Technical Interchange Symposium, 14 July 93, and the - with C.W. Larson, et al.
2. Construct spectral absorbance profiles for thermodynamic equillibrium combinations of Al, Li, AlLi, Li_2 and Al_2 employing available experimental and theoretical spectral and structural information at temperatures and partial vapor pressures appropriate to the experiments planned in the Larson PSC - with C.W. Larson, et al.
3. Assist the HEDM Evauation Project in interpretation of measured c^* values and of spectral emission data for estimating combustion efficiencies - with C.W. Larson, B. Nichols, & T. McElvey.

Topic List - Secondary Items

1. Assist in the interpretation of trapped B, N, B_2, N_2 spectra for solid H_2 detection and related low-temperature spectroscopy issues - with M. Fajardo.
2. Discuss design of high-framing-rate resonance refractive flow diagnostic experiment for development in the Larson PSC - with C.W. Larson.
3. Discuss design of volumetric solar propulsion and/or Brayton turbine experiment for possible use of the Edwards solar collector facility - with K. Laug, et al.
4. Discuss design of cavity QED experiments for production of excited-state atomic beams employing zero-point-energy modes - with F. Mead

A PRELIMINARY INVESTIGATION INTO THE NATURE OF THE GRADED PROPELLANT/INSULATION INTERFACE IN SOLID ROCKET MOTORS

Charles J. Noel

Associate Professor

Department of Textiles and Clothing

The Ohio State University

1787 Neil Avenue

Columbus, OH 43210-1295

Final Report for:

Summer Faculty Research Program

Phillips Laboratory

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D.C.

September, 1993

A PRELIMINARY INVESTIGATION INTO THE NATURE OF THE GRADED PROPELLANT/INSULATION INTERFACE IN SOLID ROCKET MOTORS

Charles J. Noel
Associate Professor
Department of Textiles and Clothing
The Ohio State University

Abstract

Conventional solid rocket motors contain composite ablative internal insulation to protect the case and payload from potential thermal damage during firing. An adhesive material is used to bond the insulation to the propellant; this so-called bond line is considered to be a mechanically weak area in the rocket motor, and numerous failures in the firing of solid rocket motors are attributed to problems with this bond line. A new generation of solid rocket motors has been proposed, in which the bond line is replaced by a graded interface between propellant and insulation. The graded interface is achieved through the use of chemically similar thermoplastic elastomers (TPEs) as the binders in both insulation and propellant. In this concept, the propellant and insulation are co-extruded, and the graded interface is obtained by shear or diffusional mixing of the molten materials as they emerge from the extruder side by side. An added advantage of this concept is that the use of TPE binders in both propellant and insulation eliminates the need for the additional curing steps necessary when conventional materials such as urethanes are used. In this study, the graded interface was simulated by shear mixing TPE based propellant with a hybrid insulation. Two different TPE binders were evaluated in the insulations which were composed of a 50/50 blend of binder with either of two silicon-containing preceramic polymers. The TPE propellant and the hybrid insulation materials were mixed at three different P/I levels, representing three points in the postulated composition continuum comprising the graded interface. The mixtures were characterized by measurement of physical properties (peak and break stress, peak and break strain, and Young's modulus), and thermal properties, using DSC, TGA, and TMA. Results obtained indicate the viability of the concept; properties tended to vary smoothly with composition suggesting a "seamless" interface between propellant and insulation could be achieved. Examination of microtomed surfaces of the P/I blends using reflected light photomicrography indicated that uniform mixes of propellant and insulation were obtained. With PEBAX-2533 as the TPE, blends with polysilastyrene were easier to process than were blends with polycarbosilane. With polycarbosilane as the preceramic, blends with Kraton D-1132 were easier to process than were blends with PEBAX-2533. The combination Kraton D-1132/polysilastyrene should be evaluated to complete the study.

A PRELIMINARY INVESTIGATION INTO THE NATURE OF THE GRADED PROPELLANT/INSULATION INTERFACE IN SOLID ROCKET MOTORS

Charles J. Noel

INTRODUCTION

Conventional solid rocket motors contain composite ablative internal insulation to protect the case from potential thermal damage during firing. This insulation consists of a variety of materials including heat resistant fibers such as asbestos, flame retardants, and carbonaceous materials such as coal, held in a crosslinked binder such as vulcanized rubber or polyurethane. Conventional solid rocket propellants generally contain fuel, oxidizer, and catalyst dispersed in a crosslinkable polymeric binder which often contains plasticizer to reduce viscosity. Both insulation and propellant need to be cured to achieve crosslinking; these steps in the process may require more than a week for completion, for large rocket motors. Since the propellant and insulation are chemically dissimilar, a layer of adhesive material is used to create a bond line between the two. Delamination at this bond line is reputed to be a major cause of solid rocket motor failure; structural integrity is lost if catastrophic unzipping of the bond line occurs under loads such as ignition pressurization and this in turn leads to flame growth along the wall, with case burn through being the ultimate result (1,2).

A new generation of solid rocket motors has been proposed in which the bond line is replaced by a graded interface between propellant and insulation (3). The graded interface is achieved through the use of chemically similar thermoplastic elastomers (TPEs) as the polymeric binders in both the propellant and the insulation. In this concept, the propellant and insulation are co-extruded, and the graded interface is obtained by shear or diffusional mixing of the molten materials as they emerge from the extruder side by side. An added advantage of this system is that the use of TPE binders in both propellant and insulation eliminates the need for the costly and time consuming curing steps needed for processing conventional propellants and insulations. Thermoplastic elastomers may be defined as materials which combine the processing characteristics of thermoplastics with the physical properties of vulcanized rubber (4). TPEs are comprised of block copolymers in a variety of forms; one of the blocks is a "soft" block with a glass transition temperature usually well below room temperature, and the other is a "hard" crystallizeable block with a relatively high melting point. Properties of the TPE can be engineered by varying both the nature, the arrangement, the molecular weights, and the relative amounts of the two blocks. Unlike vulcanized rubbers, TPEs are not crosslinked, but the aggregation of the hard blocks act

as crosslinks at use temperatures. These associations disappear at elevated temperatures, enabling the material to be processed.

Nahlovsky, Hargis, and Raterman (5) reviewed the work done in various government, industry, and university laboratories to develop solid rocket propellants based on TPE binders, both inert and energetic. Their work suggested that polyethylene block copolymers which contained polyethylene-like crystalline end blocks obtained by hydrogenation of low-vinyl polybutadiene and soft blocks obtained by hydrogenation of polyisoprene or high-vinyl polybutadiene. Palopoli *et al.* (6) report on the successful development of a solid rocket propellant using commercially available TPE binders. Propellants having up to 88% solids were obtained using Shell's Kraton D-1102 , a triblock of the SBS type, where the hard end blocks are polystyrene, and the soft middle block is polybutadiene. Over 250kg of a propellant designated TP-T-3006 was continuously processed on a twin-screw mixer/extruder in a demonstration of the feasibility of the concept. This propellant was selected as the TPE based propellant for use in this study.

Preceramic or hybrid polymers are a unique evolving class of materials which combine the favorable properties of both organic and inorganic substances (7). Upon pyrolysis, preceramic materials can be converted to a ceramic which is more resistant to to oxidation and mechanical erosion than conventional char from organic based ablative systems. Two commercially available silicon-containing preceramic polymers are polycarbosilane (PCS) and polysilastyrene (PSS). Polycarbosilane is formally poly(methylene methylsilane) but in fact is slightly crosslinked. Polysilastyrene is a linear 1/1 copolymer of dimethylsilane and methylphenylsilane, presumed to be more or less alternating. These two preceramics were both included in this study. Promising materials for future evaluation are hybrid polymers based on substituted polyhedral oligosilsesquioxanes (POSS) (7). These are currently unavailable in sufficient quantity for blending studies, but work is underway to scale up their production.

The purpose of this study was to characterize the graded interface which would be expected to form when molten TPE based propellants and insulations are co-extruded. The graded interface would be driven to form by either shear or diffusional mixing of the melts as they emerge side by side from an extruder or extruders. On one side of the graded interface one would find 100% insulation and on the other, 100% propellant. Properly formed, the graded interface would consist of a continuum of blend levels, ranging all the way from 99P/1I to 1P/99I. It was decided to "sample" this continuum by preparing three different P/I blends at 75/25, 50/50, and 25/75 by mixing presoftened propellant with presoftened insulation in a high-shear mixer. A Morton-Thiokol propellant, TP-T-3006 was chosen as the propellant. No attempt

was made to optimize an insulation formulation at this time; insulations evaluated were 50/50 blends of either of two preceramics (PCS or PSS) with either of two commercially available TPEs. PEBAX-2533 was selected because of its use in propellants. Kraton D-1102 was chosen because it is the TPE used in the selected propellant. It was intended to evaluate four insulations, combining each preceramic with each TPE; the Kraton/PSS combination could not be made because the additional PSS could not be obtained from the original supplier. By the time an alternate supplier was identified, it was too late to include this insulation in the study.

METHODOLOGY

Preparation of Insulation

The insulation materials used in this study were 50/50 blends of a TPE with a silicon-containing preceramic polymer. Two TPEs were used: PEBAX-2533 (Elf Atochem) is a block copolymer of nylon 12 (hard block) and poly(tetramethylene glycol), and Kraton D-1102 (Shell) is a SBS (styrene-butadiene-styrene) triblock polymer. Two silicon-containing preceramic polymers were also used: polycarbosilane (PCS) (Dow-Corning) and polysilastyrene (PSS) (Hüls). The materials were used as received from their suppliers. Blends were prepared by adding 50.0g TPE and 50.0g preceramic to 1000mL solvent (500mL dichloromethane and 500 mL tetrahydrofuran) in a 2.0L resin kettle. The mixture was allowed to stand for 60m; at this time the preceramic had dissolved and the TPE was either dissolved (Kraton) or swollen (PEBAX). The resulting mix was heated to the boil while being stirred with a Ross high-shear emulsifier, Mixing continued for 60m, then the solution was poured into a Teflon coated dish (11' x 9" x 5/8") and the solvent was allowed to evaporate. When the material appeared to be dry, it was treated in a vacuum oven at 80°C overnight to remove residual solvent. The resulting sheets were cut into pellets for blending with propellent. Three 50/50 blends of TPE with preceramic were prepared: PEBAX/PCS, PEBAX/PSS, and Kraton/PCS. The fourth blend, Kraton/PSS was not prepared because of difficulty in procuring additional PSS from Hüls; an alternate source was located and material has been ordered so that this blend will be prepared in the near future.

Propellant

The propellant TP-T-3006 was used as received from Morton-Thiokol. The propellant contains 88% solids and 12% TPE binder system. The reported composition of the propellant is (6):

Solids:	Aluminum	21.0%
	Ammonium Perchlorate	38.7%
	Sodium Nitrate	28.1%
	Ferric Oxide	0.2%

Binder:	Kraton D-1102	3.0%
	Shellflex 71 (Plasticizer)	7.5%
	Endex 160 (Filler)	1.5%

The propellant was cut into pellets to facilitate blending with insulation.

Preparation of Propellant/Insulation Blends

Blends of propellant and insulation were prepared using a quarter-pint Baker-Perkins mixer. Materials were preheated at 130°C for 30m prior to mixing. Blends of propellant and insulation were prepared in 120g batches at three different levels, 75/25, 50/50, and 25/75, corresponding to three points on the hypothesized continuum making up the graded P/I interface. For 75/25 P/I blends, 90.0g preheated propellent was placed in the bowl of the mixer and mixed alone for 30m at 130-135°C. The bowl was then lowered, the blades scraped down, and 30.0g preheated insulation was added. This was then mixed for an additional 30m at 130-135°C. At this time, the material in the bowl appeared to be homogeneous, so no additional mixing was done. The blended material was cast into lab-scale slabs (~4" x 3" x 0.18") using preheated Teflon molds. The 50/50 and 25/75 P/I blends were prepared in the same manner, except that in the case of the 25/75 blends, the insulation was added in three 30.0g increments, with 15m mixing for each addition of insulation.

Characterization of Materials

Mechanical properties of the materials were measured using an Instron Series 1123 Tester equipped with Series IX Automated Materials Testing System 1.15. Testing was carried out at ambient conditions (~23°C, 50%RH) using the JANNAF Standard Mini-Dog Bone Test. The 1.0" minibones were elongated at the rate of 1.0"/m. Data obtained included peak and break load, peak and break stress, peak and break strain, and Young's modulus. At least eight minibones were broken for each P/I blend.

Thermal properties were measured using a DuPont Thermal Analysis System, which included a DSC 912 module, a TGA 951 module, a TMA 2940 module, and a Model 2000 Thermal Analyst controller. Heating rates were 10°C /m for the DSC and TGA, and 3°C/m for the TMA. The TMA was operated with a 10g weight in the penetration probe mode, and a 2g weight in the expansion probe mode.

Scanning electron micrographs of materials were taken using an International Scientific Instruments Model CL6 equipped with a LaB_6 filament. Photomicrographs were taken using a Jennalumar research microscope equipped with the Datex-Man Automatic-2 photographic system, using both ASA 64 color film for slides and ASA 100 color print film.

RESULTS AND DISCUSSION

In this study, the effort was made to characterize the nature of the graded interface between propellant and insulation in solid rocket motors. The graded interface is postulated to form when thermoplastic propellant and insulation are extruded side-by-side and the two melts are driven to mix by shear forces. The graded interface consists of a continuum of compositions ranging from 100% propellant on one side to 100% insulation on the other. To characterize this interface, mixes of propellant and insulation were made at the 75/25, 50/50, and 25/75 levels, representing three points within the composition continuum. The experiment was designed to evaluate four different insulations, made by blending either of two TPEs with either of two preceramics at the 50/50 level. The two TPEs were PEBAX-2533, a block copolymer of poly(dodecanoic lactam) (nylon 12) and poly(tetrametlylene glycol), and Kraton D-1102, an SBS (styrene-butadiene-styrene) triblock copolymer. The two preceramics were polycarbosilane (PCS) and polysilastyrene (PSS), which is a 50/50 copolymer of dimethyl silane and methyl phenyl silane. Three of the four insulations were evaluated; the fourth, Kraton/PSS, will be evaluated in the near future to complete the study.

Before reporting the results of physical testing and thermal analysis, some comments on the processability of the mixes are in order. All mixes were made at 130-135°C using a quarter-pint Baker-Perkins high-shear mixer operating at 50rpm, with the exception of the 25/75 blend with PEBAX/PCS as the insulation. This material could only be mixed at 30rpm because its viscosity was apparently too high for the mixer. (One fuse was blown attempting to operate at 50rpm.) After mixing, all blended materials were hand cast into preheated Teflon molds by scooping a spatula full of material out of the bowl of the mixer and spreading the material onto the mold. P/I blends at the 75/25 level were invariably easier to spread than either 50/50 or 25/75 blends.; 50/50 blends were generally easier than 25/75 blends. The Kraton/PCS blends all spread easier than the corresponding PEBAX/PCS blends. However, all the PEBAX/PSS blends spread easier than any of the other blends. Preheated metal plates were then placed on top of the material in the mold to assist the material in filling the mold. Standard JANAFF mini-dog bones were cut from these slabs for physical testing. Since the minibones are cut with a standard die, the length and width of all the minibones are close to being the same. Cross-sectional area of the minibones is determined for the calculation of stress and modulus properties; most of the variation in the values of these areas are due to differences in the thicknesses of the individual bones. Increased thickness and increased variability in thickness resulted when material was more difficult to spread. Table 1 shows the area measurements for the P/I blends as a function of blend level and composition. The data shown

TABLE 1

CROSS-SECTIONAL AREA OF JANNAF MINI-DOG BONES
FOR PHYSICAL TESTING

Composition	P/I	Area (sq.in.)	Std. Dev. (sq.in.)	Coeff. of Var. (%)
	75/25	0.0354	0.0029	8.21
PEBAX/PCS	50/50	0.0302	0.0066	21.90
	25/75	0.0420	0.0078	18.59
	75/25	0.0349	0.0029	8.38
Kraton/PCS	50/50	0.0364	0.0010	2.78
	25/75	0.0355	0.0037	10.43
	75/25	0.0122	0.0008	6.38
PEBAX/PSS	50/50	0.0124	0.0036	29.25 (bimodal)
	25/75	0.0145	0.0016	10.68

support the qualitative observations on processability; the thinnest minibones were obtained with the PEBAX/PSS blends, the thickest with the 25/75 PEBAX/PCS , and the thickness tended to increase with insulation content. The high coefficient of variation for the 50/50 P/I blend with PEBAX/PSS is due to a bimodal distribution; five of the tested minibones had an average area of 0.0094 square inches (0.35% C.V.), while the other four tested had an area of 0.0162 square inches (1.84% C.V.).

The results obtained from the physical testing of materials are shown in Table 2. In this table are shown the results obtained for the P/I blends as well as the results obtained for the propellant and the insulations alone. The physical properties of solid rocket propellents most often reported in the literature are the Peak Stress, Break Strain, and Young's Modulus. These are given in Table 2 along with values for Break Stress, Peak Strain, Peak and Break Load, and the ratios of Peak to Break Stress (PS/BS) and Peak to Break Strain (PE/BE). Examination of the data for the P/I blends shows that Peak and Break Load and Peak and Break Stress increase with increasing insulation content in every case, and that the values of Peak and Break Stress for the P/I blends fall between the values for the propellent and the pertinent insulation. Plots of Peak Stress vs. Insulation Content show a smooth increase from 0 to 100% Insulation in all three cases, with the measured values being below values calculated by an additive law of mixing for the two compositions containing PCS, and above those calculated values for the PEBAX/PSS composition. The PEBAX/PSS insulation has the lowest Peak Stress of the three insulations studied. Peak Strain also increases with increasing insulation content in every case, and so does Break Strain with the single exception of the 50/50 blend with PEBAX/PCS insulation. This composition reached its Peak

TABLE 2

PHYSICAL PROPERTIES OF PROPELLANT/INSULATION BLENDS

Propellant : TP-T-3006

Insulation: 50/50 TPE/Preceramic

Sample	TPE/Preceram	Blend	Pk. Load (lb.)	Pk. Strs. (psi)	Pk. Strn. (in./in.)	Br. Load (lb.)	Br. Strs. (psi)	Br. Strn. (in./in.)	Yng. Mod (psi)	PS/BS	PE/BE
CJN001	PEBAX/PCS	75/25	3.85	109.6	0.039	2.075	58.8	0.272	7622	1.896	0.182
CJN002	PEBAX/PCS	50/50	7.11	238.5	0.091	5.162	173.1	1.587	7546	1.408	0.078
CJN003	PEBAX/PCS	25/75	12.46	298.1	0.307	10.558	251.5	0.369	5069	1.193	0.857
CJN004	Kraton/PCS	75/25	3.78	108.6	0.078	1.511	43.6	0.389	4043	2.778	0.211
CJN005	Kraton/PCS	50/50	7.21	197.9	1.498	6.239	170.8	1.716	3308	1.362	0.867
CJN006	Kraton/PCS	25/75	13.03	369.1	3.557	12.668	356.5	3.705	1786	1.036	0.961
CJN007	PEBAX/PSS	75/25	3.07	247.8	0.136	1.988	160.9	0.196	8686	1.531	0.708
CJN008	PEBAX/PSS	50/50	4.25	341.5	0.334	3.103	247.6	0.443	3427	1.382	0.764
CJN009	PEBAX/PSS	25/75	6.66	458.8	0.627	5.089	351.8	0.772	5912	1.325	0.814
CJN011	Kraton/PSS	75/25									
CJN012	Kraton/PSS	50/50									
CJN013	Kraton/PSS	25/75									

PHYSICAL PROPERTIES OF PROPELLANT, TPEs, AND TPE/PRECERAMIC BLENDS

TP-T-3006	(From Thiokol Report)			100				0.250	750	
TP-T-3006	(From material as received)		2.48	93.1	0.144	2.16	81.2	0.162	1046	1.152	0.893
TP-T-3006	(From remixed material)		0.639	24.9	0.028	0.488	19.04	0.036	1325	1.332	0.856
PEBAX-2533			49.39	3057	8.422	48.37	2996	8.539	581	1.022	0.986
PEBAX/PCS		50/50	13.22	816	0.558	11.87	733	0.674	8333	1.116	0.827
PEBAX/PSS		50/50	9.45	475	3.728	8.96	448	3.801	2253	1.053	0.962
Kraton D-1102			24.41	1199	6.987	23.76	1167	6.984	2132	1.028	0.986
Kraton/PCS		50/50	9.61	619	5.883	9.37	595	5.962	2175	1.025	0.986
Kraton/PSS		50/50									

Stress at a strain of 9.1% (0.091 in./in.) but did not break until a strain of 158.7% (1.587 in./in.) was reached. Plots of Break Strain vs. Insulation Content show a smooth increase from 0 to 100% Insulation in all three cases, with the single exception noted above; the measured values are slightly below values calculated by a simple additive law of mixing for the compositions containing PCS (with the one glaring exception), but were far below calculated values for the PEBAX/PSS insulation. The Young's Modulus data show no clear trend with increasing insulation content. The propellant itself, having 88% solids and only 12% binder (and only 4.5% of that 12% is polymeric) shows low Young's Modulus. Only the PEBAX shows a lower modulus; when PCS is blended with PEBAX to form insulation, the modulus increases tremendously (from 581 to 8333psi) but when PSS is blended with PEBAX to form insulation, the modulus only increases from 581 to 2253psi. Kraton has a higher modulus than PEBAX (2132 vs. 581psi); when PCS is blended with Kraton to form insulation, the modulus of the blend is the same as the modulus of the Kraton (2175 vs. 2132psi). Whatever the modulus of the insulation, all the 75/25 P/I blends have much higher modulus than does the propellant alone. There is no further identifiable trend with increasing insulation content. For PEBAX/PCS insulation, modulus may remain fairly constant as insulation increases from 25 to 100%; for Kraton/PCS insulation, the modulus may decrease over the same range; for PEBAX/PSS insulation, a downward trend over the range may be as good as any other. The modulus data show a fair amount of scatter within a sample of eight to ten minibones. Coefficients of Variation for modulus range from 10.33% (50/50 P/I blend with Kraton/PCS) to 39.11% (75/25 P/I blend with PEBAX/PCS), and average 19.73% for all P/I blends. Svob (8) reports that it is common for the C.V. to approach 15% for Young's Modulus; these data seem to be largely outside this area.

The ratios of peak stress to break stress (PS/BS) and peak strain to break strain (PE/BE) provide some information regarding the shape of the stress-strain curve. The PEBAX/PSS and Kraton/PCS insulations both tend to reach maximum stress very close to their breaking points; this peak stress is only 2-5% higher than the break stress and is reached at an extension within 4% of the extension at break. This behavior is very similar to that of the two TPEs, and these two insulations have high break strains as do the two TPEs. By contrast, the PEBAX/PCS insulation has a peak stress nearly 12% higher than its stress at break, and reaches this at only 83% of the break extension, which itself is relatively low (67%). The propellant itself is closer in behavior to the PEBAX/PCS insulation; peak stress is 15% higher than stress at break, this is reached at 89% of break extension, and the break extension is low (16%). The actual values for the peak stresses for all the insulations are much higher than the peak stress of the propellant, since the load-bearing elements in the propellant comprises only 4.5% of its mass. With a ratio of 75/25, and both PEBAX/ PCS and Kraton/PCS as insulation, the P/I blends reached peak stresses 90 to 180% above their stresses at break, and achieved these at 18-21% of break extension. The most unusual stress

strain curve was obtained for the 50/50 P/I blend with PEBAX/PCS insulation; the peak stress was 40% higher than the stress at break and was reached at an extension which, on average, was only 7.8% of the extension at break. By contrast, the 50/50 P/I blend with Kraton/PCS insulation had a peak stress 36% higher than the stress at break, but this was reached at 87% of break extension. At the 25/75 P/I blend level, the PS/BS and PE/BE ratios were similar to propellant for the PEBAX/PCS insulation, and similar to insulation for the Kraton/PCS insulation. For the PEBAX/PSS insulation the PS/BS ratio increased from 0 to 25% insulation, then decreased steadily from 25 to 100% insulation, while the PE/BE ratio showed the opposite behavior. At the 75% insulation levels, the ratios were more like those for propellant than those for insulation.

The data in Table 2 point out that the propellant is not particularly heat stable. Data are given for peak stress, break strain and Young's modulus as reported in reference 6. Minibones were cut from slabs machined from bricks of propellant which have been stored at Edwards AFB for over a year. Minibones were also cut from slabs which were cast from propellant given the same treatment as the P/I blends: preheated at 130°C for 30m, mixed alone at 130°C for 30m, then cast into preheated Teflon molds. Test results from these samples are also given in Table 2. The results obtained from the stored material suggest a slight deterioration in properties; the peak stress of 93psi for the stored material is close to the 100psi reported, but the break strain of 16% is below the 25% reported, and the modulus at 1046psi is somewhat higher than the 750psi reported. Higher modulus is consistent with lower elongation. In comparison to the stored material, the processed (remixed) material is badly degraded. The peak stress is down from 93psi to 25psi, break strain is down from 16% to only 3.6%, and modulus is up from 1046psi to 1325 psi. The polymeric TPE binder may be undergoing degradation through contact with oxidizers on storage; this degradation is accelerated by processing the propellant at elevated temperatures. It is recommended that a property profile be maintained over time for stored propellant to see how serious a problem this might be. Since the graded interface would be formed at the time the propellant is first extruded, and would not rely upon reprocessing previously formed propellant as was done in this study, the propellant degradation upon heating would not be a problem. In property comparisons discussed above, the values for stored propellant were used. Values obtained for all P/I blends may be lower than those which would actually be obtained in a graded interface, since the propellent component of the blends has undergone degradation. The TPEs in the insulation could possibly have undergone some degradation in contact with the propellant's oxidizers as well.

To evaluate the uniformity of the blends obtained, scanning electron microscopy and reflected light photomicrography of microtomed surfaces were both used. The scanning electron micrographs taken at

magnifications of 500-1250X, were difficult to interpret. Uniformity appeared good but the field of view was so small that this was hard to judge. Particulate matter was visible in both surface and cross-sectioned specimens, the amount of particulates decreased with increasing insulation level in the blends. No evidence for or against phase separation could be inferred. Photomicrographs at about 75X magnification taken by light reflected from microtomed surfaces showed the propellant and insulation to be well mixed at least at the macro level. The aluminum particles in the propellant were clearly visible, as were large and small clear crystals of ammonium perchlorate and sodium nitrate oxidizers. Bright orange crystals, presumably of ferric oxide, could be seen in some fields of view. The aluminum particles and oxidizer crystals appeared to be uniformly distributed over each specimen viewed, and were definitely in proportion to the level of propellant in the blends. Based upon these qualitative observations, the conclusion was reached that the mixing procedure used did produce fairly uniform blending of propellant and insulation.

The results obtained by thermal analysis are summarized in Table 3. This table includes results obtained by Thermogravimetric Analysis (TGA) and by Differential Scanning Calorimetry (DSC). The onset temperature determined by TGA represents the first indication of weight loss by the specimen, while the onset temperature determined by DSC represents the first clear departure from the baseline whether in an endothermic or exothermic direction. The decomposition temperature determined by TGA was obtained by constructing tangents to the initial and steepest portions of the weight-temperature curve, and determining their point of intersection; this roughly corresponded to the temperature at which specimen weight loss was 10%. The decomposition temperature determined by DSC was taken as the peak temperature of the first significant exotherm; usually there were additional exotherms at higher temperatures but these were generally smaller than the first. The glass transition temperature determined was taken as a shift in the baseline, representing a change in the heat capacity if the specimen. Glass transition temperatures by Thermomechanical Analysis (TMA) and by Dynamic Mechanical Analysis (DMA) will be measured, as will softening temperatures by TMA. Percent Char by TGA as reported in this table was the weight of the residue after heating the specimen from room temperature to 1000°C at the rate of 10°C/m, expressed as a percentage of the original weight.

The propellant TP-T-3006 showed first evidence of weight loss at 195°C, with rapid weight loss commencing at 242°C. By DSC, all samples containing any propellant showed a small but characteristic endothermic peak at 243°C, followed immediately by a significant exotherm apparantly associated with decomposition. The insulations alone all exhibit higher temperatures for initial weight loss than propellant, and much higher decomposition temperatures by TGA. The two insulations made with

TABLE 3

THERMAL ANALYSIS OF PROPELLANT, INSULATION, AND P/I BLENDS

TP-T-3006//PEBAX-2533/PCS

%Ins	T-onset (TGA)	T-decom (TGA)	%Char TGA)	Tg (DSC)	T-onset (DSC)	T-decom (DSC)
0	195	242	34.3	-92	243	300
25	171	246	43.1	-83	243	287
50	187	250	49.9	-69	243	284
75	168	238	46.9	-71	243	419
100	225	381	43.5	-67	327	425

TP-T-3006//Kraton D-1102/PCS

%Ins	T-onset (TGA)	T-decom (TGA)	%Char (TGA)	Tg (DSC)	T-onset (DSC)	T-decom (DSC)
0	195	242	34.3	-92	243	300
25	191	333	43.1	-104	243	315
50	180	398	42.8	-79	244	345
75	245	440	45.1	-78	245	354
100	238	443	38.5	-79	270	379

TP-T-3006//PEBAX-2533/PSS

%Ins	T-onset (TGA)	T-decom (TGA)	%Char (TGA)	Tg (DSC)	T-onset (DSC)	T-decom (DSC)
0	195	242	34.3	-92	243	300
25	197	266	37.1	-71	243	271
50	183	264	47.3	-91	243	270
75	226	356	42.1	-87	243	264
100	285	375	42.2	-66	235	409

TP-T-3006//Kraton D-1102/PSS

%Ins	T-onset (TGA)	T-decom (TGA)	%Char (TGA)	Tg (DSC)	T-onset (DSC)	T-decom (DSC)
0	195	242	34.3			
25						
50						
75						
100						

All temperatures are in degrees Celsius.

PEBAX had decomposition temperatures of 375-381°C while the Kraton/PCS insulation had a decomposition temperature of 443°C. TGA decomposition temperatures for the P/I blends with PEBAX/PCS insulation did not increase with increasing insulation content as expected, but TGA decomposition temperatures for P/I blends with both Kraton/PCS and PEBAX/PSS insulations did increase with increasing insulation content. Residues from P/I blends with all three insulations, reported as %Char by TGA, tended to approach or exceed the values obtained for the insulations alone, and were always higher than the residue from the propellant alone. The residue from the propellant alone was a silvery gray powder; the residues from the P/I blends and from the insulations alone were hard and sometimes brittle black ceramic-like material. Qualitative elemental analysis of these residues by EDAX indicated that the residue from the propellant alone contained aluminum and oxygen but no sodium or chlorine; residues from P/I blends contained aluminum, sodium, silicon, and oxygen, but no chlorine,; and residue from insulation alone contained only silicon and oxygen. Since the propellant contains aluminum as fuel and both ammonium perchlorate and sodium nitrate as oxidizers, it was anticipated that its residue would consist of aluminum oxide and sodium chloride; the alumina was detected but the salt was not. The preceramic polymers should convert to ceramic material by pyrolysis and oxidation, yielding silicon oxide, silicon carbide, or silicon oxycarbide. The silica was detected, but the carbon derivatives were not.

Glass transition temperatures as determined by DSC were well below room temperature, as would be expected for elastomeric materials. The values were determined by analysis of the DSC baseline in the tempperature range from -150°C to +100°C. A shift in the baseline is indicative of a change in heat capacity of the material being tested; this shift was very small in all cases where it was detected, and the shift was not detected every time (about one out of every six traces did not show a measurable baseline shift). The DSC glass transition temperatures reported are the average values of at least two detectable shifts. The values for the insulations were all lower than that for the propellant; values for the P/I blends showed no readily apparent trends with composition. Additional thermal characterization of these materials is being carried out, using both DMA and TMA equipped with an expansion probe to determine glass transition temperatures, and TMA equipped with a penetration probe to determine softening temperatures. Both TMA and DMA make use of much larger test specimens than are used in DSC (4-6mg), and may provide better measurements of glass transition temperatures than were obtained by DSC.

SUMMARY AND CONCLUSIONS

This study was undertaken to characterize the graded interface between propellant and insulation which is postulated to form when TPE-based propellant and insulation are co-extruded side by side in the making of the rocket motor of the future. The graded interface is driven to form by the shear forces present as the melts emerge from their extruder(s), and could be augmented by diffusional mixing of the melts. The graded interface is the material between 100% insulation on one side and 100% propellant on the other, having a continuously graded composition throughout. To characterize this graded interface, propellant and insulation were mixed at three different levels, 75P/25I, 50P/50I, and 25P/75I, representing a sampling of the interface's postulated composition continuum. Mixes were accomplished using a high shear mixer at elevated temperatures (130-135°C) where the TPE binders in both propellant and insulation exhibit thermoplastic behavior. The resulting P/I blends were evaluated using standard physical and thermal analysis tests. Four insulations, each a 50/50 blend of one of two TPEs with one of two silicon-containing preceramic polymers, were to be evaluated to demonstrate the effect of materials on the properties of the interface; because of the unavailability of one of the preceramics, only three were actually done. The fourth will be evaluated in the near future when additional preceramic is procured.

Photomicrographs taken using light reflected from microtomed surfaces indicate that the propellant and insulations were well mixed, at least at the macro level. Scanning electron micrographs showed no clear evidence of phase separation, but these were difficult to interpret. During mixing and casting of the mixed materials, it was noted that ease of processing of the blends followed the order PEBAX/PSS >> Kraton/PCS > PEBAX/PCS, and that within a given composition processability decreased with increasing insulation content. Support for this qualitative observation could be inferred by examination of the minibones cut for physical testing; cross-sectional area differences in the minibones are largely due to thickness differences, and thinner minibones are made from materials which are easier to process. It is recommended that measurements of viscosity and the viscosity-temperature relationship be made for these blends to quantify differences in processability. The striking improvement in processability shown by the P/I blends made with the PEBAX/PSS over those made with the insulations containing PCS may be due to the fact that the PSS is a linear polymer while the PCS, though formally linear, is actually crosslinked to some extent. The presence of a network polymer in the melt would significantly increase viscosity and make processing more difficult.

Peak and break stress, peak and break strain, and Young's modulus were determined for the materials using standard JANNAF procedures. Peak and break stress and peak strain varied smoothly through the

interface, as did break strain with one exception: the 50/50 P/I blend with PEBAX/PCS insulation had an unusually high break strain. Young's modulus showed no clear trend through the interface for any of the insulations, except that the moduli of the 75P/25I blends were all much higher than the modulus of the propellant itself; the moduli of the blends containing higher levels of insulation tended to move toward the moduli of the pertinent insulations. One point to keep in mind while examining the stress-strain data is that the TPE is the load-bearing material in the system. For the 100% propellant, there is only 3% TPE, this increases to 14.8% in the 75P/25I blends, to 26.5% in the 50P/50I blends, to 38.3% in the 25P/75I blends, and to 50% in the 100% insulations.

The ratios of peak to break stress (PS/BS) and peak strain to break strain (PE/BE) were used as indicators of the shape of the stress-strain curves of the materials. For all the materials tested, the stress reached a maximum value at a particular value of strain (extension) and then decreased with further strain until the test specimen ruptured. The TPEs used both achieved maximum stress at an extension very near the break strain, and this peak stress was only slightly higher than the break stress, i.e., PS/BS was slightly above 1.0 and PE/BE was slightly below 1.0. Two of the insulations, PEBAX/PSS and Kraton/PCS, behaved in a similar manner while the third, PEBAX/PCS, had a higher PS/BS reached at a lower PE/BE ratio. The TP-T-3006 propellant alone had ratios similar to the PEBAX/PCS, although the absolute values of the sresses and strains naturally were mich lower for the propellant. The 75/25 P/I blends all showed much higher values of PS/BS with peak stress reached at much lower values of PE/BE, compared to either propellant or insulation alone. This corresponds to the large increase in Young's modulus observed for these 75P/25I blends compared to the propellant alone. At higher levels of insulation in the P/I blends, the PS/BS ratios tend to decrease, and the PE/BE ratios tend to increase toward the values of the insulations, except for the 50P/50I blend with PEBAX/PCS insulation, where the peak stress was reached at a very low extension, as noted above. These data indicate that the insulations tend to behave like the TPEs used to make them, and that, for the most part, the behavior of the propellant/insulation blends becomes more rubber-like as the insulation content increases.

Decomposition temperatures measured by TGA tended to increase with increasing insulation content through the graded interface. Decomposition temperatures measured by DSC were lower for P/I blends than those for either propellant or insulation when the blends contained insulations made with PEBAX, but increased with insulation content for blends made with the Kraton/PCS insulation. All of the blends exhibited the endotherm characteristic of the TP-T-3006 propellant at 243°C, by DSC. The presence of insulation in the blends increased the amount of residue or char remaining at 1000°C by TGA, compared to propellant alone, and changed the nature of the residue from powder to ceramic.

One important observation made in this study is that the propellant itself is degraded by continued exposure to elevated temperatures (mixing at 130°C) and possibly by long term storage under ambient conditions. Properties measured on stored propellant showed slightly lower peak stress, lower break elongation, and higher Young's modulus than the values reported by the manufacturer. Propellant which was preheated to soften and then "mixed" alone showed much lower values for peak stress and break strain and higher values of modulus than the stored propellant. Slow degradation on storage is potentially a more serious problem than the more extensive degradation observed on heating, since it is expected that, in making solid rocket motors with graded interfaces, the propellant and insulation will be co-extruded as the propellant is formulated, and the motor will not be made by re-extruding softened previously formed propellant with molten insulation as was done in this study. It is recommended that a long term property-time profile be established for this propellant.

The work reported here shows that fairly smooth transitions in most properties are observed in blends representing a sampling of the composition continuum postulated to occur in the formation of a graded interface between propellant and insulation, i.e., that the concept of a graded P/I interface has validity. No information was developed concerning the actual dimension (thickness) of the graded interface; it is presumed that this dimension could be engineered and controlled by proper selection of materials and processing conditions. Further, no attempt was made in this study to develop an optimized formulation for the insulation in order to achieve the desired graded interface. In conventional solid rocket motors (having propellant based on thermoset binders rather than TPE binders), plasticizers are incorporated into the propellant formulations. On storage these plasticizers migrate or diffuse from the propellant through the bond line into the insulation, changing the physical properties of both. This is reputed to be a major cause of failures in the firing of solid rocket motors(1,2). Since the TPE-based propellant systems also contain plasticizer to provide proper viscosity for processing (6), it is proposed that an optimized formulation for a TPE-based insulation would also incorporate plasticizer, for two important reasons. First, the driving force for the diffusion of plasticizer in solid rocket motors is the concentration gradient which exists between the plasticizer-rich propellant and the plasticizer-poor insulation. Incorporation of plasticizer in the insulation would greatly reduce or even eliminate this gradient, and thereby minimize the extent of diffusion and the resultant property changes. Second, the presence of plasticizer in the insulation would reduce its viscosity and improve its processability, and would probably promote the shear mixing of propellant and insulation aiding in the formation of the graded interface. It should also be possible to incorporate a higher loading of preceramic polymer in the insulation, giving better char yields and insulation performance. Since plasticizer is present at the 7.5% level in the TP-T-3006 propellant, if the insulation is formulated with 7.5% plasticizer the concentration gradient would be zero,

and there would be essentially no plasticizer diffusion through the graded interface. Prerceramic and TPE could be optimized to a total of 92.5%, providing the best compromise between processability and insulation performance, using standard statistical optimization procedures.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. Sandra Slivinski for her sponsoring this project, and for providing insights into the nature of the propellant/insulation bond line. The work was carried out in close cooperation with Dr. Joseph D. Lichtenhan who, along with Drs. Tommy W. Hawkins and Kevin P. Chaffee, developed the graded interface concept, and who, with Alan G. Bolf, provided the TPE/preceramic hybrid insulations. Dr. John J. Rusek provided the TP-T-3006 propellant. Preparation of propellant/insulation blends was greatly facilitated by the work of TSgt. David E. Foxx, who translated mixing instructions into actual mixes. Physical testing was carried out by Ms. Jeri VanDyke. Mr Paul F. Jones provided instruction and assistance in the operation of thermal analysis equipment, and carried out TMA and DMA measurements. Finally the author is grateful to the researchers in the RK/FE group for many stimulating discussions and helpful insights.

REFERENCES

1. Oberth, A. E. *Principles of Solid Propellant Development*, CPIA Publication 469, September 1987.
2. Schreuder-Stacer, H. L. and Stacer, R.G. *Bonded Interface Technology Development for Solid Rocket Motors*, Report AL-TR-89-082, September,1991.
3. Lichtenhan, J. D., Hawkins, T.W., and Chaffee, K. P. *Preliminary Investigation of a Graded Interface Between Solid Rocket Motor Propellants and Insulations*, Phillips Laboratory Internal Report, July, 1993.
4. Holden, G. *Application of Thermoplastic Elastomers*, in Legge, N. R., Holden, G. and Schroeder, H. E., (Eds.) *Thermoplastic Elastomers*, Hanser Publishers, New York, 1987.
5. Nahlovsky, B. D., Hargis, I. G., and Ratermann, A. L., *Thermoplastic Elastomer Development*, Report AL-TR-89-010, August, 1989.
6. Palopoli, S. F., Hartwell, J. A., Willer, R. L., and Polleck, R. R., *TPE Propellant Development*, Report PL-TR-91-3096, June, 1992.
7. Lichtenhan, J. D., Vu, N. Q., Carter, J. A., Gilman, J. W., and Feher, F. J. *Macromolecules* 1993, 26, 2141-2142.
8. Svob, G. J. *Propellant Mechanical Properties Evaluation*, in Oberth, A. E., *op. cit.*

CONCEPTUAL DESIGN STUDY OF A SOLAR CONCENTRATOR/ SUPPORT STRUCTURE: A THREE DIMENSIONAL FINITE ELEMENT MODEL.

Rupa Purasinghe
Professor
Department of Civil Engineering

California State University at Los Angeles
5151 State University Drive
Los Angeles, California 90032 -8151

Final Report for
Summer Research Program
Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Edwards Air Force Base, Edwards, California

Focal Point
Kristi Laug

September 1993

CONCEPTUAL DESIGN STUDY OF A SOLAR CONCENTRATOR/ SUPPORT STRUCTURE: A THREE DIMENSIONAL FINITE ELEMENT MODEL

Rupa Purasinghe
Professor
Department of Civil Engineering
California State University at Los Angeles

Abstract

Under the space environment the paraboloid solar concentrators and support structures can deform and hence the focal point of the concentrators can diffuse. If this diffusion is large, energy will not concentrate on the thruster as desired. This paper addresses this aspect of pointing and accuracy analysis of solar concentrators, due to equivalent thrust loads.

The previous studies were limited to the concentrator system being modeled with a simplified finite element model that includes only the support struts and torus. The torus model was made up of several equal length beams. The simple model did not contain the paraboloid reflector, and assumes the reflector does not affect the deformation of the torus. In the present study the inflated parabolic reflector is included in the model. The results demonstrate the non uniform displacements on the reflector that confirms the reflector's potato chipping effect.

CONCEPTUAL DESIGN STUDY OF A SOLAR CONCENTRATOR/ SUPPORT STRUCTURE: A THREE DIMENSIONAL FINITE ELEMENT MODEL

INTRODUCTION:

The Solar Propulsion Concept (SPC), consists of concentrators, solar energy absorbing thruster, and a single fuel tank. The SPC has the advantage of doubling specific impulse and hence doubling payload. This system can make Low Earth Orbit (LEO) to Geosynchronous Earth Orbit (GEO) missions less expensive. A light weight paraboloid concentrator and a support structure are needed for the Solar Propulsion Concept [1]. Studies have shown that inflatable paraboloids offer great savings in payload and packaged volume compared to paraboloids mechanically erected in space [2,3]. The reflectors and support structure should be packageable within the launch vehicle and must be deployable once Low Earth Orbit is achieved. Leakage through holes caused by meteoroids is easily compensated by make up gas, due to a very low pressure requirement. Other advantages of the inflatable models are improved dynamic performance through rapid damping, non linear restoring forces, and reduction of thermal gradients [4].

Work on inflatable reflectors has been done for over 25 years. Paraboloids have been built and tested for surface accuracy and microwave performance [4]. For solar concentrators, slope error is the key accuracy parameter for measuring gross surface distortion in construction. This has been addressed by M. Thomas [5].

Another aspect of accuracy of solar concentrators under the space environment is the concentrator deflection and rotation due to thrust and temperature. The reflector deflection and rotation tends to make the focal point too diffuse. If the diffusion of the focal point is large, energy will not concentrate on the thruster as desired. This paper addresses this aspect of pointing and accuracy of focal point diffusion in concentrators.

Purasinghe [6] and Ritchie [7] used two off axis paraboloid reflectors supported on a rigidized elliptical torus to study this problem. Each torus was supported on three struts which are fixed to a turntable [Figure 1]. It is assumed that the paraboloid reflectors were attached to the torus in such a way that reflectors do not affect the torus. This system was analyzed for equivalent thrust loads in space. The deflection results from this analysis at the focal point shows diffusion at the focal point. In the current study paraboloid mirror is included to study the effect of torus/mirror interaction.

METHODOLOGY:

The finite element method was used with MSC NASTRAN [8] Program. Two views of the finite element model of the support structure is shown in Figures 2 and 3. It consists of beam elements (CBAR) for both struts and torus. The torus was modelled with straight beam elements. The struts are fixed to the turntable. The reflector component consists of a reflective membrane and a geometrically identical transparent canopy. The two form together an inflatable lens like structure which under inflation, assumes a paraboloid shape. These were modelled with CQUAD4 shell elements of NASTRAN. The inflatable structure is supported along its rim by a bending resistant torus which was described earlier. The geometry and the material properties of the system is documented in Table 1.

The focal point displacement with reference to displacement and rotation of the torus was calculated with separate three bars with displacements and twice displacements enforced at one end. The displacements at other end are the displacements of the focal point with reference to these points on the torus. A displacement analysis was done using NASTRAN Program. The model was excited with four types of loads. They are

1) Internal pressure of 10^{-5} psi of reflector/canopy chamber.

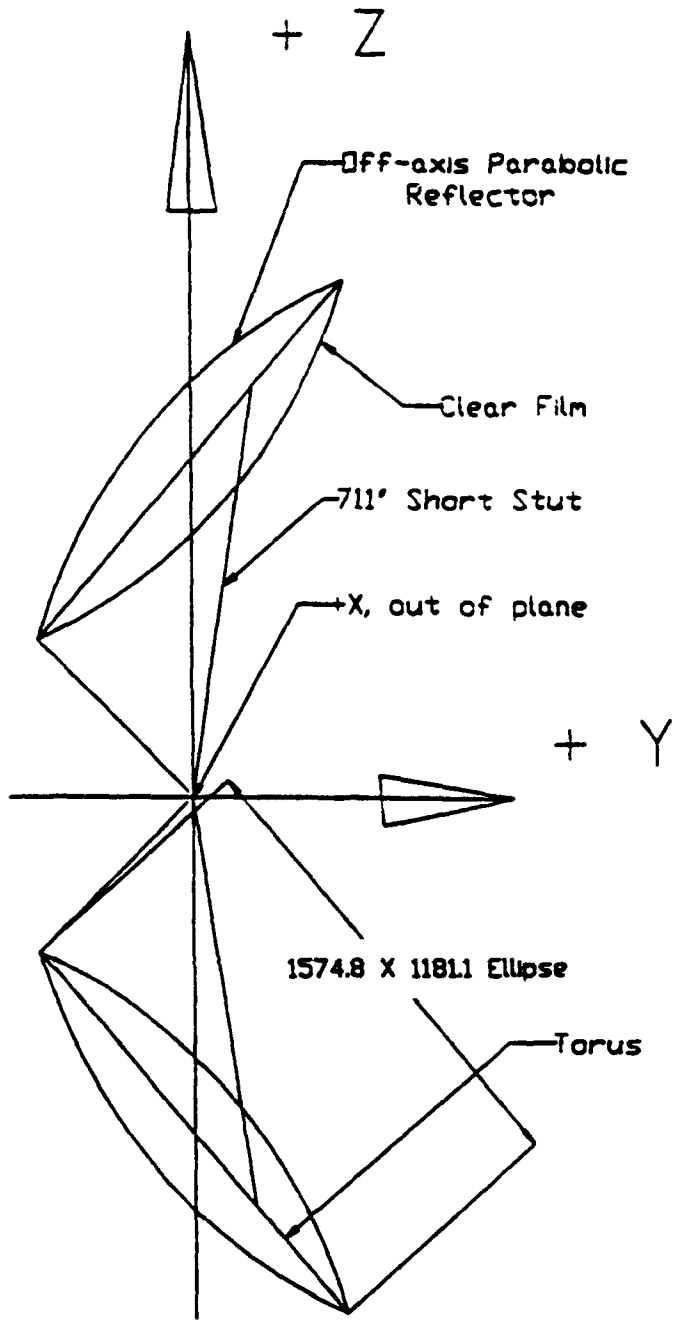


Figure 1: Solar Concentrator System

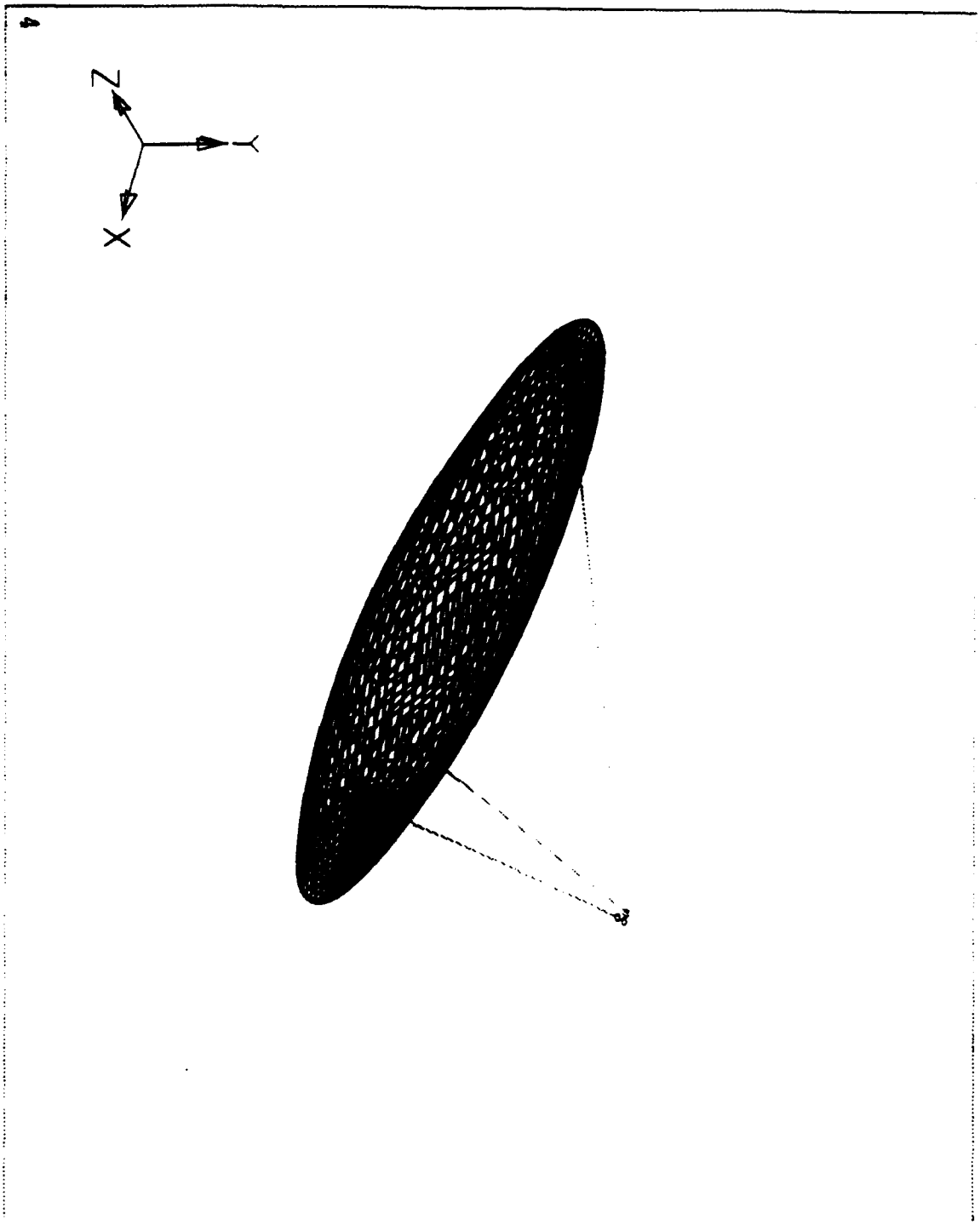


Figure 2: Nastran Finite Element Model - View 1

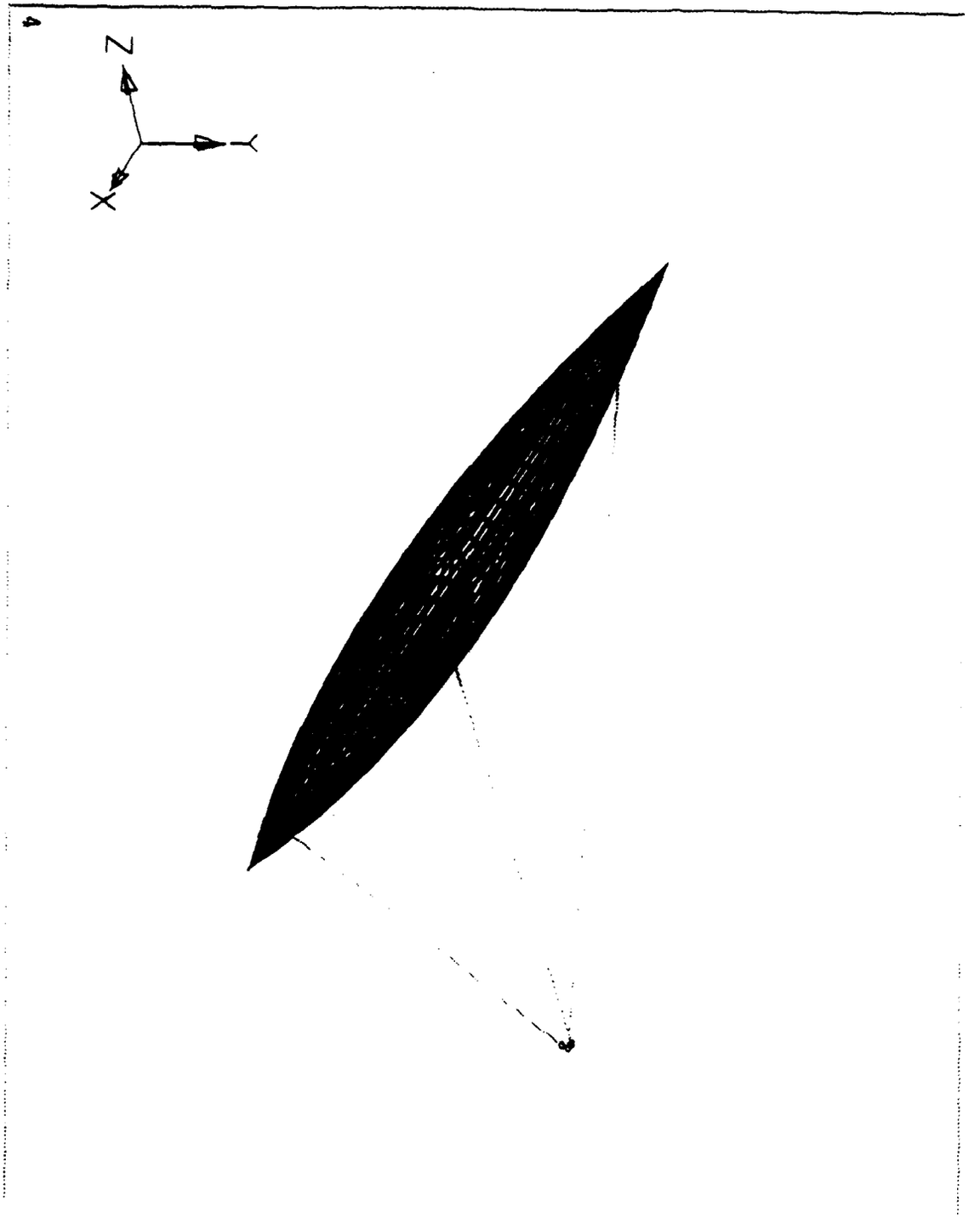


Figure 3: Nastran Finite Element Model - View 2

Table 1: Geometry and Material Properties

Nodal Coordinates of Strut 1	(0.0,-12.0,18) to (0.0 -520.8,514.2)	inches
Nodal Coordinates of Strut 2	(10.39,6.0,18.0) to (540.96,208.9,1341.7)	inches
Nodal Coordinates of Strut 3	(-10.39,6.0,18.0) to (-540.96,208.9,1341.7)	inches
Center of Torus	(0.0,0.0,1104.8)	inches
Focal Point of Paraboloid Mirror	(0.0,0.0,12.0)	inches
Materials Used	Space Cured Gr/Ep composite with M40J fibers ± 30 / 0 / 90 / 30 unidirectional tape	
Density	0.059 lb/in^3 density	
Modulus of Elasticity	17,210 ksi	
Diameter of Torus	13 in (0.015 in thickness)	
Diameter of Struts	7 in (0.015 in thickness)	
Moment of Inertia of three struts	2.02 in^4	
Moment of Inertia of torus	12.8 in^4	
Polar Moment of Inertia of struts	4.04 in^4	
Polar Moment of Inertia of torus	25.6 in^4	

2) Gravity load due to 0.002g in X direction.

3) Gravity load due to 0.002g in Y direction.

4) Gravity load due to 0.002g in Z direction.

Also the model was analyzed due to combination of internal pressure plus one of the gravity loads. The figures 4 through 7 depict the zoomed deflected shape of the reflector. It should be noted that the struts, torus and canopy are shown in its undeformed position.

ANALYSIS RESULTS:

The displacements of the end of the three struts are shown in Table 2, for gravity loads in each direction. They are less than 6 inches.

The displacement of the focal point is calculated as follows:

i) Displacement with reference to the spacial displacement of the torus.
ii) Twice displacement with reference to rotation of the torus.

This is because the torus/reflector will cause light rays to be reflected at twice the angle of torus rotation.

The Figure 4 shows the deflected shape of the reflector due to the internal pressure of the reflector/canopy chamber. The non uniform shape of displacements are apparent especially at two ends. The figure 5 depicts the deflected shape when the inflated chamber and the structural system is subjected to 0.002g acceleration in X direction. The potato chipping effect due to non uniform deflection is quite apparent in this diagram. The same effect due to

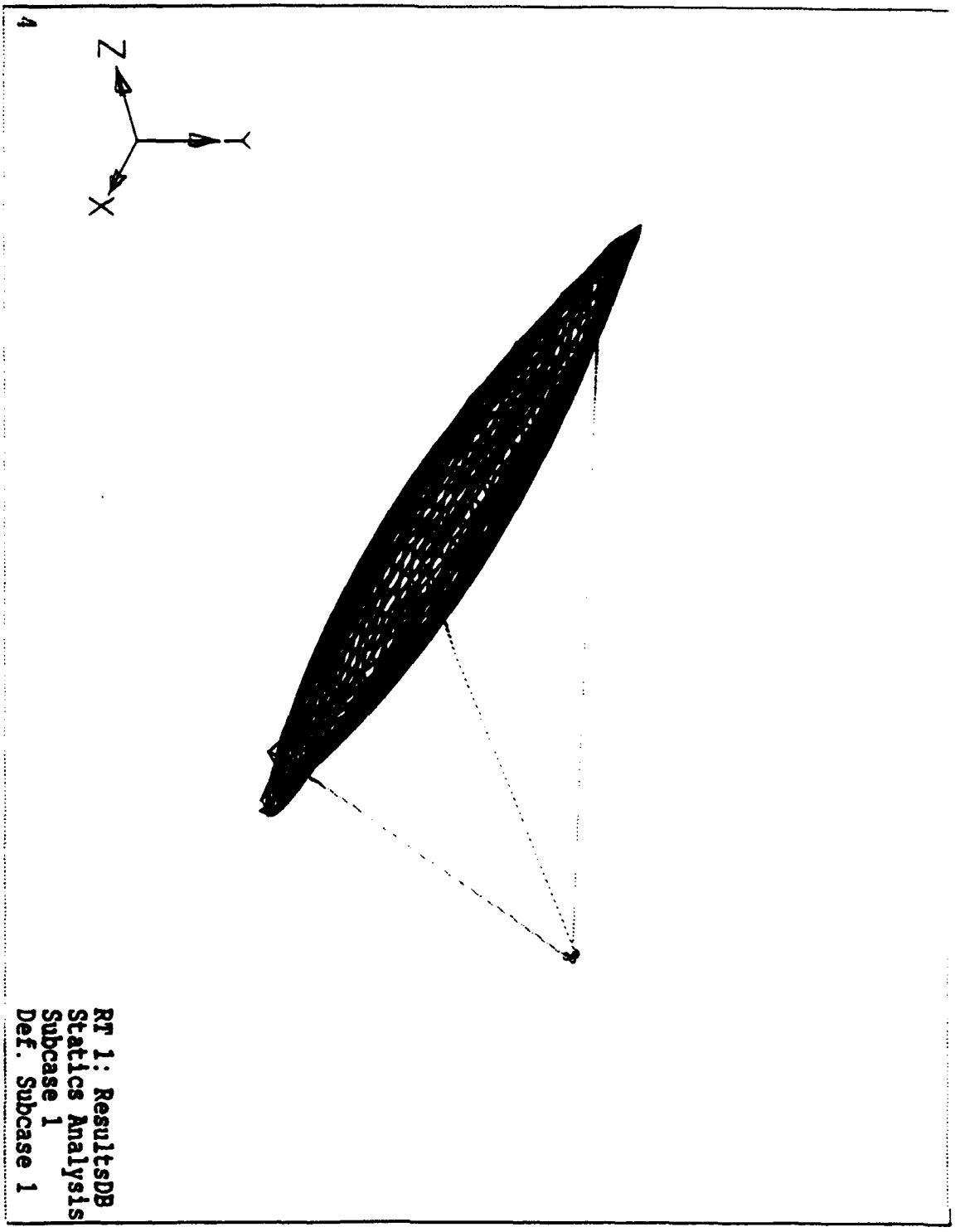


Figure 4: Deflected shape of the reflector due to internal pressure of the chamber

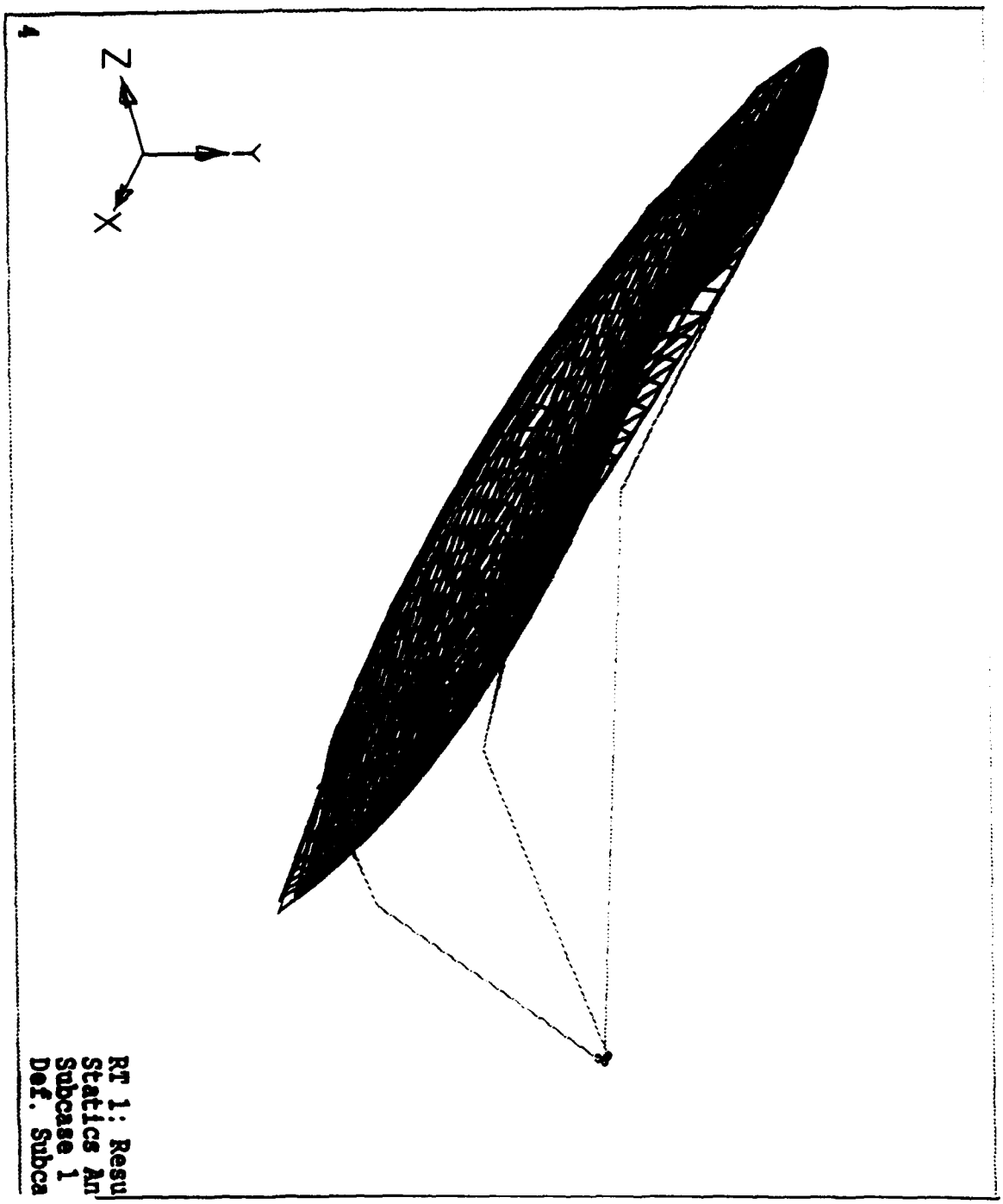


Figure 5: Deflected shape of the reflector due to internal pressure plus acceleration in X direction.

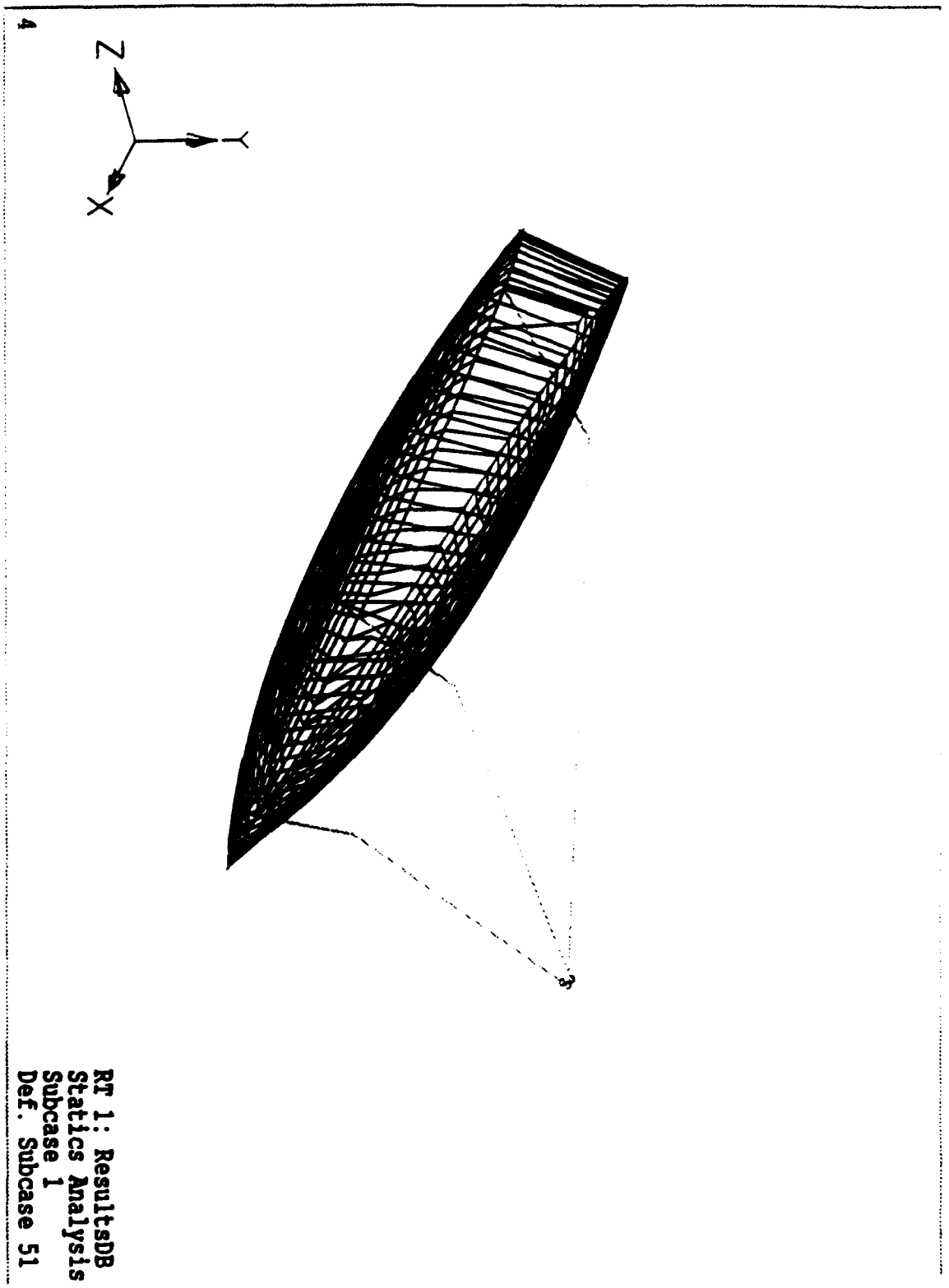


Figure 6: Deflected shape of the reflector due to internal pressure plus acceleration in Y direction.

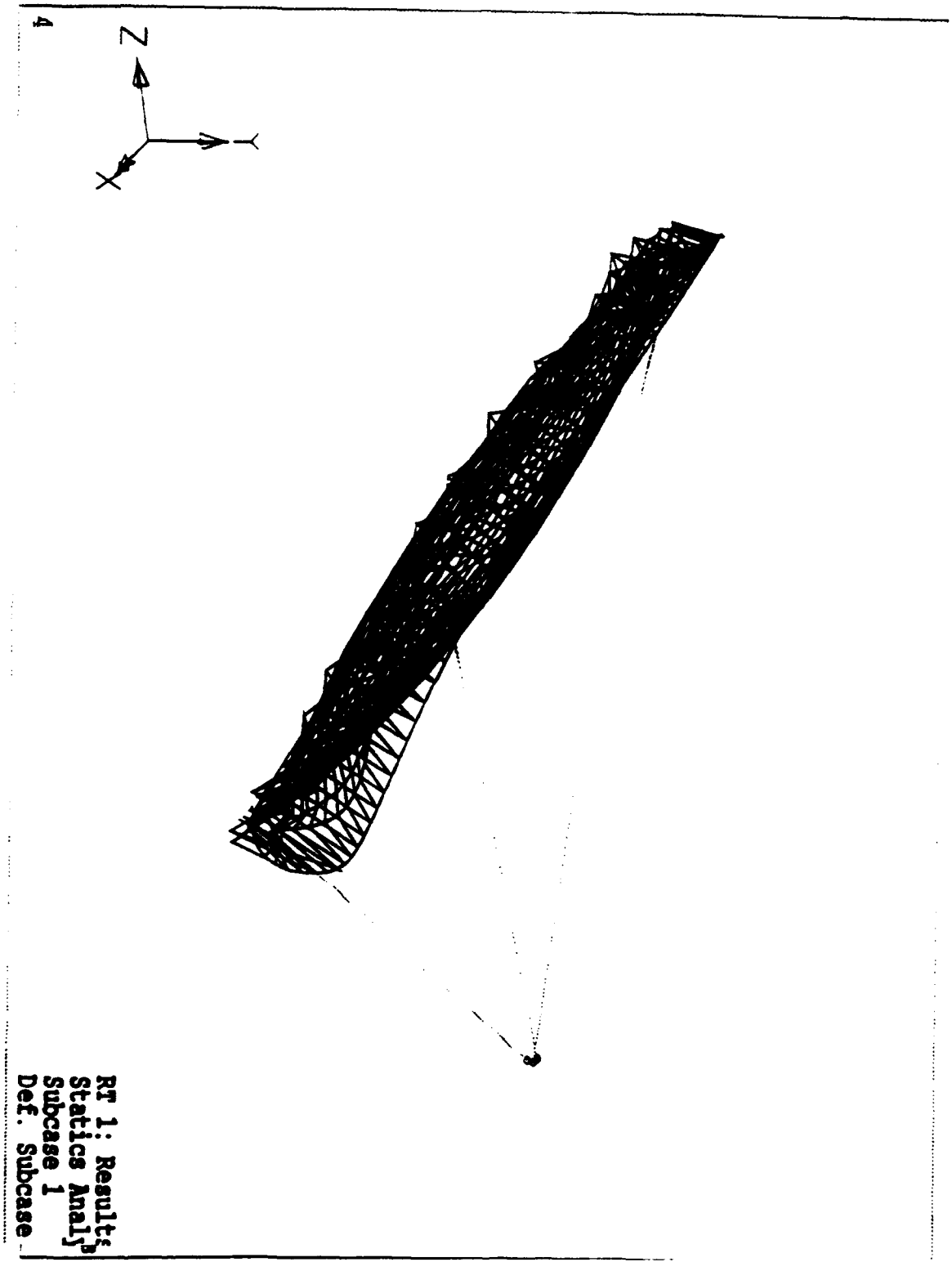


Figure 7: Deflected shape of the reflector due to internal pressure plus acceleration in $\hat{i}$ direction.

the inflation pressure and a 0.002g acceleration in the Y direction is shown in Figure 6. Similarly, the Figure 7 depicts the same effect due to acceleration in the Z direction. It should be emphasized that these figures show only the deflected shape of the reflector. The canopy and the struts are shown at their undeformed positions. In many of these plots the struts appear to have had sudden deflection near the reflector. This is just to facilitate graphic view of the deflections at the reflector.

DISCUSSION OF RESULTS:

The listed Tables II and III document the X, Y, and Z components of the displacements of the end of three support struts. All of them are less than 6 inches, which confirms small displacements of the struts. The Table 3 documents focal point displacements. The current study results were obtained by running another independent Nastran run of 3 simple struts from the end of the support struts to the focal point. The three struts end at three dummy node points at the focal point. One end of each of the struts were subjected to displacements and twice rotations obtained from the previous runs for the whole model. The displacements at the three dummy nodes are, the focal point displacements with reference to the three ends of the struts. A simple average values are documented in the Table 3, for acceleration in the X and Y directions. They are less than six inches. However, it should be emphasized that these results are for the type of materials and strut system selected. The structure can be made stiffer by replacing the three main struts with a six strut system or a built up truss system. The results are compared with the Reference[7] model for the same problem. In the Reference[7] model the ends of the three struts are connected to a non structural frame (i.e. with low EI values), which connects these three points to a triangular frame. A strut is run from the mid point of one of the struts to the focal point via center of gravity of the torus. Hence, the focal point displacement is the weighted average of the displacements of end of three struts. The comparison in Table 3 shows agreement of results of current study with that of reference [7].

Table 2 Displacements at the end of three struts: (inches)

	X displacement	Y displacement	Z displacement
Accceleration in X direction and internal pressure			
Strut 1	-4.7121	-0.5237	-1.9067
Strut 2	-4.7363	0.6694	1.8945
Strut 3	-1.0651	0.0103	0.01057
Acceleration in Y direction and internal pressure			
Strut 1	0.0446	-3.5623	0.5601
Strut 2	0.0206	-3.4828	0.5252
Strut 3	0.0424	-1.3598	-1.3943
Acceleration in Z direction and internal pressure			
Strut 1	0.0451	0.04649	0.0118
Strut 2	0.02174	0.1282	0.0289
Strut 3	0.04355	0.01517	0.0155

Table 3 Displacement of focal point: Results Comparsion (inches)

Reference 7 Results	X displacement	Y displacement	Z displacement
Acceleration in X direction	-2.35	0.0	0.0
Acceleration in Y direction	0	2.10	-0.32
Current Study Results	X displacement	Y displacement	Z displacement
Acceleration in X direction	-4.13	0.17	0.02
Acceleration in Y direction	-0.63	2.88	-0.569

SUMMARY:

This report documents the possibility of diffusion of the focal point on solar concentrators due to thrust loads. Also the non uniform deformation of the reflector documented confirms the potato chipping effect.

FUTURE WORK:

The present study documents potato chipping effect of the reflector, due to the interaction of the torus/reflector chamber. Use of a secondary concentrator system with a single chamber reflector may alleviate this problem. A pointing and accuracy analysis of a single chamber reflector is needed to study the suitability of this concept.

ACKNOWLEDGEMENT:

Financial support through the Research and Development Laboratory of AFOSR is appreciated. Thanks are due to the focal point Kristi Laug, M. Holmes, S. Cartier, Major-select D. Castillo, N. Abayakaran, and S. Lou for their help during this study.

REFERENCES:

1) K. K. Laug, " The Solar Propulsion Concept is Alive and Well at the Astronautics Laboratory," Joint Army, Navy, NASA, Air Force Propulsion Meeting, 1989, Cleveland, Ohio.

2) M. Thomas and G. J. Friese, " Pressurized Antennas for Space Radars," AIAA Collection CP 807, from the AIAA Sensor Systems for the 80's Conference, December 1980.

3) G. J. Friese, G. D. Bilyeu, and M. Thomas," Initial 80's Development of Inflated Antennas," L'Garde Report LTR-82 -GF-107, NASA Contractors Report,

4) A. J. Wendt and L. D. Surber, " Inflatable Antennas," 3rd Aerospace Expandable and Modular Structures Conference, AFAPL TR 68-17, Mar 67.

5) M. Thomas and G. Veal, " Scaling Characteristics of Inflatable Paraboloid Concentrators," Solar engineering, ASME 1991.

6) R. Purasinghe, " Pointing and Accuracy Analysis of Solar concentrators," ASME Solar Energy conference, DC, April 93.

7) G. Ritchie, "Solar Concentrator Support Structure," Solar Energy Conference, DC, April 93.

8) NASTRAN MANUAL, NASTRAN Finite Element Software.

ON THE MAXIMIZATION OF THE SEMI-MAJOR AXIS a OF A MANEUVERABLE SATELLITE MOVING INITIALLY IN A CIRCULAR ORBIT ABOVE THE EARTH

Dr. Paul Rosenthal, Ph.D.

Associate Professor

Department of Mathematics

Final Report for:

Summer Faculty Research Program

Phillips Laboratory, Edwards AFB, CA

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, District of Columbia

August 30, 1993

On the Maximization of the Semi-Major Axis a
of a Maneuverable Satellite Moving Initially
in a Circular Orbit about the Earth

Dr. Paul Rosenthal, Ph.D.
Associate Professor
Department of Mathematics
Los Angeles City College

ABSTRACT

Problem Statement: Consider a satellite moving in a circular orbit about the earth. At time $t = 0$, it's thruster is turned on and remains on for a total burn time T_{BT} then turned off. Assume one can control the direction of the thrust vector $\vec{f}$ relative to the velocity vector $\vec{V}$. (of the satellite) Assume $\alpha(t)$ is the angle between $\vec{f}$ and $\vec{v}$. We assume all vectors act thru the center of mass = center of gravity of the satellite. Within a given class of controls $\alpha(t)$, does there exist a control $\alpha(t)$ in our class that will maximize the semi-major axis a of the elliptical path followed by the center of mass of the satellite in the time frame after T_{BT} ?
We formulated then solved the above problem. We now state the results of our research in what follows:

1. $a_{max} = a_0 + \Delta a_{max}$ where,

2. $\Delta a_{max} = \dfrac{2a_0 f T_{BT}}{V_0 M_0}$ where,

a_0 = semi-major axis at $t = 0$, V_0 = velocity of the satellite at time $t = 0$, f = magnitude of the thrust vector $\vec{f}$, M_0 = mass of the satellite at $t = 0$. We assume all quantities on the right side of 2. are positive.

3. The maximizing control is $\alpha(t) = 0$ for $0 \le t \le T_{BT}$.
I.E. one keeps the thrust vector $\vec{f}$ and velocity vector $\vec{V}$ aligned during the burn.

4. $\alpha(t)$ of 3. is unique i.e. there is one and only one maximizing control and further this control is a global maximizing control.

5. Our control class also includes bang-bang controls in the competition.
(Reference 1. assumes a class of linear controls.)

On the Maximization of the semi-Major Axis a
of a Maneuverable Satellite Moving Initially
in a Circular Orbit about the Earth

Dr. Paul Rosenthal, PH.D.

Summary of the Analysis

We assume Newton's gravitational law, we assume a_o such that the drag on the satellite can be neglected, we assume a constant thrust, $M = M_o - \dot{M}\, t > 0$, M = total mass of satellite, $\dot{M} > 0$ = burn rate of thruster. We next wrote an expression for the total energy of the satellite. We then calculated the time rate of change of the total energy E_{Total}. We next derived the angular momentum H of the satellite. We utilized vector analysis in the above analysis. We next derived an expression for the angular rate of the satellite. All the above analysis assumes our time frame is $0 \le t \le T_{BT}$. We next proved

$$\dot{\Delta}\, a = \frac{2a_o f \cos(\alpha(t))}{V_o M}$$, where $\Delta a = a - a_o$ ($\dot{\Delta}\, a$ is

is the time rate of change of $\Delta\, a$.) This is for the above time frame. We next integrated $\dot{\Delta}\, a$. Then we did a maximizing argument and obtained the results 1. thru 5. stated in this report's abstract.

An Assumptional List

1. 6.7×10^6 meters $\le a_o \le 10^{12}$ meters.
2. We assume the satellite is rigid.
3. $T_{BT} \le 10^4$ seconds.
4. We assume the satellite does not rotate about it's c/m.
5. We assume $V_o = \sqrt{\frac{k}{a_o}}$, where $k = \gamma M_e$, and

   γ = gravitational constant, M_e = mass of earth.
6. We assume f = magnitude of the thrust vector $\le 10^{-1}$ newton.
7. We assume the burn rate $\dot{M} = 10^{-5}$ Kg/sec.
8. $M_o = 1000$ kg.
9. We assume at the c/m of the earth our frame of reference is an inertial frame i.e. non-rotating and non-accelerating.

(Some of the above assumptions we assume can be relaxed.)

References

1. An Analytic Method for Determining Finite Burn Effects, K. Zondervan, and S. Sturgeon. The Aerospace Corporation, El Segundo, CA, AIAA Paper # AAS 92-131, 1992
2. Modern Classical Mechanics and Control, D. Burghes.
3. Asymptotic Expansions for Differential Equations, W. Wasow.
4. NASA Report 3 CR-1005.
5. Astrodynamics, Bates, Neuville, and White.

This report is a summary of the research I performed in the time frame July 5, 1993 - August 27, 1993 while at the Phillips Laboratory, Edwards AFB, CA. Dr. Keith McFall, phd was my focal point at the Phillips Laboratory. All my reaearch I performed at the Phillips Labortory was submitted periodically in report form to Dr. Keith McFall, phd.

NONLINEAR BENDING VIBRATION OF FLEXIBLE SOUNDING ROCKETS: AN OPPORTUNITY FOR NONLINEAR IDENTIFICATION

Mark S. Ewing

Associate Professor

Department of Aerospace Engineering

University of Kansas

2004 Learned Hall

Lawrence, KS 66045

Final Report for:

Summer Faculty Research Program

Phillips Laboratory, Hanscom Air Force Base, Massachusetts

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D.C.

August 1993

NONLINEAR BENDING VIBRATION OF FLEXIBLE SOUNDING ROCKETS: AN OPPORTUNITY FOR NONLINEAR IDENTIFICATION

Mark S. Ewing
Associate Professor
Department of Aerospace Engineering
University of Kansas

Abstract

The state of the art in strength and flight stability analysis of flexible sounding rockets during launch and ascent was studied. Although interstage joints and other connections between payload segments are known to exhibit nonlinear displacements with applied moment, allowances for this behavior are not typically made. A simple model of a sounding rocket was devised comprising two flexible beams connected by a rotational spring with stiffness which is both linear and cubic in the displacement. The effect of this nonlinear spring on beam response was studied to form the basis of nonlinear identification studies. A survey of the state-of-the-art of nonlinear structural identification theory was accomplished. Methods to develop and apply nonlinear identification algorithms to beams with nonlinear connections are proposed.

NONLINEAR BENDING VIBRATION OF FLEXIBLE SOUNDING ROCKETS: AN OPPORTUNITY FOR NONLINEAR IDENTIFICATION

Mark S. Ewing

Introduction

Sounding rockets have evolved into a highly reliable vehicle to boost "science payloads" into suborbital trajectories. These payloads are used for missions ranging from atmospheric exploration to new sensor evaluation. The launch vehicles are typically multi-stage, increasingly with powerful first stage solid rocket motors (e.g., the Nike and Terrier boosters). Some now even have guidance and control systems for the boost phase, although most are still just "point and launch" devices.

As with most useful launch systems, the performance envelope with regard to payload-altitude is often tested. For example the recent launch of the Space Power Experiments Aboard Rockets Program (SPEAR III) on a Terrier Mk-70--Black Brant VC vehicle was the longest payload attempted on that launch system. Figure 1 is a schematic of the launch configuration. Concerns over the flight stability and strength of the vehicle were the subject of numerous studies [1-3] but the launch *was* successful [4]. The need for research into the subject of vehicle stability and strength was clearly indicated.

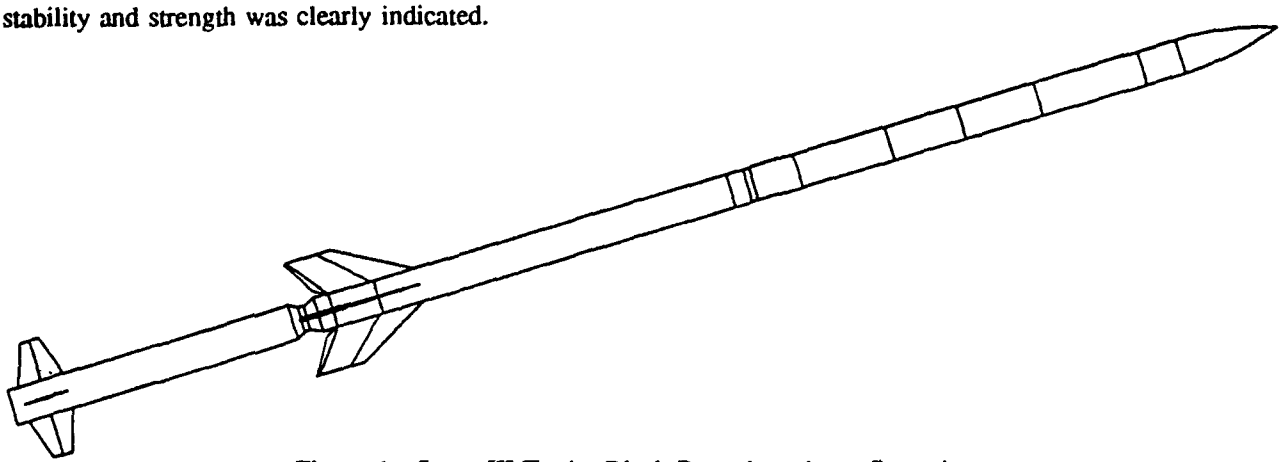


Figure 1. Spear III Terrier-Black Brant launch configuration.

With an uncommonly slender launch vehicle, the effects of large deflections of the structure during wind shear passage give pause for concern. This is especially true since the joints used to connect segments of the missile

together are known to be nonlinear in the internal moment. Figure 2 gives a cross-sectional view of a "radax" (radial-axial) joint as well as a "manacle" joint. As shown, the radax joint is based on a set of bolts set at 30 degrees to the longitudinal axis of the missile. The manacle joint is based on a "c"-shaped band which grasps circumferential flanges on the two missile segments being joined.

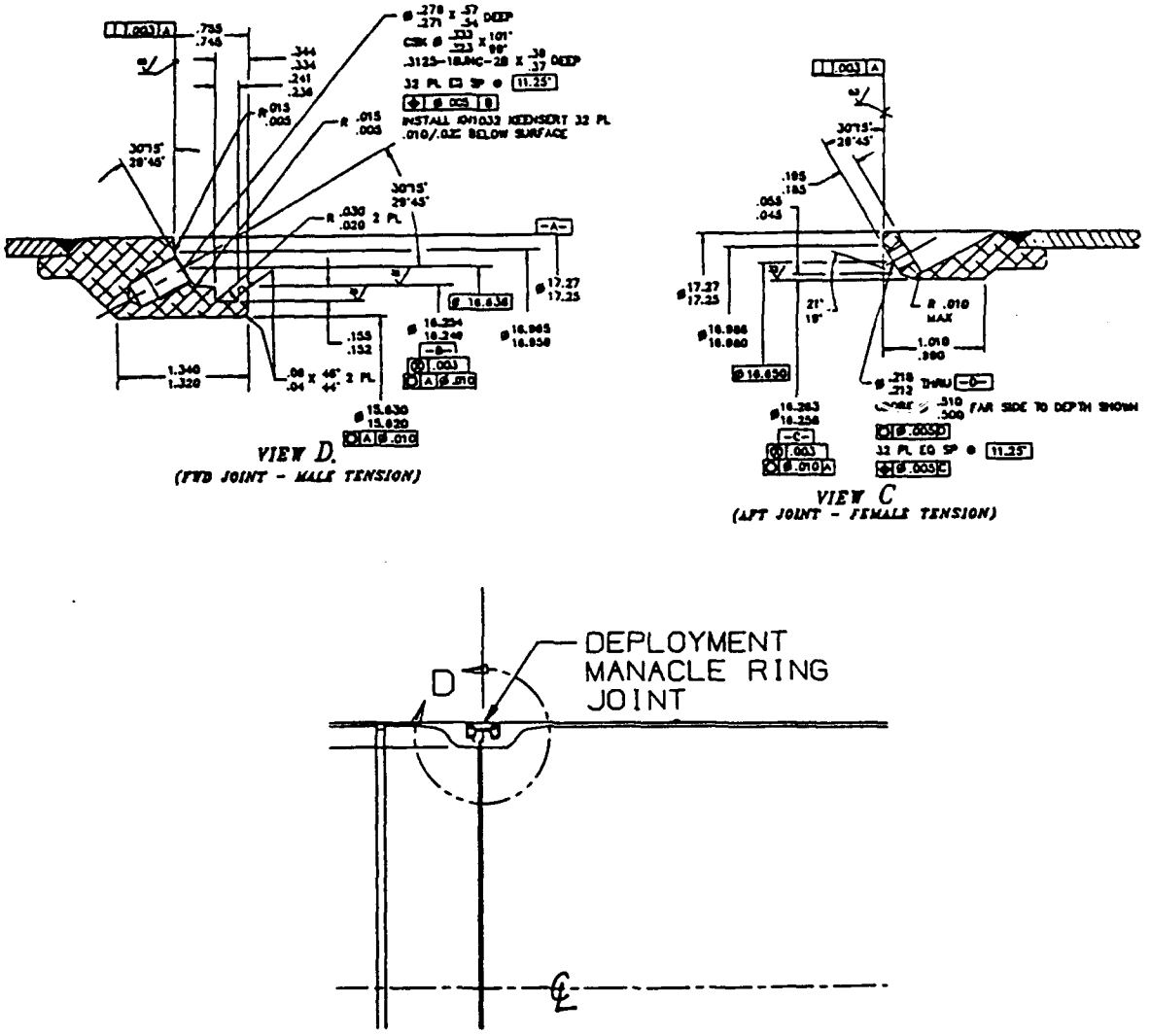


Figure 2. Radax (top) and manacle (bottom) joint details.

The results of a vertical cantilevered bending test of the upper section of the SPEAR III, which has both kinds of joints is shown in figure 3. Here the load-unload path for a point on the missile 220 inches from the tail of the payload section is recorded for a lateral load applied at a 210 inches. The "hysteretic" behavior is clearly seen. Unfortunately, no *dynamic* characterization of the joints has been attempted.

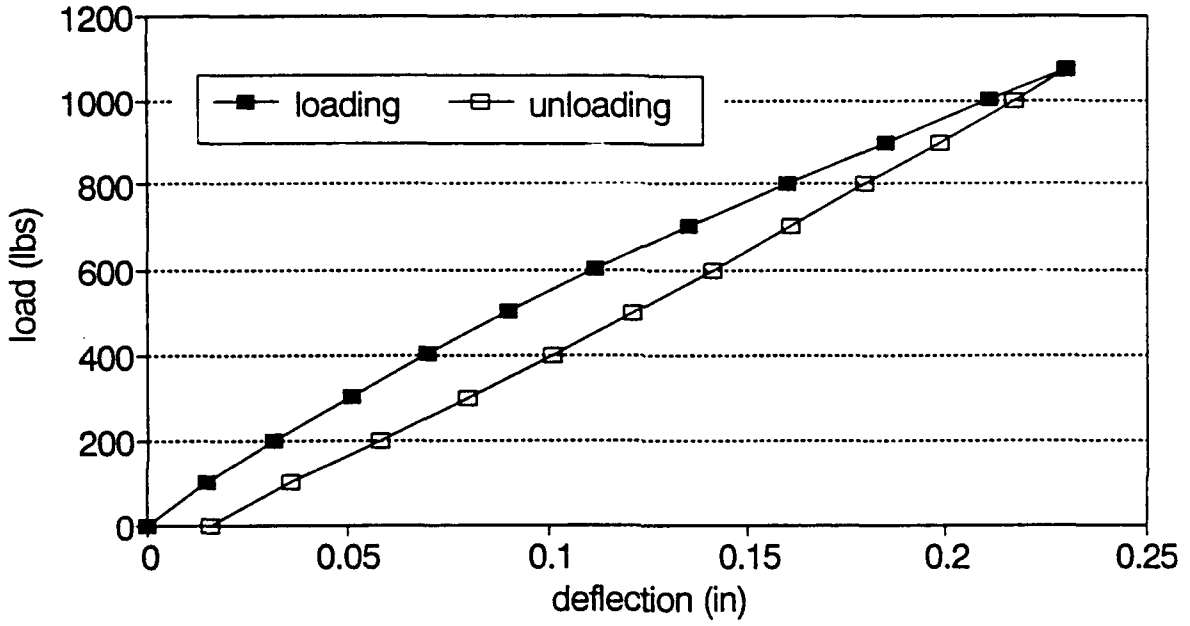


Figure 3. Load-unload curve for the Spear III payload section.

Therefore, an opportunity exists to attempt to characterize real joints using the emerging theory of nonlinear structural identification. In this context, identification refers to the determination of the values of unknown constants within a given mathematical description of a structure. Sometimes these constants are physical parameters such as stiffness or damping ratio. Structural identification of these physical parameters has been demonstrated by many researchers for linear structures using finite element models [5]. Few have successfully identified physical parameters in nonlinear systems.

State-of-the-Art in Sounding Rocket Structural Analysis

A review of the state-of-the-art in structural analysis for sounding rockets was conducted based in part on an inspection of the Spear III preflight analyses conducted for the Phillips Laboratory by Bristol Aerospace, Conatec and NASA/Wallops Island [1-3]. In addition, the analysis techniques were the subject of a day-long visit to the Wallops Flight Facility, in particular to allow discussions with Mr. Jim Lanzi of the Performance and Analysis Section.

Concerns during Launch and Ascent. The most critical strength and stability concerns during the launch and ascent phase for a sounding rocket are:

* exceedance of maximum bending moment
* inadequate static margin
* the onset of spin rate and pitch rate lock-in.

During the ascent, a sounding rocket is subjected to often serious lateral wind loads which cause directional changes as well as significant bending of the missile. The definition of the maximum bending moment a missile joint can transfer is typically done by analysis, based on earlier tests to failure on essentially equivalent missile structures. A much less exacting process is used to determine the maximum loading to be expected on the missile. In particular, a "worst case" lateral wind velocity **or** a "worst case" missile angle of attack resulting from a wind shear is typically specified. Then a flight load routine is invoked (as described later) to determine the resulting internal bending moment. The load "applied" in the analysis can be related to a certain probability of occurrence using available winds aloft probabilities. A margin of safety can then be determined by comparing the assumed bending moment to the allowable (from either analysis or test). Unfortunately, the probabilistic approach is not routinely taken and the "maximum bending moment" from analysis is simply taken as a *possible* bending moment. A more rigorous approach is, however, taken in more complicated launch environments [6].

As a stable missile is perturbed by any lateral winds, like any stable flight vehicle, it tends to recover to its original flight path (the so-called "weather vane" effect). However, for some flight conditions, the resulting forces can, instead, be destabilizing. This occurs when the center of pressure of the vehicle is located "ahead of" the center of mass. This is particularly so if the vehicle is flexible, in which case it is possible that the deflections resulting

from flight loads alter the shape of the missile so as to shift the center of pressure forward. In most cases, for sounding rockets, the center of pressure *does* shift forward and the analysis of the extent of the shift is a key stability concern. As a matter of course, the static margin is typically checked at the following flight points:

* maximum dynamic pressure
* minimum rigid body static margin
* maximum thrust

Finally, an undesired coupling between pitching and rolling may occur. This phenomena has been observed on sounding rockets, especially during the coast phase between first and second stage ignition or just after second stage ignition. During these times, the roll rate and the pitch rate may be either slowly decaying (during coast) or slowly increasing (after second stage ignition). If the rates are nearly the same, the phenomena become coupled and the missile will "cone" instead of pitching in a single plane. This phenomena has been analytically predicted using "trajectory analysis" codes (described later). Figure 4 is extracted from the Conatec preflight analysis [2] and shows the predicted pitch rate and roll rate for a rigid missile with one of the fin settings considered for launch. The flight regimes within which the pitch rate and roll rate often interact is highlighted. Although the effects of missile bending on analyses intended to predict *this "mode lock-in" may be important, this has not been attempted.*

Analysis routines. Two types of structural analysis have emerged as the popular methods of assessing flight-worthiness with regard to flight stability and strength. The quasistatic equilibrium algorithm, as encoded in the Flight Loads, Moment and Deflection (FLMD) program has been used for nearly thirty years. A transient analysis based on modal expansion theory has only recently been incorporated within a six degree of freedom "trajectory analysis" code, the General Electric Missile and Satellite Simulation (GE MASS, or GEM) program .

The quasistatic analysis within FLMD will now be described. The basis of the method is that the missile responds (deflects) in reaction to both the applied aerodynamic loads as well as the "inertial" loads "caused" by lateral and rotational acceleration. The loading considered is typically very severe due to the sudden application of a large angle of attack chosen to model wind shear. The scheme is iterative since the aerodynamic loads are a function of the shape of the missile.

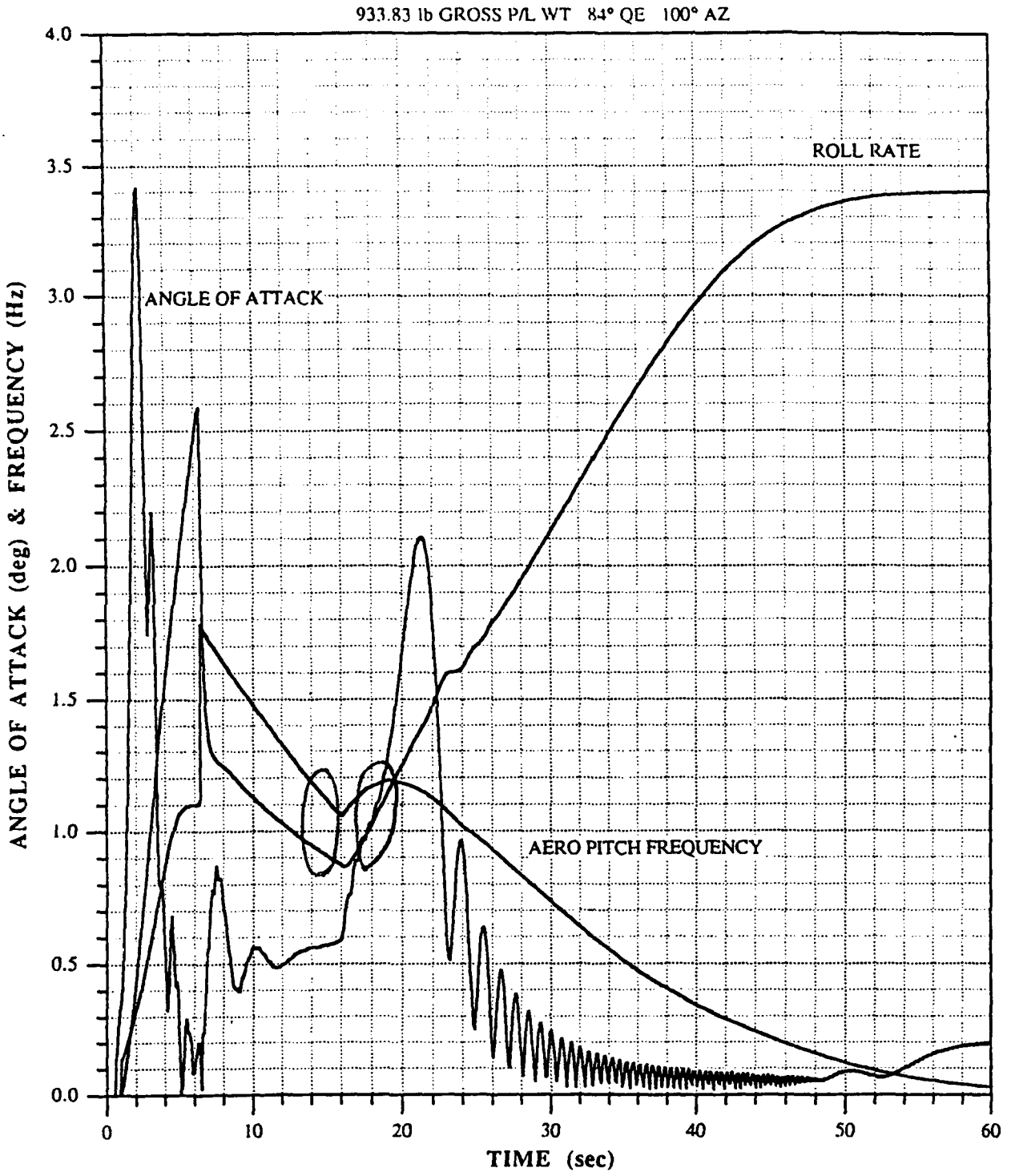


Figure 4. Aero pitch frequency (pitch rate), roll rate, and angle of attack for the Spear III as predicted by Conatec for a specific fin setting. Two areas of potential spin rate lock-in with roll rate are circled for clarity. However, in this case the two rates are sufficiently separated. Also, note that in this simulation, at the time when the two rates have the same value (t=19 seconds),the roll rate is sharply increasing while the pitch frequency has a stationary value: this is a very favorable situation.

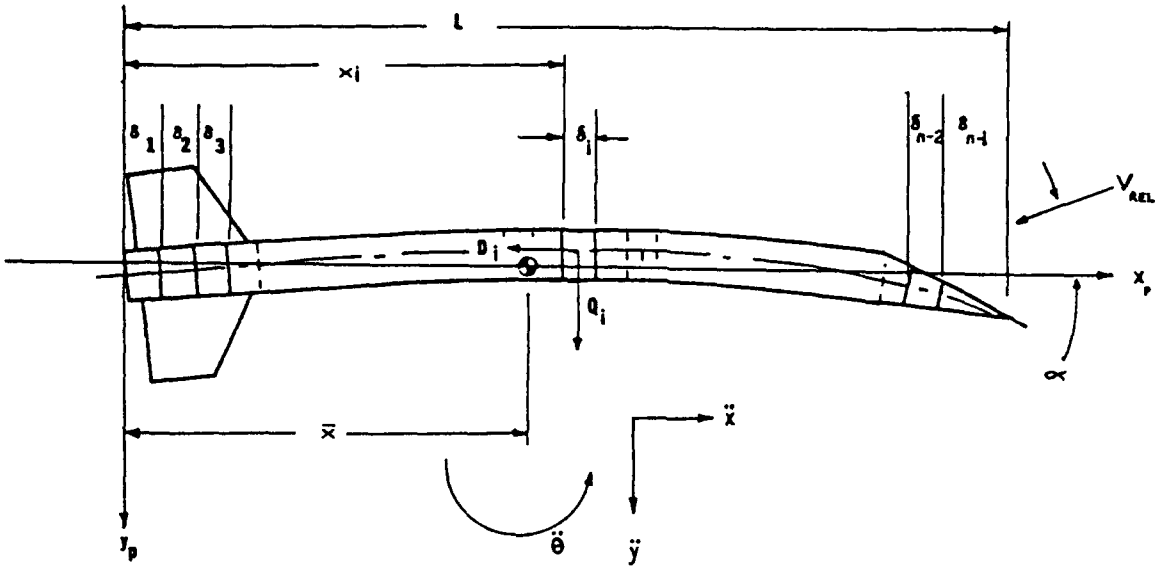


Figure 5. Coordinate system and idealization of a sounding rocket.

Using figure 5, for a given iteration of the analysis, on the ith segment, or panel, the lateral load is:

$$p_i = \left[\alpha + \frac{1}{2}(\theta_i + \theta_{i+1}) - \gamma_i - \phi_i\right] \delta_i \, q \, d \left[\frac{n}{\alpha \, q \, d}\right]_i \tag{1}$$

Here: α is the rigid body angle of attack; θ_i is the slope of the ith panel; γ_i is the fin twist angle *under loading* (user-defined) for a fin; ϕ_i is the initial fin misalignment angle for a fin; δ_i is the length of the panel; q is the dynamic pressure; d is the diameter of the missile. The tabular values of $[n/(\alpha q d)]$ have been calculated using a second order shock expansion theory [7]. If all of the lateral *aerodynamic* loads are summed, the rigid body lateral load, N_R, the location of the center of pressure, x_{cp}, and the rigid body moment, M_R, are found:

$$N_R = \sum_{i=1}^{n-1} p_i \qquad x_{cp} = \frac{1}{N_R} \sum_{i=1}^{n-1} p_i \left(x_i + \frac{\delta_i}{2}\right) \tag{2}$$

$$M_R = N_R \left(\bar{x} - x_{cp}\right)$$

Summing all the lateral loads, including the component due to thrust misalignment and the rotation of the "end" segment, θ_e, and equating them to the total mass times the rigid body lateral acceleration:

$$(\frac{W}{g})\,\ddot{y}_R = N_R + T\,(\theta_e - \beta) \tag{3}$$

Here: W is the total weight of the system; g is the gravitational constant; T is the thrust; y_R is the lateral deflection of the center of mass; and, β is the thrust misalignment angle. Similarly, solving for the mass moment of inertia times the rigid body angular acceleration:

$$I\,\ddot{\theta}_R = M_R - \sum_{i=1}^{n-1} \frac{1}{2}\,d_i\,(y_i + y_{i+1}) + T\left[y_e + (\theta_e - \beta)\,(\bar{x} - x_e)\right] \tag{4}$$

Here: I is the mass moment of inertia; d_i is the drag on the ith panel; x_e is the location of, and y_e is the deflection of, the tail of the missile. Finally, during each iteration of the analysis, the total "running load"--that is, the applied *and* inertial load--is:

$$Q_i = p_i + \frac{\delta_i\,w_i}{g}\left[\left(x_i + \frac{\delta_i}{2} - \bar{x}\right)\ddot{\theta} - \ddot{y}\right] \tag{5}$$

A static beam analysis algorithm then finds the deflections and rotations using a straightforward set of 4 equations in the moment, shear, slope and displacement. [This "state vector and transfer matrix" method is described in reference 7.] Note that equations 1-4 depend on lateral deflection and rotation, which are found based on the loads in equation 5. Therefore, equations 1-5 are used iteratively until the change in the deflections and rotations tend toward zero. Within FLMD, this is assumed to occur when the norm of the total displacement deviation "vector" from one iteration to the next is less than a user-input value.

The only non-standard capability of the static beam analysis in FLMD is that allowance is made for a rotational joint with "slop". That is, at the user's option, a moment vs. slope relationship as shown in figure 6 is allowed. The implementation is to check for rotations at a slop joint, and if the rotation is less than a user-supplied

value, no moment is transmitted through the joint. This capability was modified with an experimental version of FLMD, wherein the stiffness of a joint could be input as a piecewise linear function, that is with one slope below a user-specified value, and a (larger) slope above. This is felt to be a more accurate model of a true joint, and FLMD appeared to work well with this modification.

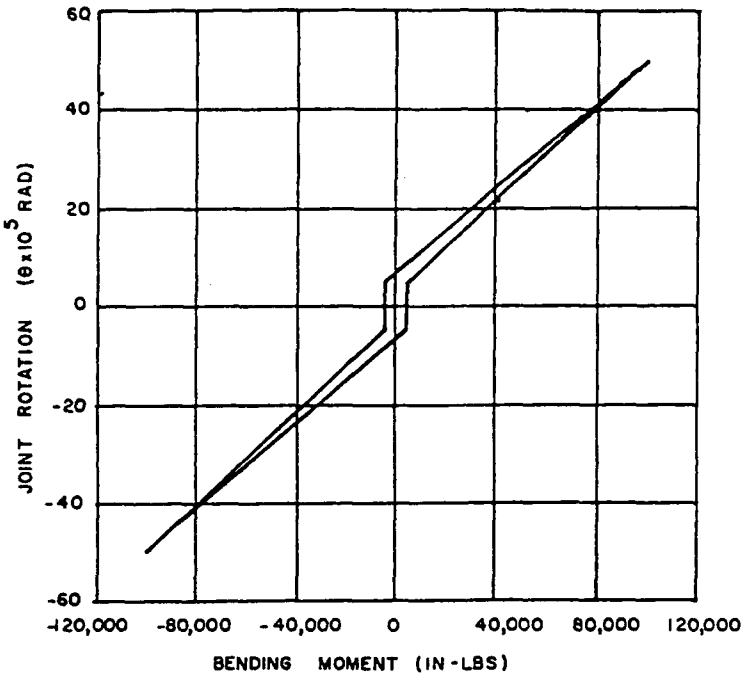


Figure 6. Bending moment versus joint rotation for a "slop" joint in FLMD.

Numerous numerical experiments were run using FLMD to investigate sensitivities to various input values. Although not described in detail here, it was found that the static margin of a missile was profoundly influenced by the assumed values of fin misalignment angle. In particular, changing this angle from $+0.2^{o}$ to -0.2^{o} changed the static margin by as much as 40%. This should probably be investigated further.

The fundamental flaw in the FLMD analysis is that a real missile takes a finite length of time to deflect. During this time, the angular acceleration resulting from a moment imbalance will either tend to reduce the angle of attack (and the loads) or increase it (and the loads). Thus, for a stable rocket, FLMD is a *conservative* routine since it does not properly address the tendency to remain stable. It should be noted that improvements on the aerodynamic load predictions within FLMD have been made using the program QUASAP [8], but the fundamental equilibrium concept is the same.

The transient vibration analysis incorporated recently into the NASA/Wallops Island version of GEM [9] is a more theoretically sound approach than encoded in FLMD, but brings with it the necessity of careful application. The basic GEM program uses the so-called "state space solution" technique. That is, the system of second-order, ordinary differential equations of motion describing the rigid body response of the launch vehicle are transformed to first order equations via a simple substitution. In equation form, if the lateral displacement of the missile center of mass is u, the substitution,

$$v = \dot{u} \tag{6}$$

transforms the equation of motion,

$$m\ddot{u} + c\dot{u} + ku = 0 \tag{7}$$

into the set of coupled equations,

$$\begin{aligned} m\dot{v} + cv + u &= 0 \\ \dot{v} &= u \end{aligned} \tag{8}$$

which are easily solved numerically.

The flexible response of the missile is also described by a set of second order, ordinary differential equations, namely, the second degree of freedom equations describing modal responses (in accordance with the modal expansion theory. For example, if the response of the jth mode is q_j, then the modal equation of motion is:

$$\ddot{q}_j + 2\zeta_j\omega_j\dot{q}_j + \omega_j^2 q_j = \frac{Q_j}{m_j} \tag{9}$$

The flexible response in the jth mode is computed (using the same state space equation solver as is used for rigid body motion) by transforming this flexible body *modal* equation of motion into a pair of first order equations. Typically only the first two bending modes of vibration in two orthogonal planes (for a total of four modes) are considered. Note that the flexible body equations of motion are actually much more complicated than is apparent in equation 9 since they are very much coupled to the rigid body equations. The coupling, through the generalized force, Q_j, and the equivalent viscous damping ratio, ζ_j, involves the orientation angles of the rigid body (i.e., angles of attack in two planes) as well as rotational rates.

Commentary. The use of the quasistatic analysis algorithm within FLMD and QUASAP is conservative and it is therefore very tempting to rely upon it. The degree to which it is conservative has not been determined. However, if future flights continue to "push the envelope" with regard to vehicle length and mass, this question may need to be answered. On the other hand, the transient analysis with a trajectory program provides a very much more detailed view of missile response. The missile is numerically "flown" through difficult flight regimes such as wind shear, and the resulting bending moment and stability margin can be calculated. However, the user then has a much more difficult task since he or she must be careful to choose a flight condition which is "appropriately severe". Fortunately, for analyses conducted within minutes of the launch, this task is eased by having an estimate of the current winds aloft, provided by a weather balloon. For earlier analyses, say during the design of the mission, the user should, ideally, assign a probability of occurrence to the flight conditions analyzed. Alternatively, a Monte Carlo analysis could be used to properly assign the probability estimates. Until such an approach is taken, the calculation of maximum bending moment will continue to be a suboptimal analysis quantity for the design process.

Regarding the overall analysis regimen for sounding rockets, there appear to be two areas of deficiency. The first, is the inability to reliably predict aerodynamic loads in the transonic regime. [This deficiency is shared by the aircraft aeroelasticity community as well!] The second is the inability to consider the effect of joint

nonlinearities on response and stability. There are many possible ways to include this effect, ranging from the use of (linearizing) "describing functions" to numerical integration of the nonlinear differential equations of motion of a missile. The methods discussed later in this report are those most amenable to the use of nonlinear structural identification, that is, the most simplistic, since the science of nonlinear identification is just emerging.

Nonlinear vibrations of two beams connected by a rotational spring

A simple model of a structure similar to a sounding rocket with nonlinear inter-stage joints has been formulated. The purpose of developing the model is to:

1. help evaluate the importance of nonlinear effects
2. provide an analytical model for comparison to finite element models.

The model is composed of two Euler-Bernoulli beams, connected by a rotational spring. This spring responds with a restoring moment proportional to both the first and third powers of relative displacements of the joined ends of the beams. See figure 7. A more detailed description and more extensive results are given in another report in this volume [10].

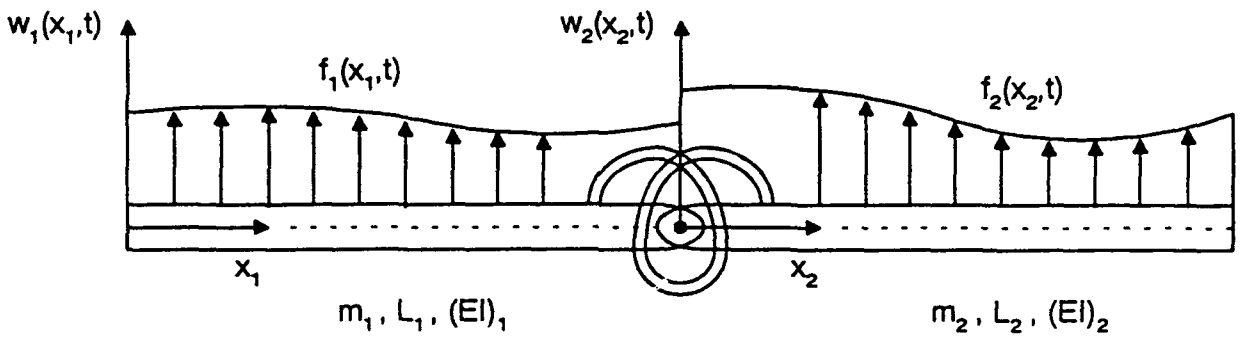


Figure 7. Idealized model of two beams connected by a nonlinear, rotational spring.

The solution for the response of the beam to lateral sinusoidal is based on an initial assumption that the response is also sinusoidal. In the course of applying boundary conditions, a displacement term is cubed, introducing

a $sin^3(\omega t)$ term. After using a trigonometric identity equating this term to two terms proportional to $sin(\omega)$ and $sin(3\omega)$, the condition of "harmonic balance" is imposed, requiring the coefficients of the $sin(\omega)$ terms to be "balanced". In this way, an expression which is a *first approximation* to the nonlinear response is obtained. The resulting, first order "harmonic balance" equation is cubic in an undetermined amplitude coefficient. This situation is identical to the mathematical situation of the single degree of freedom "Duffing" oscillator (with a purely cubic spring). As a result, near a (linear system) resonance frequency of excitation, a "jump phenomena" (or a Hopf bifurcation) occurs, wherein three possible amplitudes are in equilibrium at a given frequency. In this frequency range, then, two amplitudes are stable and one is unstable. Figure 6 in reference 10 shows a characteristic amplitude versus frequency curve for a specific value of the linear versus nonlinear spring force per radian which graphically displays the existence of three possible amplitudes.

Numerical experiments have shown that the response of the beam system can be very sensitive to small changes in physical parameters as well as to forcing frequency. In particular, as shown in figure 8, a 16% change in the ratio of the linear to nonlinear spring force per radian (6.0×10^7 to 7.0×10^7), is shown to have a dramatic effect on the response shape of the system. Such behavior, that is the sensitivity of system response on physical parameters, is a well-known characteristic of nonlinear systems.

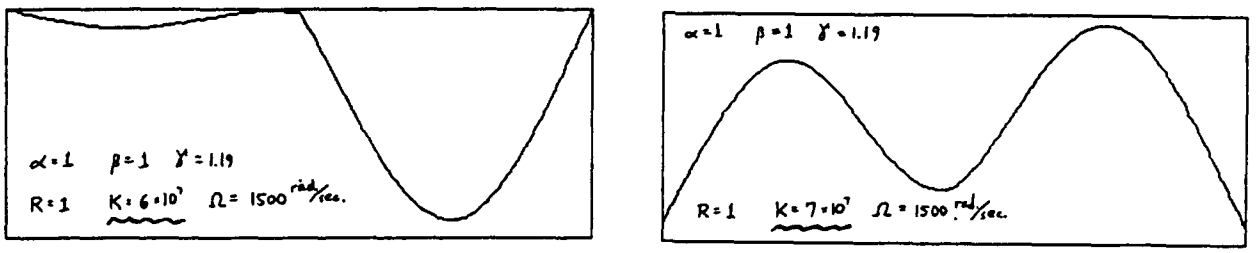


Figure 8. Response shapes for the idealized beam in reference 10 for two values of nonlinear stiffness, K.

State-of-the-art of Nonlinear Structural Identification

Identification of a structural model is easiest when a mathematical model of the nonlinearity accurately characterizes the real structure. In such cases, the identification task is to determine the values of the undetermined

parameters in the model, for example, the value of the coefficient of the cubic stiffness term in a single degree of freedom model. Unfortunately, there are many nonlinearities which are difficult to characterize. Therefore, identification is most commonly done in a *nonparametric* sense. That is, no attempt is made to identify any physical parameter; rather, the objective is to come up with a model which represents the response of a structure as well as possible without respect to any physical parameter.

Nonparametric Identification. One popular nonparametric method is to represent a nonlinear system by a set of functional series, for example a Volterra series [11]. This "functional series" method is an extension of the frequency response function (or transfer function) representation used in linear analyses. A Volterra series involves a linear representation which is augmented to include higher (than first) order frequency response functions, and is extremely computationally intensive. A similar approach characterizes the response of a nonlinear system in terms of response levels at a small number of discrete frequencies. This "describing function" method assigns response levels at a given "driving frequency", but often includes expected harmonics and subharmonics as well. [As such, a set of describing functions is a subset of a linear system's frequency response function.]

Another method is (an early) version of the "force-state mapping" technique, due to Masri and Caughey [12]. In this method, the force "input" into a structure and the full state of the structure (displacement and velocity) are used to determine the restoring forces in a structural element. These restoring forces are, in general, functions of the state (displacement and velocity) and the *history* of the state. Masri and Caughey characterized this restoring force in terms of Chebychev polynomials, which gives no direct indication of what type of nonlinearity is involved.

Parametric Identification. The most desirable form of identification technique for a structural analyst is one which is *parametric*, since knowledge of the values of physical parameters allows for better subsequent analysis and design. This is particularly true if a structure is being modified or a new structure is being designed, and a component known to be nonlinear, say a joint or attachment concept, is being considered for use in the structure.

The actual task of determining parameters of a system is dependent on the type of model being used. The most straightforward model of any structure is a set of differential equations. These may be readily available for

simple systems or be formulated in discrete form in a finite element or finite difference approach for a broader class of more complex structures. If the differential equations are known either in continuum or discrete form, the highest fidelity solutions must be determined by a numerical, time-marching solution. That is, if all of the subtleties associated with response regimes which include limit cycles or chaos are to be explored, only a direct solution will be generally satisfactory [13]. Lower fidelity solutions--which may be "good enough" in some cases--are based on simplifying assumptions. These include the method of harmonic balance [14] and the method of multiple scales (a perturbation technique) [15]. In this research effort, the harmonic balance simplification was used to solve for response to simple sinusoidal excitation [10].

Another possible approach, rather than attacking a set of nonlinear differential equations, or an approximation thereto, is to linearize the structural model. A model is first formulated with an undetermined set of physical parameters, each set for a given range of excitation level. Then, the parameter values are chosen to provide the best response prediction for a range of excitation levels. Although this may be an acceptable approach for a single degree of freedom system responding to uniform amplitude excitation, it will not work at all for excitation levels at which chaotic behavior results.

However, the only parametric identification of a nonlinear system, in terms of a nonlinear model, appears to have been accomplished using a modification of the force-state mapping (FSM) technique mentioned earlier. This technique is based on the rewriting of the equation of motion of a system,

$$m\ddot{x} + c(x,\dot{x})\dot{x} + k(x,\dot{x})x = p(t) \tag{10}$$

as :

$$f(x,\dot{x}) = p(t) - m\ddot{x} \tag{11}$$

where $p(t)$ is the forcing function, the mass is m and f is the force transmitted through the spring and damper. If the forcing function, mass and acceleration are measured, the transmitted force is known and can, in theory, be described as an analytic function of the state. Crawley and Aubert [16] and Crawley and O'Donnel [17] have assumed that the restoring force in a single degree of freedom system can be represented by simple functions which represent various types of nonlinearity. For example, for the case of a spring with both linear and cubic stiffness

as well as both viscous and dry friction (Coulomb damping), the transmitted force would be:

$$f(x,\dot{x}) = k_1 x + k_3 x^3 + c\dot{x} + F_f \, sign(\dot{x}) \tag{12}$$

where k_1 and k_3 are the linear and cubic stiffnesses, respectively, c is the viscous damping coefficient and F_f is the friction force. A least squares approach is then taken to fit measured data to the model. They have shown that linear and cubic stiffness coefficients as well as dry friction and viscous damping coefficients--all physical parameters--have been successfully derived from data carefully collected for a single degree of freedom pinned joint.

Finally, an excellent paper by Al-Hadid and Wright [18] discusses the extension of the FSM approach to include multiple degree of freedom systems. Unfortunately, this involves identification of *modal* parameters, which is not as useful to the structural analyst. However, they do formulate a method to locate regions in a model which are nonlinear.

Recommendations for further study

Three recommendations are made with regard to the analytical capabilities to support sounding rocket launch. The first is to develop a methodology to relate wind shear probabilities at launch to probabilities launch failure. The second is to characterize the nonlinear, interstage joints using a force-state mapping investigation. The third is to incorporate the nonlinearity discovered into the GEM flexible response option.

Three recommendations are made with regard to the basic science of nonlinear structural response. The first is to investigate the use of Bayesian estimation (a weighted least squares) techniques to estimate physical parameter values in the force-state mapping technique. The second is to investigate the use of Poincare section mapping from forced vibration testing to identify types of nonlinearity in a joint. The third is to experimentally characterize common interstage and inter-segment joints for dynamic analysis using the force-state mapping technique.

These recommendations are discussed in greater detail in a proposal to be submitted to Phillips Laboratory and the Air Force Office of Scientific Research.

Postscript

During the summer research period, a number of computer codes, mostly from NASA/Wallops Island, were imported and reviewed by the author. A companion report, "Computer Codes Supporting Sounding Rocket Launch System Design" [8] is being prepared for Phillips Laboratory. The report summarizes the software capabilities and describes the installation at the laboratory.

References

1. Lanzi, R. James, "Structural Analysis of SPEAR III/ Terrier Mk-70--Black Brant VC," NASA Goddard Space Flight Center, Wallops Flight Facility, Wallops Island, VA, 10 Feb 93, internal memorandum.

2. Matthews, Katherine P., "Preflight Analysis, SPEAR III Terrier Mk-70--Black Brant VC," Conatec, Inc., PO Box 456, Glenn Dale, MD 20769, R-92-59/T310, Revision A, 12 January 1993.

3. Lewak, John N., "Static Stability and Dynamic Analyses to Support the Launch of the Spear III Vehicle" (ER 93835/A), Bristol Aerospace Limited, Winnipeg, Canada, 21 April 1993.

4. Otis, John M., "Spear III Quick-look Flight Report, Launched March 15, 1993", Space Experiment Directorate, Aerospace Engineering Division, Phillips Laboratory, Hanscom AFB, MA, 10 April 1993.

5. Ewing, M.S., Chrostowski, J.D., and Hasselman, T.K., "Comparison of Popular Finite Element Model Updating Techniques", Proceedings, AIAA Dynamics Specialists Conference, Dallas, Texas, April 1992, pp 550-555.

6. Winchell, W.M., "Launch Winds Operations for the Delta Launch Vehicle", Proceedings, AIAA 30th Aerospace Sciences Meeting, 6-9 January 1992, Reno, NV (also, AIAA-92-0717).

7. Booth, Andrew W., "A Generalized Analysis of Aeroelastic Bending of a Slender Body Non-Uniform Flight Vehicle", Fairchild Hiller Corporation, Hyattsville, Maryland, NASA-Goddard Space Flight Center, Report No. ER-106, April 1967.

8. Ewing, M.S., "Computer Codes Supporting Sounding Rocket Launch System Design", report for PL/SXAI, Hanscom AFB, MA, to appear.

9. Lanzi, J., "Flexible Body Simulation Model (FLEXSM)", internal NASA Wallops Island memorandum, 4 January 1990.

10. Bard, D., "Forced Vibration of a Two-Span Free-Free Beam Joined by a Nonlinear Rotational Spring", Final Report, AFOSR Graduate Student Research Program, August 1993.

11. Gifford, S. J., and Tomlinson, G. R., "Recent Advances in the Application of Functional Series to Non-Linear Structures", Journal of Sound and Vibration, vol 135, no 2, pp 289-317.

12. Masri, S.F., and Caughey, T.K., "A Nonparametric Identification Technique for Nonlinear Dynamic Problems", ASME Journal of Applied Mechanics, Vol 46, June 1979, pp 433-447.

13. Abhyankar, N. S., Hall, E. K., and Hanagud, S. V., "Chaotic Vibrations of Beams: Numerical Solution of Partial Differential Equations", Journal of Applied Mechanics, Vol 60, March 1993, pp 167-174.

14. Porter, B., and Billett, R. A., "Harmonic and Sub-Harmonic Vibration of a Continuous System Having Non-linear Constraint", International Journal of Mechanical Sciences, Vol 7, 1965, pp 431-439.

15. Eslami, H., and Kandil, O. A., "Nonlinear Forced Vibration of Orthotropic Rectangular Plates Using the Method of Multiple Scales", AIAA Journal, Vol 27, No 7, July 1989, pp 955-960.

16. Crawley, E. F., and Aubert, A. C., "Identification of Nonlinear Structural Elements by Force-State Mapping", AIAA Journal, Vol 24, No 1, January 1986, pp 155-162.

17. Crawley, E. F., and O'Donnel, K. J., "Force-State Mapping Identification of Nonlinear Joints", AIAA Journal, Vol 25, July 1987, pp 1003-1010.

18. Al-Hadid, M. A., and Wright, J. R., Developments in the Force-State Mapping Technique for Non-Linear Systems and the Extension to the Location of Non-linear Elements in a Lumped-Parameter System", Mechanical Systems and Signal Processing, Vol 3, No 3, 1989, pp 269-290.

ANALYSIS OF A DETECTOR TEST STATION AND THE PERFORMANCE OF A VISIBLE LIGHT PHOTON COUNTER

Brian P. Beecken
Associate Professor
Department of Physics

Bethel College
3900 Bethel Drive
St. Paul, MN 55112

Final Report for:
Summer Faculty Research Program
Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D. C.

September 1993

ANALYSIS OF A DETECTOR TEST STATION AND THE PERFORMANCE OF A VISIBLE LIGHT PHOTON COUNTER

Brian P. Beecken
Associate Professor
Department of Physics
Bethel College

Abstract

There were two unrelated areas of research performed at Phillips Laboratory and summarized in this report. Accordingly, this paper is divided into two parts to reflect the different research projects.

The first part describes the analysis and evaluation of a test station designed to collect performance data for CCD's and other focal plane array detectors. The station was purchased from a vendor but had never been used. An attempt was made to make the test station operational. Unfortunately, whereas the hardware appears to be useful, the software that runs the system was determined to be essentially nonfunctional. A detailed analysis of the deficiencies is given along with suggestions for bringing the station up to operational levels.

The second area of research was testing the performance of visible light photon counters (VLPC) which have recently been developed by Rockwell International. A special purpose test station was designed and constructed. The resulting data collected shows VLPC's to be intriguing devices with possible difficulties related to quantum efficiency and response time. More data is necessary in these areas.

PRELIMINARY ANALYSIS

OF THE PULSE INSTRUMENTS TEST STATION AT PL/VTRP

1. INTRODUCTION

PL/VTRP recently purchased from Pulse Instruments a 4000 Series Low Noise Focal Plane Array Test System which is designed to test Focal Plane Arrays, Charge Coupled Devices, and IR Detectors. Already having established a strong capability in the testing of single element detectors, VTRP wishes to get into the business of testing CCD's and was attracted to the PI system by its flexibility, cost, and performance. When I arrived on the 1993 AFOSR Summer Faculty Research Program, the system had sat essentially idle for a few months because of a shortage of laboratory personnel. Having had experience with the early stages in the development of a somewhat similar test station at Texas Instruments, I was given the task of making the system operational and documenting the various nuances of its operation.

As it turns out, the system's software has major shortcomings. There were also some hardware problems, but these were manageable and can be worked around. (The hardware is discussed in a separate report.) It is expected that eventually someone else will attempt to use the PI system. The present report is envisioned as guiding the reader through the PI system in a way that he or she may avoid many of the pitfalls that ensnared me. At the very least, this report should prove to be a useful documentation of the various problems and idiosyncrasies of the PI test system.

2. SOFTWARE

2.1. General Introduction To Pidas 1.0

The manual for the Pidas software is almost worthless. The software itself has numerous bugs and many quirks which make it very difficult to use. Pulse Instruments has promised updates for the software, but in a conversation with them in July, they stated it would be several weeks before any revisions were available. This section is an attempt to be a user's guide to the existing software. Hopefully it will prove to be more useful than the manual supplied by PI.

When Pidas is first loaded, the screen will say at the top: Pulse Instruments Data Acquisition System. There are six pull-down menus available: Setup, Run, Process,

Display, Diagnose, and Misc. The last two are inactive. Apparently, one must have data before "Process" and "Display" are of any use. The logical starting place is the "Setup" menu. Here various parameters and test conditions are input to the system. Once this is done, the user chooses "Run" and starts acquiring data. I had so much difficulty getting a realistic "Setup" to "Run," that I never got a chance to actually take any data (either simulated or real). Consequently, the remainder of this report will be devoted to the precise steps needed to get a test set up correctly. Hopefully this will allow my successor to go right to work on the data acquisition part of the software.

Choosing "Setup" gives the user six active options and three inactive options on the menu. The last option is "Analysis setup." It is my belief that you need data to utilize this because it apparently deals with such issues as crosstalk, NEP, noise, rms noise, and signal. I have not attempted to pursue this option. The other five options are Test Configuration file, Timing generator, Clock driver, Bias supplies, and Device mnemonic table. Each of the next five subsections will be devoted to one of these options.

2.2. Setup

2.2.1. Test Configuration File

Choosing this option brings up the heading "Test plan definition." This process is referred to by both names. There are now four choices: New, Open, Exit, and About Editing forms. The last appears to be active, but in fact has no information in it. "Open" allows the user to choose a preexisting test configuration file. If you are just starting up, you should choose "New." Then there are two choices available: Init and Acquisition. I do not understand what "Init" is supposed to do. It is an abbreviated form of the "Acquisition" option, but it will not allow you to run the system. It is best to avoid it.

Once you choose "Acquisition," the software kicks into a DOS program. You are asked to name the "Data Acquisition Form." (Note, we now have a third name for the same thing.) Choose a name descriptive for your test (eight characters or less!), such as TEST.FRM. You must include the extension .FRM or you will not be able to access this file later! There now appears a long list of information required for the upcoming test. Most of the items have a default value which seems to work. If you want to change a default value, some items require you to type over the number, whereas other items can only be changed by hitting the right arrow key on your keyboard which toggles you through the various possible values.

When first starting out, I recommend retaining most of the default values. If TEST was chosen as the name for the "Data Acquisition Form," then here are the items that should be changed or need to have information supplied in order for the system to run:

Item	Value	Certainty of Response
Non uniformity calibration:	toggle to "none"	a reasonable choice
Device mnemonic table:	TEST.MNE	definite
Timing generator ID:	PI5800	your choice
timing pattern:	TEST.WFM	definite
Clock driver ID:	PI4000	your choice
clock pattern:	TEST.CLK	definite
Bias supplies ID:	PI4000	your choice
bias pattern:	TEST.BIA	definite
A/D Channel ID:	PI4005	your choice
CDS clamp delay	-1	your choice?
CDS clamp pulse width	0.0	your choice?
Convertion (sic) delay control:	50	your choice?
Convertion (sic) pulse width:	*100*	*your choice?*
Timing Generator program ID:	PI5800	your choice
program #:	1	must agree with timing file
output filname:	It will now read: "TEST.FRM.RAW,"but this will not work. You must delete the .FRM	definite

The choices you made, especially the ID's, must be remembered because you will need them when you answer other setup questions. I strongly suggest writing down these responses because you cannot view this file while working on another file and the exact same ID's must be used.

To leave this subroutine you must press Esc.

2.2.2. Timing Generator

Choosing the Timing Generator in Setup calls up an independent program written for DOS named "PI-Waveform Editor 1.1." This program is described fairly well in the

manual entitled "40200 System Software for the 4000 Series Low Noise Focal Plane Array Test System." (It should be noted that although both the manual and the software indicate that a mouse can be used with this software, the mouse does not appear to work.) The purpose of this software is to program the PI-5800 in order to put out the desired timing pulses. After familiarizing myself with the PI-5800, I found the use of this portion of software relatively intuitive and only occasionally needed to refer to the manual.

The first screen that comes up has a column at the far left labeled "Name." The column consists of a series of pairs: A 1, B 1, A 2, B 2, A 3, B 3, ... Each pair indicates a different channel. The number refers to the slot for a driver card in the PI-4002 Instrument Mainframe, and the letter indicates whether it is the top or bottom channel of the card. As the cursor is moved around the pattern area, its location is indicated at the top of the screen. Unfortunately, the A's are replaced by 1's and the B's by 2's. Thus, for example, the row B 2 is listed as 2 2. A pattern can be written directly by placing the cursor where desired and typing "H" for high and "L" for low. When writing a pattern, however, I strongly recommend using the F5 Edit function. This time-saving feature allows the user to write patterns with a mathematical formula.

There are several important idiosyncrasies worth noting. First, this part of Pidas was originally written as a stand-alone program before it was integrated into Pidas. You can still control the PI-5800 directly at this level by setting up your timing pattern and then hitting F4 Device. But, if Pidas has already been used to set up the PI-5800 (during a "run"), then the system will lock up (at checking GPIB configuration) and you will have to reboot the computer. This occurrence can be most annoying if you accidentally hit F4 when reaching for another key such as the F5 Edit. You will lose all your work! In order to run the PI-5800 directly from this part of the software (after Pidas has been used for a "run") you must first gain local control of the PI-5800. This task is accomplished by pressing HOME and CE simultaneously on the PI-5800 and then hitting the softkey, GO TO LOCAL.

A second very important idiosyncrasy is the naming of channels. This program allows you to rename rows (channels) that originally have cryptic default identifiers A 1, B 1, A 2, etc. However, once a name is given to a channel, then the Device Mnemonic Table file must list its Type as "PI-5800 channel" and its Address as "GPIB:1." Even more important, you cannot use the new name in the Clock Driver file. Since this is where the voltage levels of the pulse drivers are controlled, it is no longer possible to use this channel for anything but a simple TTL signal taken directly from the PI-5800. In other words, you cannot use pulse drivers for named channels!

It is my understanding that there are three very important channels for acquiring data regardless of the array being tested. The PI-4005 Acquisition Mainframe needs to have a Frame, Line, and Pixel sync. If it does not get these signals, it does not put the incoming data in succeeding bins. A "run" is not completed until the software gets a number of data bins equal to the number of pixels the array has (as specified in the Test Configuration file). These signals should be simple TTL and can be obtained directly from the PI-5800. They do not have to go through pulse drivers. Consequently, it would be appropriate to name three lines in the Timing Generator file, those corresponding to Frame, Line, and Pixel. It is hard for me to see why one would ever name more than these three channels in the timing file.

There is one more crucial thing. Apparently there is some limit to the length of a timing pattern. I tried two subpatterns of 128 bits each and Pidas crashed, giving me a system error. However, the same timing pattern can be made with three subpatterns totaling about 100 bits and it runs fine. This limitation is most puzzling because the two subpatterns of 128 bits runs just fine when it is programmed directly on the PI-5800. In fact, it runs just fine directly from this portion of the software. It is only when Pidas calls up such a timing file that a problem occurs. I found this inconsistency most difficult to track down. Because of the system error, I was receiving no error message, and initially I was quite confident that it could not be the timing file because it ran just fine separately.

2.2.5. Device Mnemonic Table

The Device Mnemonic Table tells the software the location of all the components of the system which are being used for the test. After choosing this option in the Setup menu, a list of all existing files with a .MNE extension is given. One of these files may be used or a new one created. When the file is displayed, there are ten options listed, but only the File (exit and save) and Add options are activated. When Add is chosen, the signal mnemonic name must be provided. After choosing a name, Delete, Name, and Type become activated. The first is obvious, the second allows you to change the highlighted signal mnemonic name, and the third forces you to choose a type of device from a long list. Once the device type is chosen, then the remaining appropriate options (GPIB, Mainframe, Slot, Channel, and VME) become activated.

Every test must have at least the following lines in the Device Mnemonic Table:

	Name	Type	Other Options
1)	PI4000	PI-4000 System	GPIB: 2

2)	PI5800	PI-5800 System	GPIB: 1
3)	PI4005	VME connected device	VME adr: 0000
4)	Clock Driver file names	40460 20V Driver	GPIB: 2 MF: 1 Slot: ? Chan: ?
5)	Bias Supply file names	40750 20V LN Bias	GPIB: 2 MF: 3 Slot: ? Chan: ?
6)	Channels named in the timing pattern	PI-5800 Channel	GPIB: 1

Lines 1 - 3 of the above table pick up names that were used as part of the example in Subsection 2.2.1. Obviously GPIB refers to the address of the device. Everything but the timing is controlled through the PI-4001 Control Mainframe which has a GPIB address of 2. "MF" refers to the mainframe number that is set by dip switches on the back of the PI-4002 Instrument Mainframes. At VTRP, all the clock drivers are in mainframe 1 and all the bias supplies are in mainframe 3. "Slot" refers to the card's position in the mainframe (for the clock drivers this is the same as 1, 2, 3, ... 8 in the timing pattern) and "Chan" refers to the channel of the card (for the clock drivers this is top or bottom, A or B, 1 or 2).

2.3. Run

Once the setup is complete, the test is ready to be run. Thus, "Run" is chosen in the main Pidas menu. A window will open which requests several identifying bits of information. The only question that needs to be answered is the name of the "Test Plan" to be run. The Test Plan is the file that was formed in the Test Configuration file option under Setup. It must have a .FRM extension. A list of such files in the Pidas directory is provided. Files in any other directory are not accessible. After choosing the test plan file, the operator hits the "Run" button. If the test has been setup correctly, after a few prompts the screen will display the message "data being acquired." If you want to stop the data acquisition, hit ENTER.

It is likely that an error will occur before data acquisition commences. The Pidas error messages are cryptic at best. Recently, I got Pulse Instruments to fax me a very terse one page explanation of some of the error messages. Before that, the error and warning message numbers were completely indecipherable. Application errors are still extremely difficult to track down.

Several observations about errors are useful here. A warning and an error are equally bad. Either one will stop a run cold. After receiving an error and correcting it, in

all likelihood a succeeding run will turn up the same error. This unpleasant event will occur even if the source of the error has been eliminated. Such a phenomenon will wreak havoc with trouble-shooting. The proper course of action is to close Pidas and reopen it after correcting any error. On occasion, I have even observed repeated runs stacking the error messages on top of each other. If the error message "Problem with sending bias supply file: 1" is received, reboot Pidas. This action almost always corrects this error.

The button "Run Again" apparently gives an exact repetition of the previous test. I suspect that the timing pattern and voltage levels are not even resubmitted to the mainframes. If the previous test had errors, Run Again can bypass them and give the appearance of a successful run.

2.4. General Problems with Pidas 1.0

It is my opinion that Pidas 1.0 is essentially a beta release of what is really needed to run the Pulse Instruments test station. I make this assertion despite having only used the setup portion of the software and not having attempted to actually use the more complicated data acquisition part. There are many justifications for this accusation. Some of them I have woven into the above user's guide; the rest are listed below.

One very irritating feature of Pidas is the lack of a working directory. All of the system files are in the \PIDAS directory, but there does not seem to be any way to place the data and test files in a separate directory. Each test requires establishing at least six additional files. Obviously managing files should be a priority. Besides this poor design, there are even files which contain notes between the Pidas designers mixed in with the system files.

Pidas is not really a Windows program. When creating Test Configuration files and Timing Generator files, there is no support for a mouse. Far worse than that is the inability of Pidas to display multiple files. It is most inconvenient to be making a new file and not have access to an old file for reference. Also, there is no way to compare two files to see how the tests differed.

Support for the user of Pidas is nonexistent. Not only is the manual hopeless, but there is no on-line user help. In fact, it is almost impossible to determine what went wrong when an error occurs. The messages are very cryptic and do not appear when Pidas causes an application error (a common event). Worse yet, even after an error is corrected, the same error message will continue to appear until Pidas is rebooted.

There are many small things which indicate the haste with which Pidas was written. The numerous spelling errors on menus and headings is a trivial example. A more frustrating example occurs when saving a file under a different name: the window's heading does not update. In fact, existing Test Configuration files cannot even be altered and then saved under a different name. This is not a program that was written with great care, nor was it ready to be released.

4. CONCLUSION

I believe the drive electronics of the Pulse Instruments test station is sound. I have had some experience with similar pulse drivers, timing generators, and bias supplies, and I do not think they should be an issue. It is clear, however, that the software that runs this part of the test station is significantly less than satisfactory. I do not know anything about the data acquisition part of the test station, hardware or software, because I did not get far enough to use it.

So what should be done next to the test station? It took me several weeks to sort through the setup part of Pidas. Hopefully, this report on my findings will prevent such a delay for the next person who attempts to run Pidas. That person should be able to spend a few hours and have the system up and running. Once the system is running, however, then the part of Pidas that controls data acquisition and analysis must be used. Given the almost total lack of documentation, and considering the numerous deficiencies that I encountered using Pidas, I would expect that it will take someone several weeks to have the entire system operational.

Another possibility is to wait for the software upgrade that Pulse Instruments has promised. According to Pulse, this has been in the works for a few months and will be ready in "several weeks." At least that was what they told me on July 1.

The third alternative is to have new software written at VTRP. John Hubbs is currently writing software to control a somewhat similar test station for Ball Aerospace. John thinks that it might be possible to port the software to the PI station without an inordinate amount of trouble. Tom Roberts of Nichols Research Corporation spent some time working on Pidas with me. He knows a programmer who he thinks is capable of writing the software from scratch in less than a month.

Choosing to go with the existing version of Pidas may be the most expedient choice in the short run. I think it should be possible to use Pidas to take data. But, it will not be easy to use and will result in a lot of frustration when the system is used on a regular basis. Perhaps the easiest option is to wait for Pulse to come up with a revised

version of Pidas. However, given the incredible awkwardness of Pidas 1.0, and the improbability that a future version will correct all of its problems, this is not an appealing possibility. Significant advantages do exist for writing completely new software in-house. Having someone around who knows the software inside and out is a tremendous advantage when the inevitable adjustments need to be made. Nevertheless, writing this amount of software is not a trivial task and would require a major investment. No matter what path is pursued, it will probably take at least several weeks of work before this test station is ready to perform useful tests.

PRELIMINARY ANALYSIS

OF ROCKWELL'S VISIBLE LIGHT PHOTON COUNTER

1. INTRODUCTION

A detector called the Visible Light Photon Counter (VLPC) has recently been developed by Rockwell International. VLPC's are purported to be the "most sensitive single photon counting detector in the world." The VLPC is essentially a variation on the slightly older Solid State Photomultiplier (SSPM) which is sensitive primarily to infrared radiation. At Phillips Laboratory, both the VTRP branch and the LIMI branch are interested in this promising new technology.

VLPC's are arsenic-doped silicon devices consisting of both an impurity-band conduction (IBC) region and an intrinsic region. They operate at temperatures from about 4 to 7 K. Visible photons generate electron-hole pairs in the intrinsic region. The hole is swept towards the IBC region where there is a large electric field generated by the bias applied to the device. It is in the IBC region that the hole develops enough energy so that when it collides with bound electrons they are ejected from their impurity band. The hole has sufficient energy to continue progressing through the layer, leaving behind a trail of free electrons. Each electron in turn accelerates in the opposite direction, colliding with and freeing additional electrons as it goes. When the positive space charge left behind becomes large enough, there is a local collapse of the applied field and some of the freed electrons are recombined. Thus, the avalanche is quenched. Because the avalanches are self-quenching, they tend to have well-defined sizes which cause little dispersion in gain. The result is a high-gain, low-noise, and high-quantum efficiency visible photon counter.

The primary advantage of the VLPC over photomultiplier tubes is their high quantum efficiency. Rockwell has demonstrated greater than 60% and claims that VLPC's may be optimized to over 90% for particular wavelengths. A couple of concerns for practical applications is the limited dynamic range and response time. Rockwell claims response times on the order of nanoseconds, yet in their literature they mention the existence of a "slow component of response" that is greater than a microsecond. A response that slow could severely restrict the range of applications possible for the VLPC's.

Both LIMI and VTRP wanted to verify the vendor's claims, explore other aspects of the VLPC's performance, and in general become more familiar with this new technology. Several test objectives were determined at the outset. The first objective was to determine the noise levels inherent in the device. What is the dark pulse rate and how large is the ubiquitous rms background noise? Second, what is the detector's dynamic range and how linear is its response over this range? Third, what is the VLPC's responsivity and is the quantum efficiency as high as claimed? Fourth, what kind of response times are realistic? And fifth, how do these data change with applied bias and device temperature? To meet all of these objectives obviously would require a major investment of time and effort.

2. APPARATUS

Rockwell agreed to a thirty day loan of a VLPC cassette and a cryogenic dewar. On July 23, 1993, Dick McCutcheon and Mike Petroff of Rockwell delivered a VLPC cassette to Phillips Lab, set it up, and demonstrated its use. The cassette contained 32 individual detectors, each at the end of a short length of optical fiber. The cassette was cooled simply by insertion into the liquid helium filled dewar. The fibers extended out of the dewar, could be coupled to other fibers, and then configured as desired. The top part of the cassette was at room temperature and included a QPA02 amplifier for the 32 channels. Unfortunately, the cassette and dewar arrived at Phillips lab only two weeks before the end of my research there. The resulting time constraints drastically limited the quantity of data that could be obtained and ensured that all of the above objectives could not be met.

In order to facilitate testing the VLPC's, we constructed a test box out of opaque, black plexiglass. The box was about 37 cm long and 10 cm on each side. All joints were constructed in such a way as to prevent light leakage. On one end of the box a laser diode

and its drive electronics was mounted. The other end of the box, which was a movable plate, had five optical fibers from the VLPC mounted in it. The plate and fibers could be positioned rather precisely with the use of two micrometer-style screws. Another plate was placed near the center of the box. In the center of this plate, variously sized apertures could be mounted. Small apertures were used to limit the amount of radiation incident on the end of the optical fibers. To further reduce the irradiation, neutral density filters could be slid into slots in the box between the aperture and the plate with the fibers. This arrangement allowed us to vary the intensity of the radiation over a fairly large range, to get optical alignment, and to test the performance of five detector elements almost simultaneously (by moving the position of the end plate).

A Hewlett-Packard 8082A pulse generator was used to set the timing for the laser diode's drive circuit. This circuit was custom designed and built by Dave Holmes of PL/LIMI to give minimum rise and fall times. The laser diode was a TOLD9412(SDA) made by Toshiba and operating at 650 nm with an optical laser power output between 0.3 and 3.0 mW. The diode was mounted on a large aluminum cooling plate which was continually monitored for temperature variations that could affect output power stability. Included in the laser diode package was a photodiode mounted so as to provide a monitor for the laser's optical output. The laser diode output was calibrated with a Newport 835 Optical Power Meter. Thus, the optical output power was known as a function of drive voltage and temperature. The VLPC's response was observed with a LeCroy 9450 Dual 350 MHz oscilloscope. This digital scope is capable of taking a running average of any particular number of traces. When attempting to detect small numbers of photons, we frequently utilized this feature to provide average responses and thus reduce the uncertainty in response caused by the inherent photon fluctuations.

During Rockwell's demonstration, we observed that although the QPA02 amplifier is a differential amplifier, one side of its output was being grounded when it was input into the oscilloscope. Before beginning our test, we remedied this situation by putting 50 ohms between each output and ground and then putting the signals into different channels of the oscilloscope. The scope was then set to display the difference between the channels. The resulting rejection of the common mode significantly reduced the noise, and the use of 100 ohms to terminate the ribbon cable output of the QPA02 amplifier was a better impedance match, thus increasing the signal. Shielding was also provided for the two biases required by the cassette. These modifications provided a much better signal to noise ratio than the one that had been demonstrated for us.

3. CALCULATIONS

The first thing that needs to be determined is the photon flux emitted by the laser diode. Before the laser diode was mounted in the test box, its output in continuous operation was measured with the power meter. Good correlation was observed with the data sheets provided by Toshiba. After mounting in the test box, the optical output as measured by the power meter dropped by almost a factor of two. Presumably this effect was due to reflections off of the wall the diode was mounted in, thus reducing the amount of laser light that actually got into the test box.

When the diode was pulsed, we found that the internal photodiode only responded at about 80% of the level for continuous operation. It is not clear whether this reduction was caused by a slow response on the part of the photodiode or the laser diode. The power meter, of course, will not respond accurately to short pulses. An independent photodiode gave a response slightly greater than 80%. Thus, it seems to be a safe estimate that the optical power emitted during a pulse was between 80% and 100% of that during continuous operation. We used pulses as long as microseconds and as short as 20 nsec. Any shorter and the photodiode's response was spiked rather than squarish, which would eliminate our ability to determine a narrow range of possible optical power output.

The power meter was also placed behind a 49 μm diameter aperture on the center plate of the test box. When the laser diode was operated continuously under our standard conditions, the meter gave a value of 3.0 nW. Because the detector's fiber was 20 cm from the aperture, diffraction is an important effect. The resulting Airy disk is about 6.4 mm in diameter. Since the optical fibers were measured to be about 0.78 mm in diameter, it was simple to position a fiber within the brightest part of the Airy disk using the micrometer screw adjustments. Centering was confirmed by getting the maximum response from the VLPC.

Under these circumstances, the Fraunhofer approximation is valid and it is relatively simple to calculate reasonably precisely the flux incident on the optical fiber. If we let $L(r)$ represent the fraction of total power incident on a disk of radius r, then (Born and Wolf, 1965, p.398)

$$L(r) = 1 - J_0(ka\sin(\theta))^2 - J_1(ka\sin(\theta))^2.$$

Here J_0 and J_1 are Bessel functions, θ is the half angle subtended by the optical fiber, a is the radius of the aperture, and k is the wave number. For 3.0 nW coming

through the aperture, 0.15 nW or about 500 photons per μsec will be incident on the fiber which has radius r = 0.39 mm.

4. DATA

All the data reported in this paper were taken during the six day period between our first cooldown of the VLPC and the last day of my stay at Phillips Laboratory. Because of this very brief time available for testing, it seems appropriate to view all the data reported here as preliminary. The data should be verified and extended by further tests.

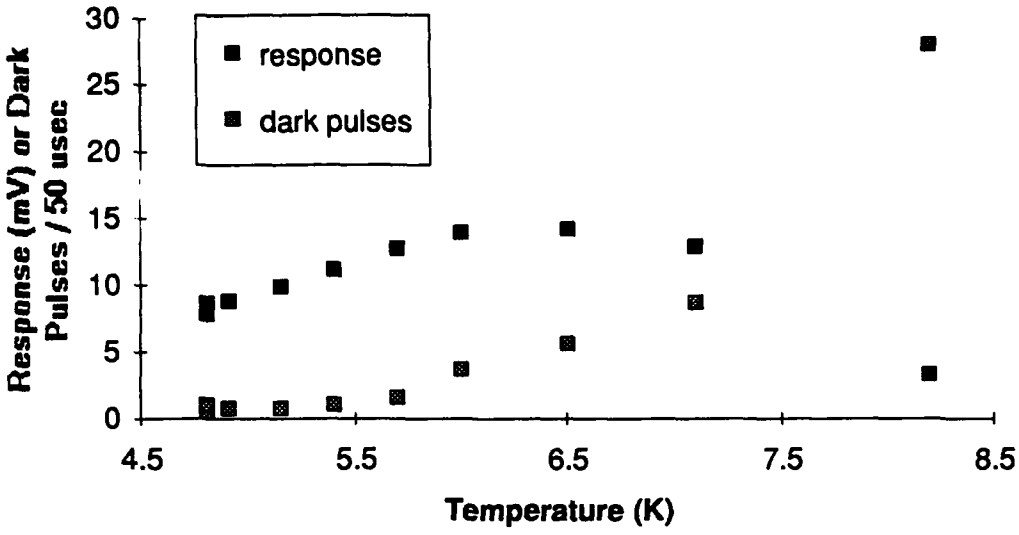
One of the first tasks was to determine the optimal operating conditions for the VLPC. The scope was used to find the rms average of the noise as we varied both the VLPC bias and the amplifier bias separately. The response to a signal was also observed as the biases were varied. We found that up to an amplifier bias of 2.3 V the signal response increased more rapidly than the noise. Further increases in amplifier bias did not improve the signal-to-noise ratio. A similar effect was observed for the VLPC bias: going below -8.0 V did not seem to help. To ensure that there was sufficient bias to be operating under the best conditions possible, we chose biases slightly greater than these threshold values. For the remainder of the testing we used an amplifier bias of 2.4 V and a VLPC bias of -8.25 V.

As stated earlier, there were 32 different VLPC elements on our cassette. Only 16 had their pinouts identified by Rockwell, and of these, five had been declared dead. We did some quick testing on five of the remaining eleven elements in order to check for uniformity of performance. Our primary criteria for choosing those five elements was the quality of polishing that their optical fibers had received. Unfortunately, all of the elements we chose happened to come from the same chip. We checked response, dark pulse rate, and rms noise. The performance of the five elements did not vary by more than 10%.

Another condition for optimal operation of a VLPC is its temperature. Data on responsivity and the dark pulse rate were taken over the available temperature range. These data are displayed in Figure 1. As expected, the dark pulses, which are believed to be thermally generated events, are observed to increase with temperature. The apparent maximum in response is a little harder to explain. An interesting and useful way of analyzing these data is to take the response to dark pulse ratio and plot it as a function of temperature (Fig. 2). The two disparate values at 4.8 K give some indication of the

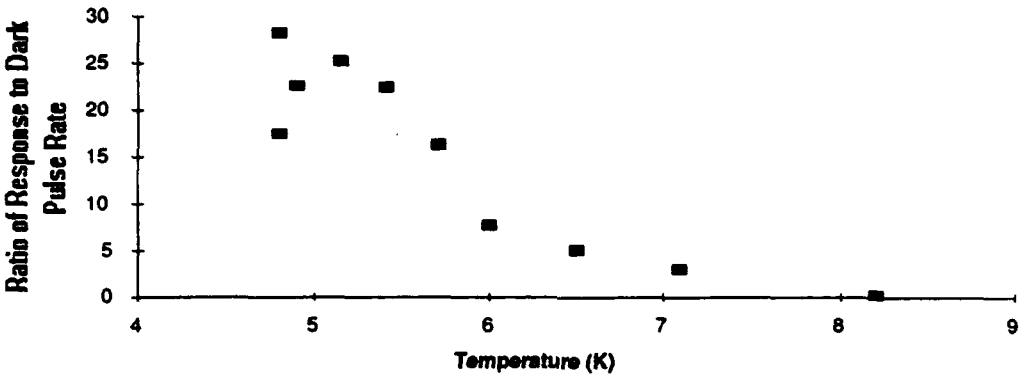
uncertainty of the data points. We did not have sufficient time to study the uncertainties of our data in a systematic fashion.

Fig. 1



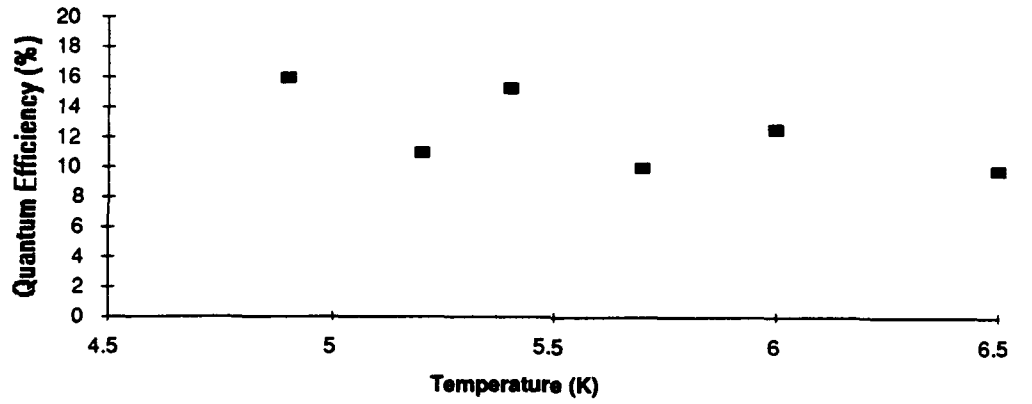
Of critical importance to the data presented here is the correct identification of dark pulses during the testing. Any scope trace of the VLPC output has a significant amount of rms noise, even when there is no optical input. The trick in determining the dark pulse rate is to distinguish between rms fluctuations and pulses. We found a number of pulses that were significantly larger than the average fluctuations and lasted up to 100 nsec, much longer than the other fluctuations. These pulses were consistently in only one direction, unlike rms fluctuations. The pulse rate was observed to increase with temperature. In fact, the rms noise, as measured by the scope, asymptotically approached a minimum value as the temperature was decreased. This trend seems to indicate strongly temperature dependent dark pulses on top of an rms background noise that is relatively independent of temperature (perhaps due to the amplifier). Confirmation of this interpretation was provided by turning off the VLPC bias, while keeping the amplifier bias at the usual 2.3 V. The dark pulses were no longer observed, and the rms noise dropped to the level that was asymptotically approached as the temperature was reduced. Further support was gained by turning the VLPC bias back on and leaking very small amounts of light into the test box. The light leakage caused a definite increase in the pulse rate, which indicates that single photons cause similarly sized pulses.

Fig. 2



If we are correctly interpreting these pulses as single events—either photonic or thermal in origin—then it should be possible to directly calculate the VLPC's quantum efficiency. When an optical pulse is incident on the detector, it will generate an average output response in mV. The response voltage divided by the average single photon pulse height will yield the average number of detected photons. The quantum efficiency is simply the percentage of photons in the incident pulse that are detected. Since we know the photon flux fairly well (we will assume 80% of continuous), we are able to determine the quantum efficiency as a function of temperature (Fig. 3).

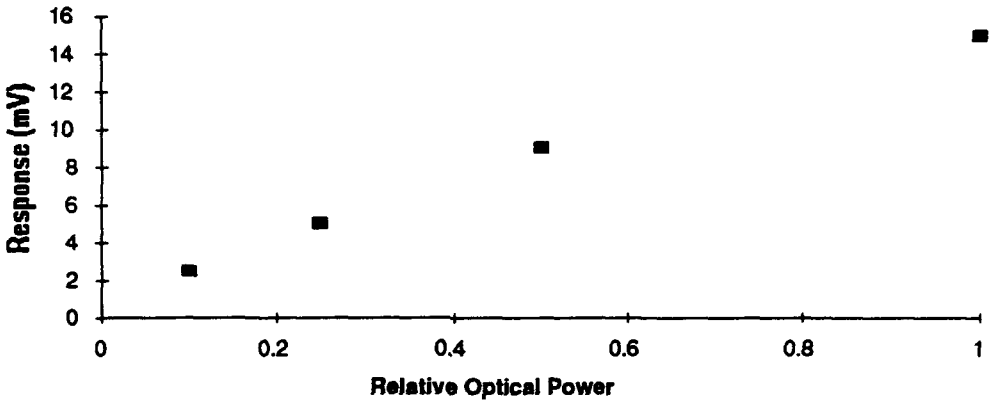
Fig. 3



The data do not indicate any significant temperature dependence. During discussions with Rockwell about this result, it was suggested that we did not examine a large enough temperature range. This may indeed be the explanation, but there was insufficient time to expand the data set. Of greater importance is the surprisingly low quantum efficiencies. Rockwell suggested part of the problem was an insufficiently

peaked VLPC bias. They also mentioned that the VLPC we were given was from an earlier batch (Histe 3) that had maximum quantum efficiencies around 50%. Another possible explanation is the rather poor connection between the VLPC's optical fibers and the fibers which attached to them and then ran to the test box. The fibers did not appear to be well polished—exhibiting many scratches. In addition, the connection was an unsophisticated butt joint. At one point optical gel was added, with no observable effect. Perhaps the best explanation for the low quantum efficiency is discussed at the end of this section.

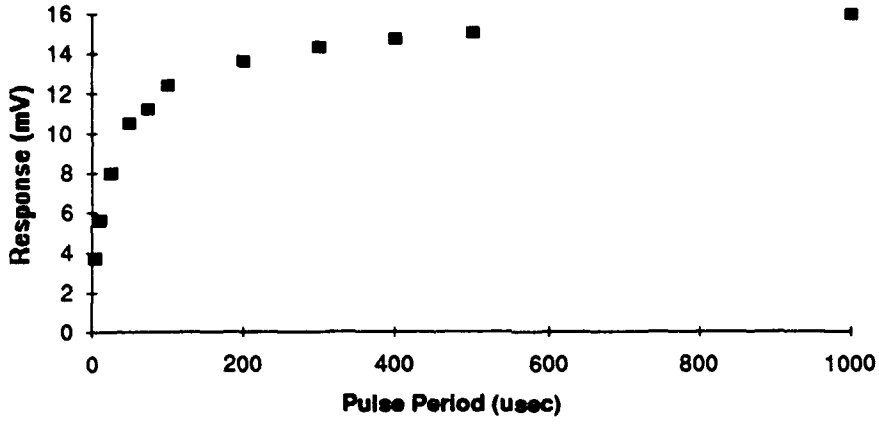
Fig. 4



A very quick check of the linearity of response was made. We did not attempt to cover the entire dynamic range. The results were reasonably linear over the limited range of investigation and are shown in Fig. 4.

Near the end of our investigation, we made a very important observation. The response of the VLPC to pulsed radiation depended strongly on the time between pulses when the period dropped below 500 μsec. The data are given in Fig. 5 for a pulse length of 2.0 μsec. This effect could be caused by two different factors. It could be a capacitative effect, because the VLPC's are capacitively coupled to the amplifier. Another possibility is the depletion of the photon absorption sites, which could cause a very slow component of response. Probably the best way to distinguish between these possible causes is to change the photon flux while holding the period and the pulse length constant. Once again, unfortunately, there was not enough time to perform this experiment

Fig. 5



Perhaps the most significant thing about this tardy observation is its possible explanation of the surprisingly low measured quantum efficiencies. That data set was taken with a period of only 100 μsec, which at the time seemed like more than enough to eliminate any slow response effects. However, from Fig. 5 it is apparent that there may have been enough of a reduction in response to explain the low quantum efficiencies. There is a crying need here for retaking the quantum efficiency data with a very long period between optical pulses.

5. CONCLUSION

Much has been accomplished, but much work remains to be done. With some effort, we were able to build a test setup that worked quite well for our purposes. We became familiar with the cooldown and operation of VLPC's. Preliminary data were obtained for responsivity and dark pulse rates as a function of temperature. We also discovered a serious duty cycle issue and attempted to measure quantum efficiency.

On the other hand, due to severe time restrictions, a significant portion of our objectives were not accomplished. The dynamic range of the detector was never investigated, and therefore, the linearity has been checked only over a part of its range. The quantum efficiency seems too low to be accurate, and it does not depend on temperature to the degree expected. We only made a start on looking into the response time, and it appears to be a significant issue because the duty cycle of the optical input pulses may have compromised our measurements of quantum efficiency. A careful and systematic study of how the VLPC's performance depends on bias was not done. Additional data should be taken in all areas to determine repeatability and uncertainty.

Particular attention needs to be paid to the response time or duty cycle issue. This issue's importance far exceeds the problems it may have caused us during the quantum efficiency measurements. If the input duty cycle really affects responsivity, then there are serious implications for possible VLPC applications. An effort must be made to determine whether this effect is due to capacitive coupling between the VLPC and the amplifier or a depletion of absorption sites within the device. In addition, the effect must be carefully measured so that its limitation on performance can be anticipated.

The other parameter that merits special attention is the quantum efficiency. It is, after all, primarily the purported high quantum efficiencies that make this device worth choosing over the well-known photomultiplier tubes. The quantum efficiency data set should be expanded over a broader temperature range and its dependence on bias must be found. An important factor here is the loss of photons at the butt joint between the two sets of optical fibers. This factor must be quantified or else the true device performance will always be uncertain.

The VLPC is an interesting device and holds great promise as an alternative to photomultiplier tubes. The primary issues that need to be addressed are a confirmation of the claimed high quantum efficiencies and the possibility of a very slow response component. Additional work is required to answer these questions.

ACKNOWLEDGEMENTS

Dave Holmes of PL/LIMI and I worked very closely together on the VLPC research. If this part of my paper had been prepared for a journal, we would have been co-authors. However, the AFOSR summer research program required the faculty member to write the report of his activities. I wish to thank Rockwell International for lending the cassette and dewar required to do this investigation. Rockwell's Dick McCutcheon and Mike Petroff were particularly helpful during their visit to Phillips Lab. Paul LeVan of PL/VTRP deserves much credit for suggesting the VLPC summer research project, coordinating it, and arranging for the loan from Rockwell. I also appreciate the efforts he made to make my summer's stay at Phillips Lab as comfortable and productive as possible. I am grateful for the good advice and counsel I received on the PI test station work from Tom Roberts of NRC and John Hubbs of Ball Aerospace.

CALIBRATION TECHNIQUES FOR BREMSSTRAHLUNG X-RAY SOURCES WITH END-POINT ENERGIES OF 60-160 KEV

Raymond D. Bellem
Associate Professor
Department of Electrical Engineering

Embry-Riddle Aeronautical University
3200 Willow Creek Rd.
Prescott, AZ 86301

Final Report for:
Summer Faculty Research Program
Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.
and
Embry-Riddle Aeronautical University
December 1993

CALIBRATION TECHNIQUES FOR BREMSSTRAHLUNG X-RAY SOURCES WITH END-POINT ENERGIES OF 60-160 KEV

Raymond. D. Bellem
Associate Professor
Department of Electrical Engineering
Embry-Riddle Aeronautical University

ABSTRACT

This work extends the calibration of a low energy (0-160 keV) bremsstrahlung X-ray source over a range of anode voltages from 50 to 160 kV. Analytical and experimental techniques have been developed which provide a high-level of confidence in the dosimetry used at these higher spectral energies. Analysis techniques using the PHOTCOEF and CEPXS electron-photon transport codes were accurate to within 30% when compared with experimental intensity measurements using a calibrated X-ray vacuum diode. PMOS FETs, used as total dose sensors, validated a scaling factor technique for computing dose rates at "high" X-ray end-point energies to PIN diode calibrated dose rates from spectra with end-point energies below 50keV. Using these techniques, the dosimetry over the entire energy range (20-160 keV) was accurate to within 10% when correlated to the ^{60}Co standard.

CALIBRATION TECHNIQUES FOR BREMSSTRAHLUNG X-RAY SOURCES WITH END-POINT ENERGIES OF 60-160 KEV

Raymond D. Bellem

I. INTRODUCTION

The Phillips Laboratory Low Energy X-ray Test Chamber (LEXR) is a key test facility in the Phillips simulated space radiation test complex. The low energy X-ray environment in LEXR provides ionizing dose rates in the range from 0.1-1000 Krad(Si)/min with spectral end-point energies between 20 and 160keV. This test facility was developed to perform research and evaluation of microelectronic and photonic devices, circuits, and subsystems designed for use in future satellite systems. LEXR is being used in conjunction with Cobalt-60 and Cesium-137 sources to do continuous dose-rate ionizing radiation over a broad range of energies and dose rates. The simulated space radiation test complex at the Phillips Laboratory is being designed to perform research and evaluation of electronic parts and assemblies in accordance with ASTM F-1467. A full analysis capability using electron-photon codes such as PHOTCOEF, CEPXS and the Integrated Tiger Series will be available to support the research projects.

The facility description and calibration techniques of LEXR for end-point spectral energies up to 50keV was reported in the Summer Faculty Research Program Final Report for 1992 and at the Hardened Electronics And Radiation Technology Conference last year and is being published in the next issue of the JRERE [1]. (Note - A copy of this paper is available from the authors or from Logicon/RDA)

The objective of the work described in this report is to extend the range of calibration of the PL X-ray facility over the full energy range of the X-ray source up to the end-point energy of 160 keV. Electron-photon transport codes are used to calculate radiation transport in the test chamber and to calculate energy deposition profiles in test objects for different radiation environments and different test structure configurations. Calibrated X-ray detectors are used to establish valid dosimetry and to provide confidence in the analysis tools being used. Dosimetry techniques were developed to meet all test conditions. All techniques will allow the use of ASTM test methods as set forth in ASTM F-1467, Standard Guide for X-ray Testers [2] and dosimetry will be traceable to NIST measurement standards wherever possible.

II. APPROACH

Experimental and theoretical analyses used in this effort were extensions of those used in the lower energy (0-50keV) work reported at the 1993 Heart Conference [1]. The theoretical work follows a three step process:

1) Calculation of the MCN 165 X-ray tube output spectrum using the TUBE electron-photon transport calculations.

2) Calculate the spectral attenuation during radiation transport and deposition in the diode detector using the PHOTCOEF and CEPXS electron-photon transport codes.

3) Calculate the diode current response using calibrated sensitivity data and the incident spectrum.

The experimental portion of this effort consists of the following steps:

1) Calibrate the detector sensitivities to a calibrated standard.

2) Measure the detector response in the X-ray chamber for a variety of tube voltages, currents and distances.

3) Correlate detector response to material dosimetry.

The SEMIRAD X-ray diode was chosen as it can be accurately calibrated and it provides excellent response over the full range of photon energies and intensities. The structure and measurement techniques of the SEMIRAD X-ray vacuum diode is described in [1]. The PIN diode used in [1] could not be used at energies above 50keV due to the inability to measure sensitivity over the full energy range.

PMOS FET's were chosen to measure total dose in SiO_2. PMOS FETs can be accurately calibrated in ^{60}Co and they provide a linear response to total dose for ionizing radiation. Device descriptions and measurement techniques are reported in [1].

III. Theoretical Calculations

The MCN 165 tube spectrum was calculated by Dozier using the TUBE electron-photon transport program as reported in [1]. Fig. 1 shows plots of the calculated spectral data for two of the spectra used in the analysis.

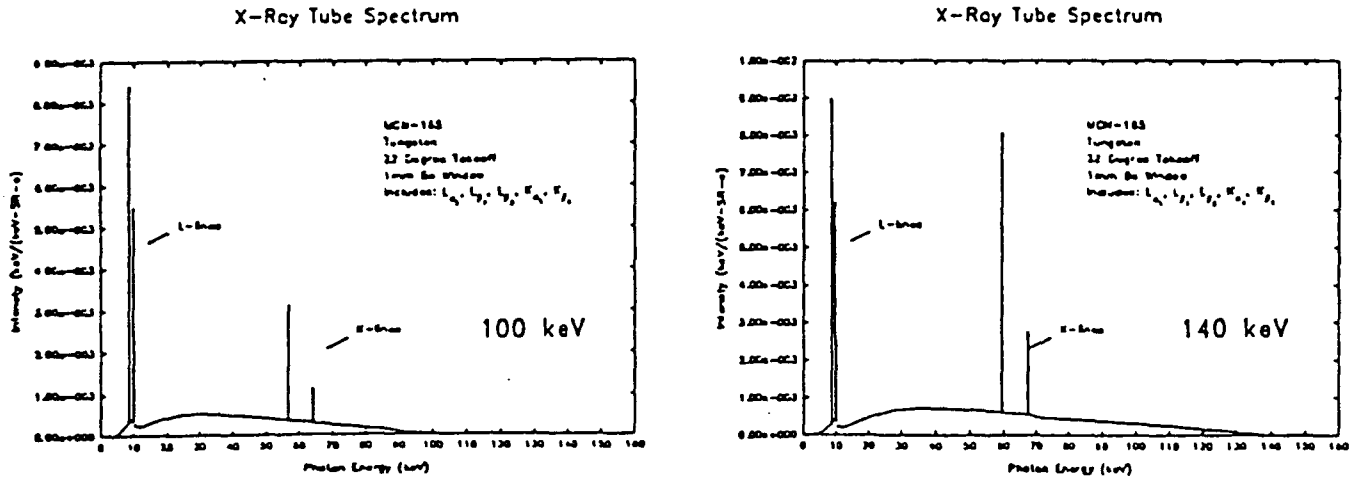


Fig 1. Calculated Tube Spectra for Anode Voltages of 100 & 140 kV

CEPXS[2] and PHOTCOEF[3] were used to calculate filtering and dose deposition in the detector structures. A comparison was made of PHOTCOEF, a semi-empirical code, and CEPXS, a discrete ordinates code, using depth dose calculations for a PMOS FET and a PIN diode structure. Less than a 20% variation in the results over the photon energies of interest was observed between the two codes.

Detector response was calculated by integrating the product of the incident spectrum with the sensitivity curve. This technique was developed in the previously reported work [1].

IV. EXPERIMENTAL DETAILS

The SEMIRAD X-ray diode was calibrated over the full energy range at the SAIC-DNA Albuquerque facility using techniques discussed in [1]. Fig. 2 is a plot of the sensitivity curve using the data provided from the calibration measurements.

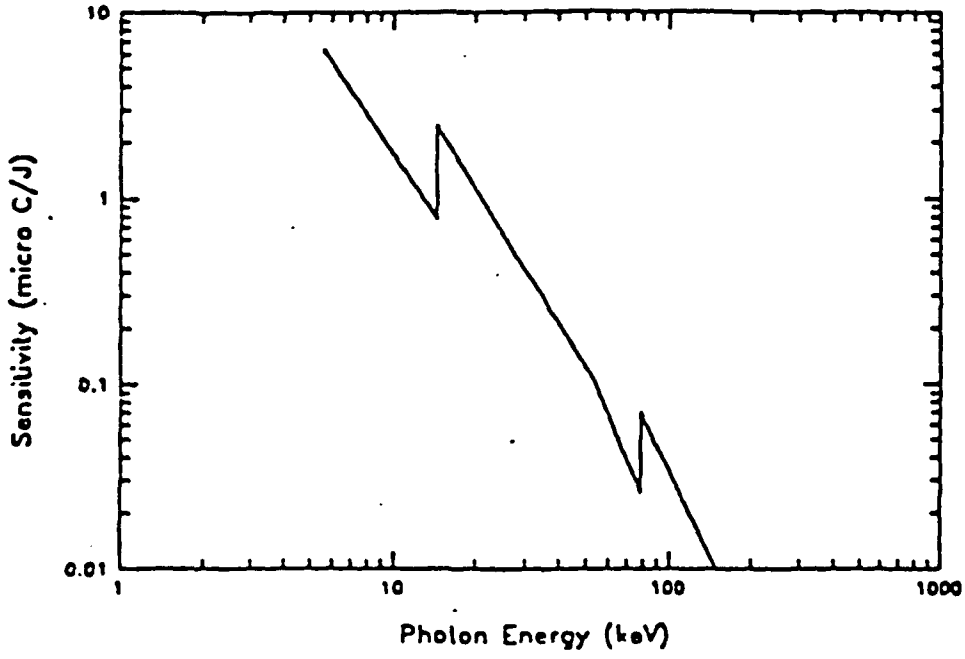


Fig 2. Sensitivity Curve

Diode current response was measured in-situ in the X-ray chamber for anode voltages and currents over the full energy range using techniques described in [1]. Diode current measured at several points is shown in Table I.

Table 1. SEMIRAD X-Ray Diode Measurements versus Calculated.

Anode Voltage (KV)	Tube Current (mA)	Distance (cm)	Calculated Diode Current (pA)	Measured Diode Current (pA)	Relative Difference (%)
30	25	25	867	915	5.5
50	25	25	1610	1957	17.7
80	25	25	2216	3020	26.6
100	25	25	2500	3440	27.6
120	25	25	2696	3780	28.7
140	20	25	2260	3210	29.6
160	15	25	1760	2530	30.4

The PMOS FET sensitivity was calibrated by measuring threshold voltage versus dose at the PL free-field ^{60}Co gamma facility. The procedures in MIL-STD 883C Test Method 1019.3 were closely followed. The PL ^{60}Co facility is calibrated annually in accordance with ANSI/ASTM 666-78 with traceability to NIST. Fig. 3 is a plot of the PMOS threshold voltage sensitivity to ^{60}Co total dose for dose rates in the range 6-20 Krad(SiO_2/min.).

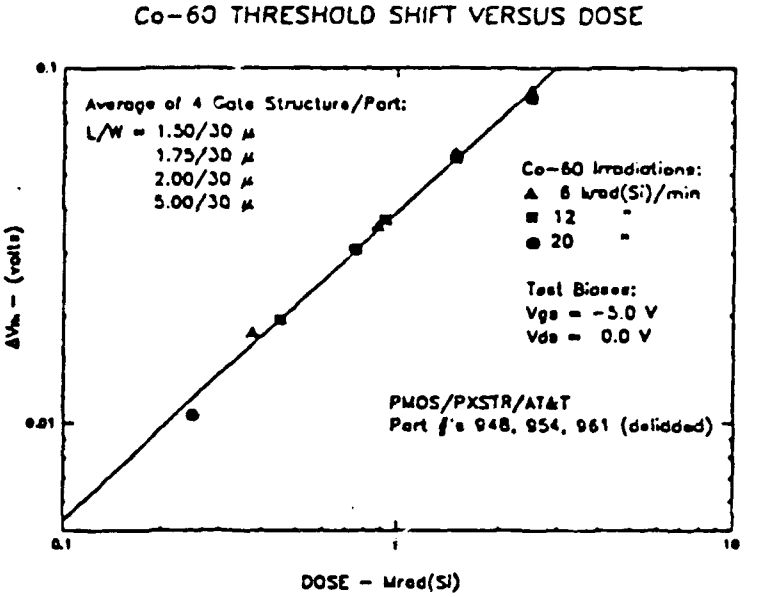


Fig 3. PMOS Threshold Sensitivity for ^{60}Co

The PMOS FETs were irradiated in-situ in the X-ray Test Facility at anode voltages of 50, 100 and 140 keV. The test chamber dose rate at the higher energies (100 and 140 kV end point spectra) was correlated to calibrated PIN data at 50kV established in [1] using the following steps;

1). Calculate the dose rate in the PIN structure using PHOTCOEF and CEPXS for the 50 kV incident radiation spectrum.

2). Calculate the dose rate in the PIN structure for the higher energy incident radiation spectra.

3). Calculate the ratio, i.e. [Dose(SiO_2/s)@100keV/Dose(SiO_2/s)@50keV] for each spectrum above 50kV.

4). Multiply the ratio from the previous step times the dose rate calibration curve for the 50kV to obtain the dose rate at the higher anode voltage being used. For the 100kV and 140kV

spectra, the ratio was found to be 1.43 and 1.487 respectively. Threshold voltage versus dose was measured for four devices for each spectrum using dose rates of 23 and 18 krad(SiO_2)/min. Details of the measurement techniques are presented in [1].

V. RESULTS/CONCLUSIONS

The SEMIRAD X-ray diode current can be directly related to the radiation intensity at any point in the test chamber. Diode currents were calculated for a number of test points in the chamber using the techniques discussed above. The diode current was then measured in the X-ray chamber at the same test points using the techniques described in [1]. Table 1 shows a compilation of the results of the calculations and measurements. Comparing the calculated values with measured will provide a means to measure the relative accuracy of the electron-photon transport codes. A combination of errors can be introduced in the calculation based on estimated errors discussed in [1]. An error analysis was done using these error estimates; 20% across the calculated spectrum, 10% across the sensitivity curve and 2% for the integration procedures. The maximum overall error due to calculations was found to be 23%. Measurement errors were shown to be within 10% [1]. The relative difference between calculated and measured values shown in Table 1 is relatively constant at approximately 30%. We draw the conclusion that the combination of these results indicate that our calculations using the transport codes fall within the overall error envelop.

The PMOS FET is being used in this effort to establish a correlation back to the calibrated ^{60}Co source as discussed in ASTM F-1467. A comparison of X-ray data at 50 kV, 100 kV and 140 kV with ^{60}Co data is shown by plotting all data overlaid in Fig 4.

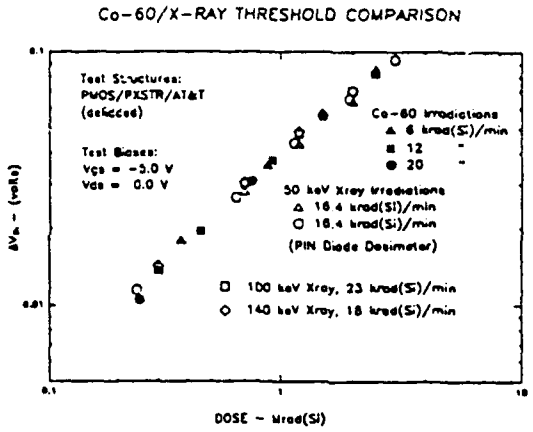


Fig 4. Overlay of PMOS Threshold Sensitivity Data

The maximum excursion (error bars) between data points at a given dose is less than 10%. This correlation is excellent for these type measurements over the broad range of photon energies being considered. These results provide confidence that the combination of techniques used to calibrate dose and dose rate in the X-ray chamber, through traceability to the ^{60}Co standard, will result in valid dosimetry over the full operating range of the X-ray chamber. A summary of the conclusions follows:

1) PHOTCOEF and CEPXS are valuable electron-photon transport codes which provide calculational support for test analysis at accuracies needed to support testing.

2) The SEMIRAD X-ray diode provides an accurate detector to measure test chamber intensities at all energies.

3) The PIN diode calibration of dose rate in Si at 50 kV scaled to higher energies (using the computations of PHOTCOEF) provides a means to measure dose rate over all energies.

4) The PMOS FET provides an accurate calibration with ^{60}Co for all X-ray energies.

5) The techniques developed by this effort and the effort reported in [1] provides the tools to test IAW ASTM F-1467.

REFERENCES

[1] R. Bellem et.al. "Application of Analytical and Experimental Dosimetry Techniques to a Low-Energy X-Ray Radiation Source," to be published in JRERE.

[2] PHOTCOEF A Nuclear Physics Utilities Program Written for IBM PC's and Compatible Microcomputers, AIC Software, P.O. Box 544, Grafton, Mass. 01519

[3] L. J. Lorence, Jr., W. E. Nelson, and J. E. Morel, IEEE Trans. Nucl. Sci NS 32, No. 6, 4416 (1985).

SYSTEM SUPPORT FOR THE SPPD PARALLEL ARCHITECTURE

Scott R. Cannon

Associate Professor

Department of Computer Science

Utah State University

UMC 4205

Logan, UT 84320

scott@cannon.cs.usu.edu

Final Report for:

Summer Research Extension Program

Phillips Laboratory / VTES

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D.C.

September 1993

SYSTEM SUPPORT FOR THE SPPD PARALLEL ARCHITECTURE

Scott R. Cannon
Associate Professor
Department of Computer Science
Utah State University

Abstract

The SPPD parallel system architecture represents an inovative approach to high-speed digital signal processing. The architecture consists of 16 Texas Instruments C30 32-bit processors with 60ns cycle times and an inovative system of global memory access busses and processor control circuitry. Efficient use of this architecture requires robust run-time software system support. This project proposed a set of operating system requirements for real-time applications on the SPPD architecture, reviewed the existing SPPD operating system (monitor) against these requirements, and finally designed and implemented an operating system prototype which addressed the shortcomings of the existing system.

SYSTEM SUPPORT FOR THE SPPD PARALLEL ARCHITECTURE

Scott R. Cannon

INTRODUCTION

The purpose of this project was to investigate providing high-level support for real-time applications on the SPPD system. The SPPD parallel system architecture represents an inovative approach to high-speed digital signal processing. The architecture consists of 16 Texas Instruments C30 32-bit processors with 60ns cycle times and an inovative system of global memory access busses and processor control circuitry. Efficient use of this architecture requires robust run-time software system support. This project proposed a set of operating system requirements for real-time applications on the SPPD architecture, reviewed the existing Rockwell SPPD operating system (monitor) against these requirements, and finally designed and implemented an operating system prototype which addressed the shortcomings of the existing system.

The Rockwell system is inadequate even for simple applications development for several important reasons. The SPADE system (including COP software) and monitor are tightly bound to the *intelect* demonstration program. It is not a general applications development system. Any changes in data format, control, I/O, or display for other applications would require changes to the host SPADE system and COP monitor. Documentation to support changes to the SPADE and COP software is poor to non-existent. The COP software resides in PROM. While COP message handling is fairly general, the run-time mode is very specific to the *intelect* application and does some intermediate processing of output messages. Software development for the COP is not supported. Although general user messages to/from the SPPD system are defined in Appendix D of the Applications Manual, the SPADE software does not recognize any messages outside the *intelect* demonstration.

The SPPD monitor is more general than the SPADE system and provides some useful real-time applications support for a) general output to the host, b) macros for simple "flag polling" for process synchronization, and c) host-controlled breakpoints and halt-state debugging. Unfortunately, interrupt vector locations are in PROM and transfer control to memory allocated to the monitor. No documentation is available for the Rockwell monitor interrupt handling routines or possible dispatch strategies. Sources for these routines are not available. This makes application use of interrupts or modification of the interrupt handling routines difficult.

The TI C30 C compiler system provides limited support for high-level software development. No scheduler support is provided -- each CPU may run only a single dedicated task. No global resource management is supported.

Developing new applications would require a significant programmer effort on all three computers; the host, COP, and the SPPD. Application use of interrupts would be difficult. Parallel software development and partitioning on the SPPD using TI C would be clumsy because of a lack of high-level OS support. All applications would be tightly bound to the architecture.

PROPOSED REAL-TIME SUPPORT REQUIREMENTS

The following general support mechanisms are proposed as a standard for high-level system support for a multi-processor environment:

1. Multi-programming capability. Multiple concurrent processes with levels of priority may be run on the same node. Scheduling overhead should occur only when a process is blocked from proceeding. Single process CPUs should have no scheduling overhead. (This models allows more freedom in developing applications and provides a more natural mapping of algorithm stages to processes. Applications become less tightly bound to architecture. While completely deterministic algorithms may not benefit from multi-programming, signal processing applications which contain adaptive or stochastic algorithms definitely will. The CPU may be more efficiently utilized; when a process is blocked waiting for an external event, another unblocked process may utilize the CPU.)

2. System management of global resources. Shared and private memory should be allocated abstractly so that application programmers need not deal with specific physical addresses. (Programs become easier to development, more portable, and less prone to bugs.)

3. Efficient process synchronization (true semaphores). Shared mutually-exclusive resources (memory, I/O ports, etc.) may be globally locked and unlocked in a manner which frees the CPU for other tasks when a resource is not available. Processes may be synchronized using blocking. (This mechanism frees the applications programmer from inefficient flag polling mechanisms and centralized CPU-specific access routines.)

4. Global timing control. Global timing functions provide system-wide alarms. Virtual timers allow more timers than are physically supported.

5. System-wide halt-state debugging. Any process should be able to halt the system. During a halt state, any CPU should provide debugger access for memory and register examination/modification. The system should be able to proceed after a halt state.

6. An inter-process message-passing system. Any process should be able to send a message to any other in a high-level manner without the programmer's knowledge of physical shared-memory addresses. Message arrival should be immediately noted without the need for inefficient polling loops.

7. Applications access to the interrupt services. Applications on each node should be able to specify independent interrupt service routines at run-time.

While the Rockwell system provides capability #5 (halt-state debugging), support for the other mechanisms does not adequately exist. Simple deterministic signal-processing applications using TI C would require at least capabilities #5 and #7.

The Tartan Ada system would provide support for all mechanisms if monitor support were available for #6 (inter-process messages) and #7 (independent interrupt service routines). Unfortunately, an adequate port of Tartan Ada would require the implementation of 3 trap interrupt-service routines (TRAP 14, 15, and 16) and a TIMER0 interrupt-service routine which would be difficult with the present Rockwell monitor and documentation. Another complication is that both Ada and the Rockwell monitor require use of TRAP-14. Adequate porting of the Tartan Ada system will require modification to the monitor interrupt handing system. Note: Some simple Ada programs can be run under the Rockwell system if several patches are used and certain constraints are followed (see Appendix A for details).

Unfortunately, inadequate documentation and a lack of monitor sources make modifications to the current monitor difficult. The purpose of this project was to a) *make prototype modifications to the SPADE host server* and b) *develop a new monitor prototype* to support mechanisms #6 (inter-process messages) and #7 (user access to independent interrupt services). With these capabilities, it should be possible to use TI C for simple deterministic applications and port Tartan Ada for applications requiring more significant high-level support.

PROTOTYPE GOALS

All messages between the SPADE host and the SPPD application go through the SPADE server (vserv.c). Two simple modification goals for the server will allow the current system to be used for initial application development:

1. The server recognizes and displays text from SPPD applications to be printed on the host.
2. The server provides a mechanism for independent application programs on the host to perform

I/O with applications on the SPPD during SPADE operation.

The new PL monitor prototype has 4 goals to facilitate the porting of the Tartan Ada system and to allow better support for applications development in C. These goals were met while preserving the interface with the host SPADE system to preserve program loading, execution, and debug control functions currently available.

1. Useful interrupts are dispatched to service routines specified by applications at runtime.
2. Each CPU has the ability to specify its own service routine for a given interrupt or trap.
3. The debugger breakpoint trap number was changed from 14 to 12 to be compatible with Ada.
4. An inter-process message system allows any CPU to send a message to any other CPU. Message arrivals interrupt the destination CPU for immediate handling. Message handling routines may be specified by applications at runtime.

USER'S GUIDE

Synchronization and breakpoints. A series of macros in *syncmacs.h* are provided for synchronization:

a) UPDATE_APPL_START This macro notifies the monitor that this particular cpu has finished with initialization code and is ready to synchronize with the other processors.

b) WAIT4_ALL_PROCS This macro causes the calling CPU to block until all other processors have finished initialization -- in other words, invoked UPDATE_APPL_START.

These last two macros perform the same function as the similarly named macros in the Rockwell SPADE Application Manual. The first macro adds one to a count, the second is a tight polling loop waiting for the count to equal the number of active processors in the application. The SPADE system determines the number of active processors from the *.prc file.

c) BREAKPOINT Halt the current application on all processors and cause all processors to return to the monitor command loop. During the command loop, the operator may interact with the debugger to examine/modify memory and registers and then perhaps continue. Any active processor may execute the BREAKPOINT macro. A word of caution: The operator should recognize that breakpoints on a parallel system are much more complex that breakpoints on a single CPU. The monitor may not recognize nearly simultaneous breakpoints on separate processors since it is not possible to

instantaneously recognize a breakpoint and halt all processors. In other words, if two processors both execute the BREAKPOINT macro at nearly the same time, the monitor may indicate to the host that a breakpoint has been reached on one of them and ignore the other. It may appear to the operator that the second breakpoint was missed.

d) INT_ON, INT_OFF	Turn interrupts on or off.
e) LOCK, UNLOCK	Lock or unlock global memory.

The original SPADE monitor system also contains a set of primitive synchronization macros which are still applicable. The user should be aware that these macros implement simple polling loops are not efficiently used with a scheduler such as that provided by Ada.

Trap and interrupt dispatching. The monitor allows each processor to specify its own set of interrupt service routines. If each CPU has its own page of low memory (where interrupts are vectored by the hardware) this would be available without effort. Unfortunately, all processors share the same low memory and it is fixed ROM. As a result, the monitor must initially handle and dispatch all interrupts. For example, Interrupt 5 for CPU 7 causes the monitor dispatcher to examine slot 7 in a vector table of 16 service routine addresses for this interrupt. If a routine address is found, control is vectored to this routine (otherwise it is ignored). So, all an application on CPU 7 needs to do is set the appropriate service routine in slot 7 for this interrupt. The dispatch table locations for each interrupt are documented in monitor.h. For example, the location of the vector table for trap 20 is TRAP20VECTBL.

Basically, hardware interrupt vector tables reside from 8300h (INT0) to 83A0h (interrupt DINT) with 16 addresses for each interrupt. All hardware interrupts are dispatched and traps numbered 10 and above. Trap vector tables are from 8510h (trap 10) to 8620h (trap 27). All processors share the same service routines for traps 0 to 9. The ROM boot code stores the CPU number of each processor at address CPUNUM. Installing an interrupt service routine is accomplished by placing the address of the routine in the appropriate vector table slot. For example;

```
int     cpunum, *loc;
void    service();                          /* an interrupt service routine */
cpunum = *((int *) CPUNUM);                 /* determine this processor's number */
loc = (int *) (TRAP20VECTBL + cpunum);/* the address of this cpu's slot for trap 20 */
*loc = (int) service;                       /* place the address of service() in this slot */
```

An interrupt service routine must save and restore all registers used and return with a RETI assembly instruction. If the routine is written in C, see the TI TMS320 C compiler manual, page 4-27. If

two processors are to share the same service routine, the routine must be re-entrant! This of course occurs if the same executable is run on more than one processor. (A routine can be re-entrant if it changes no static variables and uses a stack in local memory.)

The monitor uses interrupt 0 for the mail system. Changing table entries in the interrupt 0 vector table (at INT0VECTABLE) will make the mail system un-usable. The monitor reserves the use of traps 0 through 6, and 12.

Inter-process messages. Two monitor functions are provided to use the inter-process message system. These functions are defined in mail.h and translate to simple trap calls. When one of these functions is called, the C system places the appropriate parameters on the stack and calls the appropriate trap. Using this method alleviates the need for a special object module or library which would need to be linked to every application. If Ada is used, follow the Tartan mechanism for calling a C function and then perform an inline trap call as indicated by the mail.h definitions.

Each mail message is given a type by which categories of messages can be identified. Config_mail is called to indicate to the monitor how incoming messages are to be handled:

```
void config_mail (int type, int *addr, int *flag, int fun)
        type -- a value from 0-15 indicating the message type being configured.
        addr -- the address of memory where this type of message is to be stored (or NULL)
        flag -- the address of a flag to be set when a message of this type arrives (or NULL)
        fun -- the address of a function to be run when a message of this type arrives (or NULL)
```

The function send_mail is used to transmit a message:

```
void send_mail (int destcpu, int type, int length, int *addr)
        destcpu -- the destination cpu number (must be different from the source cpu)
        type -- a value from 0-15 indicating the message type
        length -- the message length; 0-62 words
        addr -- the address of the message to be sent (or NULL)
```

Arriving messages are formatted according to Appendix A of the SPADE Application Manual. The first word contains the message type (upper 8 bits), the source CPU number (next 8 bits), and the message length (lower 16 bits). The next word is the source context (user definable) and then the following words are the message proper. Following the message is a word reserved for a checksum value. The checksum is only used for messages which are transmitted to the host. For example, if CPU 5 wishes

to send messages of type 9 to CPU 2, the following might be used. Note that the receiving buffer must be 3 words longer than the arriving message to allow for the header/checksum overhead;

```
a) on CPU 2 (the destination);      int flag, buf[35];
                                    flag = 0;
   (prior to arriving mail)         config_mail (9, buf, &flag, NULL);
                                    ...
                                    while (!flag) delay (10);   /* wait for incoming messages... */

b) on cpu 5 (the source);           int outbuf[32];
                                    ....                        /* fill outgoing message... */
                                    send_mail (2, 9, 32, outbuf);
```

The function parameter of the config_mail call provides a versatile model of message synchronization and buffering. If for example, the receiving process would like arriving messages queued for later use, the function might be used to specify a service routine to insert messages into a circular queue. Since these functions are invoked asynchronously with the user application program as a result of arriving mail interrupts, they must follow the conventions of interrupt service routines (see the above section on traps and interrupts). Alternately, the service function might be used to actually process arriving messages and leave the other application program thread free for other lower-priority operations. In this manner, efficient use of the CPU can be achieved. When messages are not available for processing, the CPU is free to run the low priority process. When a message arrives, the higher priority function immediately receives control. Many different models can be implemented.

Printing to the host. Any messages sent to CPU 0 with type PRINT (defined in mail.h as 13h) are automatically forwarded to the host for display on the server window.

APPLICATION-SPECIFIC I/O

On the SPPD system. Messages from the SPPD to the host must follow the message definition model in Appendix A of the SPADE Application Manual. In essence, this means that each message has a 2 word header consisting of a message op code and a host context and a checksum trailer. The op code consists of a message type (upper 8 bits), the source CPU number (next 8 bits), and message length (lower 16 bits). The message length includes the header and checksum words. Sixteen output message types are defined in monitor.h as USROUT0 (10h) to USROUT15 (1Fh). Note that USROUT3 (same as PRINT) is reserved for printing messages. The checksum is simple the sum of all message words (except the checksum). In addition to the 15 output message types defined in monitor.h, 16 user error codes are also defined.

Messages can sent via the normal mail message system by sending them to CPU 0 for auto-forwarding. Also, messages may be directly sent by writing to the BIDI port on CPU 0. The port address is defined in monitor.h as BIDIWRIT. Be sure to turn interrupts off during message transmission. This is because certain circumstances may cause interrupt service routines to do output. If an interrupt occurred during transmission, the monitor could attempt to send a message within the user message resulting in confusion to the COP system and the monitor. For example, to send a message of type USROUT0 consisting of a single word;

```
int       message_word, checksum, *port, op;
port = (int *) BIDIWRIT;                    /* BIDI port write address */
checksum = 0;                               /* initialize the checksum as zero */
op = USROUT0 | (cpunum<<16) | 4             /* build the op code word -- note length of 4 */
INT_OFF;                                    /* turn interrupts off */
checksum += *port = op;                     /* send and add to checksum */
*port = 0;                                  /* send the host context as zero */
checksum += *port = message_word;           /* send the message word */
*port = checksum;                           /* finally send the checksum */
INT_ON;                                     /* and turn interrupts back on */
```

The same two methods are available to receive messages from the host. The first method is to use config_mail() to configure the mail system to normally forward messages to the appropriate buffer and set the arrival flag. Alternatively, an application could turn the interrupt 1 off and simply poll the interrupt flag. The input message is read in a same manner. The bidiport read address is defined in monitor.h as BIDIREAD. Interrupts must be turned off during reading.

```
int       checksum, op, buf[64], n, *port, input_check;
port = (int *) BIDIREAD;                    /* BIDI port read address */
checksum = 0;                               /* zero the checksum */
INT_OFF;                                    /* turn interrupts off */
asm (" PLP:       TSTB    2,IF ");          /* poll the interrupt flag */
asm ("            BZ      PLP ");
checksum += op = *port;                     /* get the op word */
length = (op & 0xFF) - 3;                   /*  and calculate the length */
for (n=0; n<length; n++)
  checksum += buf[n] = *port;
input_check = *port;                        /* the sent checksum */
asm ("            ANDN  2,IF ");           /* clear the interrupt */
if (input_check != checksum) ERROR...
```

on UNIX. Applications on the host UNIX system may also communicate with SPPD applications. User message communication with the SPADE server is via UNIX IPC queues. The reader is encouraged to review IPC queue services in the UNIX manual. Four user communication functions have been written to simplify the procedure (although general use of the IPC queues is certainly available);

```
void us_init (int *shared, int *upque, int *downque)
        shared -- returns as address of SPADE shared memory block (not covered in this guide)
        upque -- returns as the IPC queue number used to receive messages from SPPD
        downque -- returns as the IPC queue number used to send messages to SPPD

void us_format_message (Message *msg)
        msg -- a message to be formatted and a checksum calculated and inserted.

void us_send_message (int queueid, int pid, Message *msg, int length)
        queueid -- IPC queue number to use (always downque)
        pid -- process id number
        msg -- the message to be sent (formatted)
        length -- message length plus 3 for overhead.

int us_recv_message (int queueid, 0, Message *msg, int *length)
        queueid -- message queue to use (always upque)
        msg -- message structure to receive the message
        length -- integer to receive the message length
        -- returns TRUE if successful  else FALSE, blocks until a message is available.
```

To send a message to the SPPD system on a user program running on the host sun4, the host user program must attach to the IPC queues by calling us_init (). For example;

```
int     *shared,                        /* pointer to the SPADE shared memory block */
        upque, downque;                 /* up and down IPC queue id numbers */
us_init (shared, &upque, &downque);     /* attach to the shared memory block and queues */
```

The shared memory block is used for image transfer and it not covered in this brief user's guide. To send a message, the user declares a message structure, formats the message, and sends it to the "down" queue. For example, to send the integers x and y as a type 5 message;

```
int     Message msg                     /* message buffer */
int     count,                          /* message size plus 3 for headers/checksum */
        pid,                            /* UNIX process id */
        x, y;                           /* data to be sent */
pid = getpid();
msg.command = USRIN5 | (wc+3);                  /* build the command op word */
msg.context = 0;
msg.data[0] = x;
msg.data[1] = y;                        /* add the data to the message */
us_format_message (&msg);               /* format and add checksum */
us_send_message (downque, get, &msg, 5);    /* send it on down! */
```

To receive a message (after attaching to the queues with us_init());

```
us_recv_message (upque, 0, &msg, &wc);  /* wait for a message */
```

```
wc -= 3;                        /* calculate data length */
```

User programs on the host (UNIX) must link in us_util.o and reference vsh.h. User programs on the SPPD system must link in sp_cint0.obj, mutil.obj and reference monitor.h, syncmacs.h (if sync macros are used), and mail.h (if the mail message system is used). Refer to the example applications.

LIMITATIONS.

As of 9/21/93 the following limitations exist in the PL monitor prototype. It is assumed that further work will eliminate these limitations to achieve the full design goals mentioned in the sections above.

1. The INT1 interrupt should allow messages intended for users to be received and placed in the normal mail system. At present, only HALT messages from the host SPADE system are allowed. User messages must be received by direct reading with interrupts off.

2. Any user message of type USROUT0 to USROUT15 sent to CPU 0 should be forwarded to the host SPADE system. At present, only PRINT (USROUT3) messages are forwarded. User messages must be sent directly to the Bidi port.

3. The mail service function allowed in the config_mail() call has not been tested and should not be used.

4. Traps 10 through 27 should be dispatched to user-specified service routines. At present, only traps 10 through 14 and 20 are dispatched.

5. While all remaining capabilities mentioned in the user's manual have been exercised, there exists an almost infinite number of ways that the code for these operations can interleave on a parallel system. The limited testing (just a few days) should not be expected to have eliminated all bugs and quirks of the prototype monitor. In addition, since the monitor design did not follow standard design review and validation procedures normally used in a competent development cycle, it should be viewed strictly as a prototype. While it is expected to be used in the testing and evaluation of the SPPD system, the first actual application requirements specification should include requirements for a production monitor. Hopefully, the design of this monitor prototype and perhaps some sections can be re-used in this formal monitor development.

DESIGN DOCUMENTATION

The PL monitor (for Phillips Laboratory) follows the basic design of the original Rockwell monitor (as well as can be determined from the partial existing sources). The command and control message format of Appendix A of the SPADE Application Manual was retained. As a result, the PL monitor is compatible with the existing SPADE host system with the exception of the breakpoint definition. The basic goals of the PL monitor prototype (as defined in the Overview section) were to provide mechanisms for the following;

1. dispatch useful interrupts to service routines specified by applications at runtime,
2. allow each CPU the ability to specify its own service routine for a given interrupt or trap,
3. change monitor trap use to be compatible with Ada, and
4. provide an inter-process message system to allows any CPU to send a message to any other CPU or the host.

A word of caution: Since the SPPD is a parallel architecture of 16 processors, the monitor is not a single-thread system. Rather, sections may be concurrently executing on all active processors. While code for all monitor sections resides in global page-0 memory, stack and variable space resides on all active processors in local memory (809F00h and above). Many monitor code sections are *re-entrant* meaning that multiple CPU code threads may be concurrently within the same function. In addition, a monitor which supports interrupt service routines and context switching (between the application and monitor code) is by nature *unstructured*. The C30 processor does not directly support context switching (as does for example the Transputer). As a result of the above, much of the interrupt dispatching and context switching code was done in assembly.

In addition to the above complexities, a parallel system monitor must also be concerned with possible deadlock conditions involving shared memory or resources. While the prototype has not been formally tested for all possible deadlocks, care has been taken to design cooperative strategies which avoid circular wait deadlock conditions in the mail system.

Interrupt dispatching. All hardware interrupts and traps 10 through 27 are dispatched. (traps 0 through 9 are not dispatched and all processors share the same service routines.) A table is defined within monitor.h for each dispatched interrupt: INT0VECTABLE, INT1VECTABLE, etc. are the addresses for hardware interrupt tables and TRAP10VECTBL, TRAP11VECTBL, etc. are the addresses for dispatched trap tables. Each table consists of 16 slots corresponding to the 16 different CPU numbers. An address in slot N represents the beginning address of the service routine for CPU N. The master monitor code initializes each table with -1 values indicating that no service routine is present and then inserts service

routine addresses for INT0, INT1 and TRAP12 for CPU 0. The slave monitor code inserts service routine addresses for INT0 and TRAP12 for each slave CPU. All other dispatched slots are available for user application use.

Mail Service. The mail system is controlled with the 6 global memory areas specified in monitor.h and documented below:

REQ_FLAGS. This area consists of 16 words/CPU. CPU N owns the 16 words beginning at REQ_FLAGS + N*16. Setting the Kth word belonging to processor X to a non-zero value indicates that processor K has a mail message it the Kth outgoing message buffer for processor X.

MESSAGE_ADDR. This area consists of 64 words/CPU. CPU N owns the 64 words beginning at MESSAGE_ADDR + N*64. Each message area is used by the owning CPU for outgoing messages. CPU X reads a message from CPU K from the message area belonging to CPU K.

MESS_IN_USE. This area consists of 16 words. Word K is the in-use flag for the message area of CPU K. When flag K is set it indicates that CPU K has sent a message to another CPU which has not yet been read. A CPU sets its own in-use flag when a message is copied into the outgoing message area buffer for sending, but the receiving CPU clears the flag when the message has been received.

BOX_ADDR. This area consists of 16 words/CPU. CPU N owns the 16 words beginning at BOX_ADDR + N*16. A non-NULL address in the Kth word belonging to CPU X indicates that mail of type K arriving for CPU X is to be stored at this address. A user application sets the incoming buffer addresses in the processor's table by calling the config_mail() function. If an application sets a box address to NULL using config_mail(), messages of the corresponding type will not be saved.

BOX_ARRIVED. This area consists of 16 words/CPU. CPU N owns the 16 words beginning at BOX_ARRIVED + N*16. A non-NULL address in the Kth word belonging to CPU X indicates that mail of type K arriving for CPU X is to cause the flag variable at this address to be set. When a flag is set, the application is responsible for disposing of the message and clearing the flag. A user application sets the incoming flag addresses in the processor's table by calling the config_mail() function. If a box-arrived flag address is set to NULL with config_mail(), no flag will be set when mail of the corresponding type arrives.

MAILFUNCS. This area consists of 16 words/CPU. CPU N owns the 16 words beginning at MAILFUNCS + N*16. A non-NULL address in the Kth word belonging to CPU X indicates that mail of type K arriving for CPU X is to be initially processed by the function at this address. A user application sets the incoming service routine addresses in the processor's table by calling the config_mail() function. These service routines must save and restore any registers used but not turn interrupts back on. The TI C routines named c_intxx.c (where x is an integer) do save and restore all registers used, but they turn interrupts back on on exit. If a service routine is written in C using this convention, the assembly language produced must be edited to change the final RETI to a RETS. If a mailfuncs service routine is set to NULL by config_mail(), no function will be called. It is possible to configure the mail system in a variety of models. For example (other combinations are of course possible);

config_mail (N, NULL, &flag, NULL) -- ignore the message content, but set the flag on arrival.
config_mail (N, &buffer, NULL, NULL) -- copy the message, but do not set a flag.
config_mail (N, NULL, NULL, (int) func) -- ignore the message content, but run func().

Memory Map. SPPD memory is utilized by the monitor according to the following map. It should be noted that while there are gaps in the following map, all space from 8000h to FFFFh (global memory) and above 809F00h (local memory) is reserved for monitor use. Future modifications of the monitor reserve the right to use all such space.

8000h to 82FFh	reserved for future RAM vectors
8300h to 83FFh	hardware interrupt dispatch tables
8400h to 850Fh	variable space accessed by the ROM boot code. (Includes the NPROCS, APPL_START variables and the PROCESSOR_MAP).
8510h to 862Fh	trap interrupt dispatch tables
9000h to 9852h	mail system space
A000h to A0FFh	variable space accessed by the ROM boot code (initialized variables)
A100h to AFFFh	location of the master command loop code
B000h to BFFFh	location of the slave monitor code
C000h to CCFFh	location of the dispatcher routines
D000h to DFFFh	location of the master monitor code
E000h to E100h	master monitor stack space

809F00h	processor number (stored by the ROM boot code)
809F01h to 809F27h	register save area for applications (for context switching)
809F28h to 809F4Eh	register save area for the monitor (for context switching)
809F4Fh	monitor state flag
809F80h to 809FFFh	local monitor stack space

THEORY OF OPERATION.

Monitor ROM boot. The existing monitor ROM code was not changed. The source is in romreset.c. This Rockwell code resides in the low page of memory and gains control *at CPU 0* at power-up or master

reset -- all other CPU's are held. When released by CPU 0, it also gains control *at all other CPU's*. In other words, the monitor ROM boot code eventually runs on all processors utilized in a system.

It's function is to a) set system memory control registers, b) disable and clear all interrupts, c) determine the CPU number and store in local memory location 809F00h, d) if CPU 0, set the CRn control registers and load the monitor master code, and e) branch to the master (D100h) or slave code (B500h) depending on CPU number. On CPU 0, the master monitor code is loaded by responding the MMEM (modify memory) messages on the Bidi port. Receiving a LCMP (load complete) command indicates to the ROM code that the master code has been loaded. In addition to the above functions, the ROM code also defines interrupt vector locations at memory addresses 0 to 3Bh;

Master code. After being loaded by the ROM boot code, the master monitor code gains control of CPU 0. All other processors are still being held. The source for this code is master.c. The responsibilities of this code are to a) clear all global and local memory (other than page 0 where the monitor resides), b) fill all dispatch tables with -1 (indicating no service routine present), c) load the CPUs dispatch table with service routine addresses for INT0, INT1, and TRAP12, d) load the rest of the monitor (slave and master sections), e) initialize the mail system, f) release the slave processors (to run ROM boot code), reply to the host that the monitor is ready (LCRPY or load-complete reply), g) save the monitor state in the MON_REGS local memory area, and h) branch to the master command loop. The loading of monitor code is achieved by receiving MMEM and LCMP messages from the host. The monitor state saved prior to branching to the master command loop becomes the entry state each time context is switched to the monitor during a breakpoint or halt operation.

Master command loop code. This code sends the "PL monitor.." message to the host to indicate to the operator that the monitor is active. The APPL_START variable is set to zero and the NPROCS variable is set to the number of active processors. These variables allow initial synchronization of application processes through use of the UPDATE_APPL_START and WAIT4_ALL_PROCS macros. The code then enters a loop to repeatedly respond to host message commands. These commands may now consist of MMEM (modify memory -- used to load application code at this point), MREG (modify registers), XREG (examine registers), XMEM (examine memory), MRES (master reset), LCMP (load complete), and EXGO (execute application).

MREG and XREG register commands modify and examine copies of the application registers for each processor. Since the application is not active at this point, the register values displayed on the debug screen *do not reflect current processor registers.* Rather, debug displays of processor registers reflect the

values present *in the application* at the last halt or breakpoint. Processor register values are saved in local memory beginning at address REGS and this is the area modified/examined with the MREG and XREG commands.

When a MMEM/XMEM or MREG/XREG command refers to local memory or registers of a slave CPU, the master monitor must pass this command to the appropriate slave monitor thread. This is accomplished by copying the message to global memory at location COM_MESS and setting the flag at location COM_FLAG with the destination CPU number. When the slave monitor thread has processed the message, the slave clears the COM_FLAG location. Any slave monitor reply to the command is now expected to be in the COM_MESS location.

When an EXGO command is received, the master command loop sends the EXGO message to all active processors (via the above COM_MESS mechanism) and clears two state flags (at STATE and MONSTATE addresses). The flag at STATE indicates the system is not in a breakpoint state while the flag at MONSTATE indicates the system is not in the monitor state. At this point, the command loop clears any remaining Bidi port interrupts and calls trap-6 to do a context switch to the application.

Slave code. The source is in slave.c. The monitor slave code is jumped to by the ROM boot code on each non-0 CPU. This code sets local memory-control registers and clears local memory. It then sets the local dispatch table with routines for trap 12 and interrupt 0 (INT0) and saves the monitor state in the MON_REGS local memory area prior to branching to the slave command loop. The saved state becomes the entry point for the slave monitor thread at a system halt or breakpoint.\

Slave command loop. The source is slavecmd.c. This code monitors the master monitor message flag at COM_FLAG for a matching CPU number. When one is detected, it indicates a message for this monitor thread. The command message (in area COMMAS) is expected to be a MME, XMEM, MREG, or XREG command for local memory or an EXGO. The memory/register commands are responded to by placing the appropriate reply message in the COM_MESS area and clearing the COM_FLAG. The EXGO message causes the slave thread to clear the local MONSTATE flag (described above) and switch context to the application code via a trap-6 call.

Monitor trap calls. The monitor utilizes trap calls 0 .. 6 and 12 for the following functions. Monitor trap functions 0 ... 6 are re-entrant and shared by all processors. CPU 0 uses the _trap12 function in monisr0.c30 for trap 12 while all slave monitor threads share the _trap12n function in monisrn.c30. All trap functions are in monisr0.c30 (CPU 0), monisrn.c30 (slaves) or mailtrap.c30 (all mail functions).

trap 0: turns interrupts off -- can be directly called by application code
trap 1: turns interrupts on -- can be directly called by application code
trap 2: configures mail reception (called as the config_mail() function)
trap 3: sends mail messages (called as the send_mail() function)
trap 4: receive and distribute mail (called as a result of an INT0 interrupt)
trap 5: switch context to the monitor (called as a result of a breakpoint)
trap 6: switch context to the application (called by the monitor)
trap 12: breakpoint and switch context to monitor -- can be directly called by appl. code.

The trap 0 and trap 1 calls are documented on page 6-22 of the TI C30 Users Guide. The operation of the others is documented below.

Trap 2. The mail buffer and mail-received flag addresses are stored in the appropriate slots of this processor's mail control area of BOX_ADDR, BOX_ARRIVED, and MAILFUNCS (see the references to the monitor mail system above). It assumes the appropriate arguments are upon the stack as per the C calling convention.

Trap 3. The MESS_IN_USE flag for this processor is checked to verify that the area is available. If it is set, the message area still contains an un-read message. The code at this point loops waiting for the distant CPU to finish processing the message and clear the flag. During the loop, interrupts for this CPU are alternately turned back on then off to allow the local mail interrupt service routine to proceed with incoming mail. The purpose is to reduce the likelihood of circular deadlock. As soon as the flag is clear, the indicated message is formatted and copied to this processor's MESSAGE_ADDR. The MESS_IN_USE flag is set to indicate that the message area contains a new un-read message. Formatting implies that the appropriate 2-word head is appended to the message (see SPADE Application Guide, Appendix A).

Next, the appropriate REQ_FLAG for the destination CPU is checked to verify the destination CPU has completed processing of this processor's previous message. If not, a similar wait loop with interrupts on/off is entered. When the appropriate REQ_flag is available, it is set and the destination CPU is interrupted with INT0. While the interrupt is asserted and cleared, global memory is locked to prevent another CPU from interleaving another interrupt on the same PBUS.

If a message is type PRINT (USROUT3) and the CPU number is 0, the message is handled in a special way -- after copying the message to the local MESSAGE_ADDR area, it is simply sent to the host and the MESS_IN_USE is cleared locally.

Trap 4. This function is called from the INT0 interrupt service routine to check for incoming mail. The function loops through all 16 REQ_FLAG flags for this CPU. If a flag is set, the MESSAGE_ADDR area of the corresponding CPU is assumed to contain mail for this processor. The type of the message is examined and a) the message is copied to the appropriate buffer as indicated by the corresponding BOX_ADDR value and b) the appropriate flag is set as indicated by the corresponding BOX_ARRIVED value. Remember, the BOX_ADDR and BOX_ARRIVED slots are set by the config_mail() function. If a function address is found in the appropriate MAILFUNCS slot, that function is called. After processing the message, the local REQ_FLAG is cleared and the source processor's MESS_IN_USE flag is cleared.

If the flag indicated by the appropriate BOX_ARRIVED slot is still set indicating the previous message has not yet been processed, the mail system does not buffer the message but considers this to be a fatal state. A -1 is placed in the RE register and the system is breakpointed.

Trap 5. The system context is switched to the monitor from the application program. The application registers are saved in the REGS local memory area and the monitor state (registers) is restored from the MONREGS local memory area. This state and PC value correspond to the state at entry to the slave or master command loop code. *All* application registers are saved including the ST register value just prior to the trap 5 call.

Trap 6. The system context is switched from the monitor to the application. Monitor registers are not saved since the monitor will be restored to the original state (just prior to the command loop). This original state was saved in the master or slave code just prior to branching to the command loops and is never changed. The application registers are loaded from the previously saved values. If the user had INT0 and INT1 turned off, *they are turned on.* Otherwise the state is exactly restored and the PC is set to continue where the application had previously left off. When the first trap-6 is called, these register values are simply those retrieved from the corresponding SPADE .reg file.

Trap12. If the breakpoint state flag (at address STATE) has not yet been set, the flag is set and all processors are interrupted with INT0. (If the flag is already set, this is a near-simultaneous breakpoint and an interrupt is not generated.) On CPU 0, a GORPY command is sent to the host indicating a breakpoint. Finally, the context is switched from the application to the monitor via a trap-5 call.

Interrupt Dispatching. The dispatcher (with source in dispatch.c30) is responsible for vectoring interrupt control to service routines specified at runtime and specific to each processor. Actually only hardware interrupts and traps 10 through 27 are dispatched -- traps 0 through 9 utilize the same re-entrant

routines across all processors. When an interrupt or trap occurs, control is passed to the routines at the fixed memory address specified by the interrupt table at memory location 0 (in ROM code). Refer to the table in the section on the ROM boot monitor. The dispatcher has routines at each of the fixed memory addresses corresponding to an interrupt (C000h to CBF0h). These routines determine the correct CPU number for this interrupt and load the appropriate dispatch table entry. If the entry contains a -1, the interrupt is ignored (but not dismissed). If a valid address is found, control is passed to this service routine. When the dispatcher branches to the appropriate service routine, all registers are unchanged. In other words, it is as if the service routine were directly entered by hardware vector control. Since a register must be used for indirect branching, this is accomplished with a delayed branch -- the NOP instructions at the end of each dispatch segment are significant and should not be removed.

Interrupt Service Routines. All INT0 and INT1 interrupts are handled by the monitor while the application is running. (While in the monitor, the Bidi port INT1 flag is polled for commands since interrupts are turned off.) The service routine code is in monisr0.c30 (master monitor) or monisrn.c30 (slave monitors).

INT0. This interrupt occurs for two reasons -- to signal the arrival of mail messages from other processors or to signal a system breakpoint. In fact, both may occur simultaneously! The interrupt is immediately dismissed so that future interrupts are not missed. The interrupt system is of course, left off during the service routine code. First, all messages with flags set in this processor's REQ_FLAG area are handled via a trap-4 call. Next, if the monitor state flag (at MONSTATE) has already been set, the interrupt is ignored. Otherwise, if the breakpoint flag (at STATE) is set, it is assumed another CPU is causing a breakpoint: If this is CPU 0, a GORPY command is sent to the host to indicate that the system has halted. Context is switched to the monitor via a trap-5 call. If the breakpoint flag is not set, it is assumed that the interrupt was strictly for arriving mail and the routine returns to the application with interrupts re-enabled.

INT1. This interrupt is handled by the monitor on CPU 0 where it is associated with messages arriving via the host interface Bidi port. An incoming message is read and handled. If it is a user message of type (USRIN0 to USRINF) it is posted to the mail system. Otherwise only a HALT command is recognized which forces a system breakpoint via a trap-12 call.

A FEASIBILITY STUDY OF A 70-K PERIODIC SORPTION CRYOCOOLER AND THERMAL MANAGEMENT OF SODIUM SULFUR BATTERY MODULE

Amir Karimi
Associate Professor
Division of Engineering

The University of Texas at San Antonio
Division of Engineering
6900 N. Loop 1604 W.
San Antonio, Texas 78249-0665

Final Report for:
Summer Research Program
Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1993

A FEASIBILITY STUDY OF A 70-K PERIODIC SORPTION CRYOCOOLER AND THERMAL MANAGEMENT OF SODIUM SULFUR BATTERY MODULE

Amir Karimi
Associate Professor
Division of Engineering
The University of Texas at San Antonio

Abstract

This report contains the results of investigations conducted on two separate projects. The first is a thermodynamic feasibility study of a periodic sorption cryocooler, operating between a heat source at 70 K and a heat sink at 273 K. The second embodies an analysis for thermal management of a sodium-sulfur (Na-S) battery module for a GEO flight experiment. The performance objective of the cryocooler is to provide 0.5 W of cooling using a maximum thermal power of 30 W. A three-stage cascade electrochemical-sorption refrigeration cycle which consists of a heat source stage powered by an oxygen electrochemical compressor, a heat sink stage employing two sorption compressor beds, and an intermediate stage connecting the two stages is considered for thermal performance analysis. The intermediate stage, which operates intermittently, utilizes a phase change material (PCM) to facilitate energy storage for the heat dissipation of the heat source stage and uses a single sorption compressor bed powered by an external electrical heater and cooled by evaporation of the working fluid in the heat sink stage of the cryocooler. A data base containing the freezing temperature, in the range of 2-300 K, of over 500 chemical elements and compounds was created to assist in the selection of suitable working fluids and PCM for the cryocooler. It is shown that, through proper selection of working fluids, the objectives of the cryocooler are attainable by a two-stage cascade cycle. Further investigations on the vapor adsorption capacity of solid sorbents, the physical properties of gas-sorbent pair, and the sorption compressor beds' energy regeneration concepts are recommended. The thermal behavior of a Na-S battery module containing eight cells and covered with multi-layer insulation (MLI) is analyzed. Consideration is given to storage of energy in the battery during the discharge period and use of the stored energy during the charge period to maintain the batteries within the operational temperature range of 325 to 375 °C. The temperature-time history during a charge-discharge cycle indicated that without the battery, cells cannot store sufficient energy during discharge to maintain the batteries within the specified temperature range. Use of additional energy storage material is recommended to avoid requiring heaters during the charge period. The use of phase change material (PCM) for energy storage is also highly recommended.

A FEASIBILITY STUDY OF A 70-K PERIODIC SORPTION CRYOCOOLER AND THERMAL MANAGEMENT OF SODIUM SULFUR BATTERY MODULE

Amir Karimi

INTRODUCTION

This report contains the results of studies conducted on two separate projects. It consists of a thermal analysis of a 70-K periodic sorption cryocooler and a study on thermal management of a sodium-sulfur battery module. Therefore, each section of this report is subdivided to cover both topics.

A. Sorption Cooling Cycle: Reliable, low-cost, high-efficiency vapor-compression cooling/heating cycles have been the focus of considerable research for many years. Most efforts have been directed towards development of two basic concepts [1]: (i) mechanically powered vapor compression cycles, and (ii) liquid-vapor absorption systems. While the engine-driven cycles have demonstrated a high coefficient of performance, their mass, size and maintenance requirements have limited their use in specific applications such as space. The liquid vapor absorption systems have suffered from low efficiency, and the efforts to improve their efficiencies have increased the cost and size beyond economically allowable limits. The complexity, size, and gravity dependent processes of liquid absorption system also tend to limit their use in space applications.

A regenerative, solid-vapor adsorption compression cycle appears to hold great promise for overcoming the complexity, high cost, and low efficiency of these past efforts. It also offers advantages for a variety of applications since: 1) safe refrigerant may be used, 2) minimal power is required for pumping, 3) the system is composed of simple components, and 4) selection of various refrigerant/ adsorption pairs enables systems to meet a variety of applications.

Solid-vapor adsorption compression cycles satisfy many applications. Waste heat in industrial plants, exhaust gases from internal combustion engines, or the combustion products in jet engines can provide an efficient means to power sorption compressors for space cooling. Alternative energy sources such as solar energy or geothermal energy may be used to meet sorption compressor power requirements fully, or partially. A specific application is a thermal heat sink which is a vital component of the Primary Life Support System (PLSS) in the Extravehicular Mobility Unit (EMU) for space activities. NASA has been searching for a non-venting regenerative cooling device to address the EMU thermal control requirements in advanced extravehicular activities (EVA) of the Shuttle Orbiter, the Space Station EVA mission, and the Man-Mars mission [2-5]. United States' armed forces have considered various cooling devices for reducing the physiological thermal stress resulting from ground crew personnel

performing their duties in a hazardous environment [6-11]. A sorption vapor compression cycle, due to its small size and light weight, offers an ideal alternative heat sink concept for the liquid-cooled-garment (LCG).

A basic sorption cooling/heating cycle is very similar to a basic vapor compression cycle Evaporator, condenser and expansion valve are common to both cycles. The only difference between the two cycles is that the engine-driven compressor is replaced with a sorption compressor powered with thermal energy. Low-pressure refrigerant vapor (adsorbate or sorbate) enters the compressor where it is adsorbed on the internal surfaces of a microporous solid. Heat from a relatively high temperature source is transferred to the compressor bed resulting in vapor desorption and pressurization. High vapor leaves the compressor and enters the condenser.

The sorption compressor contains two beds. Each compressor bed consists of a cell containing a solid adsorbent (e.g., activated carbon charcoal). Flow of refrigerant into and out of the cells is controlled by special valves. Sorption beds are alternately heated and cooled. The bed being heated disrobes refrigerant, which pressurizes that bed. When the pressure reaches the condenser pressure, the refrigerant relief valve opens to allow the refrigerant to flow to the condenser as the bed continues to be heated to maintain high pressure. The bed being cooled adsorbs refrigerant vapor, which depressurizes that bed until the pressure reaches the evaporator pressure, at which time the relief valve from the evaporator opens to allow refrigerant flow to the bed. The bed continues to be cooled to maintain the evaporator low pressure. When the first bed (heated bed) fully disrobes the vapor and the second bed (cooled bed) fully adsorb the refrigerant, the process is reversed, allowing each bed to return to its original state, thereby, completing the cycle. Since each sorption compressor bed undergoes alternate periods of cooling or heating, incorporating a regenerator optimizes the thermal energy power input and improves the performance of the compression cycle.

The last few years have seen considerable efforts directed toward design of light weight, passive compressors. The need for reliable cryocoolers to cool infrared (IR) sensor focal planes and optics in Strategic Defense Initiative Organization (SDIO) applications, Air Force launch detection systems, and satellite surveillance devices have resulted in a renewed interest and demand for electrochemical, reversed stirling engines, and sorption compressor concepts. The emphasis has been on development of cryocooler concepts with heat source temperatures in the range of 10 to 120 K.

B. Sodium Sulfur Battery System: The battery system most frequently used for aerospace applications is the solid state nickel cadmium (Ni-Cd) battery. The operating temperature range for these batteries is 0-15 °C. The thermal control limits the temperature gradient in a Ni-Cd battery module to 5 °C to minimize the voltage differential between the cells. These batteries are being gradually phased out by higher energy capacity systems.

The nickel hydrogen (Ni-H) batteries are replacing the Ni-Cd systems in space applications. Unlike other batteries which are solid state, Ni-H batteries are liquid systems which operate at high pressure (approximately 600 psia); therefore, they must be contained in pressure vessels. The ideal operation temperature for Ni-H batteries is in the range 0-15 °C. Again, the temperature gradient between the cells should not exceed 5 °C in a Ni-H battery module.

The sodium sulfur (Na-S) battery, which has been under development during the last two decades, is a solid state, high-energy density system capable of undergoing rapid recharge, thus eliminating the trickle charge mode associated with other battery systems. During the initial period of development, focus was centered on the design of low-cost, Na-S energy storage systems for use in electrical vehicles or for employment as electrical power load leveling devices.

Only in recent years has the Na-S battery space application potential been realized. The emphasis has now shifted from cost to reliability and durability required for unmanned spacecraft. These efforts are undergoing progressive stages of development. The development of Na-S batteries has been concentrated on geosynchronous-orbit (GEO) applications. The work on development of these batteries for low-earth-orbit (LEO) applications has been extremely limited. Research in the U.S. and England has produced Na-S battery prototype configurations for future generation satellites. Detailed specifications on the British battery system, which is smaller in size than its American counterpart, were not available at the time of this study.

Hughes Aircraft Company, under contract with the US. Air Force, has been developing and fabricating Na-S battery systems for next generation satellite applications [12]. The main objective of the development program has been to design high energy density (greater than 50 Wh/lb) battery systems capable of delivering power up to 10 kW for GEO space applications. Two types of battery modules have been under consideration: 0.5-kW, 28-volt, flight test module and 10-kW, 100-volt, life test module. The flight test breadboard module contains 16 cells and the life test module embodies 64 cells. The cells are connected in series through bypass switches to prevent total battery module failure.

Sodium-Sulfur battery cells are approximately 3.5-cm in diameter and 27 cm long and weigh 500 grams. Schematic drawings of phase I and phase II designs, taken from [12]. Tubular cells are divided into an anodic reaction zone (cell core) which contains liquid sodium and a cathodic reaction zone containing liquid sulfur. The two zones are separated by a solid electrolyte, β"-alumina, toughened by zirconia impregnation. The battery reaction during the discharge cycle is exothermic, thus generating heat. It is endothermic during the initial phase of the charge cycle and becomes exothermic at later stages. The cathode container is a 0.040-inch-thick, chrome plated E-Brite in phase-I design and a 0.020-inch-thick, chrome-plated molybdenum. A 2.09-inch diameter flange is welded to the bottom of the cathode container tube

which acts as a radiator for heat dissipation during high rate battery discharge. Cells are mounted in a titanium honeycomb structure side by side in a module. The cells are thermally insulated on all sides, except the base flange, to minimize any cell to cell gradient.

The ideal operation temperature for the Na-S battery is in the range 325-375 °C. The temperature gradient between the cells in a module is limited to 30 °C band. At the beginning of life (BOL), the life Na-S test module is estimated to have 61.1 Wh/lb of specific energy to 100 percent depth of discharge (DOD). At the end of life (EOL) specific energy for the same module is projected to be 54.4 Wh/lb with no failed cells and 47.9 Wh/lb with six failed cells. Detailed specifications for Na-S battery cells are given in [12].

The Na-S battery module requires a thermal management system to maintain the heat within the battery system during charging (due to heat losses from the battery module) and to balance the dissipation during the discharge cycle. The current design for heat dissipation during high-rate discharge is a single louver which opens fully, exposing the flanged bottom of cells to deep space. A 36-layer, multi-foil, aluminum-zirconia blanket covers each battery module and louver door to minimize the heat losses. The louver door opens only during the battery discharge. Active thermal controls, such as louvers, have been identified as a problem with the Na-S battery system. One of the problems of passive thermal control is the lack of available (and well characterized) working fluids in the 300-375 °C range, which is too hot for most organics and refrigerants, yet too cold for liquid metals. Cesium heat pipes have been considered as a passive thermal control for use in both the LEO and GEO systems.

OBJECTIVES

A. Sorption Cooling Cycle: The main objective of this study is to conduct a feasibility study on an electrochemical/sorption cascade-cryocooler concept. The demonstration cryocooler under consideration is an alternative design approach to the three-stage sorption cryocooler development study in [13]. This cryocooler consists of three vapor compression cycles arranged in series. The three stages will be referred to as the "lower stage," identifying the cycle which remove heat from the cold surface; the "upper stage", referring to the cycle which dissipates heat to ambient; and the "intermediate stage", identifying the cycle which connects the lower and upper stages. The compression of vapor in each stage of the cascade cycle is achieved in a carbon-gas sorption compressor. There was an initial attempt to use mixed gases in the "upper stage" of the cryocooler. An experimental study conducted by (NIST) on mixed gas sorbent capacity showed poor results. Since the adsorption data exhibited an order of magnitude less than that required for the system to be competitive, further investigation on mixed gases was suspended [14].

An alternative design concept, as shown in Fig. 1, was adopted whereby the lower-stage compressor was replaced with an electrochemical compressor. An important feature of this cycle is the new and innovative concept of an intermittent operation in the intermediate stage of the cascade cryocooler. The precooling (condensation) process in the lower stage is achieved through energy storage in a phase-change material (PCM). Heat removal from the PCM is by a pure gas two-stage cascade system, the intermediate stage and the upper stage of the cryocooler. The intermediate stage is a single-bed, pure vapor sorption compressor cycle which removes heat from PCM in a periodic operation. The sorbent bed is cooled by the upper stage of the cryocooler and is heated by an electric heater in an intermittent fashion. The upper stage is powered by the waste heat from the electrochemical compressor and its working fluid is precooled (condensed) by heat dissipation to ambient (radiator). The working fluid loop in the lower stage can be used for heat transfer from the electrochemical compressor to the upper stage sorption compressor beds.

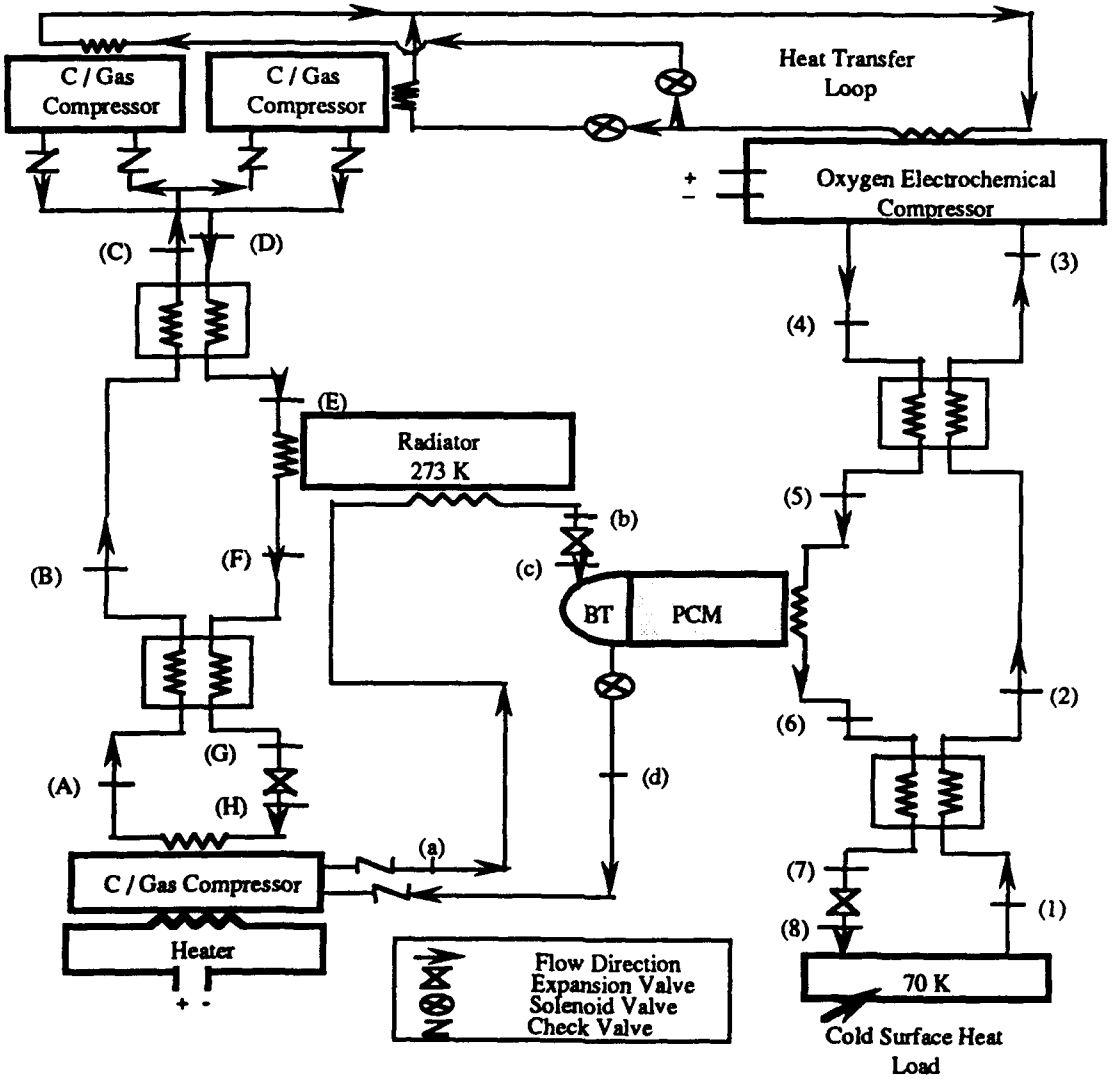


Fig. 1. Three Stage Cascade Cryocooler

The targeted goal for the coefficient of performance of the cryocooler is 0.25-0.3 W of cooling per 25 W of thermal power. The heat removal from the cold surface is achieved through evaporation of refrigerant in the lower stage at 70 K, and heat rejection occurs by precooling (condensation) of working fluid in the upper stage at 273 K. This work is limited to a thermodynamic feasibility study of the cryocooler concept described in this section.

B. Sodium Sulfur Battery Module: The main objective of this portion of the study is to evaluate the thermal control behavior of a Na-S battery module proposed for a flight experiment. The module consists of eight Na-S batteries connected in series and placed in contain The distance between each adjacent battery is one inch and the distance between the batte d the containment wall is also one inch. A specific goal of this effort is to develop a passive control system for the module that i) reduces energy losses during the charge period, ii) maintains the peak temperature during recharge within the recommended operation temperature range of 325-375 °C, and iii) reduces the need of auxiliary heaters during the charge/discharge cycle.

ANALYSIS

A. Sorption Cooling Cycle: Many factors influence the thermal performance of sorption cryocoolers. The major variables commonly considered in a baseline thermal design of a sorption refrigeration cycle include: i) selecting environmentally safe substances which meet the requirements for each stage of a cascade vapor compression cycle; ii) identifying solid sorbents with high adsorption capacity and good thermal characteristics; iii) searching for available data or conducting experiments to pick highly efficient vapor sorbate/solid sorbent pair for each stage of the cycle; iv) choosing suitable and safe phase-change materials (PCM) for the thermal storage of energy resulting from precooling the refrigerant in the lower stage; and v) establishing techniques to enhance the heat transfer processes for the compressor beds and regeneration system.

The cryocooler cycle under consideration requires a heat load temperature (evaporator temperature) of 70 K and a heat sink temperature (condenser or precooler temperature) of 273 K. Application of a single-stage vapor compression cycle requires that the working fluid (refrigerant) undergo an evaporation process at a temperature greater or equal to the triple point and a condensation process below the critical point. Although it is possible for the heat rejection to occur in the super critical range of the working fluid (no phase change), it is highly desirable that the working fluid undergo the condensation process to enhance the performance of the cooling cycle. Therefore, a single working fluid must have a freezing point (triple point) less than, or equal to, the thermal load temperature of 70 K and a critical point above the heat sink (ambient)

temperature of 273 K; thus, the pressure ratio for the compression is constrained by the critical pressure of the working fluid.

A data base containing the freezing temperature, in the range of 2-300 K, of over 500 chemical elements and compounds was created to assist in selection of suitable working fluids and PCM for the cryocooler. It contains the normal boiling and freezing points, molecular weight, critical pressure and temperature, and latent heat of vaporization for each substance. References [15-21] were used for the creation of the data base. A list of data for substances with freezing points less than 70 K are given in Table 1. Since the critical temperatures for all substances fall below the heat rejection temperature of 273 K, it indicates that a single refrigerant cannot be employed to meet both the heat source and heat sink temperature requirements of the cryocooler. Therefore, in such a cryogenic cooling cycle, it is necessary to employ a multi-stage, cascade refrigeration cycle, using a different working fluid in each stage.

Any of the refrigerants listed in Table 1 can be used as working fluids in the lower stage of the cascade cryocooler to satisfy the temperature requirement of the cold surface. However, it is desirable that the selected substance have high values for critical temperature and latent heat of vaporization at 70 K. Oxygen difluoride, fluorine, oxygen, nitrogen, nitrogen trifluoride, and carbon monoxide are good candidates for the working fluid in the lower stage. For all these refrigerants, the normal boiling point temperatures are above 70 K, indicating that the evaporator of the lower stage must operate at vacuum pressure. Oxygen was selected for use in the lower stage of the cryocooler since an electrochemical compressor is commercially available for this refrigerant.

Table 1. A list of chemical elements and compounds with normal freezing temperature below 70 K.

Refrigerant Chemical Name	Chemical Formula	M	T_f, K	T_b, K	T_c, K	P_c, bar	h_{fg}, kJ/kg	h_{if}, kJ/kg
helium (R-704)	He	4.003	1.0	4.2	5.2	2.3	20.9	
helium 4	He	4.003	3.5	4.1	5.2	2.3	20.5	5.0
hydrogen (normal) (R-702)	H_2	2.019	14.0	20.4	33.2	13.2	445.7	58.0
hydrogen (para) (R-702P)	H_2	2.019	14.0	20.3	32.9	12.8	445.4	62.8
hydrogen deuteride	HD	3.023	16.6	22.1	36.0	14.8	353.4*	
deuterium (normal)	D_2	4.032	18.6	23.5	38.4	16.5	298.18*	
deuterium (equilibrium)	D_2	4.032	18.7	23.6	38.2	16.5	303.2	
neon (R-720)	Ne	20.183	24.5	27.1	44.4	34.0	85.9	15.8
oxygen difluoride	F_2O	53.995	50.0	128.4	215.0	49.6	187.7*	
fluorine	F_2	37.997	53.5	85.0	144.4	52.2	171.9	25.6
oxygen (R-732)	O_2	31.999	54.4	90.2	154.8	50.8	212.7	13.7
air (R-729)	21%O_2, 78%N_2, 1%Ar	28.970	57.0	78.8	132.4	37.7	201.5	23.2
nitrogen (R-728)	N_2	28.013	63.1	77.4	126.3	34.0	198.6	25.8
nitrogen trifluoride	F_3N	71.002	66.4	144.4	234.0	45.3	165.4*	
carbon monoxide	CO	28.011	68.2	81.7	132.9	35.0	215.6	29.8

The oxygen electrochemical (ceramic membrane) compressor operates between 750-800 °C (1023-1073 K) and can compress oxygen up to 26 atm (26.3 bars). The compressor produces approximately 28 W of waste heat which can be used to provide thermal power for the charcoal compressor beds at the upper stage. Conduction losses are estimated to be 4.5 W and radiation losses are less than 3 W. It is still possible to operate the total system with 23 W of power for 0.5 W of gross refrigeration.

The maximum operating pressure in the compressor is constrained by the critical temperature of oxygen and the melting temperature of the selected PCM. For example, the condensation (precooling) temperature for oxygen at 26 bars is approximately 138.5 K. Therefore, the melting temperature of the selected PCM must be less than 138.5 K. For melting temperatures of PCM greater than 138.5 K the compressor must operate at a pressure higher than 26 bars. The latent heat of vaporization, h_{fg}, of oxygen decreases with increasing temperature. It changes from 230.5 kJ/kg at 70 K to 125 kJ/kg at 140 K, a 55% decrease. The value of h_{fg} drops sharply and vanishes as the critical temperature (154.6 K) is approached. Therefore, at temperatures close to the critical point, a major part of the working fluid precooling occurs through sensible heat dissipation rather than latent heat.

The precooling of the working fluid in the lower stage is achieved through heat dissipation and energy storage in the PCM canister. Therefore, the melting temperature of the selected PCM must be less than the condensation of working fluid in the lower stage. Assuming a condensation temperature of 138.5 K during the precooling of oxygen in the lower stage requires the selection of a PCM having a melting point below this temperature. Our data base contains over eighty substances with normal freezing temperatures in the range of 70 K to 140 K. Table 2 contains a partial list of the most common substances having normal freezing points in this temperature range. Preference is given to selection of a PCM which has a melting temperature close to the operating temperature of the condenser in the lower stage.

In addition to satisfying the condenser's temperature constraint, the selected PCM should have a high value for the latent heat of vaporization, h_{fg}. Table 2 indicates that isobutylene, C_4H_8, (T_f = 134.4, h_{fg} = 72.6 kJ/kg) and ethyl chloride, R-160, C_3H_5Cl, (T_f = 138.8, h_{fg} = 68.1 kJ/kg) are suitable candidates for PCM. If the condenser of the lower stage operates at a temperature above 146 K, propyl alcohol C_3H_8O (T_f = 146.0 K, hfg = 86.5 kJ/kg) can provide an additional choice as a PCM.

Table 2 can be used to assist in the selection of a working fluid for use in the intermediate and upper stages of the cryocooler. Propane (R-290) - C_3H_8 , ethane (R-50) - C_2H_6, butane (R-600) - C_4H_{10}, and pentane - C_5H_{12} are suggested for use in the upper stage and ethane, propane, and carbon tetrafluoride (R-14) - CF_4 are recommended for the intermediate stage [14]. The fluid temperature in the bladder tank (BT) of the second stage must be less than the freezing

temperature of the PCM. Since the refrigerant used in the upper stage should be able to dissipate heat into the radiator, thus its critical temperature must be greater than 273 K.

The thermodynamic analysis for each stage of the cryocooler is based on the following assumptions: i) saturated vapor leaves the evaporator, ii) saturated liquid exits the precooler/condenser, and iii) vapor at the compressor outlet is at the maximum operating temperature of the compressor. Other state property evaluations are based on the effectiveness value for the heat exchangers.

Table 2. A partial list of chemical elements and compounds with normal freezing point temperatures in the range of 80-150 K

Refrigerant Chemical Name	Chemical Formula	M	T_f, K	T_b, K	T_c, K	P_c, bar	h_{fg}, kJ/kg	h_{if}, kJ/kg
ozone	O_3	47.998	80.5	181.2	261.1	55.7	437.1*	226.0
argon (R-740)	Ar	39.948	83.9	87.3	150.9	49.0	161.6	28.1
propane (R-290)	C_3H_8	44.100	85.5	231.1	370.0	42.6	427.1	44.4
l-butene (butylene)	C_4H_8	56.108	87.7	266.9	417.2	42.8	391.0	38.1
propylene glycol (R-1270)	C_3H_6	42.082	87.9	225.5	364.9	46.2	342.0	71.4
carbon tetrafluoride (R-14)	CF_4	88.010	89.3	145.2	227.5	37.4	134.2	2.9
octafluoropropane (R-218)	C_3F_8	188.017	90.0	236.5	345.1	26.8	106.3*	
ethane (R-170)	C_2H_6	30.070	90.4	184.3	305.4	48.9	491.1	95.3
methane (R-50)	CH_4	16.040	90.9	111.7	190.7	46.4	510.4	32.6
chlorotrifluoromethane (R-13)	$CClF_3$	104.470	92.0	191.7	302.0	38.7	149.7	
krypton	Kr	83.800	104.0	120.2	209.4	55.2	108.0	10.9
ethylene (ethene) (R-1150)	C_2H_4	28.050	104.3	169.4	282.5	51.2	482.6	120.0
trifluorobromoethane (R-13B1)	$CBrF_3$	148.930	105.4	215.4	340.2	39.6	118.9	
nitric oxide	NO	30.006	109.5	121.4	180.0	64.8	460.4	76.5
chlorodifluoromethane (R-22)	$CHClF_2$	86.480	113.1	232.4	369.2	49.8	233.3	
isobutane ((R-600a)	C_3H_{10}	58.130	113.7	261.4	408.2	36.5	367.1	
73.8% R12/ 26.2%R-152a (R-500)	CCL_2F_2/ $CH_3CH\,F_2$	99.310	114.3	239.7	378.7	44.3	200.8	
dichlorodifluoromethane (R-12)	CCL_2F_2	120.930	115.4	243.4	385.2	41.2	166.0	34.4
chlorotrifluoroethylene (R-1113)	C_2ClF_3	116.469	116.0	245.3	379.0	40.5	182.0*	
trifluoromethane (R-23)	CHF_3	70.020	118.1	191.1	298.8	48.4	243.3	
vinylchloride (R-1140)	C_2H_3Cl	62.499	119.4	259.8	425.0	51.5	356.6	
vinlidene fluoride (R-1132a)	$C_2H_2F_2$	64.035	129.0	187.5	302.9	44.6	238.3	
ethyl fluoride	C_2H_5F	48.060	129.9	235.5	375.3	50.2	427.3*	
vinyl fluoride (R-1141)	C_2H_3F	46.044	130.0	201.0	327.9	52.4	368.6*	
tetrafluoroethylene (R-1114)	C_2F_4	100.160	130.7	197.2	306.5	39.4	166.3*	
methyl fluoride (R-41)	CH_3F	34.033	131.4	194.7	315.0	56.0	502.1*	
isobutylene	C_4H_8	56.108	132.8	266.2	417.9	40.0	394.4	58.8
butane (R-600)	C_4H_{10}	58.130	134.6	272.7	425.2	38.0	385.2	44.7
ethyl chloride (R-160)	C_3H_5Cl	64.520	136.8	285.6	460.4	52.7	383.0	68.1
dichlorofluoromethane (R-21)	$CHCl_2F$	102.920	138.1	281.9	451.7	51.7	238.6	
chlorodifluoroethane (R-142b)	$C_2H_3ClF_2$	100.500	142.0	263.4	410.3	41.2	215.1	
n-pentane	C_5H_{12}	72.151	143.4	309.2	469.6	33.7	358.2	
propyl alcohol	C_3H_8O	60.098	146.0	371.0			779.0	86.5

The heat exchanger effectiveness is defined as the ratio of the actual heat transfer to the maximum theoretical value for heat transfer. In a heat exchanger where both fluid streams are in the vapor phase and when the mass flow rate of both fluids are the same, the effectiveness, ε, can be expressed as:

$$\varepsilon = \frac{h_{H,i} - h_{H,o}}{h_{H,i} - h_{C,o}} = \frac{h_{C,o} - h_{C,i}}{h_{H,i} - h_{C,o}} \quad (1)$$

where, h is the enthalpy, the subscripts H and C denote the properties of hot fluid and cold fluid, respectively , and i and o refer to inlet and outlet, respectively. When one of the fluid streams is in the vapor phase and the other is in the liquid regime, the heat exchanger effectiveness is given by:

$$\varepsilon = \frac{h_{H,i} - h_{H,o}}{c_{p,\,min}(T_{H,i} - T_{C,o})} = \frac{h_{C,o} - h_{C,i}}{c_{p,\,min}(T_{H,i} - T_{C,o})} \quad (2)$$

where, T is the temperature and $c_{p,min}$ represents the smaller value of the mean specific heats for the two fluid streams. The effectiveness is a function of two parameters

$$\varepsilon = fn\ (NTU,\ c_{p,min}/c_{p,max}) \quad (3)$$

where, NTU, the number of transfer units is defined as:

$$NTU = \frac{U\,A}{(m\,c_p)_{max}} \quad (4)$$

U is the overall heat transfer coefficient based on thermal resistance between two fluid streams, A denotes the heat exchange surface area, and $(m\,c_p)_{max}$ is the larger value of the product of mass flow rate and specific heat (total heat capacity) for the two fluid streams.

The heat transfer analysis for the condensers and evaporators is based on two-phase flow models [22]. Multi-dimensional transient heat conduction equations or Heisler's charts [23] can be utilized to determine the thermal behavior of sorption compressor beds. The functional parameters for heat conduction analysis are Fourier number, $Fo = \alpha\, t/L^2$, and Biot number, $Bi = h\, L/\, k$, where α, t, L, h, k are thermal diffusivity, time, characteristic length, heat transfer coefficient and thermal conductivity, respectively.

B. Sodium Sulfur Battery System: A summary of the projected energy and thermal characteristics of phase III-Na-S battery cells, pertinent to this work, is listed in Table 3. The average discharge heat dissipation for each battery is 4.18 W at BOL and 5.34 at EOL. The energy dissipated during a charge-discharge cycle is estimated to be 4.06 Wh at BOL and 5.10 Wh at EOL for each battery. Considering eight batteries in a module, the average heat dissipation for the module would be 33.44 W at BOL and 42.72 W at the EOL. The module heat dissipation during a charge-discharge cycle is 32.48 at BOL and 40.8 W at the EOL. The cell heat capacity is estimated to be 0.0906 Wh/°C resulting in a heat capacity of 0.7248 Wh/°C.

Table 3. Phase III-Na-S Battery Cell Projected Characteristics [12]

Item	Beginning of Life (BOL)	End of Life (EOL)
Weight	443.6, g	43.6, g
Operational Life	10 years at operating temp.	10 years at operating temp.
Cycle Life	1,000 Chg./Dischg cycle	1,000 Chg./Dischg cycle
Design Capacity	25.9, Ah	27.2, Ah
Design Depth of Discharge (DOD)	10-60 percent	18-71 percent
Discharge Current	21.6, A	22.7, A
Average Discharge Current	1.93, V	1.89, V
Average Discharge Power	41.6, W	42.9, W
Charge Current	1.14, A	1.19, A
Average Charge Voltage	2.08, V	2.08, V
Maximum Discharge Time	1.2, hr	1.2, hr
Minimum Charge Time	22.8, hr	22,8, hr
Average Discharge Heat Dissipation	4.18, W	5.34, W
Max. Discharge Heat Dissipation	6.01, W	6.99, W
Energy Dissip./ Chg.-Dischg. Cycle	4.06, Wh	5.10, Wh
Cell Heat Capacity	0.0906, Wh/^{o}C	0.0906, Wh/^{o}C
Electric-to-Electric Efficiency	92.5 percent	91.0 percent
Average Specific Power Used	42.5, W/lb	43.9, W/lb
Available Energy Density	0.40, Wh/cm^3	0.36, Wh/cm^3
Design Specific Energy(to design DOD)	51.0, Wh/lb	52.7, Wh/lb
Available. Specific. Energy(to 100% DOD)	88.4 Wh/lb	78.8 Wh/lb

Each battery cell has an average diameter of 1.4-in. (3.556 cm.) and has a length of 9.44-in. Eight batteries are placed in a container having dimensions of 8.2-in. x 8.2-in. x 10.4-in. The total surface area of the container is 475.6-in^2 (0.3068 m^2) The total surface area of the battery cells exposed to the container is determined to be 343.06-in^2 (0.2478 m^2). It is assumed that the heat transfer between the battery cells and the module container walls is by radiation only. To minimize the heat losses, the container is covered with multi-layer insulation (MLI) having an effective emittance in the range of 0.005 to 0.05 [28]. The effective emittance, ε^*, is defined as:

$$\varepsilon^* = \frac{Q}{A\,\sigma\left(T_H^4 - T_C^4\right)} \qquad (5)$$

where, T_H and T_C are the hot and cold boundary absolute temperatures in K, A is the surface area, Q is the net heat transferred, and σ is the Stefan-Boltzmann constant (5.67×10^{-8} W/m^2.K^4).

Effective emittance is directly dependent upon the size of the insulated area. As a first trial, an attempt was made to determine whether it is possible to maintain the battery within the operation temperature range during a charge/discharge cycle. This requires storage of the generated heat in the module during the discharge period and use of the stored energy during the charge period.

The heat transfer analysis of the battery module is based on radiation exchange between two diffuse surfaces, namely the battery cells and the module's inside surface. The formulation is based on the assumptions that: i) each surface is opaque and gray, ii) the emission from each

surface is diffuse, iii) the reflection from each surface is diffuse, and iv) the radiosity of each surface is uniform (a uniformly irradiated and isothermal surface). It is considered that a layer of aluminum foil covers the inside surface of the module. Identifying the battery cells as surface 1 and denoting the interior of the container as surface 2, the radiosity of each surface, J_i, can be expressed as:

$$J_i = \varepsilon_i E_{bi} + (1 - \varepsilon_i) \sum_{k=1}^{2} J_k F_{ik} \quad ; \qquad i = 1, 2 \tag{6}$$

where, F_{ik}, ε_i and $E_{bi} = \sigma T_i^4$ are the shape factor, the emisivity and the emissive power of ith surface. The radiant heat transfer, Q_i, from surface i, A_i, can be obtained from

$$Q_i = \frac{\varepsilon_i A_i}{1 - \varepsilon_i} (E_{bi} - J_i) \tag{7}$$

The module's energy interaction with its environment is by radiation. Treating the environment as a black body enclosure at a temperature , T_e, and ignoring the conductive thermal resistance, an energy balance about the container wall yields

$$E_{b2} - E_e = (J_2 - E_{b2}) \left[\frac{\varepsilon_2}{\varepsilon^* (1 - \varepsilon_2)} + \frac{\varepsilon_2}{\varepsilon_o (1 - \varepsilon_2)} \right] \tag{8}$$

where, $E_e = \sigma T_e^4$, ε_o denotes the emisivity of the aluminum foil and ε^* is the effective emisivity for the MLI. Combining Eqs. (6) and (8) results in the following relations:

$$\left[\frac{1}{\varepsilon_1} - \frac{(1 - \varepsilon_1)}{\varepsilon_1} F_{11} \right] J_1 - \frac{(1 - \varepsilon_1)}{\varepsilon_1} F_{12} J_2 = E_{b1} \tag{9}$$

and

$$-\left[\frac{(1 - \varepsilon_2)}{\varepsilon_2} + \frac{1}{\varepsilon^*} + \frac{1}{\varepsilon_2} \right] F_{21} J_1 - \left\{ \frac{1}{\varepsilon^*} + \frac{1}{\varepsilon_2} + \frac{1}{\varepsilon_0} - \left[\frac{(1 - \varepsilon_2)}{\varepsilon_2} + \frac{1}{\varepsilon^*} + \frac{1}{\varepsilon_2} \right] F_{22} \right\} J_2 = E_e \tag{10}$$

Equations (9) and (10) can be solved simultaneously and the result can be used in Equation (7) for the evaluation of the heat transfer from each surface.

Treating the battery cells as a lumped capacity system (uniform temperature at any given time), an energy balance results in the following governing equation for the temperature variation during the charge/discharge cycle.

$$(Q_{gen} - Q_{rej}) = m c_p \frac{dT}{dt} \tag{11}$$

where, Q_{gen} is the rate of heat generation by the battery cells , Q_{rej} is the heat module heat dissipation, determined from Eq. (7), and $(m c_p)$ is the thermal capacity of the batteries.

DISCUSSION

A. Sorption Cooling Cycle: State points in Fig. 1 are used to describe the thermal performance of the cryocooler. Oxygen is considered to be the working fluid in the lower stage of the cascade cycle. Thermodynamic analysis of the lower stage is based on the following assumption. Saturated vapor leaves the evaporator, state 1, at 70 K, and saturated liquid exits the

precooler/condenser, state 6, at a pressure which is equal to the maximum operating pressure of the electrochemical compressor. The superheated vapor temperature at the compressor outlet is 1023 K.

Equations (1) through (4) are used to estimate the effectivness of heat exchangers and to evaluate the thermodynamic properties at state points 2, 3, 5, and 7. It is determined that, for all heat exchangers in the cycle, the value of c_{min}/c_{max} falls within a range of 0.5-1.0. The charts for counterflow or tube and shell heat exchangers display effectiveness values in the range of 0.5 to 0.8 for NTU greater than 2 and c_{min}/c_{max} in a range of 0.5-1.0. A typical value of 0.7 is used for the heat exchanger effectiveness.

Tables in [19] are used for the thermodynamic properties of each state. For oxygen, the property entries in these tables are limited to a maximum temperature of 340 K. For temperatures above 340 K, the enthalpy, h, for a superheated vapor is determined by

$$h(T) = h(T_{ref}) + \int_{T}^{T_{ref}} c_p(T)\, dT \qquad (12)$$

where, T_{ref} and $h(T_{ref})$ are the last enteries in the thermodynamic tables.

Oxygen flow rate was evaluated based on 0.5 W of gross cooling. Allowing 0.2 W of heat leak at the cold surface heat exchanger, will provide 0.3 W of net heat removal from the heat source. Using 26 atm. (2.634 MPa) for the maximum pressure of the compressor, it is determined that an oxygen flow rate of 0.00352 g/s is needed for 0.5 W of gross cooling. The specific heat dissipition by the precooler/condenser for this case is 1.3834 W.

Ethylene chloride, C_2H_5Cl, with a latent heat of fusion, h_{if} = 68.98 J/g at normal freezing temperature of T_f=136.7 K and a density of is 0.898 g/cm^3 is suggested for the phase change material to link the lower and intermediate stages of the cryocooler. During the precooling of working fluid in the lower stage heat must be dissipated to the intermediate stage. Since the baseline design is predicated on switching the charcoal compressor in the intermediate stage every six minutes, the minimum PCM required for energy storage is 7.02 grams. Using a conservative value of 7.5 g for mass, the volume needed for PCM is 6.74 cm^3.

For a gross heat removal of 0.5 W, it is determined that the mass flow rate, heat rejection and the PCM mass requirement for energy storage increases with an increase in the precooler operating pressure. The oxygen flow loop can be modified to enhance the thermal performance of the lower stage. The precooling requirements of oxygen will be reduced, if the vapor leaving the compressor is allowed to dissipate heat to the compressor beds and the radiator of the upper stage before entering the condenser. This will reduce the PCM mass requiment for energy storage.

Propane, ethane, butane, and pentane appear to be working fluid candidates for the upper stage and ethane, propane, and carbon tetrafluoride (R-14) are suggested for use in the

intermediate stage [14]. Since both stages dissipate heat to the same radiator, their critical temperature must be greater than 273 K. Therefore R-14 is not a good choice of refrigerant for the cryocooler.

Two types of commercially available solid sorbents in the market were considered for use in the intermediate and upper stage sorption compressor beds--Calgon activated charcoal and Barneby Sutcliffe carbon. Calgon activated carbon APA, pelletrized to 100/g, approximately 2 mm granules, a real density of 2.3 g/cm^3, a packing density in the range 0.4-0.7 g/cm^3, an apparent density of 0.37 g/cm^3, and an internal specific surface area of 1525 m^2/g (verbal quotation from Calgon Co.) [14]). Barneby Sutcliffe briquette is a coconut-shell base charcoal with an apparent density of 0.74 g/cm^3 and an internal specific area equal to 1400 m^2/g (verbal quotation from Barneby Sutcliff Co. [14]).

Design and analysis of adsorption compressors require a knowledge of data on adsorption capacity (mass of vapor adsorbed/ unit mass of sorbant) as a function of temperature and pressure. Physical properties of vapor-sorbent pairs play a significant role in the heat transfer analysis of sorption compressors. This requires qualitative information on properties such as thermal conductivity and specific heat for solid sorbent, and vapor/sorbent pair. It is extremely difficult to measure specific heat of gas-vapor pair used in the sorption compressors.

The working capacity of Calgon and Barneby Sutcliffe charcoals to adsorb refrigerant vapors recommended for the upper and intermediate stages was not known. The physical properties of charcoals and charcoal/vapor pairs were not available. Work was initiated to estimate the capacity of each charcoal to adsorb the refrigerant vapors.

An alternative version of the cryocooler is possible, using a two-stage cascade cycle and eliminating the intermediate stage. The compression of the vapor in the lower stage still is by an oxygen electrochmical compressor. The upper stage uses the waste heat from the electrochemical compressor of the lower for heating the sorption beds. Since the lower stage ejects heat at a temperature of approximately 138 K, there are number of refrigerants available that can meet the temperature constraints of the upper stage. Table 2 provides a list of substances with freezing temperatures less than 138 K and critical points above 273 K.

We recommend that the selection of sorbent material be based upon available information on the absorption capacity and physical properties Since the upper stage compressor beds undergo cooling and heating cycles, a regeneration concept should be integrated into the system.

B. Sodium Sulfur Battery System: Assuming 1.2 hours for discharge, the total heat generation is 40.13 Wh at the BOL and 51.26 Wh at the EOL. The heat dissipation during the charge period is estimated to be -7.64 Wh at BOL and -10.46 Wh at EOL. Therefore, to complete a charge/discharge cycle without requiring heaters, heat dissipation of 1.7 W could be allowed for the module. The shape factor for the surfaces of the battery module was determined using I-

DEAS thermal analysis package. Identifying the battery cells as surface 1 and the container as surface 2, it was determined that F_{11} = 0.22552, F_{12} = 0.77448, F_{21} = 0.55865, F_{22} = 0.44135.

Using emisivity of 0.18 for the battery cells and 0.0035 for MLI, and assuming that the interior surface ot the container is covered with aluminum foil having an emisivity of .04, Eqs. (6)-(10) were used to evaluate the heat losses from the battery module. It was assumed that the environment enclosing the battery module was at 300 K. The module heat losses as a function of (T-300), where T is the cell temperature, is shown in Fig. 2. The influence of the emisivity of MLI on heat losses from the module is displayed in the figure.

Using the calculated values for the module heat loss, Q_{rej}, a value of 33.44 W for Q_{gen} during the discharge ($0 < t < 1.2$ hours) and - 0.3352 W ($1.2 < t < 24$ hours) during the charge period, and assuming 0.7428 Wh/°C for the heat capacity of batteries, Eq. (11) was solved numerically using the Euler's scheme. The resulting temperature time history of the battery cell is shown in Fig. 3. The figure also exhibits the influence of the effective emisivity of the MLI on the temperature - time history of the battery. It is shown that during the discharge period the battery can be maintained within the operation temperature range. However, the battery temperature falls far below the lower limit of operation temperature . Therefore, it is not possible to maintain the batteries, under the conditions described, within the operation temperature range in a charge/discharge cycle without using external batteries.

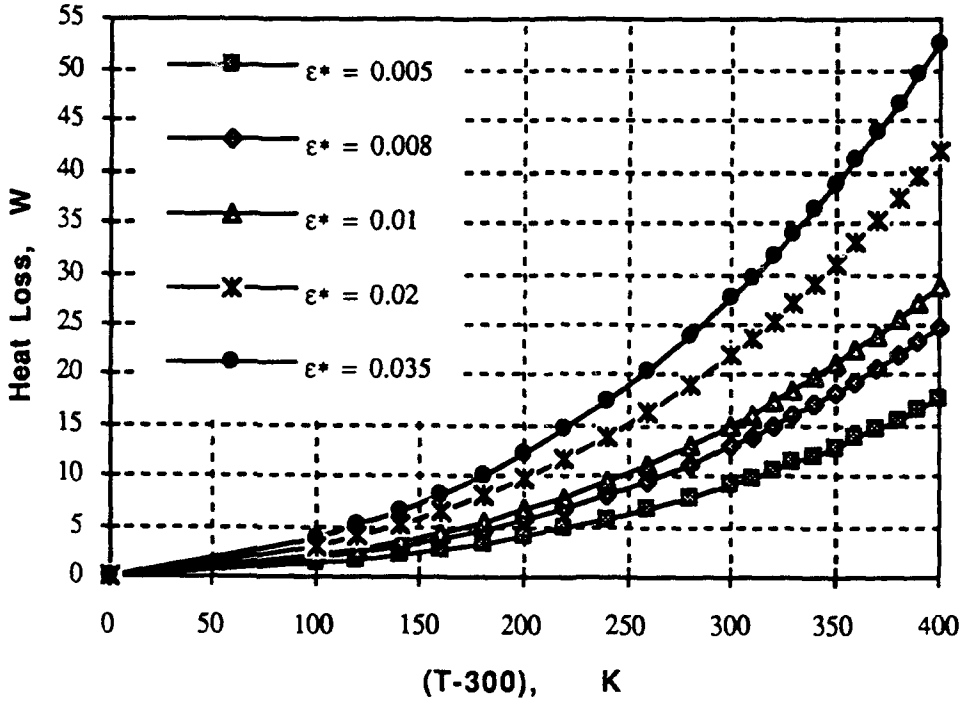


Fig. 2. Heat losses from the Na-S battery Module

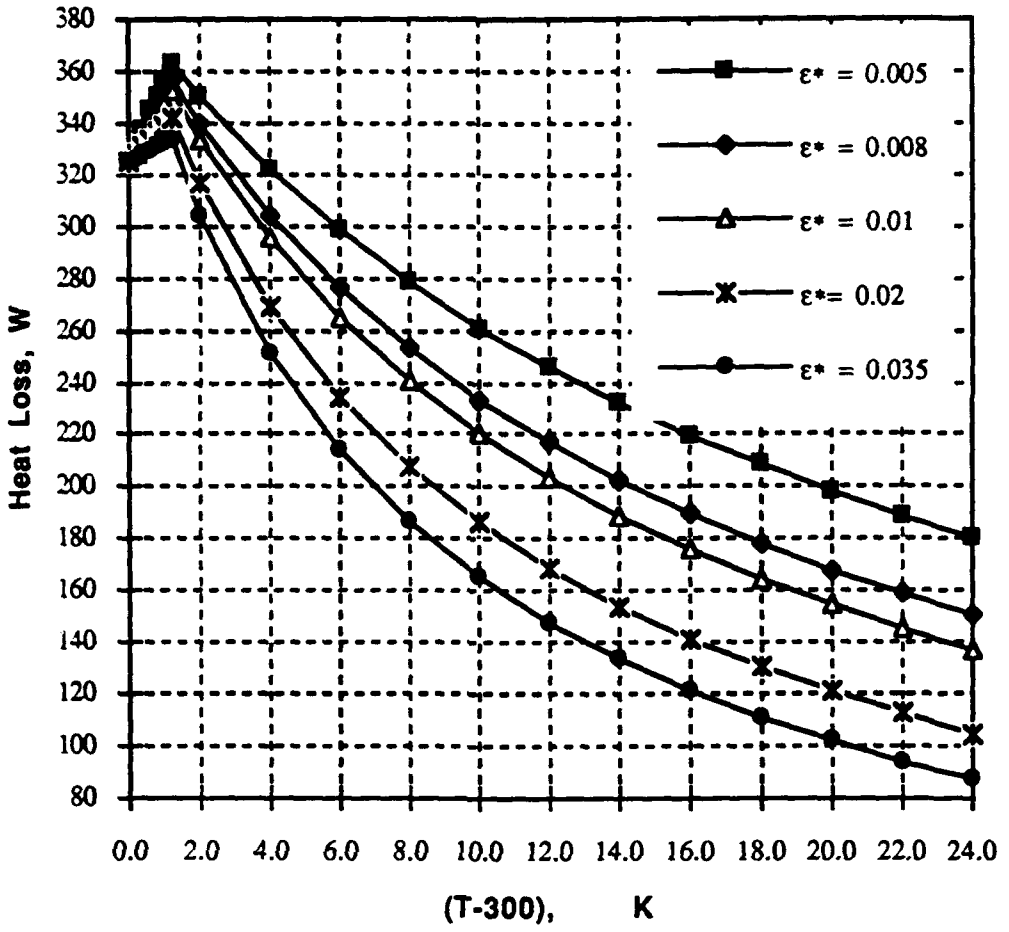


Fig. 3. Temperatur-Time History of Na-S Battery

We recommend the use of a thermal storage medium to enhance the thermal behavior of the battery module. The thermal storage can be achieved through sensible heat or by latent heat of fusion. Our data base provides a list of materials that can be used for sensible energy storage. To minimize the additional weight, the selected material should have a high specific heat. If the volume is a design criteria then the product of density and specific heat should have maximum value. The selected material should also be a good conductor in order to enhance the heat transfer process.

The list of material available for energy storage through latent heat is limited. Most organic compounds have a melting point far below the stated temperature range of operation for the batteries. Few metals and alloys have a melting point temperature in the range of 325 to 375 ^{o}C. A data base was created which contains a selected list of the phase change materials for Na-S battery thermal control applications. Selected PCM for the energy storage should have a high value for heat of fusion. Many alloys in the data base are eutectic or binary systems. This means

that the melting process occurs at a range of temperature dependent upon the concentration of each element in the alloy. The heat transfer analysis of binary or eutectic alloys is complex and is recommended for further study.

CONCLUSIONS AND RECOMMENDATIONS

A. Sorption Cooling Cycle: The performance objective of the cryocooler is to provide 0.5 W of cooling using a maximum thermal power of 30 W. A three-stage cascade electrochemical-sorption refrigeration cycle which consists of a heat source stage powered by an oxygen electrochemical compressor, a heat sink stage employing two sorption compressor beds, and an intermediate stage connecting the other two stages is considered for thermal performance analysis. The intermediate stage, which operates intermittently, utilizes a phase change material (PCM) to facilitate energy storage for the heat dissipation of the heat source stage and uses a single sorption compressor bed powered by an external electrical heater and cooled by the evaporation of working fluid in the heat sink stage of the cryocooler. A data base containing the freezing temperature, in the range of 2-300 K, of over 500 chemical elements and compounds was created to assist the selection of suitable working fluids and PCM for the cryocooler. It is shown that, through proper section of working fluids, the objectives of the cryocooler is attainable by a two stage cascade cycle. Further investigatio. ~n the vapor adsorption capacity of solid sorbents, the physical properties of gas-sorbent pair, and the sorption compressor beds energy regeneration concepts is recommend.

B. Sodium Sulfur Battery System: A thermal analysis was conducted on a sodium-sulfur battery module for flight experiment. Consideration was given to storage of energy in the battery during the discharge period and use of the stored energy during the charge period to maintain the batteries within the operational temperature range of 325 to 375 °C. It was shown that the batteries cannot store sufficient energy during the discharge in order to maintain the batteries within the specified temperature range. It is suggested that consideration be given to use of additional energy storage material to avoid requiring heaters during the charge period. The use of phase change material (PCM) for energy storage is highly recommended. Many of the PCMs that can be used in the specified range of temperature are eutectic or binary system alloys. Since the evaluation of thermal behavior of these alloys require complex analysis, further studies are recommended.

References

[1] Shelton, S.V., Wepfer, W.J., and Patton, J.S., "Solid/Vapor Adsorption Heat Pumps for Space Application," SAE Technical Paper Series No. 881107, !8th Intersociety Conference on Environmental Systems, San Fransisco, California, Jully 1988.

[2] "Development of a Prototype Thermal Control Subsystem for Advanced Extravehicular Activity Portable Life Support System (Regenerable Nonventing Thermal Sink - RNTS I),", United Technologies, Hamilton Standard, Final Report, NAS9-16609, DRL No. T-1700, March 1986.

[3] "Subsystem Conceptual Study Report for the Regenerable Non-venting Thermal Sink (RNTS)," United Technologies, Hamilton Standard, Final Report, Contract No. NAS9-16609, December, 1982.

[4] "The Advanced Nonventing Extravehicular Activity Cooling System (RNTS II),"United Technologies, Hamilton Standard, Final Report, NAS9-17301, DRL No. T-1833, January 1988.

[5] Karimi, A., "Re-examination of METMAN, Recommendations on Enhancement of LCVG, and Development of New Concepts for EMU Heat Sink," Final Report, NASA-NGT-44-005-803, 1990, pp 10.1-10.15.

[6] Karimi, A. "A Thermal Evaluation of 'LSSI' Liquid-Cooled System: An Engineering Perspective", USAF Summer Faculty Program, Technical Report, Vol. 1/2, Report No. 48, September 1983.

[7] Karimi, A., "A Thermal Evaluation of a Portable Battery-Operated Vapor Compression Cooling System", Final Report UES/F49620-85-C-0013, September 1985, 25 pp.

[8] Karimi, A., "A Thermal Evaluation of the `LSSI' Liquid-Cooled System: An Engineering Perspective", USAF Summer Faculty Program, Technical Report, Vol. 1/2, Report No. 48, September 1983, 23 pp.

[9] Karimi, A.," An Experimental Study of the Portable Liquid-Cooled System", Final Report, SCEEE/RPI Subcontract No. 83 RPI 20, December 1984, 35 pp.

[10] Karimi, A., and Chen, T.Y.,"Design and Evaluation of Intermittent Liquid Cooling System for Ground Crew Personnel", Final Report, Brooks AFB, USAF-SAM-TP-90-2, 1990.

[11] Karimi, A., "Thermodynamic Analysis of Liquid Cooling Technology", Final Report, SCEEE/ARB/85-63, May 1986, 52 pp.

[12] Semka, R.P., Prince, L.S., Taenaka, R.K, and Zellmer, D.D., "High Energy Density Rechargeable Battery," Interim Report, WRDC-TR-89-2131, Wright Research and Development Center, WRDC/POOS-2, 1990.

[13] Bard, S., Hughes, R., Jones, J., Mon, G., and Rogers, D., "65 K Sorption Cryocooler Development Study,", JPL Final Report, PL-TR-1047, Aug. 1992.

[14] Dunn, S. B., "Regenerative Electrochemical Demonstration Cryocooler Notebook, " Phillips Laboratory, Space Thermal Technology Branch, Kirtland AFB, NM.

[15] 1993 ASHRAE Fundumentals Handbook , ASHRAE Publications, Atlanta, Ga., pp. 19.4 - 19.6. 1993.

[16] Handbook of Physical Chemistry, 64th Edition, CRC Press, Cleveland, Ohio, 1984.

[17] Handbook of Tables for Applied Engineering Science, Second Edition, CRC Press, Cleveland, Ohio, 1976.

[18] Heat Exchanger Design Handbook, Volume 5, Hemisphere Publishing Co., New York, 1983, 1984, pp. 5.5.1.4-5.5.1.39.

[19] Stewart, R.B., Jacobson, R.T., and Penoncello, S.G., Thermodynamic Properties of Refrigerants, ASHRAE, Publications, Atlanta Ga.

[20] Reid, R.C., Prausnitz, J.M., and Poling, B.E., The Properties of Gases and Liquids, Fourth Edition, McGraw Hill Pub. Co. New York, 1987.

[21] Reid, R.C., Prausnitz, J.M., and Sherwood, P. K, The Properties of Gases and Liquids, Third Edition, McGraw Hill Pub. Co. New York, 1977.

[22] Carey V. P., Liquid-Vapor Phase-Change Phenomena: An Introduction to Thermophysics of Vaporization and Condensation Processes in Heat Traqnsfer Equipment, Hemisphere Publishing Corporation, Washington, 1992

[23] Heisler, M.P., "Temperature Charts for Induction and Constant Temperature Heating," Trans. ASM, vol. 69, 1947, pp. 227.

SPIN EVOLUTION OF THE LAGEOS SATELLITE

Arkady Kheyfets
Assistant Professor
Department of Mathematics

North Carolina State University
Raleigh, NC 27695–8205

Final Report for:
Summer Research Program
Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1993

SPIN EVOLUTION OF THE LAGEOS SATELLITE

Arkady Kheyfets
Assistant Professor
Department of Mathematics
North Carolina State University

Abstract

LAGEOS (LAser–ranged GEOdynamic Satellite) is a dense spherical satellite covered with a total of 426 corner-cube LASER retroreflectors, which allow its tracking to extremely high accuracy. Such accuracy yields a medium term (years to decades) inertial reference frame, which may be determined via relatively inexpensive observations and which can be used as an adjunct to more difficult and more data intensive absolute frame measurements via VLBI (Very Long Baseline Interferometry). LAGEOS will lead to significantly better position determination and timing accuracy. It will allow bench marking of a global-positioning system against a surface-based coordinate system. There is a substantial secular precession of the satellite line of nodes consistent with the classical, Newtonian precession from the non spherical Earth. There has been a suggestion (Ciufolini) of launching an additional satellite (LAGEOS–3) that would experience equal and opposite classical precession than LAGEOS–1. LAGEOS–3, besides providing a more accurate real–time measurement of the rotation rate of the earth and its orientation parameters, it will provide the first direct measurement of the magnetic component of gravity (*the gravitomagnetic or frame–dragging force*) predicted by Einstein's 1915 and still standard theory of gravitation. Of the five dominant error sources in this experiment, the largest one (potentially) involves surface forces on the satellite and its consequent impact on the orbital nodal precession. The surface forces are a function of the orientation and spin rate of the satellite. Consequently, we have undertaken a theoretical and experimental effort to model the spin dynamics of the LAGEOS satellite. We present in this report our preliminary results.

SPIN EVOLUTION OF THE LAGEOS SATELLITE

Arkady Kheyfets

THE SPIN OF LAGEOS AND THE LAGEOS-3 MISSION

The Laser Geodynamic Satellite Experiment (LAGEOS-3) is a joint USAF, NASA and ASI proposed program to measure. for the first time, a new quasi-stationary property of the earth — the gravitational magnetic dipole moment of the earth (gravitomagnetism) predicted by Einstein's theory of general relativity. Just as an electrical current of mass (e.g. the spinning motion of the earth) gives rise to a gravitational dipole field. This gravitomagnetic field causes the local inertial frames to be dragged around with the earth at a rate proportional to the angular momentum of the earth and inversely proportional to the cube of the distance from the center of the earth. This will cause the line of nodes of LAGEOS-3 to precess eastward at 32 *mas/yr*. Although this frame dragging effect is small compared to the torque on the orbital plane due to the oblateness of the earth, it is an essential ingredient in the dynamics of accretion disks. binary systems and other astrophysical phenomenon.

Today, almost eighty years after Einstein introduced his geometric theory of gravity, we have just begun to measure — to verify — his gravitation theory. Of no less stature than the "tide producing" $-M/r^2$ "electric component" of gravity is the inertial-frame defining "magnetic component" of gravitation $-J/r^3$. To see this force in action; first, inject a satellite into a polar orbit about an earth-like mass idealized as not spinning with respect to the distant quasars. The satellite will remain in orbit in a continuous acceleration towards the center-of-mass of the attracting body under the influence of the Newtonian $1/r^2$ force. and its orbital plane will remain fixed in orientation with respect to distant quasars. Second, spin this central body, give it angular momentum and follow the trajectory of the satellite. Its orbital plane will experience a torque along the body's rotation axis. The orbital plane will undergo a precessional motion in the direction of the central body's rotation. The mass in motion of the central body or "mass current" produces a magnetic field of gravity — the gravitomagnetic field. In the case of a satellite orbiting at two earth radii, the orbital plane will precess about the body axis of the earth at approximately 28 *mas/yr*.

This force has never been directly measured. A measurement of this gravitomagnetic force can be cast in parallel to the pioneering work of Michael Faraday on the measurement of a magnetic force

between two current–carrying wires. However. the laboratory setting for this gravity measurement will be the 4-dimensional curved spacetime (Kerr) geometry enveloping the earth. The idea behind this gravity measurement is simple. The Everett–Fairbanks experiment (Gravity Probe-B) proposes a polar orbiting gyroscope. The Ciufolini LAGEOS-3 experiment[1] *uses the the orbital plane as a gyroscope.* In particular, in 1976 NASA launched the LAGEOS-1 satellite. It is a totally passive satellite, a $60cm$ diameter ball of aluminum with 426 retro–reflecting mirrors embedded in its surface. There are 25 globally–located laser tracking stations to observe LAGEOS–type satellites. LAGEOS-1 was injected into a two earth–radii circular orbit at an inclination of 110 deg. The orbital plane rotates due to the oblateness of the earth at a rate of $126\,\mathrm{deg}/yr$. This torque can only be modeled to 450 mas/yr — which is not accurate enough to measure the 32 mas/yr gravitomagnetic force. The idea of Ciufolini: launch another LAGEOS satellite (LAGEOS-3) into an orbit identical to that of LAGEOS-1 except that its inclination is supplementary ($70\,\mathrm{deg} = 180\,\mathrm{deg} - 110\,\mathrm{deg}$). This proposed orbital plane will rotate in the opposite direction, $-126\,\mathrm{deg}/yr$. The intersection of the two (LAGEOS-1, LAGEOS-3) orbital planes will sweep out a "tandem–generated gyro plane" — a plane inertial enough to provide a measurement, accurate to five percent, of the "magnetic component" of gravity.

In this report we discuss the motion of an oblate spheroidal metallic satellite in orbit about the earth. In particular, we are concerned with the effects of the earth's gravitational and magnetic field upon the spin vector of the satellite. The motivation behind this study is the LAGEOS-3 satellite experiment mentioned above. The success of this experiment depends upon the detection of a 32 milliarcsecond per year eastward drift of the line of nodes of the two satellites. The utilization of two satellites forms what we refer to as a tandem-generated gyroscope (see Fig. 1), canceling out many of the much larger precessional effects due to mass eccentricities of the earth. A strong effort is underway to model the orbital and spin dynamics of these satellites, and make an assessment of the uncertainties these will add to the desired measurements. There are five major classes of errors in this experiment: geopotential (other than even zonals) & tides, earth radiation pressure, uncertainty in other relativistic effects, earth and solar-induced thermal forces, and even zonal geopotentials (per $0.1\,\mathrm{deg}$ *inclination injection error*). *Due to the recent* GEM-T1 *improvements in* the earth's zonal harmonics, the errors due to solid-earth tides are now smaller than those due to surface effects. The largest source of error on the satellite such as Yarkowsky thermal drag, neutral and charge particle drag, and others[2,3,4], depend crucially upon the orientation and spin rate of the satellite[5]. It is

the surface forces, in particular the Yarkowsky thermal drag, also referred to as the Rubincam effect, that cause a change in the nodal precession of the satellite, thus contributing to the largest error source in the LAGEOS-3 measurement of the gravitomagnetic effect. It is for this reason that we undertake in this report a theoretical model of the spin dynamics of the LAGEOS-type satellites, and compare it with observational results. One of the important effects considered is the Yarkowsky thermal drag, caused by differential heating and delayed reradiation. This effect is fundamentally related to the nature of the spin vector of the satellite.

Previous studies on the spin dynamics of the LAGEOS satellite[6] were valid only for spins much larger than the orbital period. However, today the spin of the satellite, decaying with a three-year time constant, is rapidly approaching the orbital period. Therefore, we require a low-frequency regime analysis. This low frequency regime, we show, exhibits complex behavior, and the asymptotic state of this forced, damped system, appears to be robust. In this report, we will examine the spin-orbit resonance phase, discuss the asymptotic state of the spin of the satellite, and examine the Aponov exponents. It is the goal of this research to provide an optimum strategy for the measurement of the spin dynamics of the LAGEOS satellites in support of the proposed gravity measurement. In addition, we use our theoretical model to propose an optimum orbital injection procedure for LAGEOS-3. It is our opinion that the LAGEOS-3 satellite should be injected into the orbital plane with as large a spin rate as possible. We demonstrate in this report that the LAGEOS-1 satellite will be sufficiently predictable to support the gravitomagnetic measurement. The results in this report provide the first analysis of the asymptotic spin dynamics of the LAGEOS satellite. Previous calculations were unable to analyze the spin-orbit resonances of the satellite, which will play a crucial role in the experiment.

It is rather interesting that after 36 years the spin dynamics of satellites are once again important in the field of astrodynamics. In 1957 Vinti[7] analyzed the spin dynamics of a non-ferromagnetic spherical satellite in the earth's magnetic field which was of critical importance to the alignment of antennas. Today, we perform the same analysis on a slightly oblate satellite, of critical importance to the first measurement of the magnetic component of gravity, as predicted by Einstein. That the uncertainties induced by the surface forces on LAGEOS are on the order of 4% percent out of a 6% experiment make the theoretical modeling paramount.

There are many factors to consider when analyzing the dynamics of the spin vector of an oblate, metallic satellite. The most prominent effect is the torqueing due to the gravitational field of the earth. This

arises due to the oblateness of the satellite, with the oblateness of the earth producing negligible contributions than can be added to our calculations as an adiabatic correction. If the satellite's (bulging) equatorial plane lies in the plane of orbit, no such torques are possible. However, when the satellite is not placed exactly in such a position, gravitational torques will arise. In an effort to model these torques, we consider the situation of an oblate spheroid in orbit around a point mass. As was done by Bertotti, we parallel the development in Goldstein[5].

However, we note that the Bertotti and Iess analysis of the effects of gravity on the spin dynamics of the oblate satellite, that resulted in predicting the chaotic spin dynamics in the future, is not appropriate for small rates of spin (when the spin frequency is of the order of the orbital frequency). Their prediction is based on the "Hipparcos" formula for the rate of precession ω_p of the oblate satellite in the inhomogeneous gravity field of the Earth

$$\omega_p = \frac{3}{2}\,\Delta\,\frac{\omega_0^2}{\omega_3}\cos\theta \tag{1}$$

where ω_0 is the orbital angular velocity, ω_3 is the satellite spin rate, θ is the obliquity angle of the satellite (the angle between $\vec{\omega}$ and the normal to the orbital plane) and

$$\Delta = \frac{I_3 - I_1}{I_3} \tag{2}$$

is the satellite oblateness. Here I_3 and $I_1 = I_2$ are the principal moments of inertia (the principal direction corresponding to I_3 is assumed to be that of $\vec{\omega}$ and coincides, by an assumption, with the symmetry axis of the oblate satellite).

It is argued by Bertotti and Iess that, since $\omega_p \propto 1/\omega_3$, the gravitational precession in the asymptotic limit of small ω (when ω_3 becomes of the order of ω_0) becomes very fast and may make the spin dynamics chaotic.

The latter conclusion is based on a misunderstanding. It has been shown by one of us[9] via a careful analysis of assumptions underlying the "Hipparcos" formula that even in the approximation commonly used in deriving the formula (averaging of the gravitational potential over the satellite orbit, dipole cutoff of the multipole decomposition, etc.) a care should be exercised. Equation (1) can be used only when

$$\frac{6\omega_0^2\Delta\cos^2\theta}{\omega_3^2} \sim \frac{\omega_0^2}{\omega_3^2} \ll 1 \tag{3}$$

i. e. when the spin rate of the satellite is much greater than its orbital angular velocity. The latter restriction is very easy to overcome. The corrected equation for ω_p is

$$\omega_p = \frac{1}{2}\frac{\omega_3}{\cos\theta}\left(1 - \sqrt{1 + \frac{6\omega_0^2\Delta\cos^2\theta}{\omega_3^2}}\right) \tag{4}$$

This equation obviously imposes a bound on $|\omega_p|$

$$|\omega_p| < \omega_0\sqrt{\frac{3}{2}\Delta} \tag{5}$$

which makes it clear that the rate of precession cannot grow enough to cause the chaoticity of the satellite spin dynamics. The subsequent qualitative investigation[10] has shown that when magnetic forces are included in the picture the precession rate remains bounded and should be much smaller than ω_0. Another conclusion reached in the course of this analysis[9] has been that, when both gravitational and magnetic forces are taken into account, the nutation, although bounded in its amplitude, does not disappear completely even in the asymptotical limit. The value of these results, however, is limited by the fact that they do not tell one what are the bounds exactly. Neither they provide any information on the time scale of reaching the asymptotic limit. However, they show clearly that the chaoticity of the LAGEOS spin dynamics caused by an unbounded growth of gravitationally induced precession cannot occur.

Another important factor governing the evolution of the spin vector is the interaction of the metallic core of the satellite with the magnetic field of the earth. The LAGEOS satellite is composed of two aluminum hemispheres bolted together, with a brass cylindrical core along an axis (the original axis of spin)[11]. The spinning of the metallic object in the magnetic dipole field of the earth (and the motion through that field) will cause eddy currents within the satellite, which will in turn cause dissipation (through Joule heating) and a slowdown of the spin, and furthermore will cause torques on the spin vector. These torques can be understood as the interaction of the magnetic dipole caused by the induced eddy currents with the magnetic field of the earth. In modeling this effect, we have treated a uniform, spherical object in the orbit of a perfect magnetic dipole (as the reader will recall, we are concerned with qualitative results). We have allowed the orbit to be placed at an arbitrary inclination and our results in this report correspond to $I \sim 78\,\text{deg}$. The problem of a spinning metallic sphere in a constant magnetic field has been treated by Landau and Lifshitz, and we avail ourselves of their results. For our purposes, we ignore the torques caused by the changing

magnetic field due to the orbit (as opposed to spin) of the satellite. These torques can be shown to be negligible until asymptotically late stages of motion, and have no qualitative effects upon the dynamics.

The source of many of the difficulties in doing analyses of such orbiting, spinning bodies lies in the involved coordinate systems needed to describe their motion. Thus, it is important at this point to give a brief description of the coordinates we will use in this report. In our analysis of the spin dynamics of the LAGEOS satellite we found it convenient to introduce the following four coordinate systems:

1) *The earth–centered inertial (ECI) reference frame* $\{x_2, y_2, z_2\}$. Here the z_2-axis is aligned with the body axis of the earth. The x_2-axis lies in the earth's equatorial plane at zero degrees longitude, and the origin coincides with the center of the earth.

2) *The orbit–centered inertial frame (OCI)* $\{x_1, y_1, z_1\}$. Here the z_1-axis is oriented along the normal to the orbital plane of the satellite with its $\sim 22\,\mathrm{deg}$ coinclination. The x_1-axis is defined to be the intersection of the orbital plane and the equatorial plane of the earth, and the origin is the center of mass of the earth. We have assumed here that this frame is inertial as we have not included the secular drag of the line–of–nodes of the orbital plane due to the oblateness of the earth. This $\sim 126\,\mathrm{deg}/yr$ precession can be included at the end of our analysis as an adiabatic correction.

3) *The body frame (non–inertial)* $\{x_b, y_b, z_b\}$. Here the z_b-axis is aligned along the body axis of the satellite which is assumed here to be a slightly oblate ($\sim 3\%$) spheroid of brass. The origin is at the center of the satellite. In our calculations, the body axis is related to the orbit–centered frame through the three Euler angles θ, ϕ and ψ. The nutation angle θ is the angle between z_b and z_1, while the angle of precession ϕ is the angle between the $x_1 - axis$ and the line of nodes. Finally, the spin angle ψ is the angle between the line of nodes and the x_b-axis.

4) *The Landau–Lifshiz (non inertial) coordinate system* $\{x^{ll}, y^{ll} z^{ll}\}$. The x^{ll}-axis is aligned along the instantaneous angular momentum vector of the satellite ($\vec{\omega}$), and the instantaneous magnetic field ($\vec{B}$) at the satellite lies in the $x^{ll} - z^{ll}$ plane. Note that z^{ll} need not be aligned with the body axis of the satellite, and in fact the asymptotic behavior of the satellite they are vastly different.

SPIN DYNAMICS OF LAGEOS: THE EQUATIONS

Spin dynamics of the axially symmetric satellite is determined by Euler's equations

$$\begin{aligned} I_1\dot{\omega}_1 - \omega_2\omega_3(I_1 - I_3) &= N_1 \\ I_1\dot{\omega}_2 - \omega_3\omega_1(I_3 - I_1) &= N_2 \\ I_3\dot{\omega}_3 &= N_3 \end{aligned} \tag{6}$$

where ω_1, ω_2, ω_3 are the components of the satellite angular velocity in the body frame, $I_1 = I_2$, and I_3 are principal moments of inertia. and N_1, N_2, N_2 are the components of the torques in the direction of the satellite principal axes. After substituting the expressions for ω_1, ω_2, ω_3 in terms of the Euler angles

$$\begin{aligned} \omega_1 &= \dot{\phi}\sin\theta\sin\psi + \dot{\theta}\cos\psi \\ \omega_2 &= \dot{\phi}\sin\theta\cos\psi - \dot{\theta}\sin\psi \\ \omega_3 &= \dot{\phi}\cos\theta + \dot{\psi} \end{aligned} \tag{7}$$

the Euler equations become

$$\begin{aligned} \ddot{\theta} &= \ddot{\theta}_{free} + \frac{N_1\cos\psi - N_2\sin\psi}{I_1} \\ \ddot{\theta}_{free} &= \left(\frac{I_1 - I_3}{I_1}\right)\dot{\phi}^2\cos\theta\sin\theta - \frac{I_3}{I_1}\dot{\psi}\dot{\phi}\sin\theta \end{aligned}$$

$$\begin{aligned} \ddot{\phi} &= \ddot{\phi}_{free} + \frac{N_1\sin\psi + N_2\cos\psi}{I_1\sin\theta} \\ \ddot{\phi}_{free} &= \left(\frac{I_3 - 2I_1}{I_1}\right)\frac{\cos\theta}{\sin\theta}\dot{\theta}\dot{\phi} + \frac{I_3}{I_1}\frac{\dot{\theta}\dot{\psi}}{\sin\theta} \end{aligned} \tag{8}$$

$$\begin{aligned} \ddot{\psi} &= \ddot{\psi}_{free} + \frac{N_3}{I_1} - \frac{N_1\sin\psi + N_2\cos\psi}{I_1}\frac{\cos\theta}{\sin\theta} \\ \ddot{\psi}_{free} &= -\left(\frac{I_3 - I_1}{I_1}\right)\frac{\cos^2\theta}{\sin\theta}\dot{\theta}\dot{\phi} + \frac{\dot{\theta}\dot{\phi}}{\sin\theta} - \frac{I_3}{I_1}\frac{\cos\theta}{\sin\theta}\dot{\theta}\dot{\psi} \end{aligned}$$

The torque components N_1, N_2, N_3 are due to gravitational and magnetic forces acting on the satellite

$$N_i = N_i^{(g)} + N_i^{(m)}, \quad i = 1,2,3 \tag{9}$$

Gravitational torques in the body frame are given by

$$\begin{aligned} N_1^{(g)} &= -\cos\psi\frac{\partial V}{\partial\theta} - \frac{\sin\psi}{\sin\theta}\frac{\partial V}{\partial\phi} + \frac{\cos\theta\sin\psi}{\sin\theta}\frac{\partial V}{\partial\psi} \\ N_2^{(g)} &= \sin\psi\frac{\partial V}{\partial\theta} - \frac{\cos\psi}{\sin\theta}\frac{\partial V}{\partial\phi} + \frac{\cos\theta\cos\psi}{\sin\theta}\frac{\partial V}{\partial\psi} \\ N_3^{(g)} &= -\frac{\partial V}{\partial\psi} \end{aligned} \tag{10}$$

The gravitational potential V is (the standard dipole approximation is used)

$$V = \frac{GM(I_3 - I_1)}{2R^3}(3\gamma^2 - 1) \tag{11}$$

where γ is the direction cosine between (1) the radial vector from the satellite center of mass to the center of the Earth, and (2) the symmetry axis of the satellite. It is related to Euler's angles via

$$\gamma = \sin\theta \sin(\eta - \eta_0 - \phi) \tag{12}$$

so that

$$V = \frac{GM(I_3 - I_1)}{2R^3}(3\sin^2\theta \sin^2(\eta - \eta_0 - \phi) - 1) \tag{13}$$

and

$$\begin{aligned} -\frac{\partial V}{\partial \theta} &= -\frac{3GM(I_3 - I_1)}{R^3}\sin\theta\cos\theta\sin^2(\eta - \eta_0 - \phi) \\ -\frac{1}{\sin\theta}\frac{\partial V}{\partial o} &= \frac{3GM(I_3 - I_1)}{R^3}\sin\theta\sin(\eta - \eta_0 - o)\cos(\eta - \eta_0 - \phi) \\ \frac{\partial V}{\partial \psi} &= 0 \end{aligned} \tag{14}$$

Equations (10) and (14) lead to the following final expressions for the components of the gravitational torque in the body frame

$$\begin{aligned} N_1^{(g)} &= \frac{3GM(I_3 - I_1)}{R^3}\sin\theta\sin(\eta - \eta_0 - \phi)\left\{\begin{array}{c} -\cos\theta\cos\psi\sin(\eta - \eta_0 - \phi) \\ +\sin\psi\cos(\eta - \eta_0 - \phi) \end{array}\right\} \\ N_2^{(g)} &= \frac{3GM(I_3 - I_1)}{R^3}\sin\theta\sin(\eta - \eta_0 - \phi)\left\{\begin{array}{c} \cos\theta\sin\psi\sin(\eta - \eta_0 - \phi) \\ +\cos\psi\cos(\eta - \eta_0 - \phi) \end{array}\right\} \\ N_3^{(g)} &= 0 \end{aligned} \tag{15}$$

The components of the magnetic torque in the body frame $N_1^{(m)}$, $N_2^{(m)}$, and $N_3^{(m)}$ can be figured out starting from the torque components acting on a conducting ball of radius a spinning with the angular velocity $\vec{\omega}$ in an external magnetic field $\vec{B}$ in the Landau–Lifshitz frame[12].

The Landau–Lifshitz frame is determined by vectors $\vec{\omega}$ and $\vec{B}$. Its axis z^{II} is picked to coincide in its direction with $\vec{\omega}$. its axis y^{II} is oriented in the direction of $\vec{\omega} \times \vec{B}$, and the axis x^{II} is orthogonal to both y^{II} and z^{II} with its direction chosen so that $x^{II}y^{II}z^{II}$ forms a left–handed coordinate system. We introduce notatons $\hat{x}^{II}$, $\hat{y}^{II}$, $\hat{z}^{II}$ for the vectors of the orthonormal frame associated with these axes. These vectors form

the Landau–Lifshitz frame. If we introduce an arbitrary rectangular coordinates x, y, z with its associated orthonormal frame formed by unit vectors $\hat{x}$, $\hat{y}$, $\hat{z}$ such that in this frame $\vec{\omega}$ and $\vec{B}$ are represented as

$$\begin{aligned}\vec{\omega} &= \omega_x\,\hat{x} + \omega_y\,\hat{y} + \omega_z\,\hat{z} = \langle \omega_x, \omega_y, \omega_z \rangle \\ \vec{B} &= B_x\,\hat{x} + B_y\,\hat{y} + B_z\,\hat{z} = \langle B_x, B_y, B_z \rangle\end{aligned} \tag{16}$$

then the transition between the two frames is given by

$$\begin{pmatrix} \hat{x}^{ll} \\ \hat{y}^{ll} \\ \hat{z}^{ll} \end{pmatrix} = \begin{pmatrix} \frac{[(\vec{\omega}\times\vec{B})\times\vec{\omega}]_x}{\omega|\vec{\omega}\times\vec{B}|} & \frac{[(\vec{\omega}\times\vec{B})\times\vec{\omega}]_y}{\omega|\vec{\omega}\times\vec{B}|} & \frac{[(\vec{\omega}\times\vec{B})\times\vec{\omega}]_z}{\omega|\vec{\omega}\times\vec{B}|} \\ \frac{(\vec{\omega}\times\vec{B})_x}{|\vec{\omega}\times\vec{B}|} & \frac{(\vec{\omega}\times\vec{B})_y}{|\vec{\omega}\times\vec{B}|} & \frac{(\vec{\omega}\times\vec{B})_z}{|\vec{\omega}\times\vec{B}|} \\ \frac{\omega_x}{\omega} & \frac{\omega_y}{\omega} & \frac{\omega_z}{\omega} \end{pmatrix} \begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} \tag{17}$$

where $\omega = |\vec{\omega}|$.

The components of the magnetic torque in this frame is given by

$$\begin{aligned} N_x^{ll} &= V\alpha'' B_x^{ll} B_z^{ll} \\ N_y^{ll} &= V\alpha' B_x^{ll} B_z^{ll} \\ N_z^{ll} &= V\alpha''(B_x^{ll})^2 \end{aligned} \tag{18}$$

where $V = \frac{4\pi}{3}a^3$ is the volume of the ball, B_x^{ll}, B_y^{ll}, B_z^{ll}, are components of the magnetic field in Landau–Lifshitz frame. and the coefficients of magnetization are,

$$\begin{aligned} \alpha' &= -\frac{3}{8\pi}\left[1 - \frac{\delta}{a}\frac{\sinh\left(2\frac{a}{\delta}\right) - \sin\left(2\frac{a}{\delta}\right)}{\cosh\left(2\frac{a}{\delta}\right) - \cos\left(2\frac{a}{\delta}\right)}\right] \\ \alpha'' &= -\frac{9\delta^2}{16\pi a^2}\left[1 - \frac{a}{\delta}\frac{\sinh\left(2\frac{a}{\delta}\right) - \sin\left(2\frac{a}{\delta}\right)}{\cosh\left(2\frac{a}{\delta}\right) - \cos\left(2\frac{a}{\delta}\right)}\right] \\ \delta &= \frac{c}{\sqrt{2\pi\sigma\omega}} \end{aligned} \tag{19}$$

Here c is the speed of light and σ is the specific conductivity of the material forming the ball. At small values of ω, α' and α'' can be approximated by the expressions

$$\begin{aligned} \alpha' &\approx -\frac{4\pi}{105}\frac{a^4\sigma^2\omega^2}{c^4} \\ \alpha'' &\approx \frac{a^2\sigma\omega}{10c^2} \end{aligned} \tag{20}$$

Components of the magnetic field can be evaluated easily in ECI in the dipole approximation. In the spherical coordinate representation of the ECI frame the magnetic field components are

$$\begin{aligned} B_{r_1} &= -\frac{d}{R^3}\cos\theta_1 \\ B_{\theta_1} &= \frac{d}{R^3}\sin\theta 1 \\ B_{\phi_1} &= 0 \end{aligned} \tag{21}$$

In the rectangular coordinate representation of the ECI frame (index 1 is used everywhere for quantities in ECI)

$$\begin{aligned} B_{x_1} &= B_{r_1}\sin\theta_1\cos\phi_1 + B_{\theta_1}\cos\theta_1\cos\phi_1 \\ B_{y_1} &= B_{r_1}\sin\theta_1\sin\phi_1 + B_{\theta_1}\cos\theta_1\sin\phi_1 \\ B_{z_1} &= B_{r_1}\cos\theta_1 - B_{\theta_1}\sin\theta_1 \end{aligned} \tag{22}$$

The magnetic field components in OCI frame are

$$\begin{aligned} B_{x_2} &= (B_{r_1}\sin\theta_1 + B_{\theta_1}\cos\theta_1)\cos\phi_1 \\ B_{y_2} &= (B_{r_1}\sin\theta_1 + B_{\theta_1}\cos\theta_1)\sin\phi_1\cos\xi + (B_{r_1}\cos\theta_1 - B_{\theta_1}\sin\theta_1)\sin\xi \\ B_{z_2} &= -(B_{r_1}\sin\theta_1 + B_{\theta_1}\cos\theta_1)\sin\phi_1\sin\xi + (B_{r_1}\cos\theta_1 - B_{\theta_1}\sin\theta_1)\cos\xi \end{aligned} \tag{23}$$

where ξ is the colatitude angle and index 2 is used for the quantities in the OCI.

The angles θ_1 and ϕ_1 are determined by the satellite position on its orbit. Indeed, the satellite coordinates in OCI are

$$\begin{aligned} x_2 &= r\cos(\eta - \eta_0) \\ y_2 &= r\sin(\eta - \eta_0) \\ z_2 &= 0 \end{aligned} \tag{24}$$

where

$$\eta = -\frac{2\pi}{T_{\text{orbit}}}t \tag{25}$$

In ECI we have

$$\begin{aligned} x_1 &= R\cos(\eta - \eta_0) \\ y_1 &= R\sin(\eta - \eta_0)\cos\xi \\ z_1 &= -R\sin(\eta - \eta_0)\sin\xi \end{aligned} \tag{26}$$

Hence,

$$\begin{aligned} \sin\theta_1 &= \sqrt{\cos^2(\eta - \eta_0) + \sin^2(\eta - \eta_0)\cos^2\xi} \\ \cos\theta_1 &= -\sin(\eta - \eta_0)\sin\xi \\ \cos\phi_1 &= \frac{\cos(\eta - \eta_0)}{\sin\theta_1} \\ \sin\phi_1 &= \frac{\sin(\eta - \eta_0)\cos\xi}{\sin\theta_1} \end{aligned} \tag{27}$$

The three Euler equations (Eq. 8) with the magnetic and gravitational torques included give us the vehicle to analyze qualitatively the spin dynamics of the LAGEOS-1 satellite. We present our numerical results of integrating the Euler equations in the next two sections.

SPIN DYNAMICS OF LAGEOS: THE INITIAL-VALUE DATA

We solved the Euler equations (Eq. 8) using a fourth-order Bulirsch-Stoer algorithm. We integrated the equations for 2×10^9 seconds as we wanted to (1) reproduce the experimentally-measured spin rates (launch through 17 yrs.), (2) examine the spin-orbit resonance (~29 yrs. after launch), and (3) reveal the asymptotics of the spin dynamics (~41 yrs. after launch). The experimentally-measured exponential decrease in the spin rate imposed a constraint on our theoretical model linking the "effective" radius of the satellite (a) with the satellites "effective" conductivity (σ).

$$\sigma a^5 \sim 2.685 \times 10^{22} cm^5/s \tag{28}$$

The satellite was modeled as an $11.9591cm$ radius spheroid of brass ($\sigma = 1.098 \times 10^{17} 1/s$). The moments of inertia of the satellite's body axis was chosen to be $I_{z_b} = 1.1516 \times 10^6 g - cm^2$, while the moment of inertia perpendicular to the body axis was $I_{x_b} = I_{y_b} = 1.084 \times 10^6 g - cm^2$ which corresponds to approximately 4% deviation from sphericity.

We started the integration at $t = 0$ with the satellite in a 68 deg inclination orbit at a radius of 1.227×10^7 cm. The satellite was initially over the equator at 0 deg longitude ($\eta = 0$). We inserted the body axis of the satellite into the orbital plane with the angular momentum vector parallel to the OCI Y_1-axis. The initial spin of the satellite was determined by the best fit to the experimental measurements and was set to 10.472 rad/sec:

$$\begin{aligned} &\upsilon(t=0) = 0, \quad \theta(t=0) = \frac{\pi}{2}, \quad \phi(t=0) = 0 \\ &\dot{\theta}(t=0) = 0, \quad \dot{\phi}(t=0) = 0, \quad \dot{\psi}(t=0) = 10.472 \end{aligned} \tag{29}$$

Finally, our magnetic field was assumed to be a perfect dipole field of moment $M = 7.9 \times 10^{25} gauss/cm^3$ and aligned along the z_2-axis of the ECI frame.

SPIN DYNAMICS OF LAGEOS: RESULTS

We have identified three distinct phases in the spin dynamics of the lageos satellite (Fig. 2). The first phase is characterized by an exponential decline in the spin rate ($\dot{\psi}$) of the satellite with nutation and precession evolving as a small amplitude bobbing of the body axis in and out of the orbital plane. This

persists until the angular velocity approaches the orbital angular velocity ($t \approx 8 \times 10^8$ *sec*). The onset of the spin–orbit phase is preceded by an exponential increase in amplitude of the nutation and precession oscillation. The resonance drives the spin of the satellite out of the orbital plane. This transition occurs rather abruptly taking only $\approx$ 1 year. The angular momentum of the satellite remains orthogonal to the orbital plane for only $\approx$ 1 year and then transits to the third phase — the asymptotic phase. This third phase is characterized by a drift of the angular momentum vector $\approx$ 40 deg off the orbital plane. The nutation amplitude is increases in this transition. The asymptotic state of the spin vector is rather interesting (Fig. 2). The angular momentum vector sweeps out a cone centered on orthogonal to the orbital plane. The cone has an opening angle of $\approx$ 100 deg and is swept out every eighteen orbits. However, the nutation has an amplitude of $\approx$ 5.2 deg. The period of this nutational motion is close to the orbital period. In this asymptotic phase the body axis is quite independent from the angular momentum vector. We were rather surprised that the angular momentum vector is so well behaved as the body axis is much less predictable. Although we have not performed a complete survey of phase space for the robustness of this asymptotic state, we have reproduced this state using a substantially different initial condition — an initial condition where the initial spin vector was directed $\approx$ 45 deg out of the orbital plane. This deserves a more detailed investigation.

CONCLUSIONS

Of the five largest sources of error identified in the LAGEOS-3 experiment, the earth and solar–induced surface forces are the largest. The anisotropic heating of the satellite and subsequent reradiation gives the satellite a thermal rocketing perturbation (referred to as the Rubincam effect or the Yarkowski thermal drag) which tends to degrade the experiment. To model this effect requires, in part, a detailed knowledge of the behavior of the angular momentum of the satellite. Toward this end, we derived and solved numerically a simplified set of Euler equations that evolves the angular momentum vector for a slightly oblate spheroid of brass orbiting an earth–like mass idealized as being a perfect sphere and having a perfect polar–oriented dipole magnetic field. The Euler equations included both the tidal gravitational torques, the eddy–current torques as well as the resistive damping torques modeled by complex magnetization coefficients[12]. Using this rather simplified model we identified here three phases of the rotational dynamics — an exponential damping phase, a locking phase and a forced oscillation phase (Fig. 2). We also identified an error in the previously established model[6]. The error has led to confusion and, in an attempt to reconcile the data with the model, has led others to hypothesize an erroneous model for the moments of inertia of LAGEOS-1[13].

Our results have led us to formulate the following unresolved questions: (1) Why are the eddy currents damped primarily in the cylindrically–shaped brass core of the satellite and not in the aluminum shell?; and (2) What is the physics behind the three distinct phases of the LAGEOS spin dynamics? We are addressing these questions by (1) introducing a more realistic model of the satellite and earth into our simplified calculations[14], and (2) by exploring more of phase space by way of Poincare sections. Our results presented here provide us with clues, now we must piece them together to reveal the physics behind the complex motion we observe.

Acknowledgments

All work presented in this report was accomplished in a close cooperation with Warner Miller. Both of us wish to thank Stirling Colgate, Douglas Currie, Christopher Fuchs, Richard Matzner, John Ries, and James Thorne for many helpful discussions. We also wish to acknowledge the hospitality of the Astrodynamics Group at the Phillips Laboratory (PL/VTA).

REFERENCES

[1] Ciufolini I., Measurement of the Lense–Thirring drag effect on LAGEOS and another high altitude laser ranged satellite, *Phys. Rev. Lett.*. **56**, 278 (1986).

[2] Rubincam D. P., Drag on the LAGEOS satellite, *J. Geophys. Research*, **95**, 4881 (1990).

[3] Rubincam D. P., LAGEOS orbit decay due to infrared radiation from earth, *J. Geophys Res.*. **92**, 1278-1294 (1987).

[4] Rubincam D. P., The LAGEOS along track acceleration: a review, Paper presented at the First William Fairbanks meeting on relativistic gravity experiments in space, Rome, Italy, September 10–14. (1990).

[5] Rubincam D. P., Yarkovski thermal drag on LAGEOS, *J. Geophys. Research*, **93**, 13805 (1988).

[6] Bertotti B., Iess L., The rotation of LAGEOS, *J. Geophys. Research*, **96**, 2431 (1991).

[7] Vinti J. P., Theory of the Spin of a Conducting Satellite in the Magnetic Field of the Earth, Defense Technical Information Center, BRL-1020, 1957.

[8] Goldstein H., *Classical mechanics*. Addison–Wesley Publ. Co., Reading, MA (1981).

[9] Kheyfets A., Spin Dynamics of Lageos Satellite, Final Report, 1992 Air Force Summer Research Program. July 1992.

[10] Fuchs C., Lagrangian formulation of LAGEOS's spin dynamics, Final Report. 1992 Air Force Summer Research Program. August 1992.

[11] Johnson C. W., Lundquist C. A., Zurasky J. L., The LAGEOS satellite, Paper presented at International Astronautical Federation XXVII Congress, Anaheim, CA, October 10–16, (1976).

[12] Landau L.D., Lifshitz E. M., *Electrodynamics of continuous media*, Pergamon Press, Oxford (1984).

[13] Scharroo, R., Wakker, K. F., Ambrosius, B. A. C., and Noomen, R., On the along–track acceleration of the LAGEOS satellite, *J. Geophys. Res.*, 96 729–40 (1991).

[14] Halverson, R. P. and Cohen H. Torques on a spinning hollow sphere in a uniform magnetic field, *IEEE Trans. Aerosp. Navig. Electron.*, ANE–11 118–22 (1964).

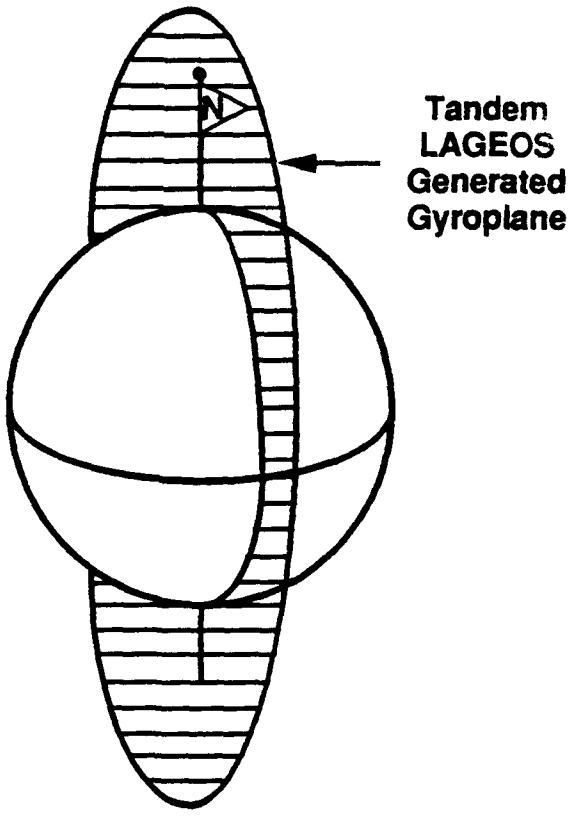


Fig. 1: The tandem-generated gyro plane and the LAGEOS-3 satellite experiment: The idea is to launch another LAGEOS satellite (LAGEOS-3) into an orbit identical to that of LAGEOS-1 except that its inclination is supplementary (70 deg = 180 deg − 110 deg). This proposed orbital plane will rotate in the opposite direction, $-126\,\mathrm{deg}/yr$. The intersection of the two (LAGEOS-1, LAGEOS-3) orbital planes will sweep out a "tandem-generated gyro plane" — a plane inertial enough to provide a measurement, accurate to five percent, of the "magnetic component" of gravity.

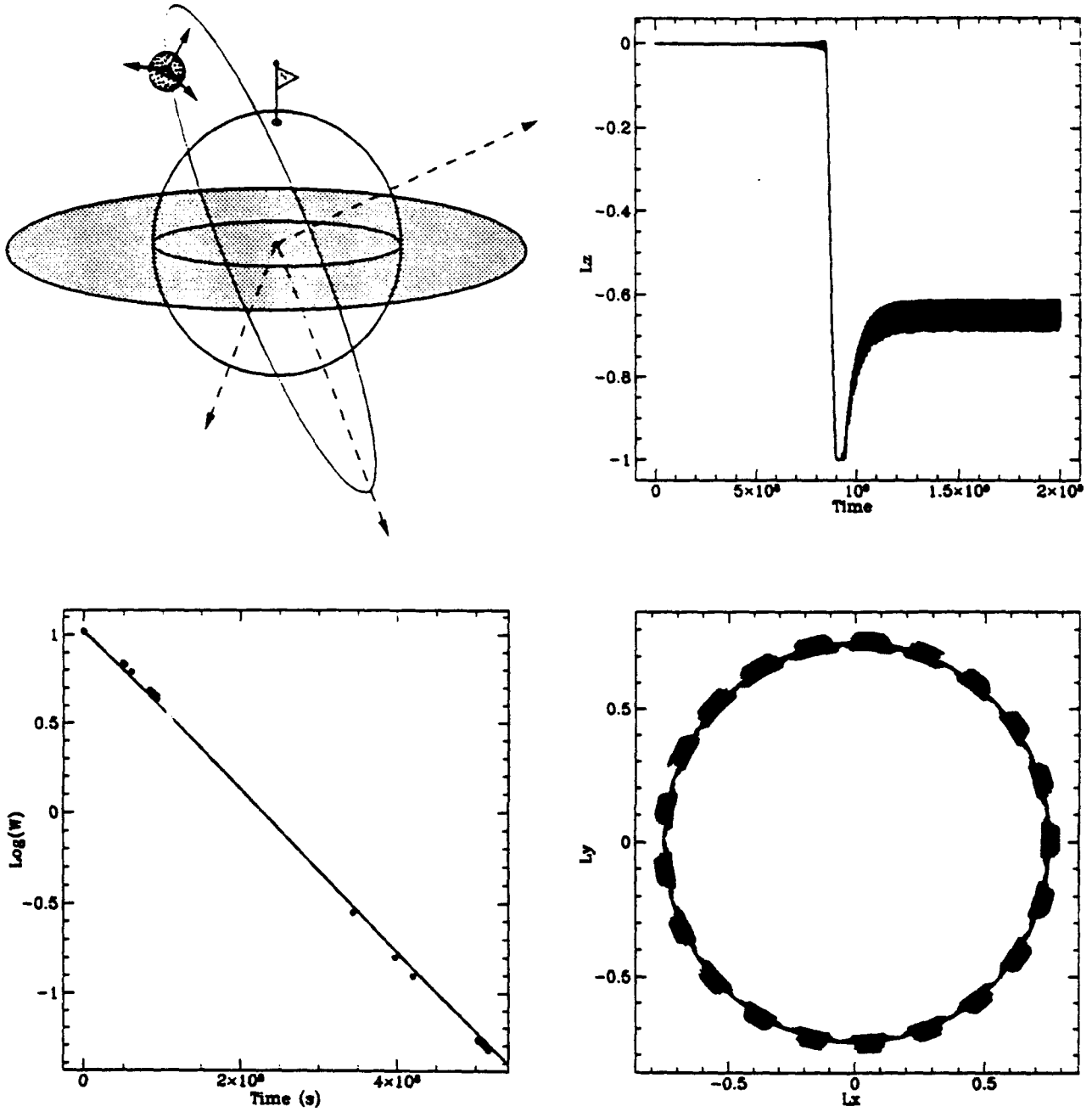


Fig. 2: The spin dynamics of LAGEOS: Numerical Results: We present here results from our numerical simulation of the dynamic evolution of the LAGEOS satellite (top left figure). The evolution of the satellite's angular momentum reveals three unique phases as demonstrated in the plot of the component of angular momentum orthogonal to the orbital plane (top right). The first phase is characterized by an exponential decrease in spin of the satellite with negligible nutation and precession. In the first phase, the decay in spin for a 17 cm radius sphere of brass agrees with the data (bottom left). When the angular velocity decays to a value comparable to the orbital angular velocity the nutation and precession increase and the angular momentum vector lifts off the orbital plane and settles down orthogonal to the orbital plane. The third phase is characterized by a drift of the angular momentum vector ≈ 40 deg off the orbital plane. It sweeps out a cone centered on the orthogonal to the orbital plane with an opening angle of 100 deg. All characteristic angular velocities are of the order of the orbital angular velocity.

INVESTIGATION OF DEEP LEVELS in MBE grown $Hg_{1-x}Cd_xTe$ with DLTS EXPERIMENTS

Samuel Lakeou

Professor

Department of Electrical Engineering

University of the District of Columbia

4200 Connecticut Ave NW

Washington, D.C. 20008

Final Report for:

Summer Faculty Research Program

Sponsored by:

Air Force Office of Scientific Research.

Bolling Air Force Base, Washington, D.C.

July 1993

Investigation of deep levels in MBE grown $Hg_{1-x}Cd_xTe$ using DLTS experiments

Samuel Lakeou

Professor

Department of Electrical Engineering

University of the District of Columbia

Abstract

Deep levels in MBE grown $Hg_{1-x}Cd_xTe$ were investigated for the first time using DLTS experiments. The experiments included a state-of-the-art data acquisition system which provided better control over a wide range of temperature values. Also, the experimental setup is made to include the CCDLTS option which is known to complement the basic DLTS experiments for further investigation of major properties of the sample. After an initial calibration of the experimental setup with an AlGaAs sample, experiments performed on $Hg_{.724}Cd_{.276}Te$ samples indicate the existence of traps at four levels. However, the interpretation of the experimental results was far from exhaustive.

Due to the narrow gap of the $Hg_{1-x}Cd_xTe$, it is generally recommended to modify the DLTS/CCDLTS experiments so as to account for voltage dependent effects (i.e. appearance of shoulder on the DLTS scan and Poole-Frenkel effect). These effects are not investigated in this study but the "shoulder" predicted in DLTS scans in narrow gap materials can be identified on the DLTS scans.

Investigation of deep levels in MBE grown $Hg_{1-x}Cd_xTe$ using DLTS experiments

Samuel Lakeou

I. Introduction

Starting in 1959, when W. D. Lawson et al. in England published some details on mercury cadmium telluride (HgCdTe), $Hg_{1-x}Cd_xTe$ was ranked among the infrared detectors of choice for terrestrial and various other applications. In both medium and far infrared detection (3-5 and 8-14μm) $Hg_{1-x}Cd_xTe$ is in fact considered as the most important detector material [1,2,3]. However, due to its materials properties and the diversity of materials defects (grain boundaries, Te precipitates, high dislocations etc.) very few deep level studies have concentrated on $Hg_{1-x}Cd_xTe$. Deep level transient spectroscopy (DLTS) was however utilized to characterize trapping centers [7,8,9] but to our knowledge no attempt was made to use this technique on MBE grown HgCdTe. Our attempt is to investigate deep levels using DLTS technique on MBE grown $Hg_{1-x}Cd_xTe$ materials provided by Rockwell Science Center (Thousand Oaks , Ca).

The DLTS/CCDLTS (Constant capacitance DLTS) experiments were equipped with a basic setup from SULA[1] technologies. Data acquisition was done with a state-of-the-art data acquisition software (Labview[2]) which made the DLTS/CCDLTS scan fully automated both for the acquisition of data and the control and monitoring of the temperature variation. Substantial time was spent in the calibration and the proper programming of the data acquisition apparatus. The program is tailored to accommodate both DLTS and CCDLTS experiments. The current study is therefore a compilation of results both on the calibration AlGaAs sample and Rockwell's $Hg_{1-x}Cd_xTe$ samples. Further analysis of the observed experimental data should provide more insight on the nature of the traps.

II. Deep Level Transient Spectroscopy (DLTS)

II.1 Brief description of DLTS

DLTS is one of the most recent and popular techniques for studying deep levels in semiconductors. In the areas of deep level detection, this technique has progressively supplanted other material characterization techniques (photo conductivity, optical absorption, luminescence and Hall measurements) which are generally sensitive to shallow defect levels.

---

[1] Sula Technologies (Palo Alto, Ca)

[2] Labview is a licensed product of National Instruments Corp

As first outlined by D.V. Lang [6], this technique detects deep levels through the monitoring of the change in junction capacitance transient of a P-N diode resulting from a monitored change of the bias condition of the diode. Although the technique is applicable to random doping conditions, it is more useful for Schottky type barriers (i.e. $n^- - p$ or $p^- - n$ diodes). As the bias voltage is toggled between its original value -Vb0 to -Vb1 as in Fig. 1, the space charge zone at the junction, which was devoid of free charges (swept away by the built-in junction potential) reduces in length, thus allowing temperature dependent majority carrier emission and capture by dopants/traps. The junction capacitance will subsequently vary. Similarly, if the bias voltage is such that it forward biases the junction (+Vb2), minority carriers will be injected predominantly to the least doped side of the junction thus leading to a variation in the junction capacitance.

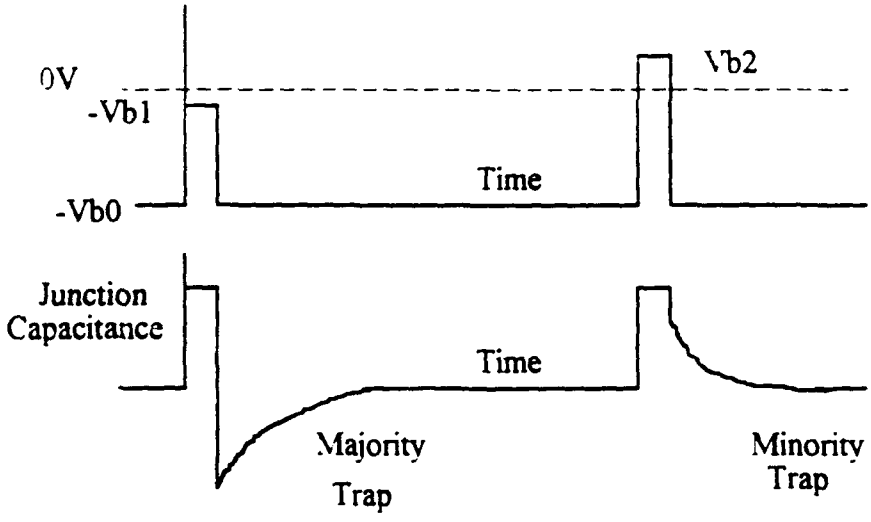


Fig. 1 Junction capacitance transients due to majority (Vbias <0, reverse bias) and minority (Vbias>0, forward bias injection pulse) traps.

In the junction area the emptying /filling of the traps with electrons/holes is governed by the bias condition. It can be shown [6] that the trap concentration can be derived directly from the change in the value of the junction capacitance. For an $n^- - p$ junction the trap concentration is:

$$N_T = 2(\frac{\Delta C}{C})(N_A - N_D), \qquad (1)$$

with N_T being the trap concentration, ΔC the relative change of the junction capacitance (at t=0) with respect to the steady state capacitance C and $N_A - N_D$ the net acceptor concentration.

Furthermore, due to the carrier emission rate of traps as expressed in equations (2), the capacitance transients are temperature dependent and also on recombination processes. It is therefore important to find a temperature

value for which the transient is the most pronounced in order to be able to extract from the transient valuable information of trap energy levels.

$$e_1 = \frac{(\sigma_1 < v_1 > N_{D1})}{g_1} e^{\left(\frac{-E}{kT}\right)},$$

$$e_2 = \frac{(\sigma_2 < v_2 > N_{D2})}{g_2} e^{\left(\frac{-E}{kT}\right)} \qquad (2)$$

$e_{1,2}$ are the minority/majority emission rates. $\sigma_{1,2}$ are the minority/majority capture cross sections. $\langle v_{1,2} \rangle$ are the mean thermal velocity of minority/majority carriers. $g_{1,2}$ are the degeneracy of the trap level . $N_{D1,2}$ are the effective density of states in the minority/majority carrier band and ΔE is the energy difference between the trap level and the minority/majority carrier band)

The thermal emission of traps must sometimes be modified to account for the effect of external fields (Poole Frenkel effect) and can be expressed as follows [7]:

$$\frac{e}{e_0} = \left(\frac{kT}{\beta\sqrt{F}}\right)^2 \left\{1 + \left[\frac{(\beta\sqrt{F})}{kT} - 1\right] e^{\frac{\beta\sqrt{F}}{kT}} + \frac{1}{2}\right\} \qquad (3)$$

where e_0 is the trap thermal emission under no external field. F is the average electric field at the junction and

$\beta = \left(\frac{q^3}{\pi\varepsilon}\right)^{\frac{1}{2}}$, where q is the carrier charge and ε is the dielectric constant. This correction is found to be generally mandatory for a proper understanding of defect behavior particularly in narrow gap materials as the $Hg_{1-x}Cd_xTe$.

The best alternative to the tedious task of monitoring the temperature scan of the junction capacitance transients for determining trap levels is offered by the DLTS technique which sets an emission "rate window" which allows the DLTS experimental setup to detect transient response peak only if the trap emission rate falls within the rate window. Altough several rate window emission techniques were exploited in the past. the exponential "correlator" seems to provide high sensitivity for transient peak detection. It is also this technique that we have used in our DLTS experiments.

As first proposed by Miller et al. in 1975. the exponential correlator uses an exponential weighting function w(t) which is multiplied by the baseline restored transient signal . The product is then integrated to provide a response that will display maxima at transient peak locations when a temperature scan of the diode is performed. The basic principle of this technique is depicted by Fig. 2.

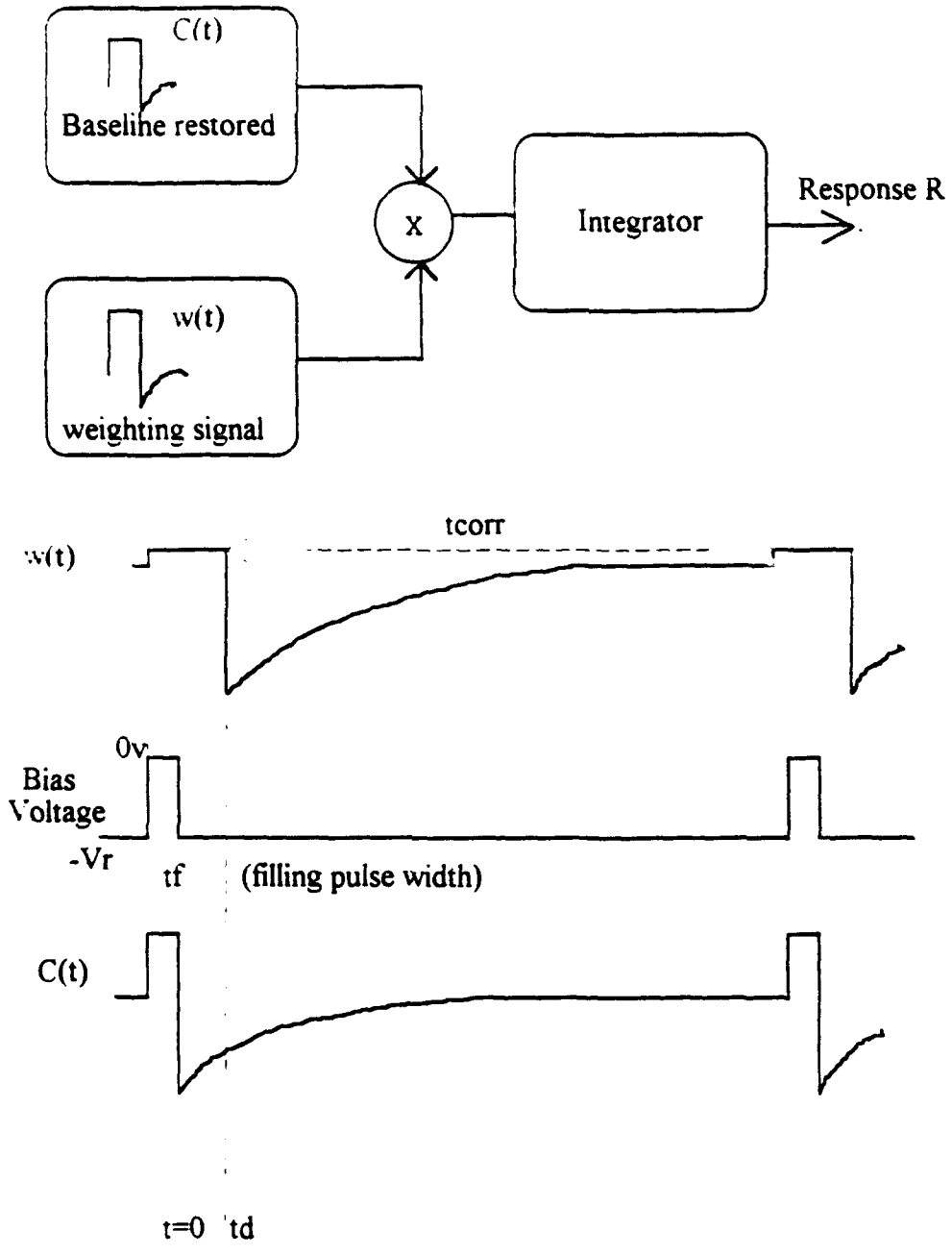


Fig. 2 Basic description of the exponential correlator. Note that the weighting signal and the bias voltage are synchronized

Letting $w(t) = Ae^{\frac{-t}{\tau_c}}$, $C = C(0)e^{\frac{-t}{\tau}} + C(\infty)$ and tcorr=$\beta\,\tau_c$ it can be shown mathematically that the response R will take the following form [15] :

$$R(\tau) = \frac{A}{\beta}\left\{\frac{\tau}{\tau+\tau_c} - \frac{\tau_c}{\tau+\tau_c}e^{\left[\frac{-\beta(\tau+\tau_c)}{\tau}\right]} - e^{\frac{-\beta\tau_c}{\tau}}\right\} \qquad (4)$$

By a proper adjustment of the time delay td and a good selection of the time constant τ a maximum R will be obtained at $\tau = \tau_e$. The DLTS apparatus used in our experiments is based on the correlator technique and is described later in this report.

A good DLTS based temperature scan will therefore provide a series of capacitance transient extrema which identify the existence of majority (minima) or minority (maxima) traps as described in Fig. 3.

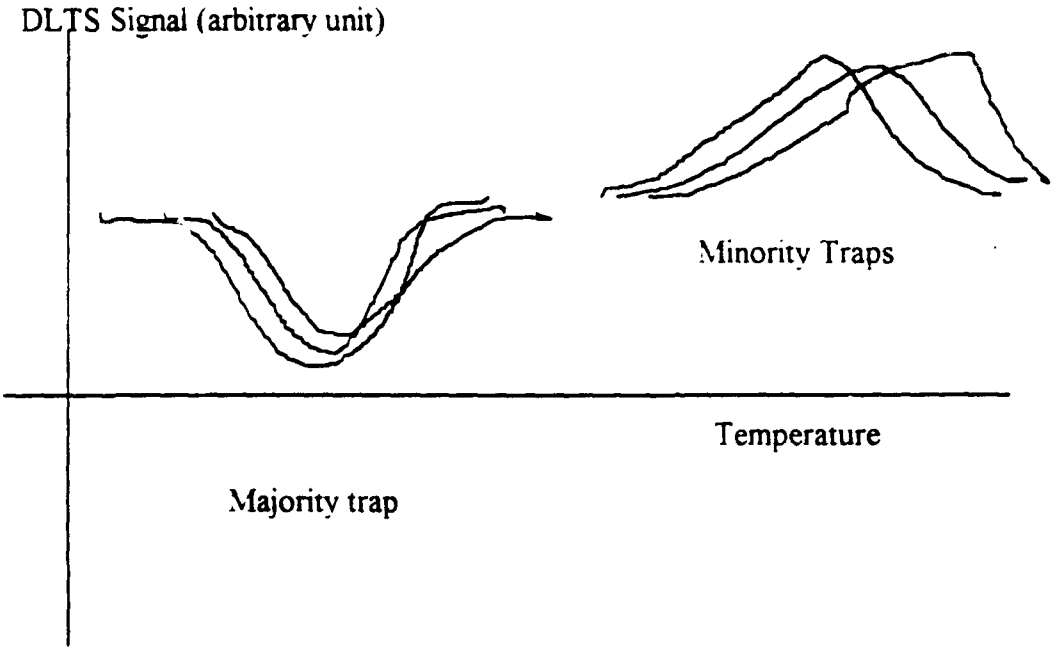


Fig.3 Typical extrema obtained during DLTS.

The basic experimental setup for performing DLTS is depicted in Fig 4.

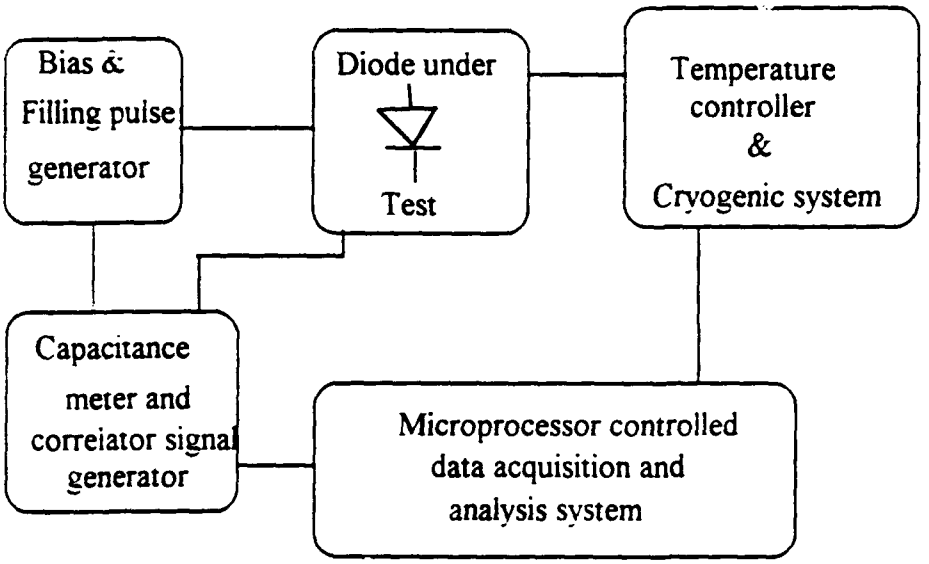


Fig.4 Block diagram of a DLTS apparatus

Given some parameters (carrier concentration, effective carrier mass, x: fraction of cadmium in the $Hg_{1-x}Cd_xTe$) pertaining to the samples under test, DLTS experiments will provide the following main characteristics of the sample:

1. **Trap concentration** $N_T = 2(\frac{\Delta C}{C})(N_A - N_D)$, as per equation (1)

2. **The trap energies** E_T (or trap activation energy) which can be evaluated by evaluating the slope of the plots:

$$\ln(T_{peak}^2 \tau_{peak}) \text{ vs } (\frac{1000}{T}) \qquad \text{known as Arrhenius plots} \qquad (5)$$

where T_{peak} and τ_{peak} represent the temperature and time constant respectively observed on the DLTS scan results. (5) results from the following formula [11]:

$$T^2 \tau = \frac{e^{\frac{E_T}{kT}}}{\sigma\left(\frac{2k^2 m^* g}{\pi^2 \hbar^3}\right)} \qquad (6)$$

where T is the temperature, τ is the trap emission rate , E_T is the trap energy, σ is the carrier capture cross section, m^* is the carrier effective mass and g is the trap degeneracy.

The final trap energy provided by the DLTS experiment will be $E_{DLTS} = E_T - E_\sigma$, where E_σ is the activation energy for cross section evaluated as in the following section (3.). This correction is essential due to the temperature dependence of the capture cross section.

3. **The capture cross section** σ which can be calculated from the following equation [11]:

$$\frac{1}{\tau_f} = \sigma \langle v \rangle n \qquad (7)$$

where τ_f represents the time constant for filling the trap evaluated by studying the variation of the capacitance transient magnitude as a function of the fill time (pulse width of the filling pulse), <v> is the average thermal velocity and n is the carrier concentration (case of an n-type sample).

The temperature dependence of the capture cross section can be expressed as:

$$\sigma(T) = \sigma_\infty e^{\frac{E_\sigma}{kT}} \qquad (8)$$

where σ_∞ can be evaluated from the Arrhenius plots at T=∞ .

For a given trap level , one can first evaluate σ from equation (7) and find E_σ from equation (8).

II.2 Brief description of the CCDLTS.

The Constant Capacitance DLTS is a technique introduced [12,13,14] as yet another way of determining the trap

concentration and energy level. The technique calls for adjusting the bias condition of the diode in order to maintain a constant capacitance across the junction during a temperature scan . Unlike the DLTS experiment outlined in the previous section. the transient to be monitored is the bias voltage transient instead of the capacitance transient.

A typical $p^+ - n$ barrier model with impurities is shown in Fig. 4

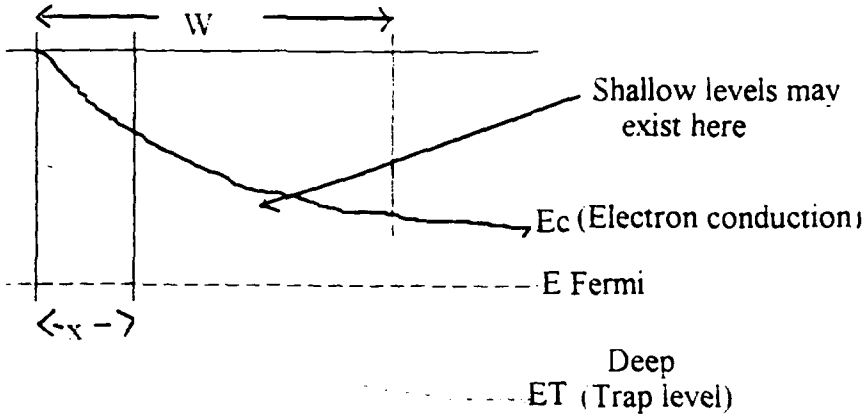


Fig.4 Energy diagram of a p-n diode with impurities. The ionization of the deep level impurities takes place after their energy have passed over the Fermi level. W is the width of the depletion region . x represents the ionized portion of the deep level impurities. Shallow trap levels can however exist in the area λ=W-x).

It can be shown that the trap concentration . the junction capacitance and the bias voltage can be related as follows:

$$\overline{N_T}(x_\lambda)e^{\frac{-t}{\tau}} = \frac{C^2(W)}{\frac{1}{2}\varepsilon q A^2\left(1-\lambda(W)\frac{C(W)}{\varepsilon A}\right)^2}\Delta V_B(t) \quad (9)$$

where $\Delta V_B(t) = V_B(t) - V_B(\infty)$. difference between the bias at initial bias-switching and at an infinite time. $\overline{N_T}$ represents the average trap deep level trap concentration. C(W) is the junction capacitance. A the junction area . From the evaluation of λ (W) which just before application of a filling pulse can be given

by $\lambda(W) = \sqrt{\frac{2\varepsilon(E_F - E_T)}{q^2 N_S(W)}}$ where $\frac{E_F - E_T}{q}$ = Minimum bias voltage producing a discernible transient and

$N_S(W)$ is the shallow level density which just before the application of a filling pulse can be equated to

$\frac{2 \cdot Vbias \times C^2(W)}{\varepsilon q A^2}$. the trap concentration can be derived from equation (9). Also. by plotting $\frac{\Delta V_B(t)}{\Delta V_B(0)}$ vs time . the time constant τ can be evaluated after a proper voltage scan.

III. Experimental setup

III-1 Hardware

The basic hardware used in the DLTS experiment is based on the standard scheme of Fig. 3. The capacitance meter and the correlators block are from SULA Technology. This same block can also accommodate the CCDLTS with some circuit connection modifications. A GPIB-compatible RMC temperature controller/cryogenic system is essential for the temperature scan with good accuracy. The outputs of the four correlators are connected to a data acquisition board compatible with the Labview data acquisition and analysis software (see next section for detail). The board is then connected to a Labview data acquisition board mounted on a slot of an IBM compatible computer.

A wide range of time constant selection is provided (from .02 to 100ms) which is very welcome in the traditionally "trial and error" operation of selecting the right time constant which provides a peak in the DLTS signal.

The outputs of the correlators, which are in effect outputs of an integrator, provide voltage levels which in turn can be converted into a the magnitude of the differential capacitance in picofarad (pf) or trap concentration for DLTS or into trap concentration in the CCDLTS operation. Also, multiple (1 to 1000) pre-amplification gain factors are provided for amplifying weak DLTS signals.

III-2 Software.

The software used to control the experiments is based on state-of-the-art data acquisition package from Labview. The advantage of the version of the software used in our experiments is its ability to avoid any high level programming (C, Basic) for writing the test programs. The concept of "virtual instruments" (VI's) which constitutes the premise of this package makes it possible to create icons for all sorts of operation including meters and x-y plotters. A hierarchical construction of an entire data acquisition system with graphical interface for entering test parameters on the fly can be achieved with good dependability.

The final front panel of the program written for the experiments is highlighted as an attachement to this report. The program is capable of performing the following:

- Setting of the initial and final temperature levels for the scan;
- Setting of all the parameters of the temperature controller ;
- Digital readout of the temperature during scan;
- Graphical readout of the DLTS scan;
- Indicators for testing if the acquisition of data is occurring within the prescribed error range;
- Digital readout of the trap concentration levels for the CCDLTS experiment;
- A switch selectable (Exp 2 switch) option to directly scan and display graphically the bias voltage transient; and
- To transfer to floppy or hard disk all acquired data for spreadsheet applications.

IV Experimental results

IV.1 Preliminary experiments

Preliminary calibration of the experimental was first conducted using a randomly selected AlGaAs sample. The data acquisition program run on this sample produced the following DLTS scan of Fig.5:

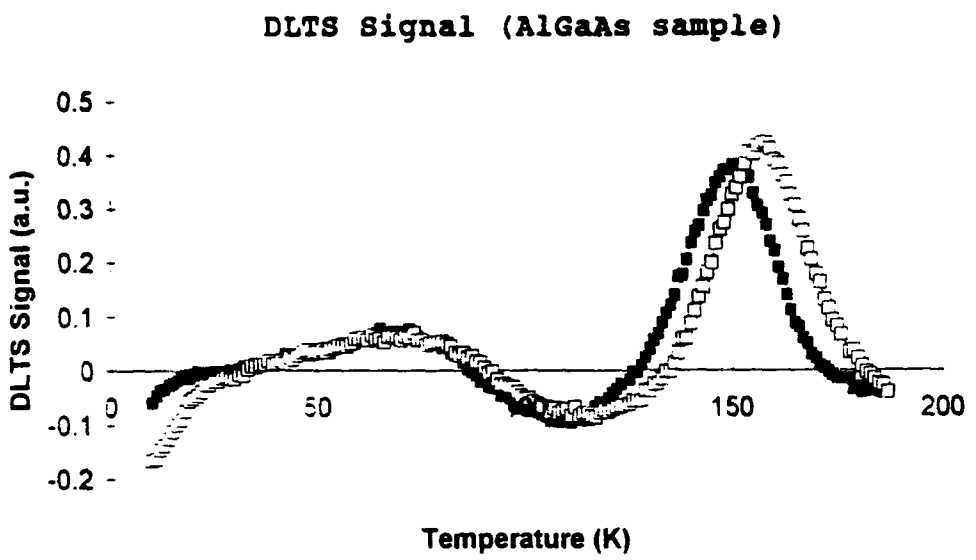


Fig.5 Example DLTS signal of an AlGaAs sample

Also, the $Hg_{1-x}Cd_xTe$ sample diodes received for testing were evaluated through their I-V characteristics. From the samples received from Rockwell as illustrated in Fig.6

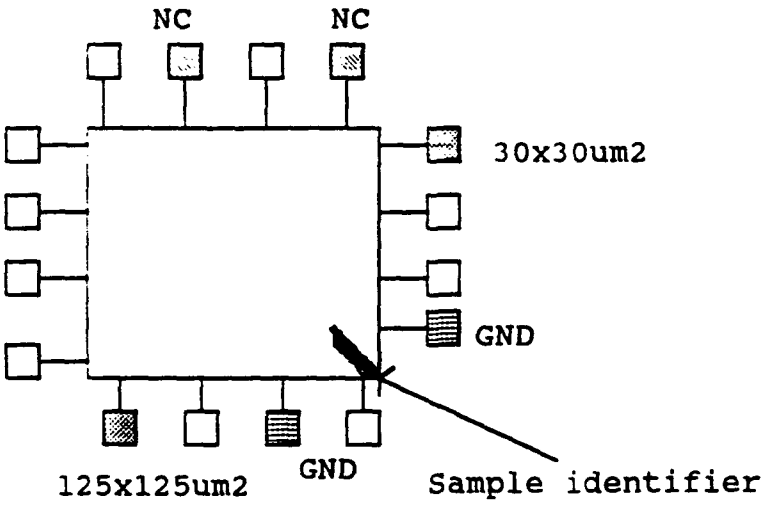


Fig. 6 $Hg_{.724}Cd_{.276}Te$ samples selected for evaluation. Only the samples 125x125um2 and 30x30um2 had a good I-V characteristic.

The I-V characteristic of the 125x125um2 sample is depicted by Fig.7

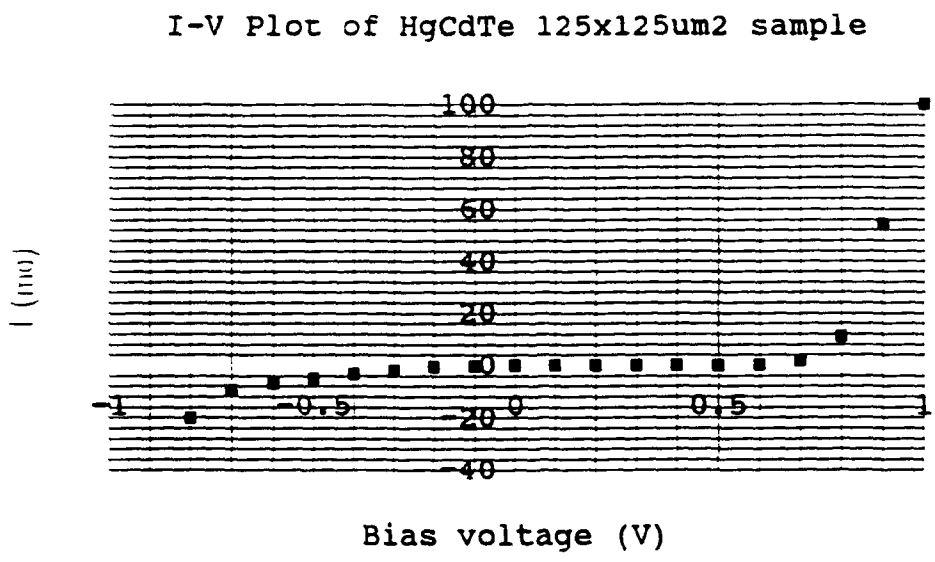


Fig. 7 I-V Characteristic of the 125x125um2 HgCdTe diode.

IV-2 DLTS scan results

The DLTS scan results identified four distinct traps (two majority (electron) traps and two minority (hole) traps)

For each trap level the following experiments and measurements were conducted:

- Temperature scan of the DLTS signal;
- Ploting of the corresponding Arrhenius praph;
- Evaluation of the magnitude of the DLTS signal vs trap fill time; and
- (depending on smple parameter availability) Corrected trap energy level and concentration.

IV-2-1 Results for trap T_{H1} (hole trap at 250K)

The obtained results are compiled in the following figures (Fig. 8.9.10)

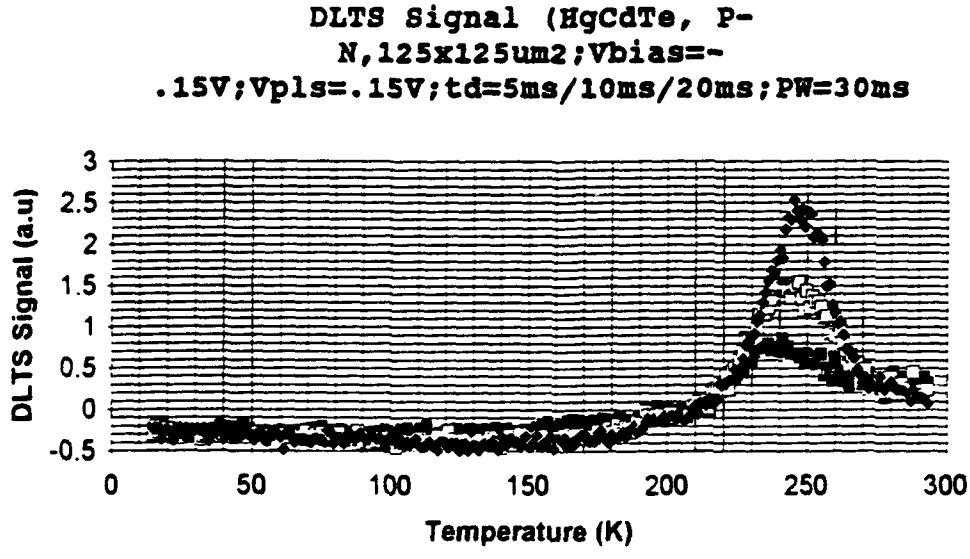


Fig.8 DLTS signal of the 125x125um2 sample (trap T_{H1})

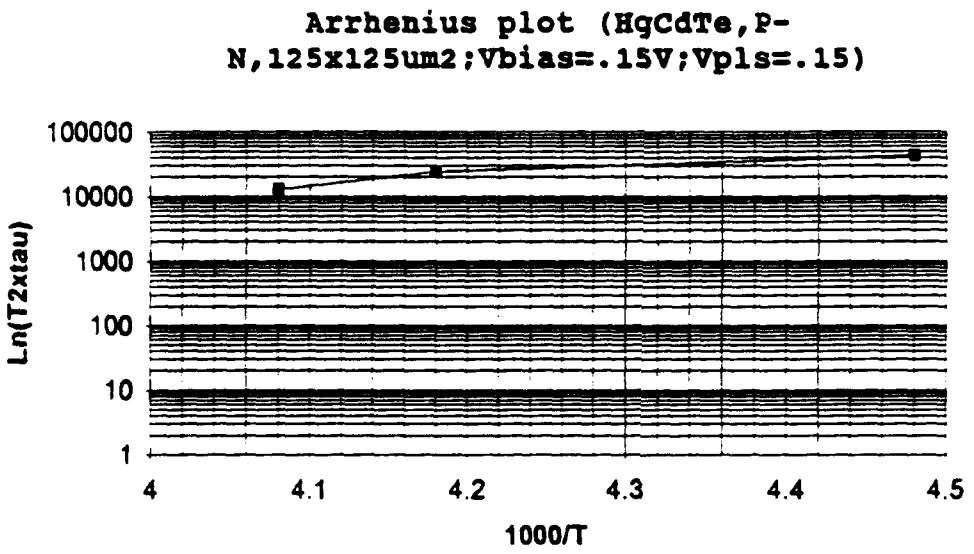


Fig. 9 Arrhenius plot for trap T_{H1}

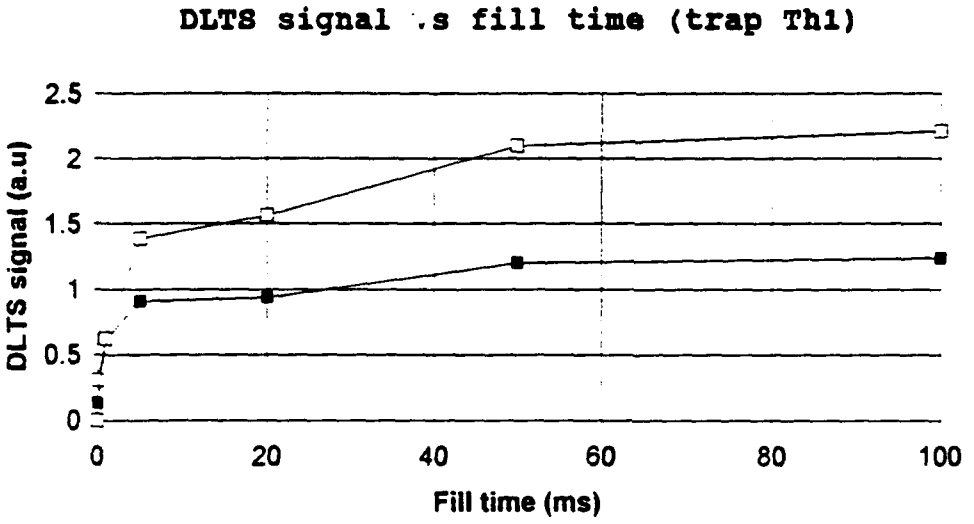


Fig. 10 DLTS signal as a function of the fill time for trap T_{H1} (Vbias=-.15V;Vpls=.15)

IV-2-2 Results for trap T_{H2} (hole trap at 144K)

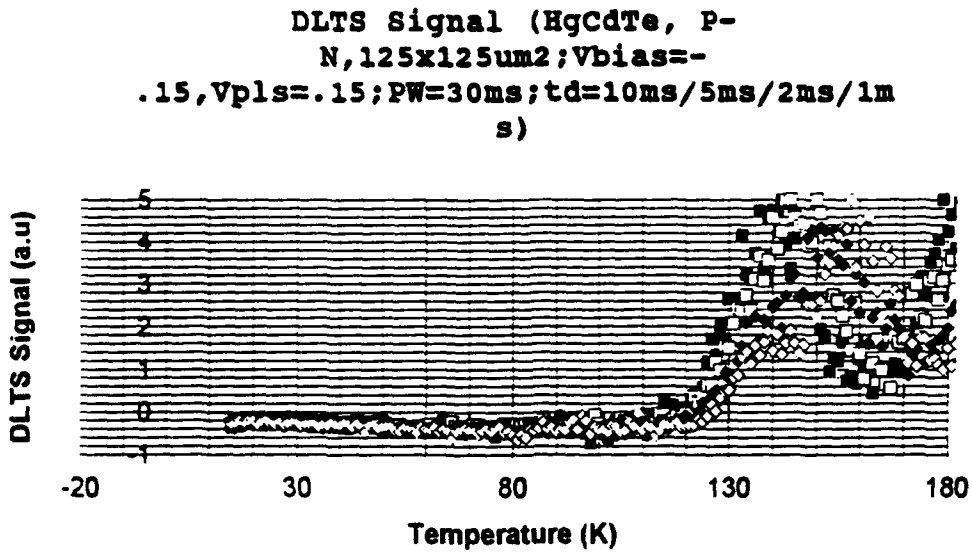


Fig 11 DLTS signal for trap T_{H2}

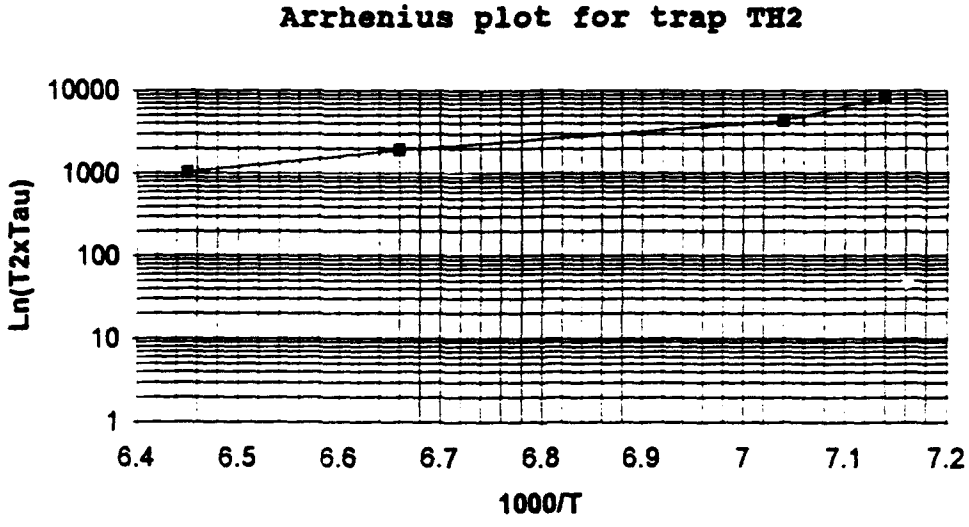


Fig 12 Arrhenius plot of trap T_{H2}

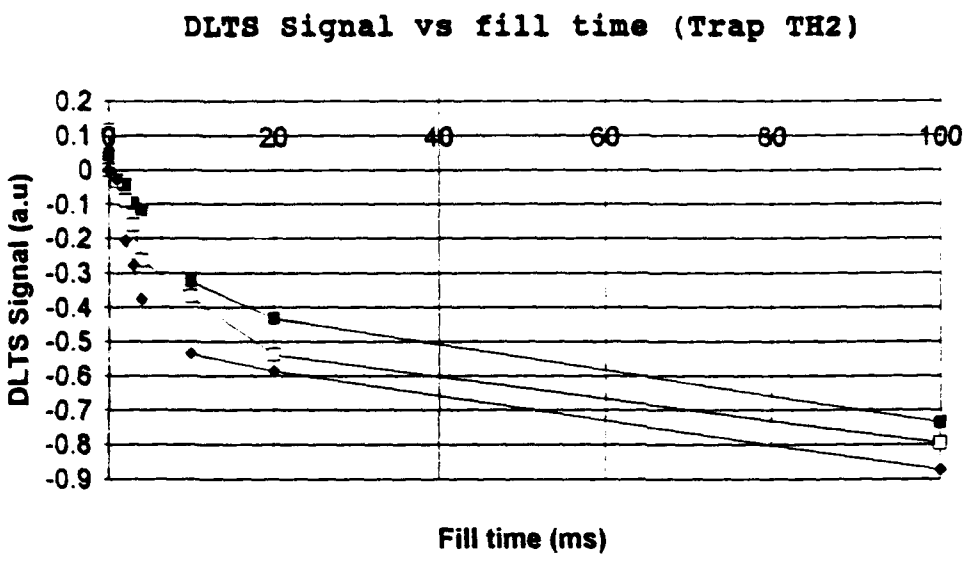


Fig. 13 DLTS signal as a function of fill time for trap T_{H2} (Vbias=-.15;Vpls=.15)

IV-2-3 Results for trap T_{E1} (Electron trap at T=45K)

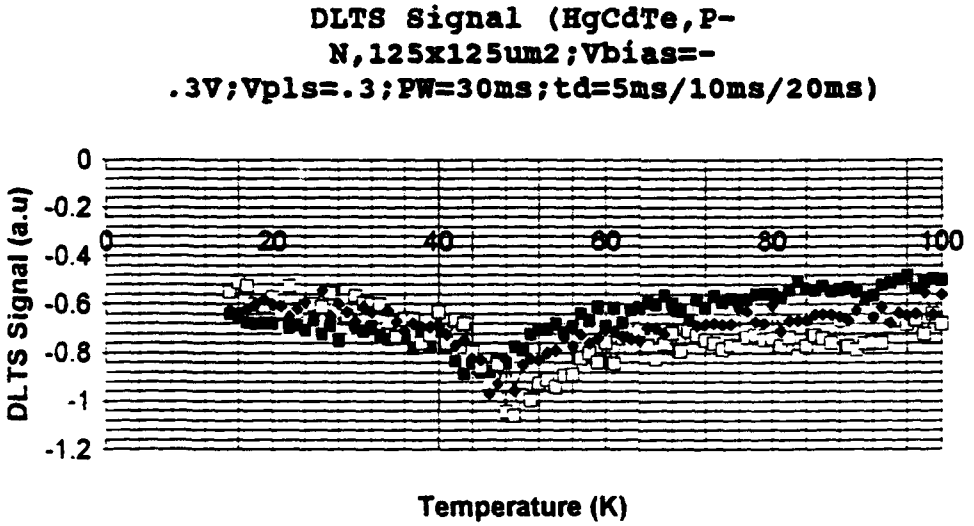


Fig . 14 DLTS signal for trap T_{E1}

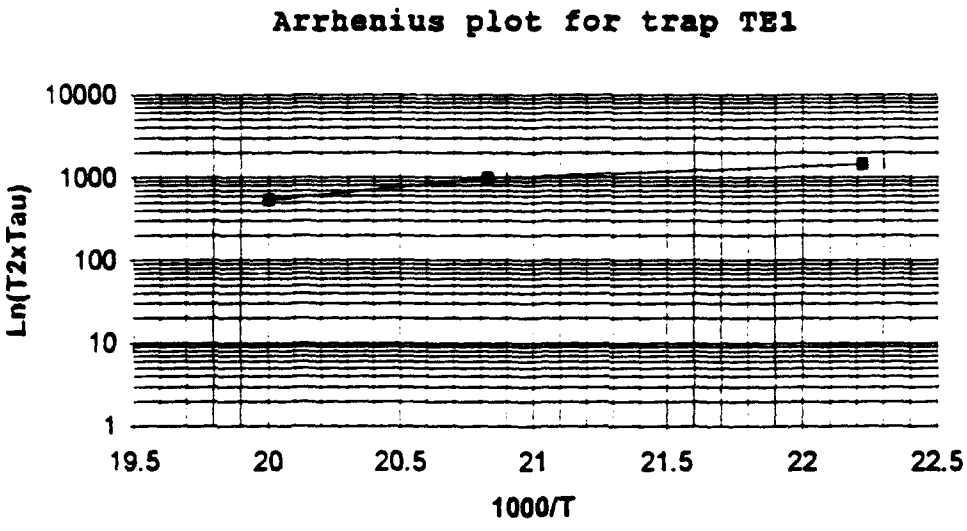


Fig. 15 Arrhenius plot of trap T_{E1}

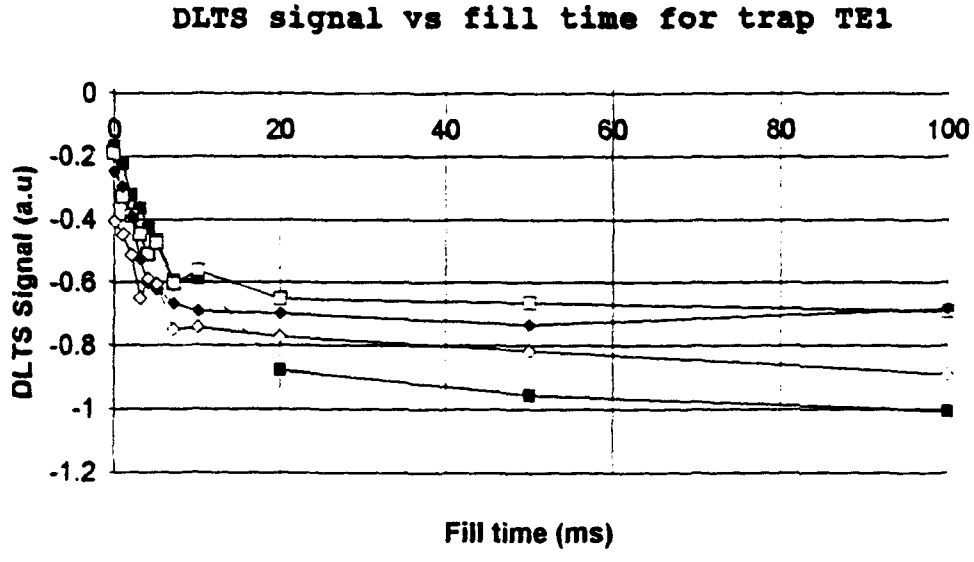


Fig. 16 DLTS Signal vs fill time for trap T_{E1}

IV-2-4 Results for trap T_{E2} (Electron trap at T=290K)

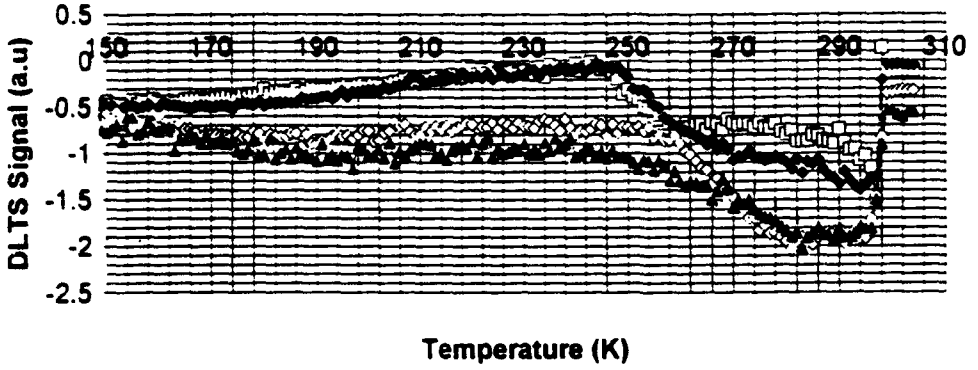
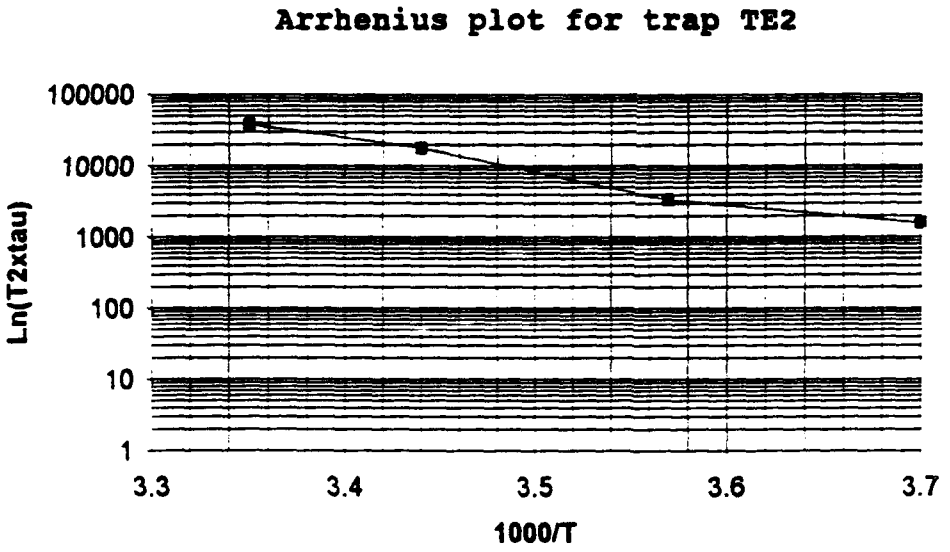


Fig. 17 DLTS scan for trap T_{E2}



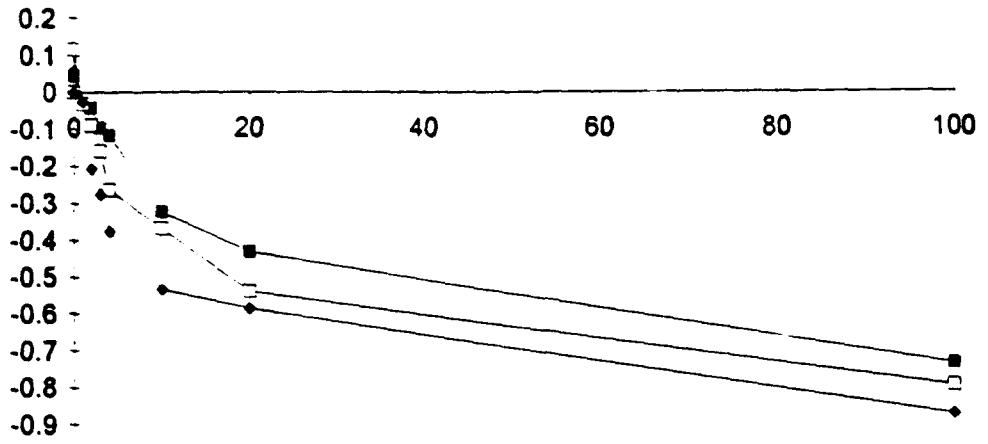


Fig. 19 DLTS signal vs fill time for trap T_{E2}

IV-3 Summary of results.

The following table contains a summary of the experimental observations and calculations for all the traps:

	T (K)	C0 (pf)	C (pf)	NT(cm -3)	ET (eV)*	v (cm -s-1)	sig0(cm 2)	sig(cm 2)	Esig(eV)	Edlts(eV)*
TH1	250	0.15	15	7.00E+13	0.26	1.05E+07	3.36E-24	4.80E-11	0.16	0.1
TH2	144	0.3	8	2.62E+14	0.56	1.38E+07	3.36E-25	6.05E-11	0.4	0.12
TE1	50	0.06	7.1	5.91E+13	0.19	3.36E+07	2.47E-23	5.31E-22	0.05	0.14
TE2	290	0.09	38	2.10E+13	-0.5	8.12E+07	1.23E-25	2.19E-22	-0.18	0.13

* Numbers not final.

The sample parameters used for the calculations are:

$N_D = 3.5 \times 10^{15} cm^{-3}, n_i(77K) = 1.8 \times 10^{10} cm^{-3}, m_n^* = .017 m_0, m_p^* = .5 m_0$ as provided by Rockwell.

In the table C0 represents ΔC(0). v represents the thermal velocity $\langle \upsilon \rangle \langle v_{n,p} \rangle = \sqrt{\frac{8kT}{\pi m^*_{n,p}}} \langle v_{n,p} \rangle = \sqrt{\frac{8kT}{\pi m_{n,p}}}$ as in equation (7), sig(0) is the magnitude σ_{∞} as in equation (8). sig is the capture cross section σ and Esig is the activation energy of the trap E_σ.

Remarks:

- All the experimental results are obtained from the Labview program labeled **CCDLTS5.VI**
- The CCDLTS loop adjustment of the SULA equipment could not provide a stable constant capacitance loop. Therefore the CCDLTS experiment was put on hold.
- For a good scan it is recommended that error detection of the set temperature value be set between +.3K and -.7K for low temperature values (T<50K). For higher temperature values the error margin can be resticted to between +.3K and -.3K.

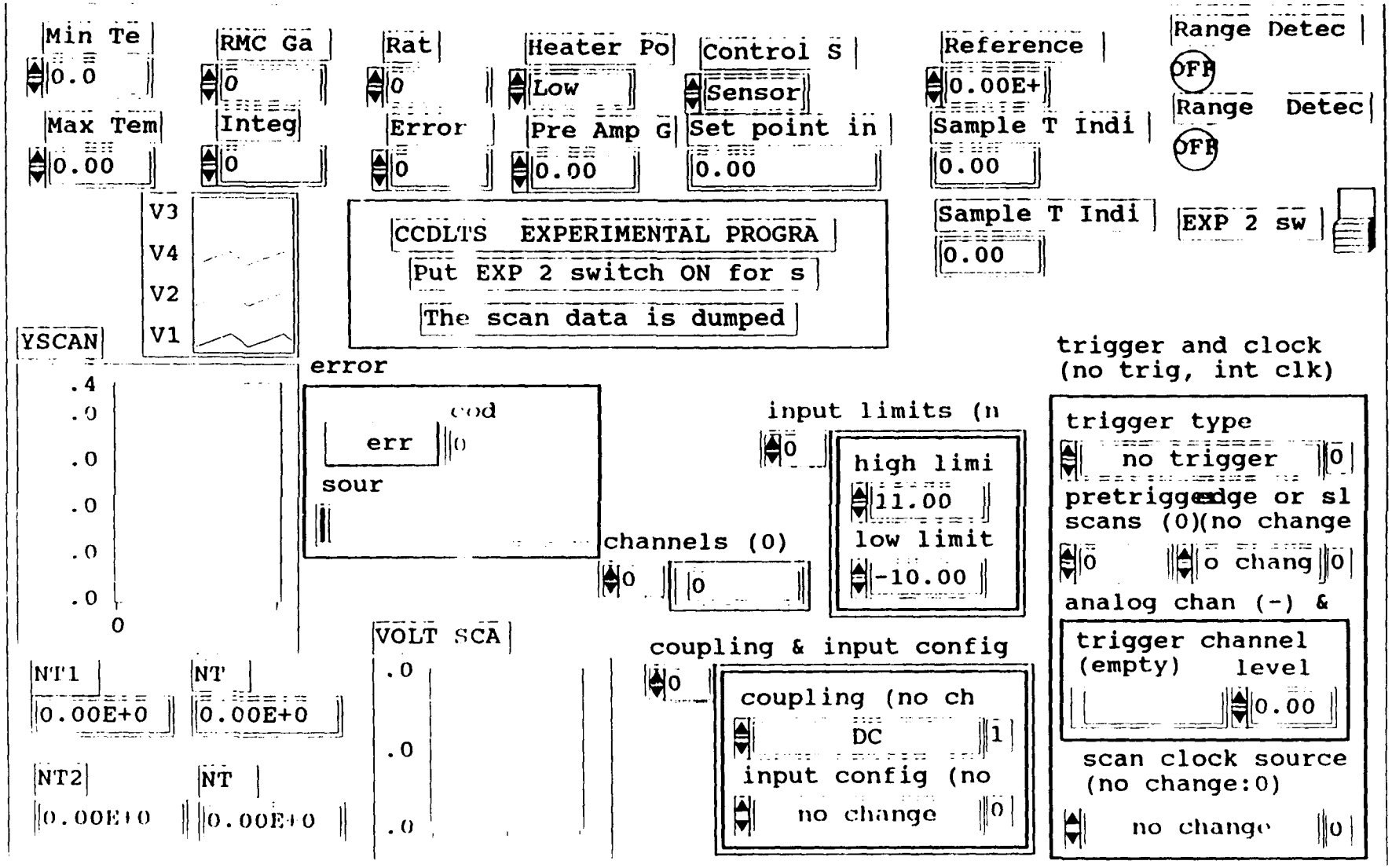
- Further studies of the MBE grown HgCdTe samples should provide more insight on the trap conditions.

Acknowledgments.

I am grateful to the Office of Scientific Research of the Air Force for having given me the opportunity to perform this work under its sponsorship. I wish to thank Dr. Jose Colon for having devoted much effort in helping me throughout my work.. I am also very grateful to all the technicians. engineers. scientists and other members of the sensors research team at the Phillips Laboratory for their support.

References

1. R.E. DeWames. J.M Arias. L.J. Kozlowski . SPIE Vol. 1735. 1992
2. R. Balcerak. Lynette Brown. J. Vac. Sci. Technology. B 10(4), 1992
3. W. Tennant. C.A. Cokrum. M.A Kinck. M.B. Reine, R.P. Ruth. J. Vac. Sci. Technology B 10(4) .1992
4. K.H Herrmann. M. Happ. H. Kissel. K.P Mollmann and J.W. Tomm. J. Appl. Phys. 73 (7), 1993
5. S. Zachaman. Eliezer Finkman and G. Bahir
6. D.V Lang, J. Appl. Phys. Vol. 45. No. 7. 1974
7. D.L Polla. C. E. Jones. J. Appl. Phys. 52(8). 1981
8. H.S. Tan. M.K Han and S.C. Ng
9. C.E Jones and D.L. Polla. J. Vac. Sci. Technology, 21(1). 1992
10. E. Finkman. J. Appl. Phys. 54 1883-6. 1983
11. G.L. Miller. D.V. Lang. and L.C Kimmerling, Ann. Rev Mater Sci. 7. 377. 1977
12. G. Gotto. S. Yanagiasawa. O. Wada and H. Takanashi. J. J. Appl. Phys., Vol. 13. No 7. 1974
13. Z. C. Huang and C.R Wie. J. Appl. Phy. 73 (9), 1993
14. J.J. Shian. A. L. Fahrenbrenbruch and R. H. Bube. S. Sta. Elec. Vol. 30. No. 5. 1987
15. P. Blood and J.W Orton "The Electrical characterization of semiconductors: majority carriers and electron states". Academic press. 1992


Min Te
0.0
RMC Ga
0
Rat
0
Heater Po
Low
Control S
Sensor
Reference
0.00E+
Range Detec
OFF
Max Tem
0.00
Integ
0
Error
0
Pre Amp G
0.00
Set point in
0.00
Sample T Indi
0.00
Range Detec
OFF
V3
V4
V2
V1
CCDLTS EXPERIMENTAL PROGRA
Put EXP 2 switch ON for s
The scan data is dumped
Sample T Indi
0.00
EXP 2 sw
YSCAN
.4
.0
.0
.0
.0
.0
0
error
cod
err
0
sour
input limits (n
0
high limi
11.00
low limit
-10.00
trigger and clock
(no trig, int clk)
trigger type
no trigger
0
pretrigge edge or sl
scans (0)(no change
0
o chang
0
analog chan (-) &
trigger channel
(empty)
level
0.00
scan clock source
(no change:0)
no change
0
channels (0)
0
0
VOLT SCA
.0
.0
.0
coupling & input config
0
coupling (no ch
DC
1
input config (no
no change
0
NT1
0.00E+0
NT
0.00E+0
NT2
0.00E+0
NT
0.00E+0

(Attachment)

A PRELIMINARY DESIGN OF A DUAL SOURCE FOR SIMULATING THE AM0 SPECTRUM

V. SIVARAM MURTY
ASSOCIATE PROFESSOR
DEPARTMENT OF COMPUTER SCIENCE AND PHYSICS

TEXAS SOUTHERN UNIVERSITY

3100 CLEBURNE

HOUSTON, TX 77004

FINAL REPORT FOR

SUMMER FACULTY RESEARCH PROGRAM
PHILLIPS LABORATORY
KIRTLAND AIR FORCE BASE, NEW MEXICO

SPONSORED BY:
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
ROLLING AIR FORCE BASE, WASHINGTON ,D.C.

SEPTEMBER 1993

PRELIMINARY DESIGN OF A DUAL SOURCE FOR SIMULATING THE AM0 SPECTRUM

V. SIVARAM MURTY
DEPARTMENT OF COMPUTER SCIENCE AND PHYSICS
TEXAS SOUTHERN UNIVERSITY
HOUSTON, TX 77004

Abstract

The use of multijunction photovoltaic cells in space applications has required sophisticated testing procedures using sources which can faithfully reproduce the AM0 solar spectrum. In the past commercial simulators using xenon lamps have been used. However they present a problem in that there are atomic lines in the near infrared spectrum of xenon, which severely hamper accurate testing in that spectral. region. This work presents some very preliminary design for a dual source using an Oriel lamp simulator and a tungsten lamp to produce the AM0 spectrum.

Introduction

Until recently multijunction solar cells have not been available in the efficiency ranges that would make them suitable for extensive practical use. However there have been many developments that have led to the design and production of several multijunction solar cells, many of which have found a wide acceptance in space applications.. The advent of these high efficiency cells has made it necessary to devise new test procedures in the laboratories to characterize these cells. This in turn has needed the use of a solar simulator that can simulate the Air Mass Zero solar spectrum in the laboratory. There have been many apparatuses on the market that simulate the AM0 spectrum. Most notables are the Xenon lamp simulators used with a variety of filters.

While these simulators do a good job of simulating the solar spectrum in the ultra violet and the visible region, they present a problem in the near infra red region., because of the extremely intense atomic lines of xenon in this wavelength range.. The AM0 spectrum as obtained by the simulator manufactured by Spectrolab Inc. for example does not tackle this issue. The problem thus presented is that, that these intense atomic lines which cannot be successfully filtered out, cause introduction of thermal energy into the cells when the higher energy band gaps of the multijunction device are being utilized.

It therefore becomes necessary that there be a simulator with two different sources with different spectral characteristics, one that simulates the region in the UV-visible and the other source that can provide the spectral characteristics in the near infra red region. Obviously the Xenon lamp properly filtered can be used for the UV-Visible region, while for the near infrared region a source which essentially has a continuous blackbody spectrum in that region is needed. These radiation from these two sources can then be mixed to produce the source necessary for the simulator..

The present work describes preliminary attempt to select two such sources and necessary filtering optics that will produce the required AM0 spectrum in the entire region from 200nm to 2000 nm.

Procedure

The spectrum obtained from the xenon lamp source simulator manufactured by Oriel Corporation and the AM0 spectra were digitized at the Aerospace Corporation, El Segundo, CA. In order to select the proper filters the ratio of the intensities of the Xe lamp spectrum and the AM0 spectrum were calculated at the different wavelengths. This quantity is defined as the transmission factor. Wherever the ratio exceeded unity, its value was set at unity. The complement of the transmission (1.0-Transmission) was also calculated and tabulated.

The information obtained f rom the spectra and the calculations are ploteed in the two charts that accompany. The chart marked CHART4.XLC shows the Xe lamp spectrum, the AM0 spectrum and the transmission . The chart marked CHART7.XLC shows the absorbance at different wavelengths.

Data Analysis

Both CHART5 and CHART7 show very expected results, namely that the spectrum in the region from 200nm to approximately 850 nm resembles the AM0 spectrum. The nature of the AM0 spectrum for wavelengths beyond 850nm is essentially a blackbody spectrum. Wien's Law considerations dictate a source at a temperature of about 2900 degrees Kelvin.

Therefore the following items are needed to design the necessary simulator (a) A source similar to the source of the Oriel AM0 simulator with (b) a filter or cold mirror that will absorb the infra red component

of the radiation from this source; and transmit/reflect most of the UV-visible component ;and (c) a tungsten lamp operating at approximately 2900 degrees Kelvin with a filter or cold mirror that will transmit/reflect the infrared component from the tungsten lamp and absorb the UV-Visible parts of the radiation from this tungsten lamp.. In addition it is suggested that bifurcated randomized fiber optic bundles be used to mix the properly selected radiation from these two sources to be used in conjunction with the solar cells.

Preliminary Selection of the Optical Components.

A search for appropriate filters, cold mirrors and optical fiber bundles available on the market led to the selection of the following components.:

(a). Oriel MCS series cold mirror at 0 degrees angle of incidence . This mirror absorbs all wavelengths from 200 nm to 800nm. and reflects 97 % of the energy at higher wavelengths. This mirror can be used in conjunction with the tungsten lamp to for the high wavelength portion of the simulator.

(b)Oriel 51960/59060 filter that filters out the infrared component and transmits 90% of the UV-visible component. This filter can be used in conjunction with the Oriel or other simulators using the Xenon lamp for producing the UV-Visible component of the overall simulator radiation

(c) For randomly mixing the radiation from these two sources., an Oriel fused silica bifurcated fiber optic bundle with low attenuation in the overall region of operation is to be used.

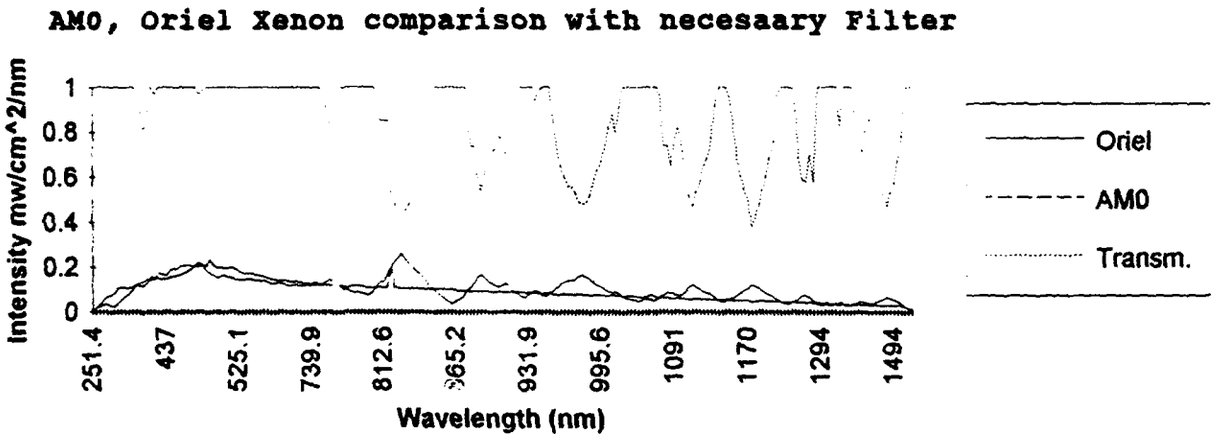
Recommendations for Further Work.

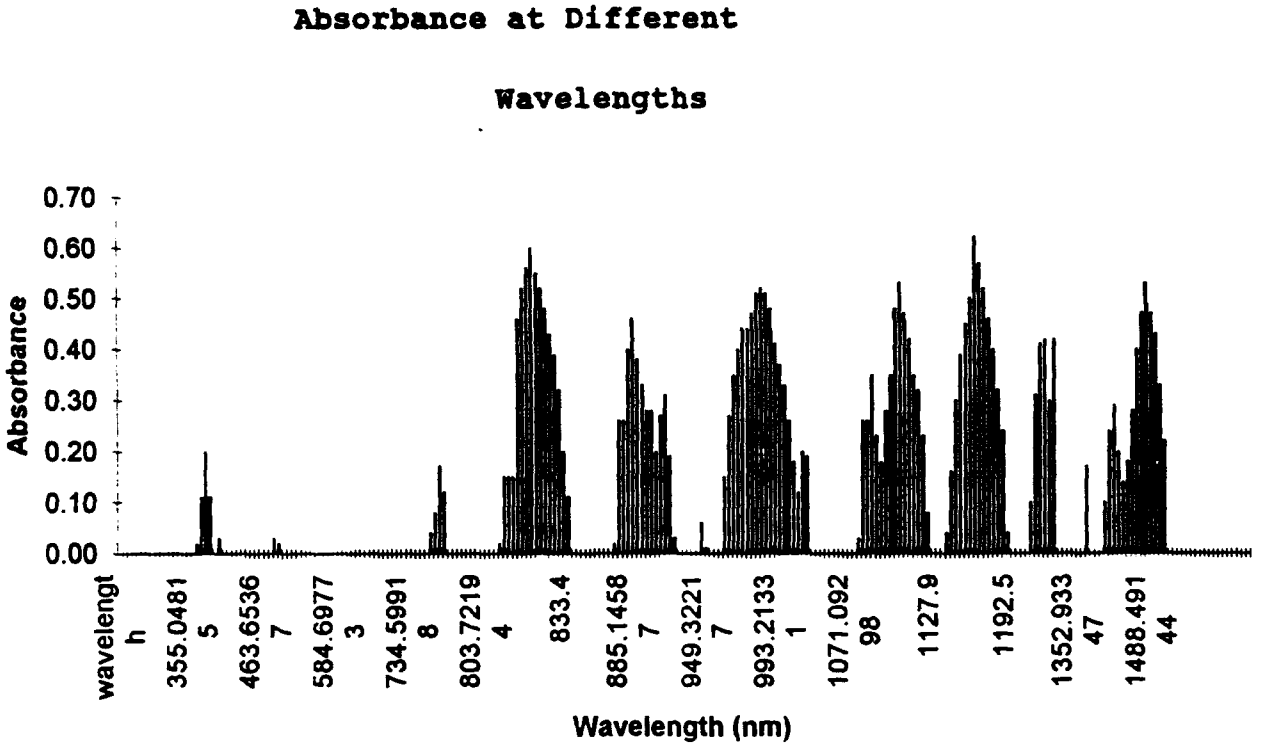
The future mission of the Cinventional Energy Branch of the Phillips Laboratory lists the design and construction of a AM0 simulator spectrum for use in the entire wavelength region, as a top priority item. Following his initial stab at a design of such a simulator, the next step in this venture would be to design the necessary apertures and stops that are needed to obtain the proper intensities at the different wavelengths. and to obtain the total integrated intensity of 1650 W/m2.

Therefore a detailed analysis of the spectrum obtained from this preliminary design is needed. and an intensive analysis of the spectrum i needs to be performed.

Acknowledgments

This work was supported by United Sates Air Force Office of Scientific Research, Bolling AFB. Washington DC. The author is grateful to this office for the award of a Summer Faculty Research Participation Associateship. The work was performed at and for the Phillips Laboratory, Kirtland AFB, NM . The author would like to thank Lt. Col. Ernest Hererra, Dr Inara Kuck, Captain Kelly Gaffney and Mr. Ralph James of Phillips Laboartory for providing him the facilities for the work and for invaluable discussions on this project. The author would also like to thank Dr. Dean Marvin and Mr. John Nocerino of Aerospace Corporation who made available the facilities to obtain the spectra. that were needed for this work.


AM0, Oriel Xenon comparison with necesaary Filter
Intensity mw/cm^2/nm
1
0.8
0.6
0.4
0.2
0
251.4
437
525.1
739.9
812.6
865.2
931.9
995.6
1091
1170
1294
1494
Wavelength (nm)
Oriel
AM0
Transm.


Absorbance at Different Wavelengths
Absorbance
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
wavelength
355.04815
463.65367
584.69773
734.59918
803.72194
833.4
885.14587
949.32217
993.21331
1071.09298
1127.9
1192.5
1352.93347
1488.49144
Wavelength (nm)

References

1. Measurement Techniques For Multijunction Cells, Gary F. Virshup, *Proceedings of the 22 nd IEEE Photovoltaic Specialists Conference* (IEEE, New York, 1990)

2. Mixing Optical Beams For Solar Simulation, Bhushan L. Sapori et al., *Proceedings of the 22nd IEEE Photovoltaic Specialists Conference* (IEEE, New York, 1990)

3. Two Source Simulator For Improved Solar Simulation. M. Bennett and R. Podleany, *Proceedings of the 20 th IEEE Photovoltaic Specialists Conference* (IEEE, New York, 1990)

4. Private Communication. Keuth Emery (1993)

DESIGN OF AN ACTIVE VIBRATION CONTROL SYSTEM FOR ISOLATION OF AN OPTICAL BENCH

Dr. Richard Patton
Assistant Professor
Mechanical Engineering Department

Mississippi State University
P. O. Drawer ME
Mississippi State, MS 39762

Final Report For:
Summer Faculty Research program
Phillips Lab
PL/LIAA
3500 Aberdeen Ave. , SE
Kirtland AFB, NM 87117-5776

Sponsored by:
Air Force Office of Scientific Research,
Bolling AFB, Washington DC

July, 1993

DESIGN OF AN ACTIVE VIBRATION CONTROL SYSTEM FOR ISOLATION OF AN OPTICAL BENCH

Dr. Richard Patton
Assistant Professor
Mechanical Engineering Dept.
Mississippi State University
Drawer ME
Mississippi State, MS 39762

Abstract

The isolation of an optical bench mounted on an airplane using active noise cancellation was studied. The most difficult problem is the necessity for very high levels of isolation at low frequencies (0.1 Hz). At these frequencies, rapid separation of the base motion into its rigid body component and its vibration component becomes a problem; low pass filters take too long for satisfactory performance. Failure to properly separate the signal can cause the bench to run to the isolator stops, causing performance degradation. An active noise cancellation filter using strain gauges (which have no rigid body component) as the reference signal was simulated with several different algorithms. The algorithms were tested under a variety of different frequency inputs and different modal frequencies of the body. A recursive least squares filter was found to give the fastest convergence and best results. A least means squares, on the other hand, gave unsatisfactory performance.

DESIGN OF AN ACTIVE VIBRATION CONTROL SYSTEM FOR ISOLATION OF AN OPTICAL BENCH

By: Dr. Richard Patton

Introduction

The isolation of an optical bench mounted on an airplane was studied. For the purposes of this problem, passive isolation systems were inadequate due to the very small vibratory motions permitted (nanoradian range). The modal frequency of the isolator is 0.1 Hz. The problem is further complicated by the physical problems that an airborne mounting poses. Airplanes are subject to turbulence (large random inputs), have very low frequency modes (0.5 Hz.) and have a base acceleration which is unknown. Unlike a bench mounted to ground, which has a 1 g acceleration in the vertical direction, a bench mounted on an airplane has a variable base acceleration. This creates the difficult problem of separating the base motion from low frequency vibrations. Figure 1 demonstrates the problem. In this figure, a constant acceleration base motion, and a cosine (0.1 Hz) base motion are compared. As can be readily seen, the cosine motion does not appreciably differ near zero from the constant motion. Simple filtering cannot begin to distinguish between the two until they separate, which requires a minimum of 1/2 second. After that time, the filter will slowly separate out the two signals. Unfortunately, by that time the bench will have moved substantially, and possibly reached the limit of its motion. It is thus necessary to distinguish between the two using a different approach. The approach chosen was adaptive noise cancellation (ANC) [1].

Isolator Design

The isolators utilized for the mounting the optical bench have the following two components:

1) a pneumatic isolator, providing passive vibration control through isolation
2) a linear motor, providing active vibration cancellation.

The provision of the linear motor means that the force exerted by the motor will be under computer control. The motor necessitates the use of an active vibration control scheme. The entire system contains the following major components:

1) optical bench
2) 4 or more isolators
3) an indeterminate number of sensors placed in the body of the aircraft
4) an indeterminate number of sensors placed on the bench
5) computer and appropriate interfaces to sensors and actuators

Active Isolator Design Problems

Insofar as the active control portion of the isolator is concerned, the following problems are paramount:

1) caging the bench.

As the airplane maneuvers, it creates a variable g-load on the bench. The bench, in order to remain caged, must have a force exerted upon it which reflects this g-load. This force is exerted by the linear motors in the isolators. Alternatively, when the plane is flying level, the pneumatic isolators exert a force which creates 1 g of acceleration at the bench, which is the correct amount. In this case, the linear motor should not force the bench out of alignment. In both cases, the problem is the same: a part of the system must be concerned with distinguishing between the rigid body acceleration of the aircraft and the flexural acceleration at the base of the isolators.

While it may seem that the rigid body acceleration can merely be filtered out of the signal, a short reflection will show that this is not

easy to accomplish. The isolator will have a modal frequency at 0.1 Hz. As such, one cycle would take 10 sec. to complete. Since any filter would take at least 1/8 of a cycle to determine the rigid body acceleration, this translates into a 1.25 second wait for convergence. The bench could easily slip out of its cage by then. Even worse, the age of the data which would be used for active vibration cancellation would be decidedly old. A filter that slow is simply unacceptable.

2) prediction of bench motion

Indirect adaptive control requires the use of a predictor. Based on the results of the predictor, the active portion of the control loop exerts a force on the bench designed to keep it caged and to reject unwanted noise. This is the feed forward portion of the loop.

The effects which the predictor must address are the following:

1)base motion excitation (through pneumatic isolator)
2) persistent bench vibration (from prior causes)
3) bench excitation.
4) isolator linear motor force

Items (1), (3) and (4) are inputs to the system, while item (2) is the dynamic response of the bench to past excitation. Thus, the X-vector to the system must include each of these items. For the purposes of this project, it is assumed that the bench excitation (3) is zero. To cage the bench, the base excitation motion excitation must be separated into its rigid body and flexural components. This requires an active noise cancellation (ANC) filter.

The persistent base excitation is measured by the strain gauge-accelerometer system used to determine the rigid body motion. In fact, the system can utilize this as an input directly. The bench excitation motion can be measured by an accelerometer located on the bench. The control input into the linear motor is known directly; the time history

of this value goes into the X-vector also. Thus, the X-vector can be derived from these two measurements, plus knowledge of the time history of the control input.

Characterization of the system:

It is assumed that there is no knowledge of the bench system. Thus, the responses to the inputs must be found adaptively. While this may be the ideal, the fact is that there will be an enormous number of problems associated with finding the response in real-time. Two schemes are offered:

1) Take a large amount of data off-line, and use various adaptive schemes to message it (e.g. RLS or LMS). Choose which algorithm is the best in terms of fast convergence and available computing power, and utilize it in real time. The chosen parameters can be used as a starting point.

2) Begin in the same way as above, but do not use a full adaptation scheme when it comes to the real time algorithm. Many of the components and parameters will not change from run to run. In recognition of that fact, it simpler scheme can be used. For example, a simple adaptive scale factor may provide sufficient adaptation and performance. This would be much simpler to implement, and can sometimes give excellent results with a minimum of computing power.

Rigid-body motion:

The rigid body motion estimate obtained from the strain-gauge-accelerometer system can now be used. The rigid body motion essentially becomes the desired output which the control system must attain. It is then a simple matter to compute the correct input, u. The controller is then very similar to a model reference adaptive controller, with the rigid body motion used as the reference model.

Active Noise Cancellation Filter

The design of the active noise cancellation filter (ANC) consists of utilizing a strain gauge output as a reference signal to an ANC filter. The accelerometer output is the raw input to the filter. The output of the ANC is the rigid body motion of the system.

Simulation Model:

The simulation model consists of two masses:

mass 1) - a base, or rigid body mass
mass 2 - an airplane flexural, or modal mass

The motion of mass 1 represents the rigid body motion of the plane. The motion of mass 2 represents the motion of the plane at the base of the bench. The two masses are connected by a spring and damper. The modal mass was varied to create different modal frequencies. The frequency varied from 0.5 to 5 Hz. A small amount of damping was used in association with the spring; the viscous damping rate was 1% of the spring rate, reflecting the low internal damping of metal. Forces were exerted on the modal mass at 1,2,3 and 5 Hz. It was assumed that a strain gauge was installed in the plane; this corresponds to a displacement measurement between mass 1 and mass 2. Acceleration was measured at the base of the bench (mass 2). The objective was to develop an adaptive filter utilizing the accelerometer output and the strain gauge output, which outputs the base motion.

Results

Figure 2 shows a test case with sinusoidal motion of the rigid body at .16 Hz. with a 5 Hz. mode. A mismatch in the initial conditions causes the 5 Hz. ringing in the motion of mass 2. The ringing gradually dies out due to damping.

Figure 3 illustrates the problem of dimensions. A strain gauge corresponds to a relative displacement, whereas accelerometers measure acceleration. It was found that using a strain gauge as the reference signal in a Least Mean Squares (LMS) adaptive filter did not prove to be very satisfactory in filtering the accelerometer output. Apparently, the problem is that the difference in the two signals causes the problem. Thus in Figure 3, while the ringing is largely filtered out of the estimated rigid body acceleration, so is the rigid body motion (indicated by the smooth, sine curve). Thus, utilizing the strain gauge as a reference signal did not work in this case.

Figure 4 shows a partial resolution of this problem. In this figure, a numerical derivative of the strain gauge output was used as the reference signal. It was used to filter the velocity, obtained by numerically integrating the accelerometer output. An adaptive LMS filter was used. A numerical derivative of the filter output was taken as the estimated rigid body acceleration.

The results are encouraging. In this case, despite the presence of much ringing, the rigid body acceleration is not lost, although it would be desirable for the ringing to be reduced. However, the LMS filter was found to have a serious flaw; it required different filter constants (mu) for successfully filtering different modal frequencies. Because a range of modal frequencies is present in the airplane, this algorithm was discarded as unsatisfactory.

A recursive least squares (RLS) was then tried, with the numerical derivative of the strain gauge as the reference signal filtering the numerical integral of the acceleration. The results are shown in figure 5. The RLS in general gave much better results, with less ringing, than the LMS algorithm. It did not require many states to give good results; the RLS used in figure 5 has just 3 states. It also gave good results when used with different modal frequencies.

Conclusions

An adaptive noise cancellation using recursive least squares and strain gauge output as the reference signal has been successfully used to filter the output of an accelerometer. The filter output is an estimate of the base. rigid body motion of the system. This output will be used as input to the active isolation system which is under development.

Two problems arose from using this system. The first was that an LMS algorithm gave unsatisfactory results as compared to an RLS. The second was that the reference signal had to be of the same type as the filtered signal. For this reason, the input signal (acceleration) was integrated to yield velocity, and the reference signal (relative displacement) was differentiated to yield velocity.

References

[1] Widrow, B. and Stearns, S. D., "Adaptive Signal Processing," Prentice-Hall, Englewood Cliffs, NJ, 1985.

FIGURE 1

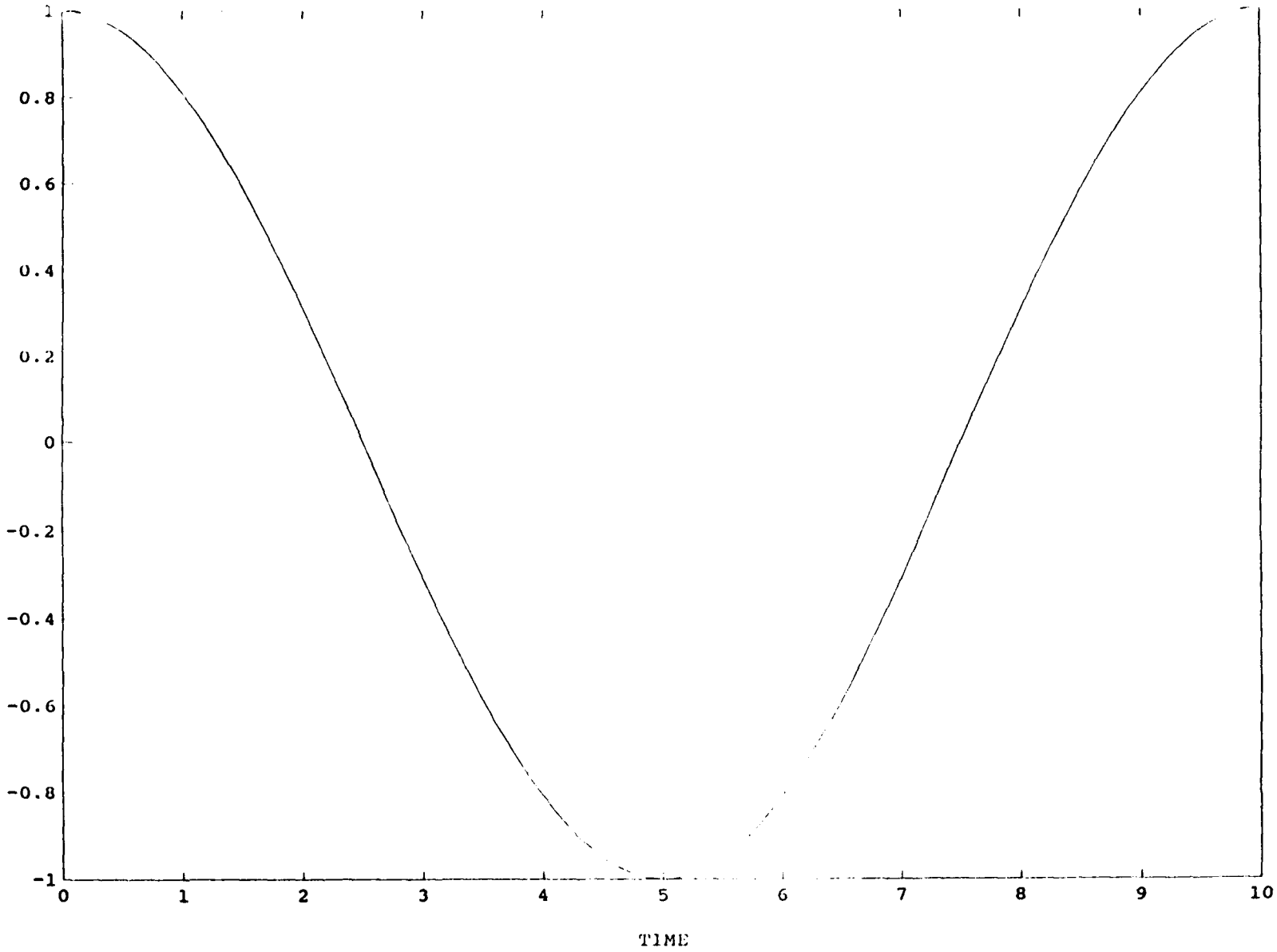


FIGURE 2 - ACCELERATION OF MODAL MASS

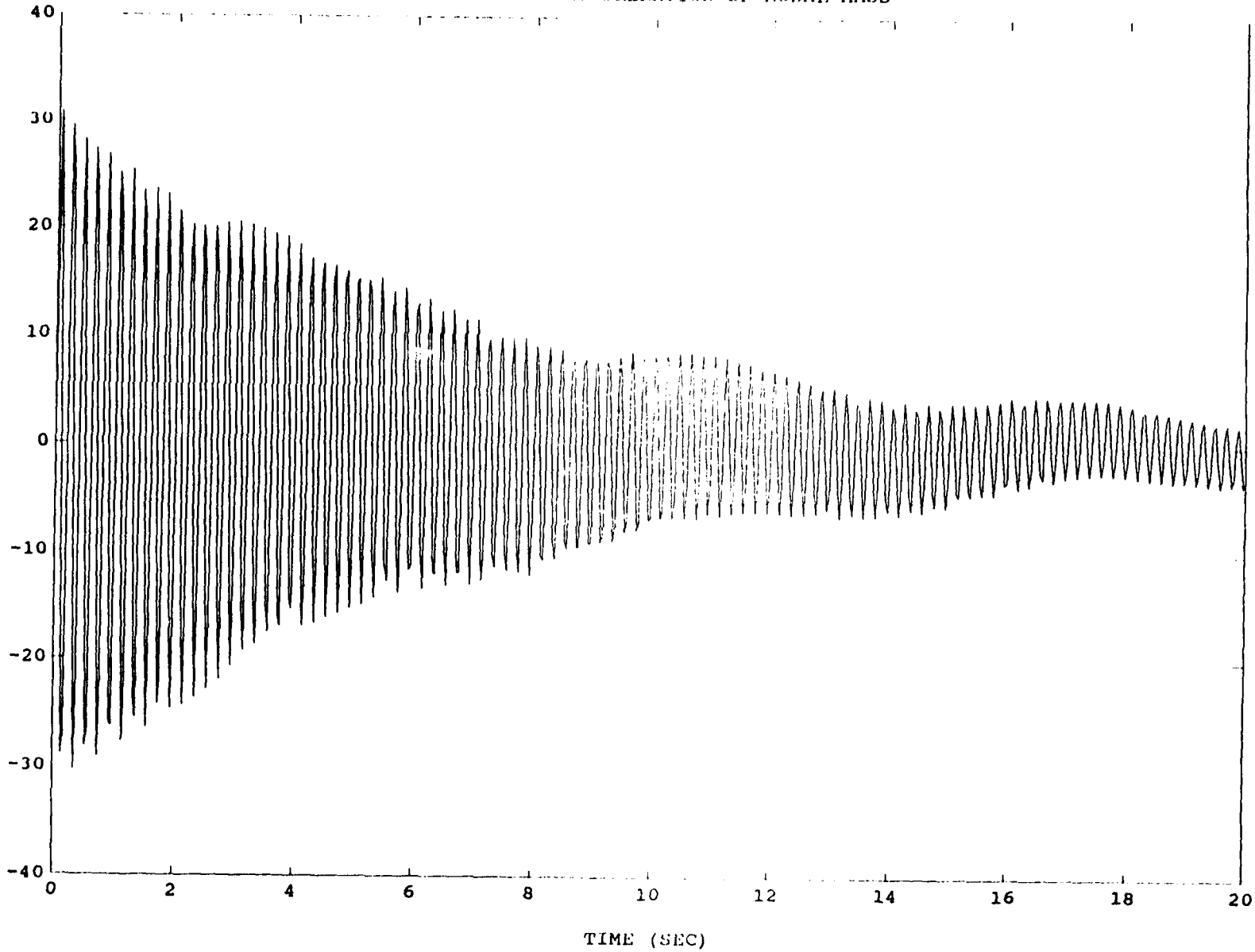


FIGURE 3 - ESTIMATED AND TRUE RIGID BODY ACCELERATION

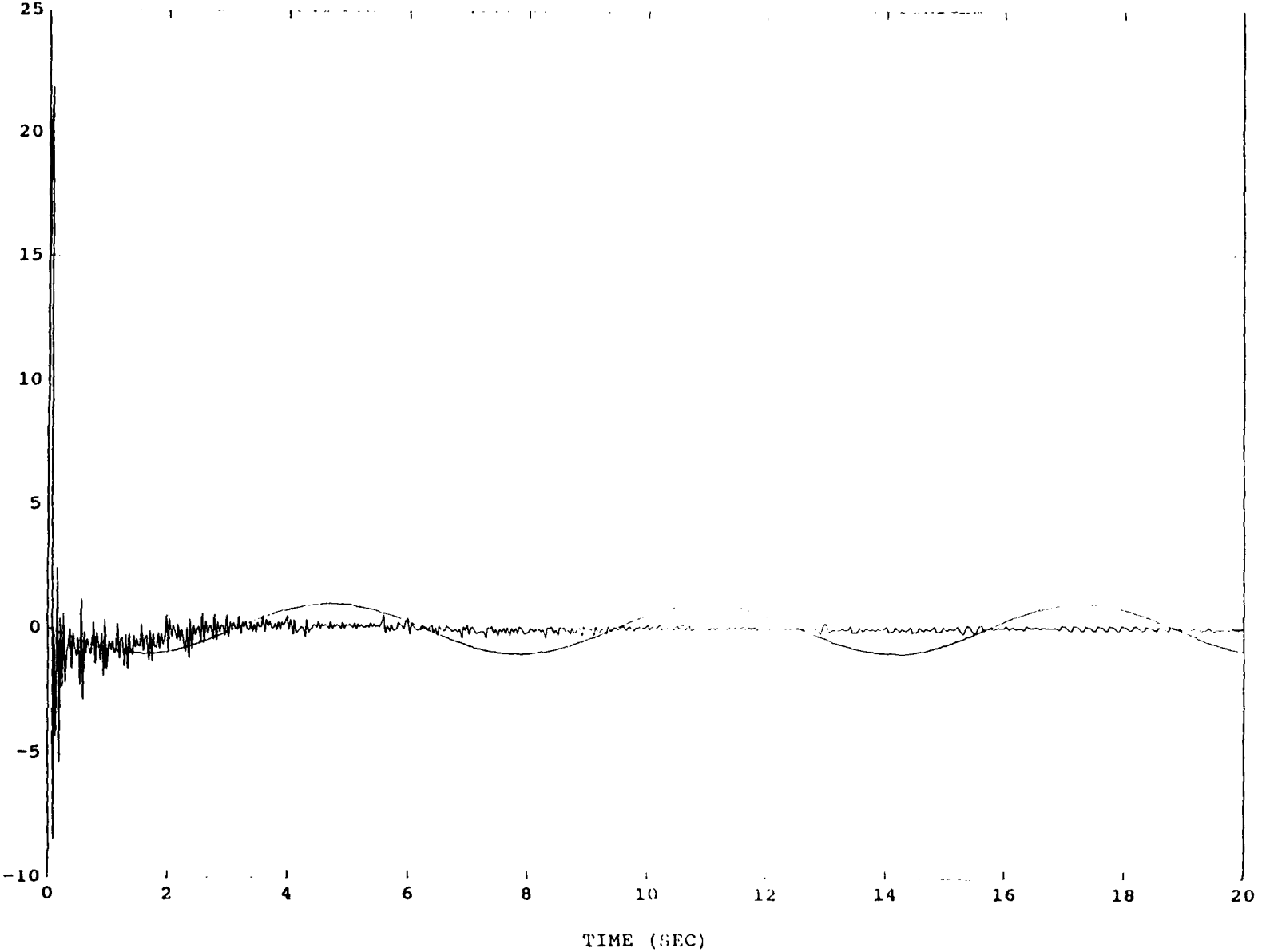


FIGURE 4 - ESTIMATED AND TRUE RIGID BODY ACCELERATION

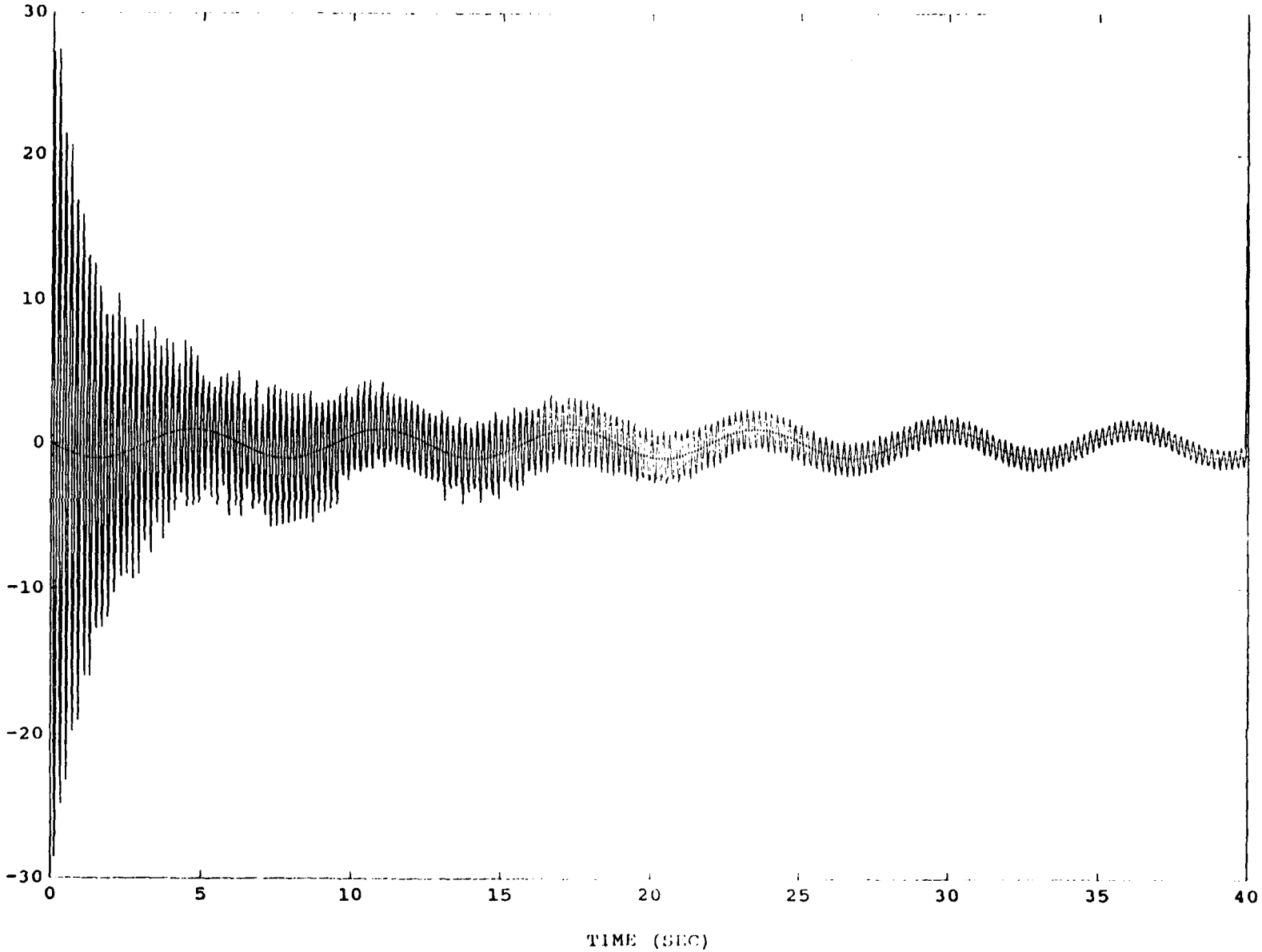
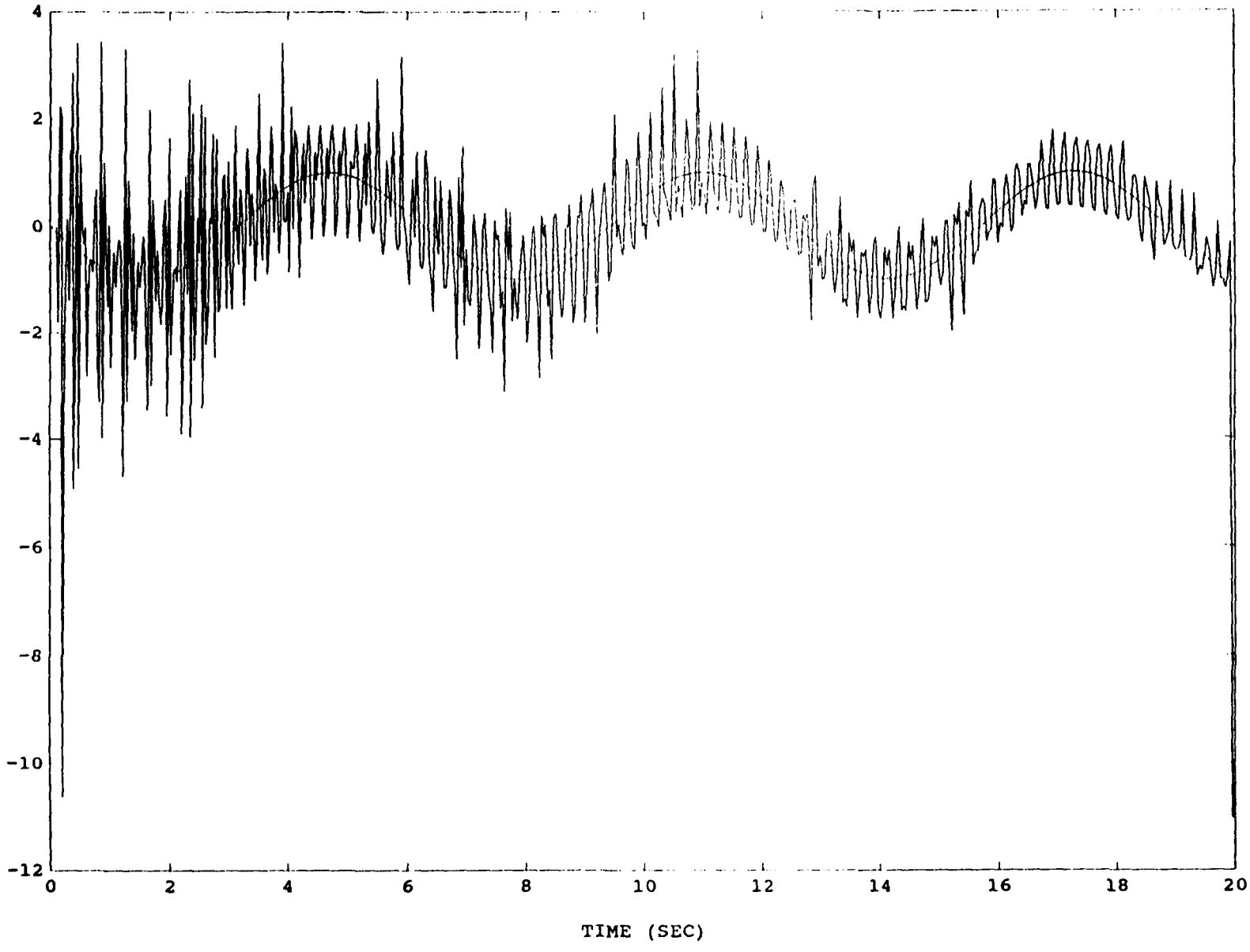


FIGURE 5 - ESTIMATED AND TRUE RIGID BODY ACCELERATION



WARPING OF FLAT COMPOSITE ISOGRID PANELS DURING CURE

Dr. Christopher A. Rotz
Associate Professor
Department of Manufacturing Engineering and Engineering Technology

Brigham Young University
435 CTB
Provo, Utah 84602

and

Thomas D. Kim
Phillips Laboratory
Edwards Air Force Base, California

Final Report for:
AFOSR Summer Research Program
Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Edwards Air Force Base, California

September 1993

WARPING OF FLAT COMPOSITE ISOGRID PANELS DURING CURE

Thomas D. Kim
Phillips Laboratory

Dr. Christopher A. Rotz
Associate Professor
Department of Manufacturing Engineering and Engineering Technology
Brigham Young University

Abstract

This report describes the work done to determine why nominally flat composite isogrid panels designed for the Clementine satellite were warping during autoclave curing. Analytical models indicated that the largest factor was the mismatch in the coefficients of thermal expansion of different parts of the tooling. The next most important factor was the mismatch between the skin and rib materials of the isogrid itself. New tooling was designed to eliminate the first factor, and successfully used to make a ribs-only panel with no warping. An analytical model developed in the project predicted that warping caused by the rib-skin mismatch would be 0.13 in. over a length of 58 in. A panel with skin was manufactured and found to have a warp of 0.15 in. Further improvements in producing flat panels will depend on finding appropriate materials for the rib and skin.

WARPING OF FLAT COMPOSITE ISOGRID PANELS DURING CURE

Dr. Christopher A. Rotz
Thomas D. Kim

1. INTRODUCTION

The Phillips Laboratory at Edwards Air Force Base, California has developed a technique for filament winding fiber reinforced composite isogrid structures. Isogrids would have a shorter lead time for design and manufacture, be faster and less expensive to produce, and be lighter in weight than their aluminum or honeycomb counterparts (1, 2). Potential areas of application include rocket payload shrouds and fairings, aircraft fuselages, and other aerospace structures.

The Phillips Lab has successfully manufactured flat isogrid panels approximately 2 ft x 2 ft in size (3). Recently, the Laboratory began a project to design, fabricate and test composite isogrid support panels, approximately 3 ft x 5 ft in size, for the solar arrays on the Clementine Satellite. The panels had to be flat within a tolerance of 0.01 in over their entire length. This was not thought to be a problem since the small panels had all met this constraint. Unfortunately, the large panels warped during autoclave curing. Several changes in materials and tooling were tried and some improvements were made, but none were successful in meeting the flatness requirement.

The purposes of this project were to determine the causes of the warping, and to devise a method for producing flat panels within the tolerances required by Clementine. This paper will present a summary of the theoretical and experimental work done to understand the problem, a proposed method for manufacturing flat panels, and the experimental work done to verify that the proposed solution was correct.

2. BACKGROUND

To avoid needless duplication of effort, the first step in this project was to gather all available information on the prior attempts to manufacture the large isogrid panels at Phillips Lab. It was hoped that some conclusions about the causes of the warping could be drawn once the data were organized and examined.

Figure 1 shows two of the Clementine panels: an inboard panel on the left and an outboard panel on the right. (The inboard panel shown was intentionally manufactured without a skin layer as part of an experiment that will be explained later.) The inboard panel was approximately 46 in long by 18 in wide, while the outboard panel was approximately 46 in by 25 in. Both panels were manufactured oversize and trimmed around the outside edges to their final dimensions.

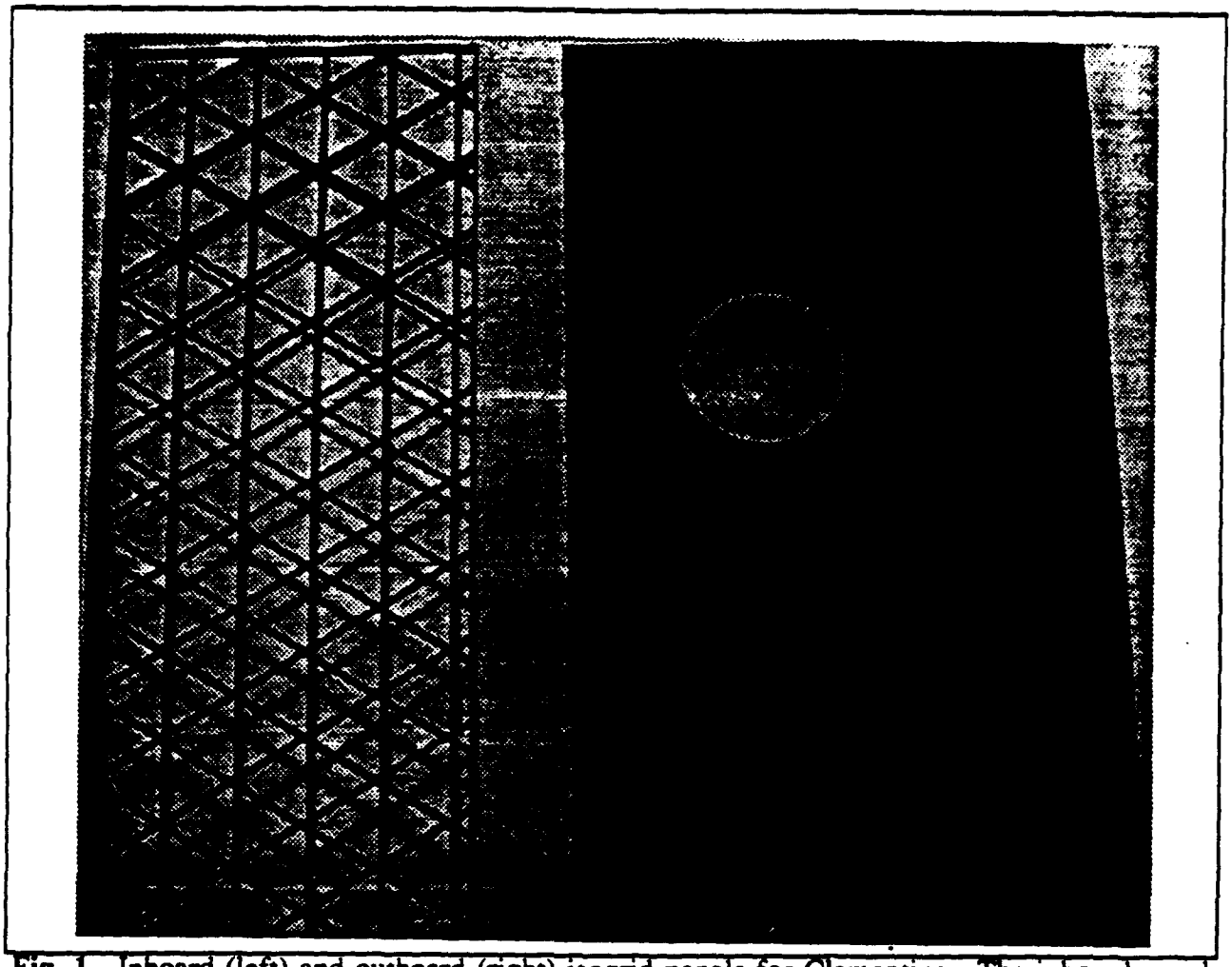


Fig. 1. Inboard (left) and outboard (right) isogrid panels for Clementine. The inboard panel shown was manufactured without a skin layer. The inboard panel was approximately 46 in x 18 in, while the outboard panel was 46 in x 25 in.

Figure 2 shows some important dimensions for the ribs. Figure 3 shows the basic tooling used to manufacture the first isogrids. The base plate was a 0.5 in thick K-100 aluminum plate. The "mold" was made of silicone rubber as described in Ref. 3. The steel borders (0.375 in high by 1.0 in wide) were only loosely bolted to the base plate. In some cases the borders were removed before curing. There was a 0.25 in expansion gap between the border and the mold, as shown in the figure. The caul plate was a 0.5 in thick precision ground steel plate. The isogrid part was fabricated by winding between 35 and 40 prepeg tows into grooves in the rubber mold to form the ribs. Prepreg tape was laid over the mold and ribs to form the skin.

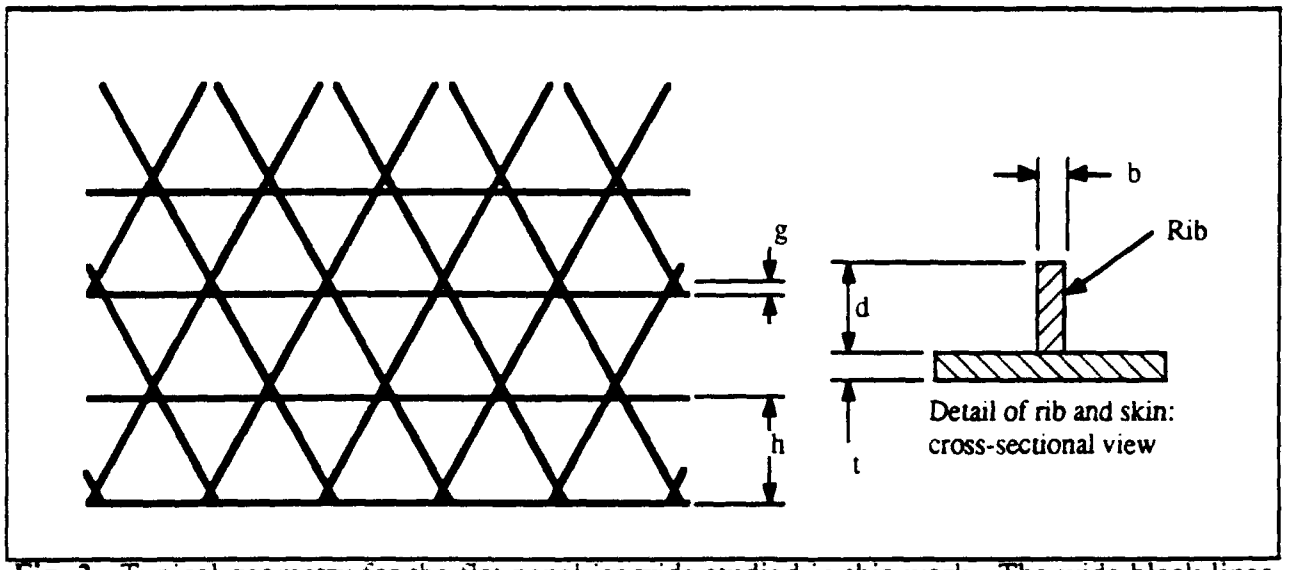


Fig. 2. Typical geometry for the flat panel isogrids studied in this work. The wide black lines are the ribs. The dimensions were: d = 0.52 in, b = .052 in, h = 3.2 in, and g = 0.3 in. The dimension t depended on the number of skin plies (see Table 1).

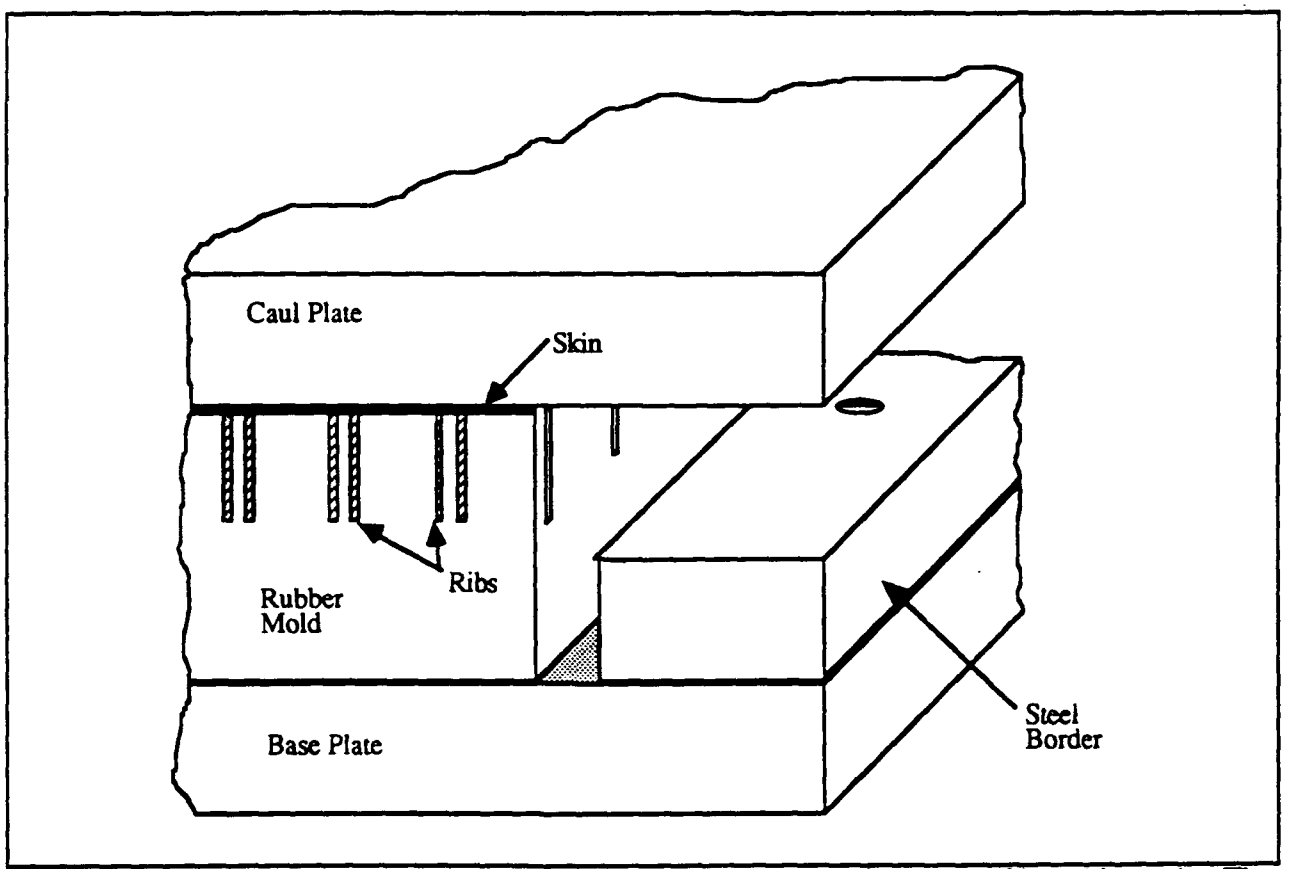


Fig. 3. Schematic diagram of basic tooling used in original manufacturing of isogrid panels. The ribs were formed by winding prepreg tow into slots in the silicone rubber mold. Prepreg tape was then laid over the ribs in the mold to form the skin. Note the 0.25 in gap between the steel borders and the rubber mold. Bagging material and release sheets are not shown.

Alternate tooling systems used for fabricating some of the panels are shown in Figures 4 and 5. The "barrier-frame" setup, shown in Figure 4, was similar to the original tooling, except that the steel borders were replaced by a frame made from steel bars (0.25 in x 1.375 in in cross-section), as shown. The bars were pinned together but were not attached to the base plate. The frame fit snugly against the rubber mold. The "two-rubber mold" setup, shown in Figure 5, had the same base plate and rubber mold described previously, but the caul plate was a sheet of 0.060 in thick stainless steel. A second rubber mold and an upper tooling plate were placed on top of the caul plate, as shown. No prepreg material was placed in the second rubber mold.

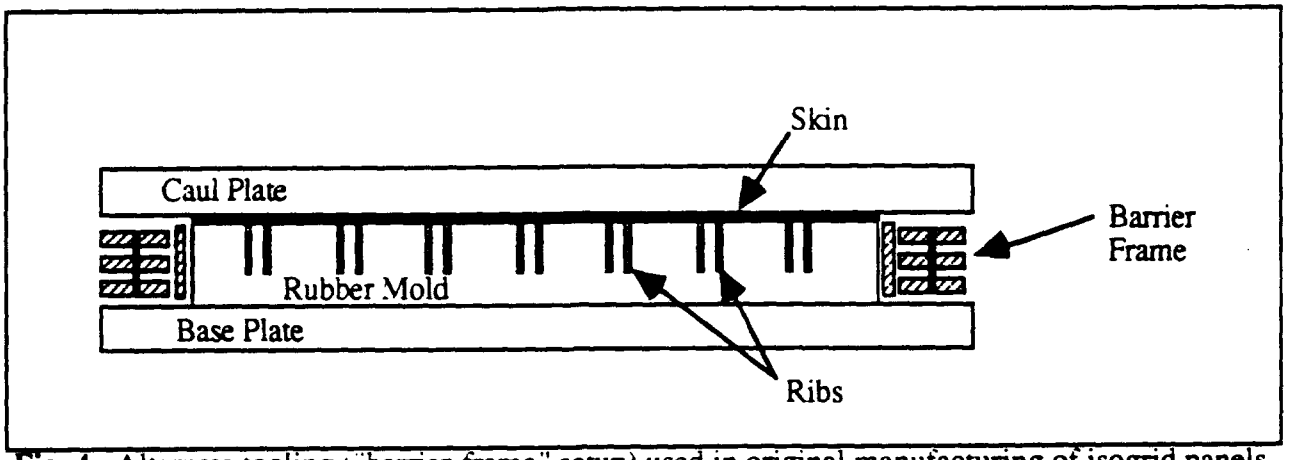


Fig. 4. Alternate tooling ("barrier-frame" setup) used in original manufacturing of isogrid panels. The base plate was 0.5 in thick K-100 aluminum and the caul plate was 0.5 in thick tool steel. The barrier frame completely surrounded the rubber mold and was made of 0.25 in thick by 1.375 in wide steel strips, shown in cross-section above. The strips were pinned together but were not attached to either of the plates.

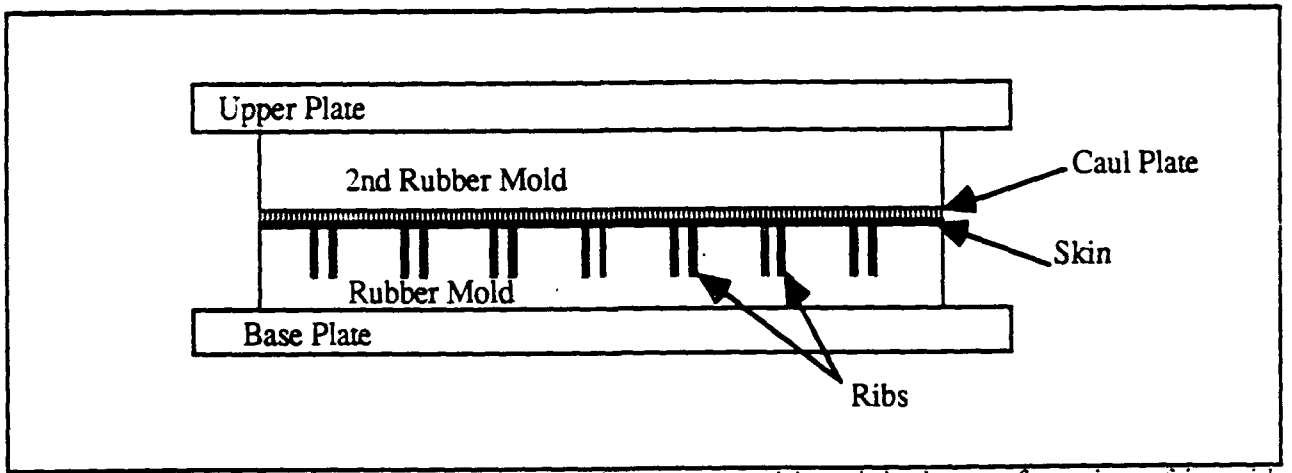


Fig. 5. Alternate tooling ("two-rubber mold" setup) used in original manufacturing of isogrid panels. The upper and lower plates were 0.5 in thick K-100 aluminum. The caul plate was 0.060 in thick stainless steel. The grooves in the second rubber mold are not shown. No prepreg material was wound into or laid up on the second rubber mold.

A summary of the tooling and processing conditions, as well as materials and lay-up sequences, used to fabricate the panels is shown in Table 1. The isogrids were all made from carbon fiber reinforced epoxy (ICI Fiberite 977-2) or carbon fiber reinforced dicyanate ester (DCE) in the form of prepreg tow and unidirectional tape. The fibers were either IM7 or P75. In some cases panels were fabricated without skins ("ribs-only"). In all cases a release sheet was placed between the skin and the caul plate and the tooling completely enclosed in a vacuum bag. All curing was done in an autoclave for 2 hours at 350°F at a pressure of 80 psi.

3. STUDY OF PREVIOUSLY-MADE PANELS

As shown in Figure 6, the maximum distance between the bottom of a straightedge laid on a panel and the panel's upper surface was used as a measure of the amount of warping. When the straightedge was laid parallel to the long side of the panel the maximum distance was called the longitudinal warping, δ_l. When the straightedge was oriented diagonally, it was called the diagonal warping, δ_d. Note that the values of δ_l and δ_d depend on the length of the straightedge. In some cases, the panels had been trimmed down and were shorter than the straightedge. To make all the warping values comparable it was assumed that the warped shape was a circular arc. The raw warping numbers were normalized to an effective straightedge length of 58 in for δ_l and 67.2 in for δ_d. The results are shown in Table 1. Note that every panel, including those made without skins, warped so that the "skin side" was concave, as shown in Figure 6.

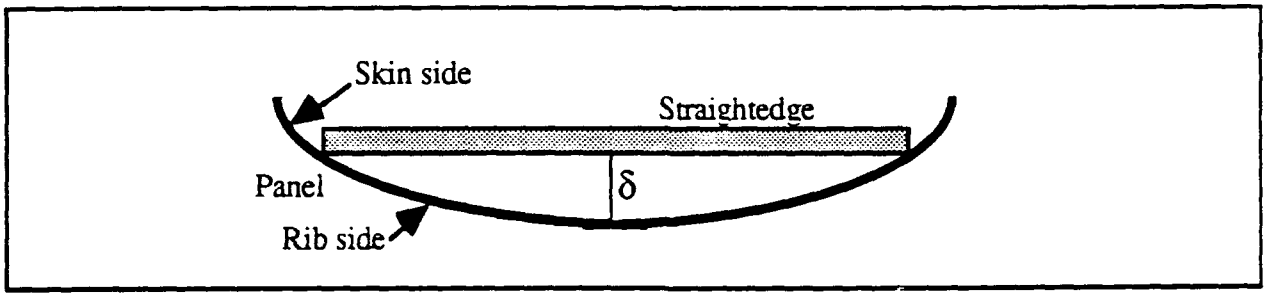


Fig. 6. The amount of warping (δ) in each isogrid panel was measured with a straightedge, vernier caliper (not shown), and/or micrometer (not shown). Measurements were made with the straightedge parallel to the long side of the panel (longitudinal warping) and along the diagonal direction (diagonal warping).

A ribs-only panel (#6) was studied in more detail by mapping out its warping as shown in Figure 7. The panel was placed on a precision flat granite slab and the distances from the slab surface to the tops of the ribs were measured with a hole depth micrometer. The rib height (dimension d in Figure 2) was subtracted off, leaving the warping distance as the result. The data were then input to the contour plot generator of MATLAB.

Table 1. Warping results for isogrid panels manufactured prior to the beginning of this project.

Part No.	Ribs		Skin		Panel Type	Tooling, Bagging, & Cure Cycle						Normalized Warping δ_l (δ_d)	Misc. Comments
	Mat'l	Geom	Mat'l	Geom		Base plate[1]	Rub. mold[2]	Caul plate[3]	2nd rubber mold[2]	Upper plate[1]	Notes		
1	IM7 DCE	39 tows	P75 DCE	6 plies .027 in (60, -60, 0, 0, -60, 60)	Out	#2	#3	.060 in ss	#2	0.5 in K-100	No borders	0.31 in (0.44 in)	
2	IM7 DCE	39 tows	none	none	In	#1	#1	0.5 in steel	none	none	Rubber filler blocks, loose borders	0.41 in (0.45 in)	
3	IM7 977-2	39 tows	IM7 977-2	4 plies .026 in (0,90,90,0)	Out	#2	#3	.060 in ss	#2	0.5 in K-100		0.64 in (0.75 in)	
4	IM7 DCE	39 tows	none	none	Out	#2	#2	.060 in ss	#3	0.5 in K-100		0.02 in? (.01 in)	Bag broke, ribs not consolidated
5	IM7 977-2	42 tows	IM7 977-2	4 plies .025 in (0,90,90,0)	In	#1	#1	0.5 steel	none	none	"Old" cure cycle, loose borders	0.25 in (0.35 in)	Bag broke/lost vac
6	IM7 DCE	39 tows	none	none	Out	#2	#2	0.5 steel	none	none	Barrier frame	0.36 in (0.49 in)	
7	IM7 DCE	39 tows	IM7 DCE	?	Out	#2	#2	.060 in ss	#3	0.5 in K-100	"Old" cure cycle	0.66 in (0.95 in)	
8	IM7 DCE	57 tows	IM7 DCE	3 plies (60, 0, -60)	Out	#2	#3					?	Cut up for testing
9	IM7 DCE	42 tows	IM7 DCE	3 plies (60, 0, -60)	In	#1	#4					?	Shipped to NRL

1. K-100 Aluminum, 0.5 in thick. Plate #1: 56.550 x 32.250 in. Plate #2: 56.500 x 35.625 in.

2. Rubber molds:
   #1: 56.25 x 32.25 x 1.250 in, inboard panel (had a molded-in composite stiffening panel)
   #2: 56.55 x 35.55 x 1.450 in, outboard panel (had a molded-in composite stiffening panel)
   #3: 55.75 x 35.13 x 1.557 in, outboard panel
   #4: 55.83 x 31.88 x 1.125 in, inboard panel

3. Steel plate: 57.125 x 42.25 x 0.5 in

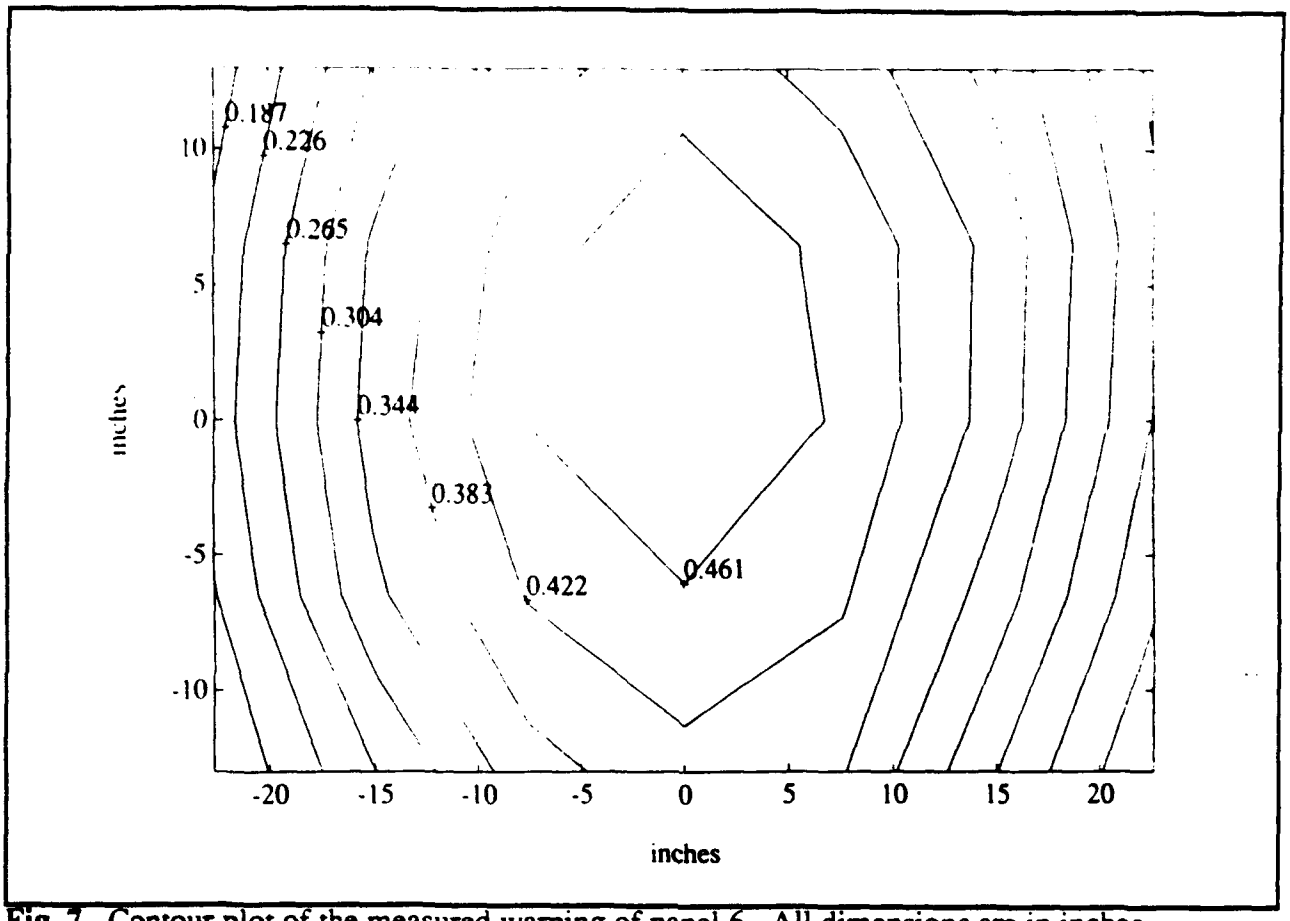


Fig. 7. Contour plot of the measured warping of panel 6. All dimensions are in inches.

It is difficult to draw any firm conclusions from the warping results for the panels listed in Table 1. Too many variables were changed from one attempt to the next to determine exact cause and effect relationships. Some of the critical data that might make such an assessment possible are missing. The panels with the fewest number of manufacturing differences between them are 2 and 4, 3 and 5, and 2 and 6. Comparing 2 and 4 might indicate that warping was reduced by the use of the two-rubber mold setup, or by an accidental break in the vacuum bag. Comparing 3 and 5 might indicate that warping was reduced by not using the "old" cure cycle, by not using the two-rubber mold setup, or by an accidental break in the vacuum bag. Comparing 2 and 6 might indicate that warping was reduced by using the barrier-frame setup. Even though such comparisons do not conclusively identify the cause of the warping problem, the following seem to be the most likely explanations:

Possible Causes of Warping

1. Since the vacuum bag completely surrounds the tooling, it effectively forces the steel, rubber, and aluminum to move together in bending. Since these materials have different thermal expansion coefficients the tooling must therefore warp out-

of-plane at elevated curing temperatures. The part is thus cured in a curved mold and retains its curved shape on returning to room temperature.

2. The thermal expansion of the rubber is much higher than any other material used in the process, and thus may cause warping during cure since it is not restrained on its sides.

3. The skin and rib materials have different thermal expansion coefficients, and thus even if the mold remains perfectly flat during cure, the part would warp on returning to room temperature.

4. Warping is caused by residual stresses in the parts due to non-uniform curing.

If explanation #3 were the only cause of warping, then the ribs-only panels should have come out flat. Since they did not, it means that other causes must also be acting, and probably have a greater effect. Explanation #4 was tested by heating a ribs-only panel to 350°F in an oven. Since no change in its warped shape could be detected, it was concluded that residual stresses were not a major factor. It is postulated that warping of the panels comes from a combination of explanations #1, 2, and 3 above, and that either 1 or 2 are the dominant factors.

4. ANALYTICAL MODELING

4.1 Finite Element Model

Finite element analysis of possible warping of the tooling during cure (explanation #1) was carried out for several different cases. Since a vacuum bag completely surrounded the tooling and part during cure, the imposed 80 psi autoclave pressure may have been high enough to prevent slip between the layers. Thus it was assumed that the tooling, ribs, and skin all deformed as a multilayer laminated plate under the thermal load imposed during cure. The properties and dimensions of each layer in the model were the same as those of the actual parts, except for the rubber mold. The prepreg tows wound into the rib slots change the effective properties of the mold, so the mold was modeled as the two-layer laminate shown in Figure 8. The upper layer was assumed to have the properties of a quasi-isotropic fiber-reinforced composite material with a matrix of rubber and the tows acting as the fibers. This layer is referred to hereafter as the "rubber and rib composite". The bottom layer had the properties of the rubber alone.

The various models studied are shown schematically in Figure 9. Case AAA was a plate with 3 layers: rubber, rubber and rib composite, and steel. Case BBB was a plate with 4 layers: K-100 aluminum, rubber, rubber and rib composite, and steel. Case CCC was a plate with 6

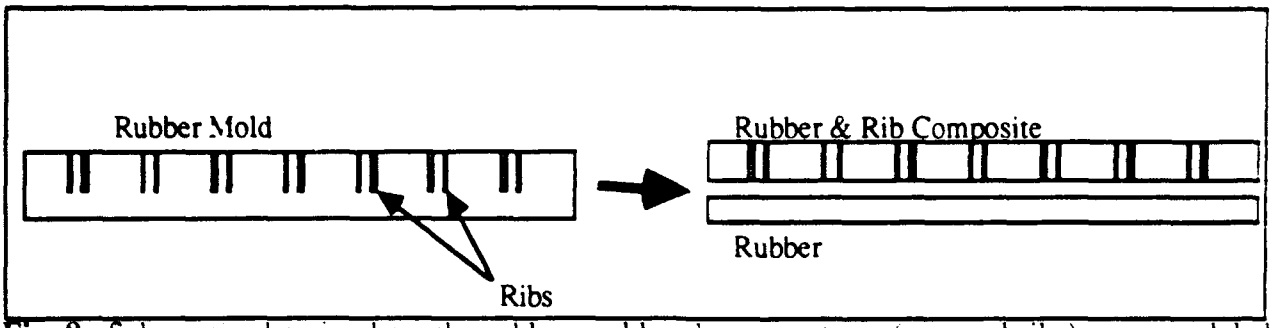


Fig. 8. Schematic showing how the rubber mold and prepreg tows (uncured ribs) were modeled by a two-layer laminate. The upper layer had the properties of a quasi-isotropic fiber-reinforced composite where the matrix was the rubber material and the fibers were the tows. The bottom layer had the properties of the rubber alone.

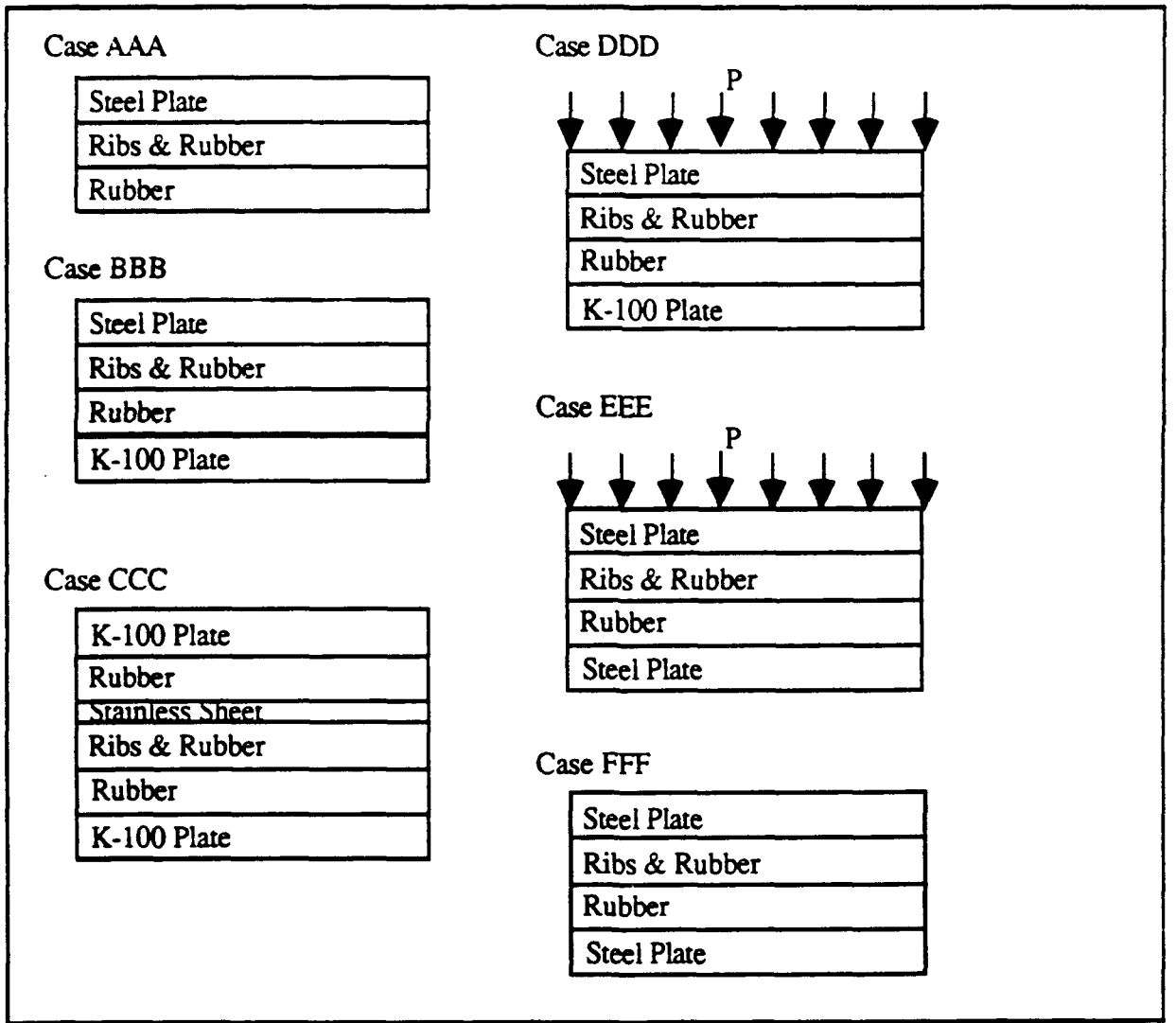


Fig. 9. Schematic diagrams of the different warping models studied with finite-element analysis. In all cases the temperature load was 280°F (representing an ambient temperature of 70°F and a curing temperature of 350°F). The pressure load P was 80 psi, where indicated.

layers: K-100 aluminum, rubber, rubber and rib composite, stainless steel, rubber, and K-100 aluminum. Case DDD was the same as case BBB with the addition of an 80 psi pressure load acting on the upper surface only. Case EEE was the same as case DDD except the bottom layer of K-100 aluminum was replaced by steel. Case FFF was the same as case BBB except the bottom layer of K-100 aluminum was replaced by steel. Note that case BBB models the basic tooling shown in Figure 3. Case CCC models the two-rubber mold tooling shown in Figure 5. Cases DDD and EEE model the consequences of bagging the tooling in such a way that the pressure load acts only on the top surface. Case FFF models a proposed new tooling setup to eliminate the CTE mismatch between the steel and aluminum plates.

The properties used in the analysis are listed in Table 2. For simplicity, all the layers were assumed to be 36 in wide and 58 in long. Because of symmetry, only a quarter section of each laminated plate was modeled. The layers were divided into 20 node quadrilateral elements with 40 elements in each layer. A temperature load of 280°F was imposed in each case (70°F ambient temperature with a cure temperature of 350°F).

The CTE values for the rubber and the composite materials were taken from data measured at Aerospace Corporation, shown in Figure 10. The ribs were all uniaxial lay-ups while the skin was a 6-ply laminate (+60, -60, 0, 0, -60, +60). No 4-ply data were available. Note that the slopes of the curves are not constant across the range of temperatures encountered during curing (70° to 350°F). The data were extrapolated to 350°F and an effective "secant" CTE was used in all the calculations. Data for the silicone rubber are not shown, but its CTE was found to be constant from 70° to 350°F.

Table 2. Properties used in the theoretical analyses.

Property	Material				
	K-100 Aluminum	Rubber	Rubber & Rib	Steel	Stainless Steel
Thickness	0.5 in	1.0 in	0.5 in	0.5 in	0.060 in
Modulus	10 Mpsi	.015 Mpsi	0.48 Mpsi	30 Mpsi	28 Mpsi
Poisson's Ratio	0.3	0.45	0.3	0.285	0.3
CTE	13 ppm/°F	130 ppm/°F	5 ppm/°F	6.5 ppm/°F	9.6 ppm/°F

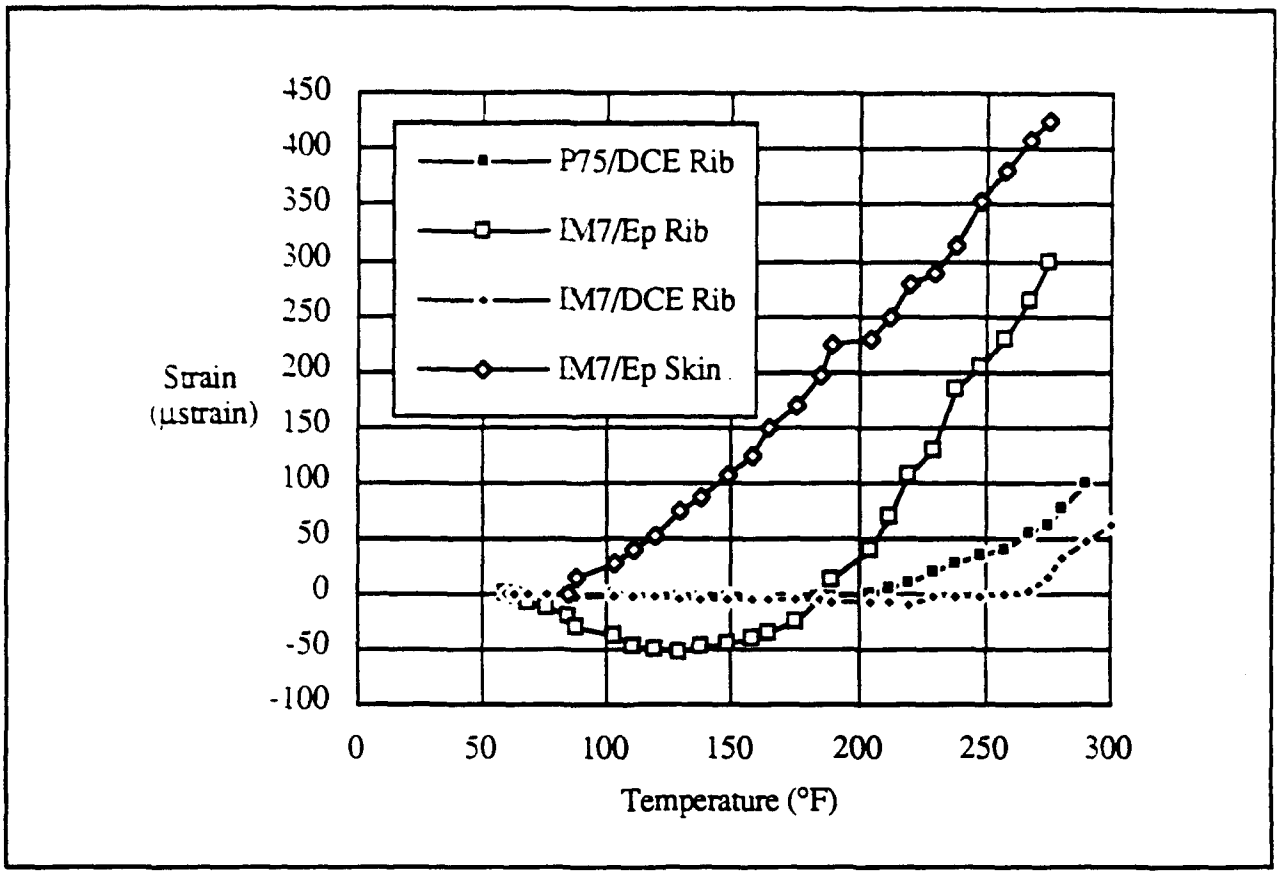


Fig. 10. Results of thermal expansion tests conducted at Aerospace Corporation.

4.2 Laminated Beam Model

The tooling and part were also modeled as a laminated beam subjected to temperature loads. It was hoped that such a simple model would be useful for quick and easy estimating of the warping in the panels. Such a model could also be used to which of all the design and tooling parameters had the greatest effects on warping. Equations for the bending deflections of such beams can be easily derived (see Ref. 5) and so will nor be repeated here. Such equations were programmed into a spreadsheet, which was used for all the beam model calculations. Properties and dimensions used in the model were identical to those reported in the previous section.

4.3 Warping Caused by Tooling Deflections

The results from the theoretical models are summarized in Tables 3 and 4. Measured warping values for the actual panels at room temperature are included for comparison. A deformed geometry plot from the finite element analysis for case AAA and contour plots of the warping for cases BBB and CCC are shown in Figures 11, 12, and 13 as representative examples of the results.

Table 3. Predicted warping of the tooling at 350°F, compared with measured warping of the panels at room temperature.

Case	Predicted Warping of Tooling at 350°F		Measured Warping of Panel at 75°F
	Finite Element Model	Beam Model	
AAA	3.15 in	2.4 in	n. a.
BBB	0.65 in	0.38 in	0.48 in
CCC	0.091 in	0.00 in	0.25 in
DDD	0.073 in	n. a.	n. a.
EEE	0.073 in	n. a.	n. a.
FFF	0.16 in	0.00 in	n. a.

Table 4. Predicted and actual panel warping values (predictions based on warping of tooling during cure).

	Panel (see Table 1)							
	1	**2**	**3**	**4**	**5**	**6**	**?**	**7**
Measured values								
δ_l (in)	0.31	0.41	0.64	0.02	0.25	0.36	0.56	0.66
δ_d (in)	0.44	0.45	0.75	0.01	0.35	0.49	0.63	0.95
FEA model predictions								
δ_l (in)	0.46	0.46	0.05	0.05	0.46			
δ_d (in)	0.65	0.65	0.09	0.09	0.65			
Laminated beam model predictions								
δ_l (in)	0	0.38	0	0	0.38	0.41	0.38	0

The theoretical predictions from the finite-element model agree fairly well with the measured data except for panels 3 and 7. The beam model does well except for panels 1, 3 and 7. Note that these are all cases where the two-rubber mold setup was used. In a beam model it is assumed that plane sections remain plane during the deformation. Since the modulus of the rubber is so much smaller than that of the other materials used, it may be that this assumption is invalid. The beam model was used to study the effects of various parameters on the predicted warping. It was found that the biggest factor influencing warping was the difference in thermal expansion between the aluminum and steel tooling plates (case BBB). The model also indicated that warping

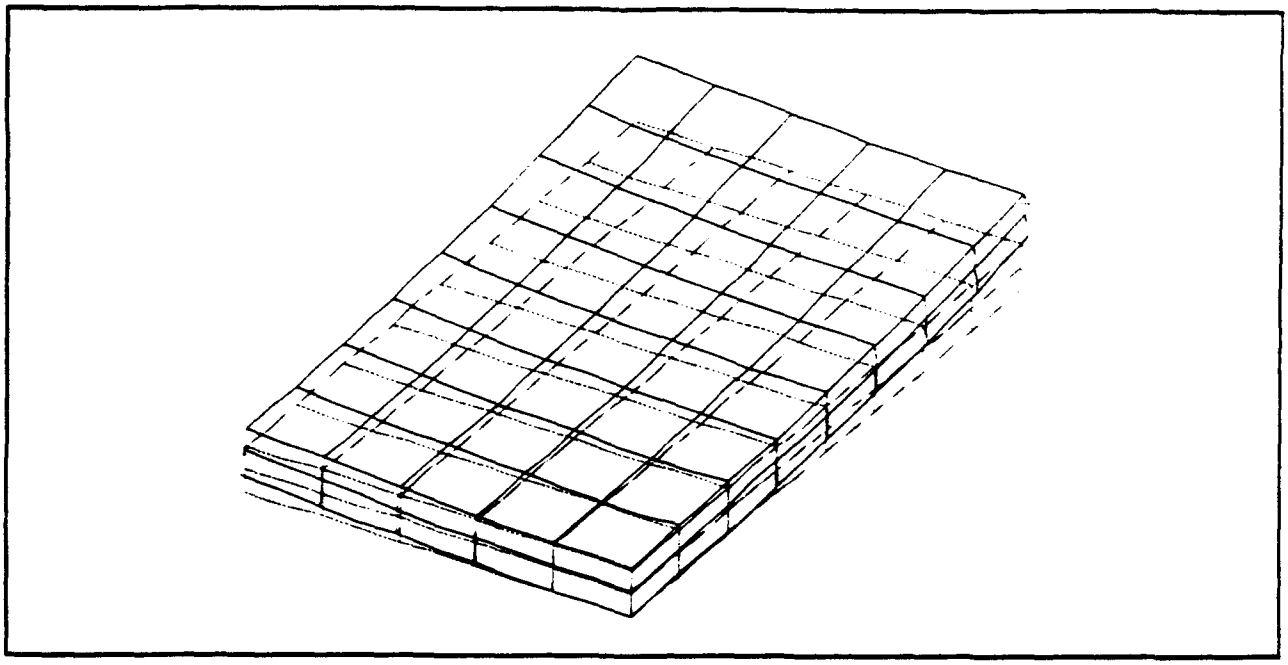


Fig. 11. Deformed geometry plot for case AAA.

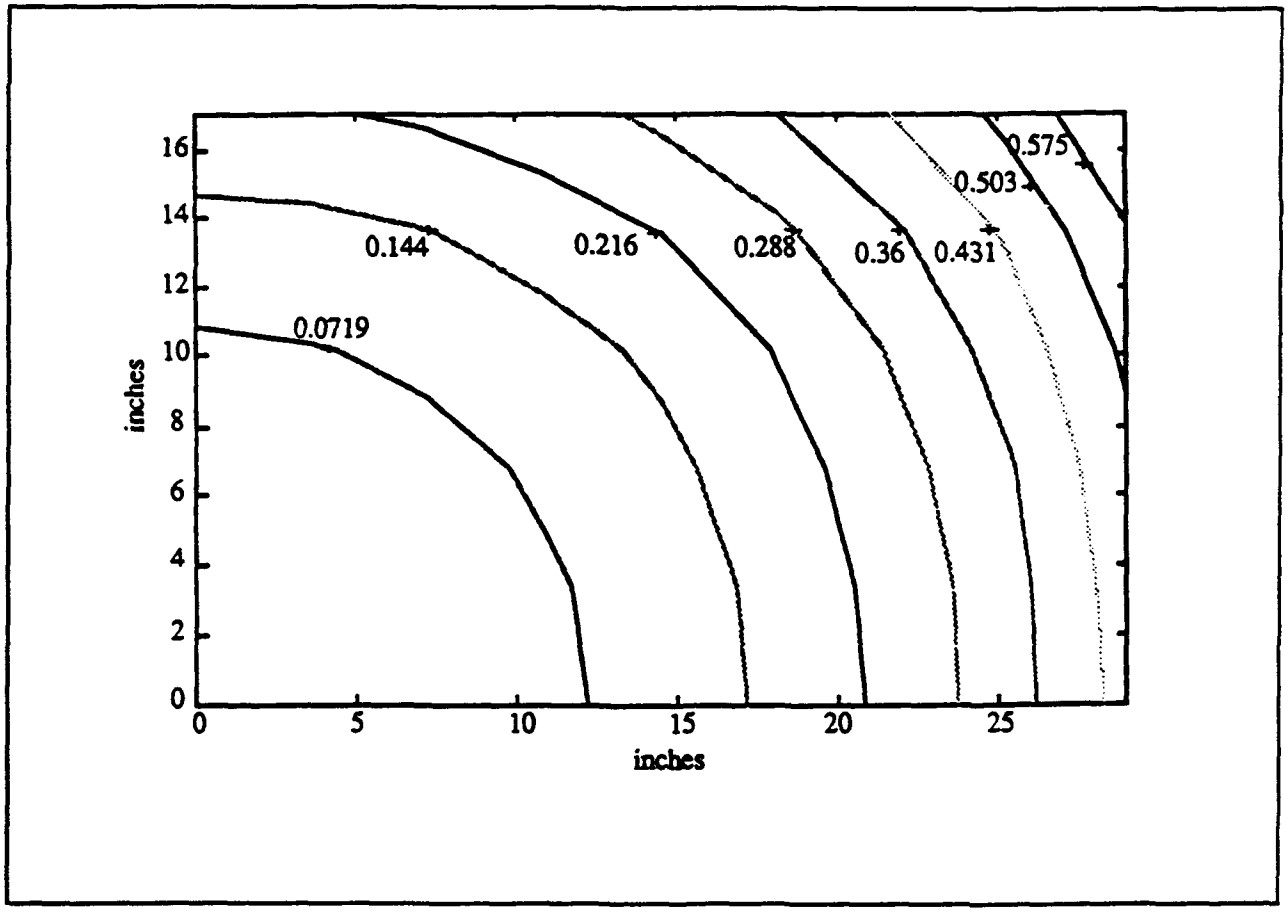


Fig. 12. Contour plot of the NASTRAN output for case BBB. All dimensions are in inches.

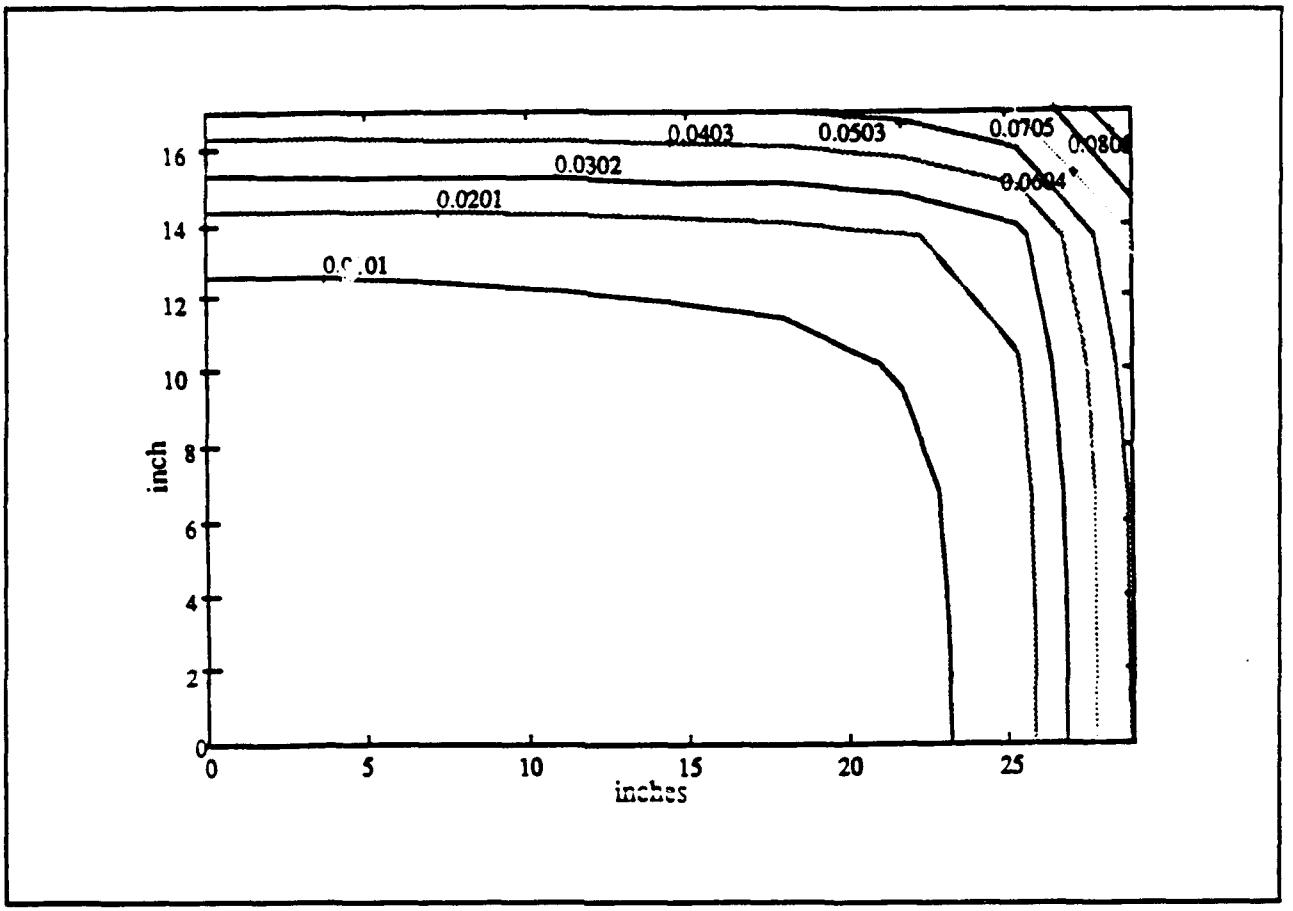


Fig. 13. Contour plot of NASTRAN output for case CCC. All dimensions are in inches.

was much less sensitive to the properties of the rubber and the rubber-rib layer. Both models predict that if steel plates are used both top and bottom (as in case FFF), that the warping caused by bending of the tooling would be eliminated.

4.5 Warping Caused by Thermal Expansion Mismatch Between the Skin and Ribs

Once the problems with the tooling are eliminated, the panel may still warp on cooling down to room temperature because of the differences between the skin and rib materials. The beam model was used to predict the warping for this case, and the results are shown in Figure 14. Note that the amount of warping depends most strongly on the difference between the CTE of the rib and that of the skin. For a panel made entirely of IM7/977-2 prepreg with a 6-ply skin, the predicted warping caused by this effect is 0.13 in.

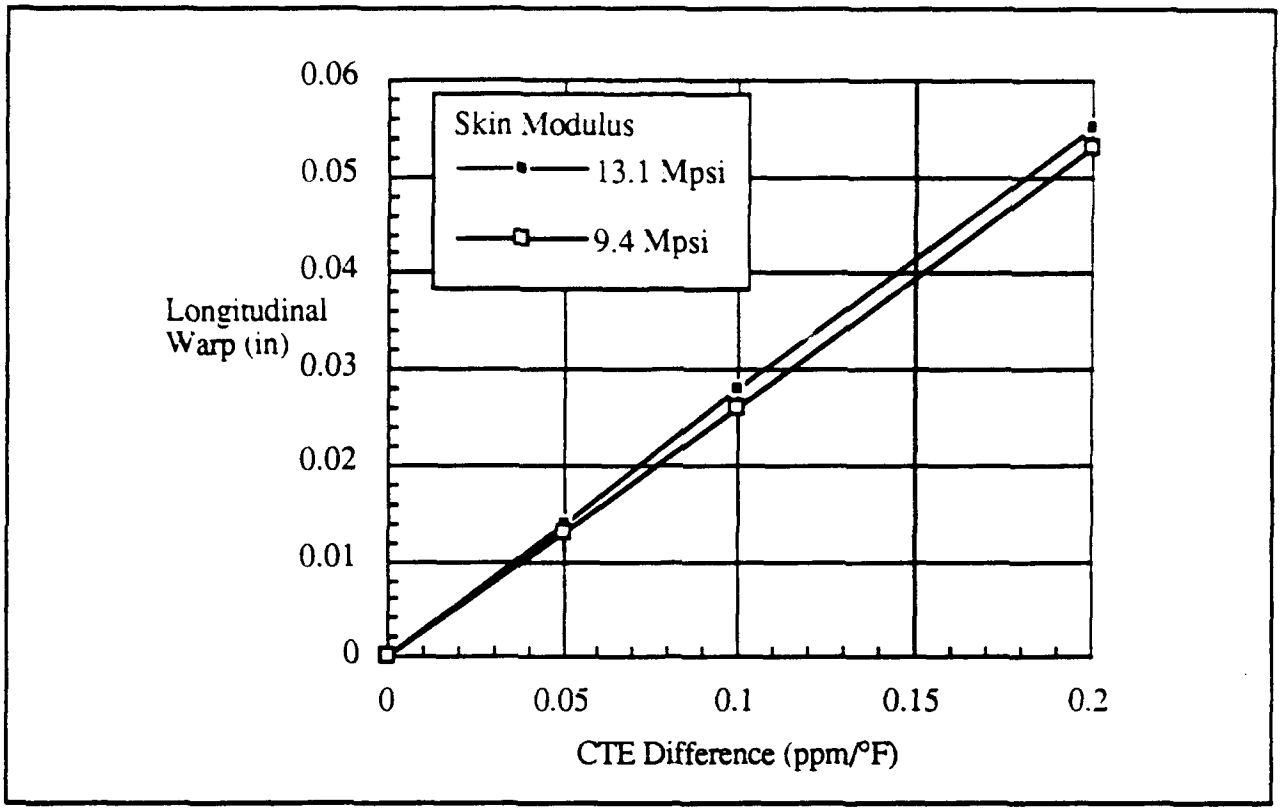


Fig. 14. Predicted longitudinal warp, δ_l, for a large isogrid panel during cool down from the cure temperature (350°F) as a function of the difference between the CTE of the rib and that of the skin. The rib modulus used in the calculations was 25.1 Mpsi.

5. PROPOSED TOOLING CHANGES

Based on the above results, the tooling was redesigned as shown in Figure 15. Both the base plate and caul plate are 0.5 in thick steel, which should eliminate the warping problem caused by the combination of aluminum and steel in the original setup. Steel borders 2.0 inch wide by 0.5 inch thick are rigidly mounted to the base plate with 3/8-in bolts located about 6.75 in apart. The borders fit snugly against the mold and completely surround its periphery. The tops of the borders come up to the bottom of each rib. Their purpose is to confine the mold and prevent it from expanding laterally during the cure process. The barrier-frame setup (Figure 4) had attempted to do this previously, but since the frames were not anchored to the base plate, they were too flexible to resist the expansion of the rubber.

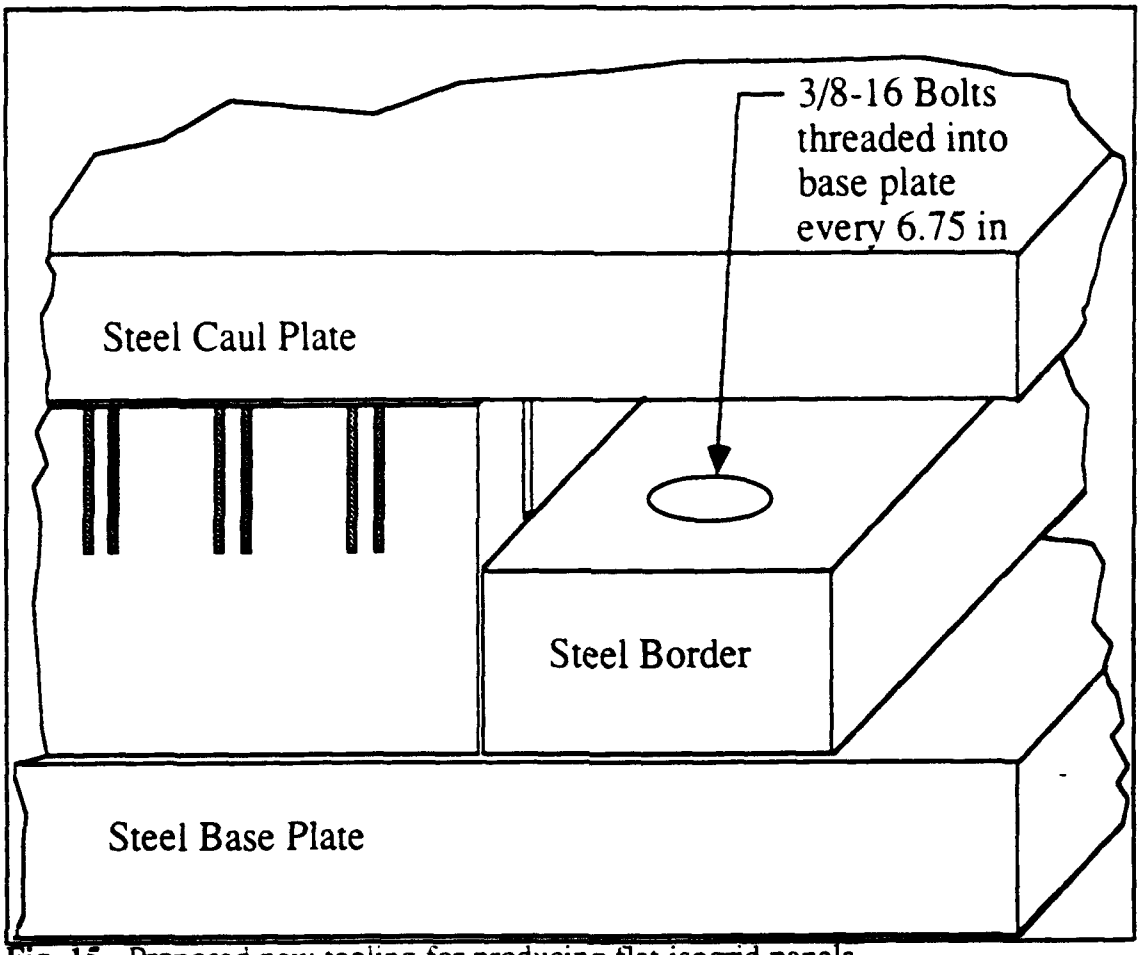


Fig. 15. Proposed new tooling for producing flat isogrid panels.

6. EXPERIMENTAL VERIFICATION

An isogrid panel was fabricated without a skin, using the above new tooling. Leaving off the skin would eliminate any warping caused by the rib-skin mismatch, making any improvements secured by the change in tooling easier to judge. 35 tows of IM7/977-2 material were used for the ribs. The tooling was vacuum bagged and cured at 350°F and 80 psi for two hours. (The entire cure cycle took 20 hours, including heat up and cool down times.) The resulting ribs-only panel had no detectable warping, as predicted.

Next, an isogrid panel with skin was fabricated using the same tooling. IM7/977-2 material was used for both the skin and the ribs. The rits were 37 layers of unidirectional tow, and the skin was a 6 ply lay-up [+60, -60, 0, 0, -60, +60] of prepreg tape. After bagging, the part was cured at 350°F and 80 psi for 150 min. The 80 psi was applied when the temperature of the composite reached 200°F. The part was held for 2-hours at that temperature to insure good resin flow and compaction in the skin and the ribs. The total cure cycle took 16.5 hours.

The resulting panel had a longitudinal warp of about 0.12-0.15 in. This warping was caused by the mismatch in the CTE for the rib and skin. Recall that the predicted value was 0.13 in. Also note that the previous best result for a panel with a 6-ply skin was 0.31 in (panel 1).

7. CONCLUSIONS

Most of the warping in the original isogrid panels was caused by differences in the coefficients of thermal expansion of the aluminum, steel, and rubber. A new tooling design which replaced the aluminum base plate with steel and which rigidly confined the rubber mold between steel borders eliminated the warping caused by these effects. An unwarped ribs-only panel was successfully made with the new tooling. The secondary cause of warping was the difference in the thermal expansions of the ribs and skin. A panel with skin was manufactured with the new tooling and found to have a warp of 0.15 in. A theoretical model developed in the project had predicted a value of 0.13 in. Further improvements in producing flat panels will depend on finding appropriate materials for the rib and skin.

8. REFERENCES

1. A. Reddy, R. Valisetty, L. Rehfield, *Continuous Filament Wound Concepts for Aircraft Fuselage Structures*, J. Aircraft, Vol. 22, No. 3., March 1985.

2. F. Rehfield, *Continuous Filament Advanced Composite Isogrid: A Promising Structural Concept*, AFOSR Contract No. F4962077-C-0077, 1979.

3. J. Harvey, T. Kim, J. Koury, J. Tracy, *Continuous Filament Wound Isogrid Structures for Space Applications*, ASM, Long Beach, 1993.

4. A.C. Ugural, *Stresses in Plates and Shells*, McGraw-Hill, Inc., 1981.

5. R. R. Roark and W. C. Young, *Formulas for Stress and Strain*, 5th Edition, McGraw-Hill, Inc. 1975.

EQUATION OF STATE FOR HYPERVELOCITY COLLISIONS WITH HOMOGENEOUS MATERIALS

by

Jon H. Shively
Civil and Industrial Engineering and Applied Mechanics Department

School of Engineering and Computer Science
California State University at Northridge
18111 Nordhoff Street
Northridge California 91330

Final Report for:
Summer Faculty Research Program
at

Phillips Laboratory
Kirtland Air Force Base

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

and

California State University at Northridge

September 30, 1992

EQUATION OF STATE FOR HYPERVELOCITY COLLISIONS WITH HOMOGENEOUS MATERIALS

Jon H. Shively
Civil and Industrial Engineering and Applied Mechanics Department
California State University Northridge

ABSTRACT

Shock data from collisions of hypervelocity particles with homogeneous materials were evaluated. A master equation for shock that relates the reduced pressure and reduced particle velocity was established:

$P/(\rho_0 C^2) = (U_p/C)(1 + sU_p/C)$

where P = Shock Pressure,

ρ_0 = density,

C = Bulk Sound Velocity,

U_p = Particle Velocity, and

s = the Slope of the U_s Vs U_p line.

A universal slope for the linear relation between the shock velocity U_s and the particle velocity U_p was also established. The slope was estimated to have a value of 1.333 for all of the materials evaluated. Comparisons of the slopes were made for different crystal structures, face-centered-cubic versus body-centered-cubic, and for different crystal orientations of single crystals of NaCl. All slopes were found to conform to previously established universal slope for shock. Comparisons to another liquid also exhibited conformance to the master equation above and to the universal slope. The linearity and universality of the slope is related to the shape of the potential energy equation between chemical species in a homogeneous material. The universal value was calculated using the Lennard-Jones interactive potential between chemical species in a homogeneous material. The theoretical value calculated from the Lennard-Jones equation was between 1.4 to 1.6 which compares favorably with the value of 1.333 obtained from the data analysis.

EQUATION OF STATE FOR HYPERVELOCITY COLLISIONS WITH HOMOGENEOUS MATERIALS

Jon H. Shively

1.0 INTRODUCTION

2.1 Hypervelocity Collisions of Space Debris and Shock

A collision by a piece of space debris or a micro meteorite traveling at hypervelocities (1 to 15 km/s) with a spacecraft will produce a shock waves in the spacecraft material. Shock wave phenomena have been the subject of numerous investigations and theoretical studies. Most notable has been the extensive studies of shock phenomena in solids and liquids by Los Alamos National Laboratory that has been a leader in this field for more than 50 years.(1,2) Hypervelocity shock data have been compiled by Los Alamos, (3) by the Air Force Weapons Laboratory (now the Phillips Laboratory)(4), and by Sandia National Laboratory (5).

Attempts have been made by researchers to relate the observations from shock experiments to theory (1-2). Several attempts have been made to establish a universal equation of state describing the thermodynamics of shock (6-11). The equations developed, to date, contain at least two empirically determined material parameters, the Gruneisen constant and the slope of the shock velocity versus particle velocity line. Thus, the equation of state is not universal but contains parameters which depend on the material.

The quest for a universal equation of state while elusive has value in predicting shock behavior. The open literature data on shock of condensed materials are limited in applicability to the space debris problem because the upper bounds on the data lie in the range of 3 to 5 km/s. The space debris particles can potentially be traveling at velocities in the range of 10 -15 km/s. In addition, much of the data derived from experiments at Los Alamos and Sandia Laboratories are limited to materials that are not of interest for spacecraft designers. As a result, important data on the shock behavior of spacecraft materials are needed to order to predict the behavior of materials in low-earth orbit. Unfortunately, that shock data that is applicable is for the low-end of the hypervelocity regime (1 to 5 km/s). However, if a universal equation which describes shock behaviors of all materials could be established, then the spacecraft designers could make better estimates of spacecraft reliability. Not only could such a universal equation be used to predict shock behavior at the high-end of the hypervelocity regime (5 to 15 km/s) but the equation could be used to predict performance for untested materials or materials for which the data are sparse. Moreover, the need for extrapolation of existing data could be validated with the results of a few critical experiments rather than a whole series of expensive shock experiments at hypervelocities. In response to this need, a study was undertaken this Summer with the hope that a single universal equation of state for shock in the regime of 1 to 15 km/s could be established

and that this equation would describe the phenomena of shock wave propagation in homogeneous solids.

This report presents work that is aimed at finding a universal equation for shock for which the slope of the shock velocity, U_S, versus the particle velocity, U_p, is a constant and is universal for all materials. The main objective of the effort this Summer is to establish a basis for the constancy and universality of the slope of the U_S Vs U_p line. Also, the current effort is follow-on to that performed last Summer as a Faculty Researcher Program and from the Research Extension Grant awarded for FY 1992. During last FY 1992, the shock data was found to fit a universal equation that contains the slope of the U_S Vs U_p curve for metals, polymers, ceramics, and water. Moreover, during FY 1992, the slope was established to be a constant and universal for all of the materials studied. This report then summarizes the previous results and combines them with the results of the effort this Summer 1993 at the Phillips Laboratory as part of the AFOSR Faculty Researcher Program.

1.2 Review of Shock Physics

A collision of a hypervelocity projectile with a target produces shock waves in the target material. The relation between the pressure induced by the shock, the volume and the energy is called the Hugoniot (12). The most common expression for the Hugoniot is the so-called Mie-Gruneisen compression equation of state:

$$P - P_H(V_H) = [(\Gamma(V)/V)][(E - E_H(V_H))] \quad (1)$$

where P_H, V_H, and E_H are pressure, volume, and energy, respectively, along the Hugoniot and $\Gamma(V)$ is the Gruneisen constant for the material (13).

Assuming that the shock wave propagates as a plane wave, the Hugoniot can be expressed in measurable terms of mass, momentum, and energy as in the following conservation equations:

$$\rho U_S = \rho_0(U_S - U_p) \qquad \text{Conservation of Mass} \quad (2)$$

$$P - P_0 = \rho_0 U_S U_p \qquad \text{Conservation of Momentum} \quad (3)$$

$$PU_p = 1/2\,\rho_0 U_S U_p 2 + \rho_0 U_S(E - E_0) \qquad \text{Conservation of Energy} \quad (4)$$

where

ρ = density

U_S = shock velocity

U_p = particle velocity

P = pressure

E = energy

and subscript. o. refers to the initial state (undisturbed).

Substituting Equations 1, 2, and 3 into Equation 4 yields

$$E_H - E_O = 1/2\,(P_H - P_0)(V_0 - V_H) \tag{5}$$

The experimental determination of the Hugoniot requires measuring any two of the following three quantities: U_S, U_p, or P (see Equations. 2 and 3). There are several other important relationships that define various aspects of the shock phenomena. For example, most materials exhibit a linear relationship between the shock velocity and the particle velocity, although the reason for this behavior is not understood.

$$U_S = C + sU_p \tag{6}$$

where C ≈ bulk sound velocity = shock velocity at zero pressure. The bulk sound velocity can be determined from a U_S - U_p plot where C is the intercept. The bulk sound velocity in solids is given in Reference 4

$$C^2 = C_L^2 - (4/3)\,C_S^2 \tag{7}$$

where

C = bulk sound velocity

C_L = longitudinal speed of sound

C_S = shear speed of sound

Because the shear velocity in liquids is zero, C equals C_L. However, for solids the shear velocity is significant and Equation 7 applies. Extrapolations to U_S for U_p = 0 correspond closely to the bulk sound velocity. If the shock velocity is linearly related to the particle velocity, then the following equation is obtained which is convenient for representing shock data:

$$P/(\rho_0 C^2) = (U_p/C)(1 + sU_p/C) \quad (8)$$

In the past, many workers in shock testing have associated the slope, s, with the Gruneisen constant, Γ, which is used in Equation 1. The works of Slater (13) and Dugdale and MacDonald (14) lead to equations for Γ in terms of the slope, s. The prediction in Reference 13 is s = (Γ/2) + 1/2 and the prediction in Reference 14 is s = (Γ/2) + 1/3.

2.0 REVIEW OF EQUATIONS OF STATE

2.1 Literature Review

The phenomena of shock are characterized by an equation for the Hugoniot given by Equation (1). The shock Hugoniot is a trace on an adiabatic surface of the equation of state of the materials under adiabatic conditions. Therefore, the isothermal equation of state must be modified for adiabatic conditions, such as Equation 5, which uses the Gruneisen parameter, Γ. Jeanloz recently reviewed some selected equations of state as applied to shock in materials (9). The equations reviewed were developed by Rose, et.al,(15.16), Grover, et al, (17), and Jeanloz et.al (9). They seem to fit the shock data better than equations developed by Slater (13) and Birch (7). The equations developed by Prieto el.al.(6), while not included in Jeanloz comparisons, have been shown by Preito (18) to be equivalent to both the Rose and the Jeanloz equations of state. All the above attempts to develop equations of state are dependent on measured bulk quantities. They are intended to apply to both isothermal and adiabatic processes.

Shock caused by hypervelocity collisions is governed by conventional thermodynamic variables and is essentially an adiabatic process. If the phenomena are similar in different materials, one might expect a single equation of state to describe the processes. However, Equation 1 suggests that the Hugoniot is controlled by the value of the Gruneisen constant, Γ, which varies from material to material.

2.2 Principle of Corresponding States and Shock

Preito et.al., (19) developed their equations based upon The Principle of Corresponding States which broadly states that fluids behave alike when in corresponding states; i.e., when n thermodynamic variables bear a constant ratio to n critical values of those same thermodynamic variables. Consider as an example, the Van der Waal's equation of state for a gas. The p, v, T relation for a given gas can be represented in terms of reduced variables p_r, v_r, and T_r, which are ratios of p, v, and T to p_c, v_c, and T_c, respectively. The subscript c refers to a critical value of the variable for a particular gas. However, any gas can be represented by the reduced form of the Van der Waal's equation of state as long as p_c, v_c, and T_c are known for that gas.

Therefore, the Van der Waal equation is considered to be universal providing the critical values of p, v, and T are measured.

All gases then are said to behave identically when in their corresponding states determined by the critical values, p_c, etc. While corresponding states may not seem to be precise, this principal has had wide use in the fields of equilibrium and transformation properties of pure substances (20). The principle of corresponding states has been used to describe incompressible flow behavior where for dynamic similarity, the Reynolds number is frequently used.

Prieto et.al., (19) has developed an equation that relates the shock thermodynamic variables (P, V, and E) to experimental parameters, i.e., particle velocity, shock velocity, compressibility, and density. In so doing they assumed a linear relation between shock velocity and particle velocity. They defined the compression, Z, by

$$Z = 1 - \rho/\rho_0 = \Delta V/ V \tag{9}$$

where ρ is density and V is volume.
From this it also follows from Equations 2 - 4, that

$$P = \rho_0 C^2 Z(1 - sZ)^{-2} \tag{10}$$

Using the principle of corresponding states

$$P_1/[(\rho_1 C_1 2)/s_1] = P_2/[(\rho_2 C_2 2)/s_2] = p_r \tag{11}$$

where 1 and 2 refer to two different materials characterized by s_i,
and rewriting Equation 10 we get

$$p_r = v_r (1 - v_r)^{-2} \tag{12}$$

Thus, Prieto succeeded in expressing pressure and volume in terms of reduced variables. A proof of the validity of Equation 12 is the prediction of the reduced pressure using experimental values of s. Prieto, et.al., made a few comparisons for a few metals which have nearly the same values of s_i .They obtained a very close fit to that data using Equation 12 for shock (6). As is discussed below, s_i, is not unique to each material but now appears to be a universal constant for all homogeneous materials.

3.0 EVALUATION OF SHOCK DATA

3.1 Summary of FY 1992 Work on Data Correlations

There is a significant amount of shock data available for evaluating theories about the equation of state in the range of 0.1 to 5 km/s (3,4) . Recently, it has become possible to produce particle velocities exceeding 5 km/s. Some very expensive experiments using an electric gun have achieved particle velocities exceeding 15 km/s. In addition, experimental techniques have improved with the use of laser interferometers and streak cameras to measure velocities. Instrumentation of pressure transducers has also improved. However, most of the data predates these experimental improvements.

Shock data are available for a wide variety of materials and in the range of 0.01 to 5 km/s. The data for the most part have been generated by flyer plate tests in which the plate velocity is measured using streak cameras and the shock velocity is determined through momentum transfer to the target. In a few cases, the pressure is measured during shock by means of transducers. Other variables are usually derived from the physics of shock (Equations 2 to 4). Selected data from the above reference data banks were electronically scanned into a computer and entered into a computer-based spread sheet program. The result of this work was published this Spring in a Phillips Laboratory Report (21).

The Los Alamos and Phillips Laboratory data for shock were manipulated in a special way; viz. the pressure was normalized by dividing by the bulk modulus of sound, $B_S = \rho_0 C^2$, and the particle velocity was normalized by dividing by the bulk sound velocity, C. The results are presented in Figure 1 for different materials. The results show that the selections for this normalization of pressure and velocity, while not obvious selections, correct the data to form a master curve. If either normalization step is omitted, the data do not follow a master curve. In retrospect, it is easy to see the validity of the normalization when one considers Prieto's use of Principle of Corresponding States as a means of describing the shock process.

There are two ranges of reduced particle velocities of interest U_p/C, <0.1 and U_p/C,>0.1, From Equation 8 one can show that at low ratios of U_p/C, <0.1 with s in the order of magnitude of 1, the second factor on the right hand side of the Equation (8) is of the order of one. This means that the normalized pressure approximately equals the normalized particle velocity. Therefore the relationship $P/(\rho_0 C^2)$ to (U_p/C) should be linear with a slope of 1. This appears to be the case as can be seen in Figure 1 for $U_p/C < 0.1$. However, as U_p/C approaches 0.1, the curve begins to bend upward as the second factor in Equation (8) rapidly increases above one. In essence, Equation (8) is a quadratic in U_p/C in which the squared term dominates at values of $U_p/C > 0.1$. Thus at ratios of $U_p/C > 0.1$, the normalized pressure increases non-linearly with normalized particle velocity as can be seen in Figure 1.

What makes Figure 1 more remarkable is that eight different materials fit on one curve including one liquid. Examination of Equation (8) shows that s, the slope of the U_S Vs U_p line, is a parameter of the material; i.e., the slopes vary from one material to another. However the correlation shown in Figure 1 suggests that materials have the same slope. In fact, a careful review of the literature on shock and the data indicate that homogeneous materials generally have slope values which range from 1 to 4 (9).

The correlation shown in Figure 1 was tested further. Additional materials were checked against Figure 1. Ceramic materials such as quartz, corundum, and LiF were compared and each of these ceramics fits the curve. Three other metals were evaluated by normalizing the pressure and particle velocity and were found to superimpose on Figure 1. Polymers, such as epoxy, polysulfone, Mylar, Lucite, and Plexiglas were compared to the master curve and fell on top of the curve. Shock data obtained on iron powders ranging in densities from 4.7 g/cm^3 to the theoretical density 7.8 g/cm^3 were compared to the master curve in Figure 1. These comparisons included correction for both the density and sound velocity changes in iron as a result of the initial powder compaction. The iron powders data superimposed on the master plot. In short, all homogeneous materials evaluated last Summer exhibit the same normalized pressure, $P/(\rho_0 C^2)$, response to a collision by a particle traveling at a normalized velocity, U_p/C.

Comparisons of the same materials were made using the shock velocity and the particle velocity. The results are shown in Figure 2. As can be seen, the slopes of the materials within the scatter in the data appear to follow a single slope, but each has a different intercept. Again within experimental error the intercepts are equal to the bulk sound velocity for each material. One issue related to the slope is the problem of phase changes during the shock wave propagation. Shock induced phase transformation complicate the use of Equation (8) that uses a single sound velocity and a constant initial density. Phase changes will definitely affect both the density and the sound velocity of a material. Therefore the correlation is not completely valid for shock induced phase transformations. Quartz is an example where the shock induced transformations are predicted to occur in the solid state. These phenomena could be reflected in the quartz data that has a slight anomaly in the slope.

3.2 Additional Data Correlations for Summer 1993

The effort this Summer has focused on reasons for the linearity and universality of the slope of the U_S Vs U_p graph. Several additional attempts were made to find exceptions to the master curve shown in Figure 1. This included evaluating crystal structure, crystal orientation, density effects on the slope, and the slope for other liquids.

The role of crystal orientation on the slope of the U_S Vs U_p graph was evaluated in some detail. Data were available for Na Cl with three different crystal orientations (3) ; [100],[110], and [111]. The single crystals of three different orientations were subjected to hypervelocity

collisions in one direction. The shock velocity and particle velocity were measured. The results are shown in Figure 3. As can be seen in the figure the shock velocity in each of the three directions in a crystal superimpose on each other. Surprisingly shock waves travel in salt in identically in different directions with no evidence of anisotropy. Sound waves would not be isotropic. Moreover, the shock behavior is consistent with the master curve, Figure 1, and the. master equation, Equation 8. This strongly suggests that the slope of the U_S Vs U_p graph is independent of direction the shock wave travels through the sodium chloride crystal.

The majority of the data available for shock in metals has been evaluated by Jeanloz (9) and Roulf (11). They concluded that the slopes of the U_S Vs U_p graphs range from 1.2 to 1.5 and are constant. One minor observation missed by them is that FCC metals have slopes that are slightly larger than the slopes for BCC metals. The FCC metals have slopes that are reported to be from 1.3 to 1.6 and BCC metals have slopes reported to be from 1.1 to 1.4. As can be seen in Figure 4 the slope of tungsten falls slightly below that of gold. Atom coordination could explain these very slight differences. Gold with 12 nearest neighbors might respond slightly differently to a compression wave than tungsten with only 8 nearest neighbors. However, the effect must be very small when you consider that the master plot covers a host of different crystal structures for ceramic materials with much lower coordination numbers than metals which tend to be close-packed. It is difficulty to tell whether the difference between FCC and BCC is real without a detailed evaluation of the experimental errors.

One of the major points found during previous work is that the slope of the U_S Vs U_p curve for water is very similar to all solids. This is shown in Figure 2. The analysis of data this Summer was extended to other liquids. So far it appears that <u>all</u> liquids behave the same as solids under shock. As an illustration of these results, water and benzene are compared in Figure 5. Both liquids have the same slope and therefore behave the same with respect to shock wave propagation.

Shock in low density materials such as foams and woods was also examined in light of the master equation. It was found that the behavior of woods is an apparent exception to the current correlations as can be seen in Figure 6. Iron conforms to the master plot, Figure 1, but the woods do not conform. While the wood data do not superimpose, they do converge at the master plot at high values of the reduced pressure. The failure of a wood to overlay on the master plot is due lack of a correct value for the bulk sound velocity. For the reduced pressure and reduced particle velocity a value of the bulk sound velocity is needed which corresponds to a theoretically 100 % dense material, no voids. Whereas the sound velocity correction was possible for iron powders because the bulk sound velocity for 100% dense iron is known, a similar correction is not possible for the woods.

4.3 Potential Functions and the Slope of the U_S Vs U_p Curve

Shock behavior must somehow be related to the nature of the mechanical interaction when two species are squeezed by the compression wave as it travels through the material. If materials respond in such a way that the slope of the U_S Vs U_p line are the same, then this mechanical interaction between the electron clouds must be independent of the material. King (25) has established that the slopes of the U_S Vs U_p curve, to a third order approximation, is independent of the material and depends only on the shape of the potential energy function with respect to radial separation of the chemical species. Thus, all materials would have an identical slope and that slope would be essentially constant over the range of compressions. If the interactive potential energy in the compressive region can be expressed as the following function,

$$V(r) = A\Phi(r/r_0) \tag{13}$$

where $V(r)$ = Interactive potential energy in the compressive region $r/r_0 < 1$,

A = Materials constant

r = Radial separation of chemical species

r_n = Radial equilibrium separation

then it follows that

$$-s = d^3\Phi / 4d\Phi^2 \tag{14}$$

where s = slope of the U_S Vs U_p curve

$\Phi = \Phi(r/r_0)$

The Lennard-Jones equation (22) is a popular potential energy function for interactions between two components (atoms, ions, or molecules).

$$V(r) = A/r^m + B/r^n \tag{15}$$

Equation 15 is plotted in Figure 7 for two values of m. The value of m dominates the compressive region of the potential curve and reflects the interaction between the electron clouds surrounding atoms, ions, or molecules. This equation was extended by Kittel (24) and Lennard-Jones (23) to a crystal with coordination number of 12 where A and B have specific

values. If we redefine equation (15) by dividing r by r_0, we get

$$V(r/r_0) = C[12.13(r/r_0)^{-m} - 14.45(r/r_0)^{-n}] \quad (16)$$

which is of the desired form $V(r) = A\Phi(r)$. The usual values taken for m, n are 12 and 6 respectively. Although there is no theoretical basis for the value of $m = 12$, there is general agreement that n should equal 6 for most ionic solids.

According to King (25) and Pastine (8) the slope of the U_S Vs U_p line is related to the ratios of the third derivative of the potential to the second derivative of the potential as given by Equation (18). Since the functional of r/r_0 of Equation (15) is the same for Equation (16), the ratios of the derivatives can be calculated and the values of the constants A and B in Equation (15) cancel. As a result, the slope is predicted to be independent of the material or its crystal structure, except that r_0 varies from one material to another. Even if Equation (16) were modified for other coordination numbers, the constants would cancel. Thus, the slope is predicted to be totally independent of the material but depends only on the normalized, (r/r_0), functional shape (m,n) of the potential energy curve. The constants m and n in Equation (19) determine the magnitude of the slope of the U_S Vs U_p line. Moreover, the slope would be universal for all materials if m and n are the same. In the analysis by King, the application of Equation (18) to the prediction of the slope of the U_S Vs U_p line, higher order terms were neglected, i.e., the expression for the potential is assumed to be a power series. Therefore, the use of a simple expression such as the Lennard-Jones Equation may not be a totally accurate representation of the potential energy.

A comparison of the predicted values of the slope using Equations (14) and the expression for the potential energy given in Equation (15) is shown in Figure 8. Two values of m were used, 10 and 12 respectively along with $n = 6$. As can be seen in the figure, the predicted slope is nearly constant in the range of shock compressions of $V/V_0 = (r/r_0)^3 < 0.7$. Also, for the materials evaluated the universal value of the slope is 1.333. The agreement with the calculated slope (1.4-1.6) is remarkable considering the uncertainly in the applicability of the Lennard-Jones equation to the compressive region of the potential curve.

The Lennard-Jones equation has been used extensively to predict the behavior of gases and condensed vapors. It is used to predict hydrodynamic applications , transport phenomena, and critical phenomena for gases and liquids.(20). It has been used successfully in predicting shock behavior and detonations in gases. (20). Moreover, the Lennard-Jones equation using reduced variables (Principle of Corresponding States) has led to good agreement with the observation of liquids and their vapors at or near the critical point. Thus the Equation while

empirical. is nevertheless useful for predicting equilibrium and non equilibrium phenomena using the Principle of Corresponding States (20). The predicted universality and almost linearity of the slope in Equation 8 implies that liquids and solids would behave the same under shock waves. These predictions using the Lennnard-Jones equation for the potential energy are consistent with the evaluation of shock data for which the slope is apparently universal and constant for homogeneous materials. Moreover because the Lennard-Jones equation has worked very well in making predictions for liquids and their vapors and for ionic solids, it appears that it may provide an accurate description of the potential energy in the compressive region for all dense solids.

4.0 OTHER TASKS DURING THE SUMMER OF 1993

Considerable effort was devoted to the review of the open literature on the equation of state for shock . A bibliography containing 11 books and 153 references from the open literature was made. Most of the above journal articles and reports were reviewed carefully for application to the question of the linearity and universality of the slope of the U_S Vs U_p line.

In addition to the review process, two reports were issued this summer by Phillips Laboratory (21, 26) which summarize the work performed during the last two summers on the AFOSR Faculty Researchers Program and on the AFOSR Summer Extension Grant for 1992. (20, 26). Finally, presentations of the equation of state analysis were made to the Phillips Laboratory and to Los Alamos National Laboratory. In addition, The author attended the American Institute for Physics Conference on High Pressure Science and Technology on June 28, 1993 in Colorado Springs where the above work was informally discussed with other workers in the shock physics field.

5.0 SUMMARY AND CONCLUSIONS

The work performed this summer as part of the AFOSR Program of Faculty Researchers has led to additional understanding of the shock behavior caused by hypervelocity collisions with homogeneous solids . The results this summer produced further confirmation of the correlations of normalized shock pressure with normalized particle velocity and its concomitant master equation. Moreover, the universality was further confirmed for other solids, single crystals, and other liquids. A method was developed for predicting the slope of the U_S Vs U_p line from an equation for the potential energy between chemical species in a material. This represents a major step towards greater understanding of shock phenomena in homogeneous spacecraft materials. This new theoretical step supports the data evaluations that were discovered and reported previously as part of this Faculty Researchers Program, viz. a single universal equation of state for shock, Equation 8. Thus the major objectives were fulfilled (1) to establish a universal equation of state for shock, Equation 8, and (2) to rationalize a constant universal value of the slope of the U_S Vs U_p graph of shock data.

6.0 ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. Charles Stein for his critical comments and encouragement. Also, Tsgt Robert Robertson is thanked for his analytical support in the analysis of the shock data. Finally, the support from the AFOSR is greatly appreciated. The author found this experience intellectually stimulating and worthwhile. It is hoped that the results presented herein will enable the science of hypervelocity collisions to move forward.

7.0 REFERENCES

1. Walsh, John M. and Christian, Russell H., "Equation of State of Metals from Shock Wave Measurements," The Physical Review, 97, No. 6, pp. 1544-1556, March 15, 1955.

2. McQueen, R.C. and Marsh, S.P., "Equation of State for Nineteen Metallic Elements from Shock-Wave Measurements to Two Megabars," Journal of Applied Physics, 31, No. 7, pp. 1253-1269, July 1960.

3. Marsh, S.P., Los Alamos Shock Hugoniot Data, University of California Press, Berkeley, 1980.

4. Kohn, Brian J., Compilation of Hugoniot Equations of State, AFWL-TR-69-38.
Air Force Weapons Laboratory, Kirtland AFB, NM, April 1969.

5. Hermann, Walter, Sandia Laboratories Shock Data Base, Sandia Laboratories, Albuquerque NM, 1992.

6. Prieto, Fernando and Renero, Claude, The Equation of State of Solids, J. Phys.Chem. Solids, 1976, 37, pp. 151-160.

7. Birch, F., Journal of Geophysical Research, 83, 1978 p 1257.

8. Pastine D. Interpolation Formula for the Relationship Between Shock Velocity, U_S, and Particle Velocity, U_p, in Solids, Communications, p 440, 1968.

9. Jeanloz, Raymond, "Shock Wave Equation of State and Finite Strain Theory," Journal of Geophysical Research, 94, No. B5, pp. 5873-5886, May 10, 1989.

10.Rodean, H. Relationship for Condensed Materials Among Heat of Sublimation, shock-wave Velocity and Particle Velocity, J Chem. Phys., 49, No. 9, p 4117, 1968.

11.Rouff, A., Linear Shock-Velocity-Particle-Velocity Relationship, J Appl. Phys. 38, No. 13, p 4976, 1967

12. Zel'dovich, Ya and Raizer, Yu, Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena, Volume 2, Chapter XI, Academic Press, 1966

13. Slater, J.C., Introduction to Chemical Physics, Chapter 14, Mc Graw-Hill, New York, NY, 1939.

14. Dugdale, J.S. and MacDonald, D.K.C., "The Thermal Expansion of Solids," The Physical Review, 89, p 832, February 1953

15. Rose, J.; Smith J.; and Ferrante.J., Universal Features of Bonding in Metals, Phys. Rev. B, 28, No. 4, p 1835, 1983.

16. Vinet, P.; Ferrante, J.; Smith J.; and Rose J., A Universal Equation of State, J Phys. C: Solid State Phys., 19, L467-L473, 1986.

17. Grover, R., Getting, I.C., and Kennedy, G.C., "Simple Compressibility Relation for Solids," Physical Review B: Solid State, 7, No. 2, pp. 567-571, January 1973.

18. Prieto, F. and Renero, C., Equation for the Shock Adiabat, J. Appl. Phys., 41, No. 9, p 3876, 1970.

19. Prieto, Fernando, A Law of Corresponding States for Materials at Shock Pressures, J. Phys. Chem. Solids, 35 pp. 279-286, 1974.

20. Hirchfelder, Joseph; Curtis, Charles;, and Bird, R. Byron, Molecular Theory of Gases and Solids, Chapter 4, pp. 234-328, Wiley, 1954.

21. Shively, J.H., Stein, C., and Robertson, R., An Equation of State for Shock in Homogeneous Materials and Comparison to Shock Data, PL-TR-92-1060, Phillips Laboratory, space and Missiles Directorate, Air Force Materiel Command, Kirtland AFB, NM 87117.

22. Lennard-Jones, J and Devonshire, A. Critical Phenomena in Gases I, Proc. Roy. Soc., A163, p 53, 1937.

23. Lennard-Jones, J . The Equation of State of a Gas, Proc. Camb. Phil. Soc., 22, P 105, 1931.

24. Kittel, Charles, Introduction to Solid State Physics, 5th Ed, Chapter 3, pp. 76 - 102, John Wiley, New York, NY, 1973.

25. King, Thomas, Boeing Aircraft, Renton Washington, Private Communication

26. Shively, Jon H. and Roybal, Robert, Chemical and Structural Effects Caused by Hypervelocity Impacts to Polysulfone Graphite Reinforced Composite and Its Resin, PL-92-1076, Phillips Laboratory, Space and Missile Directorate, Air Force Materiel Command, Kirtland AFB, NM 87117, June 1993

8.0 FIGURES

See following pages.

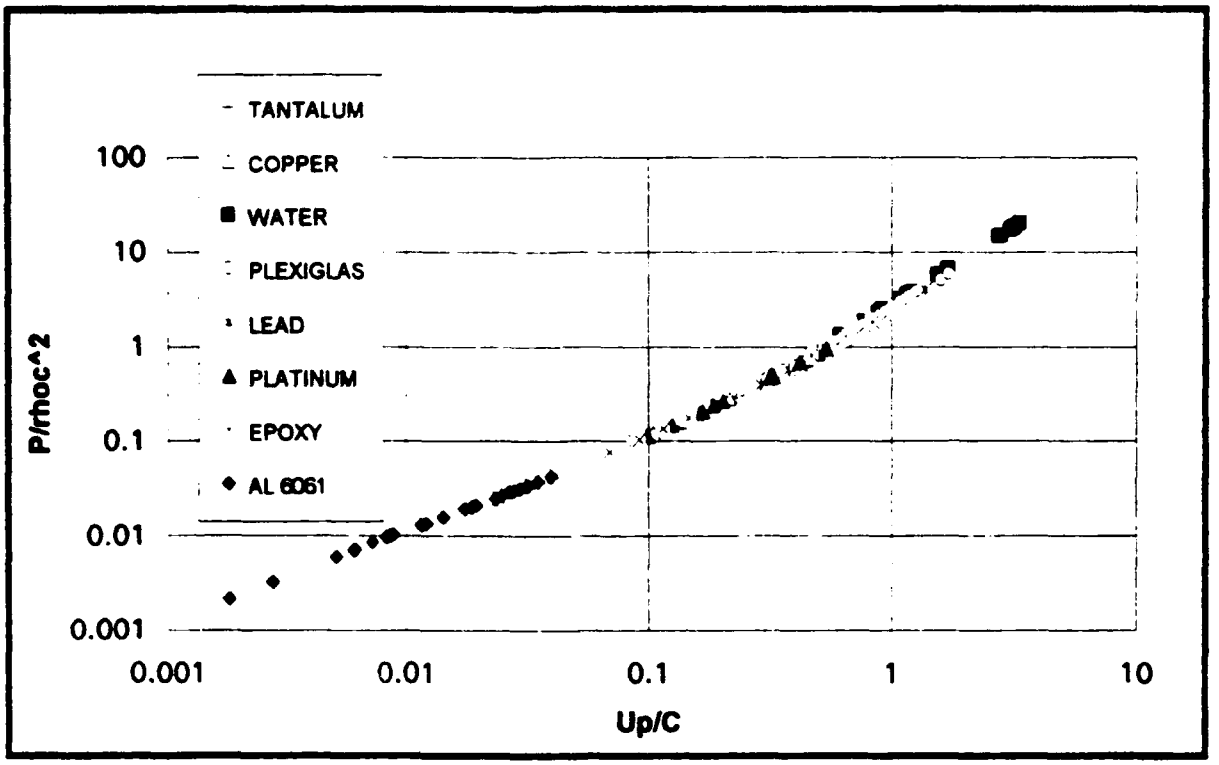


Figure 1. Master curve of normalized pressure for hypervelocity collisions as a function of normalized particle velocity for eight materials.

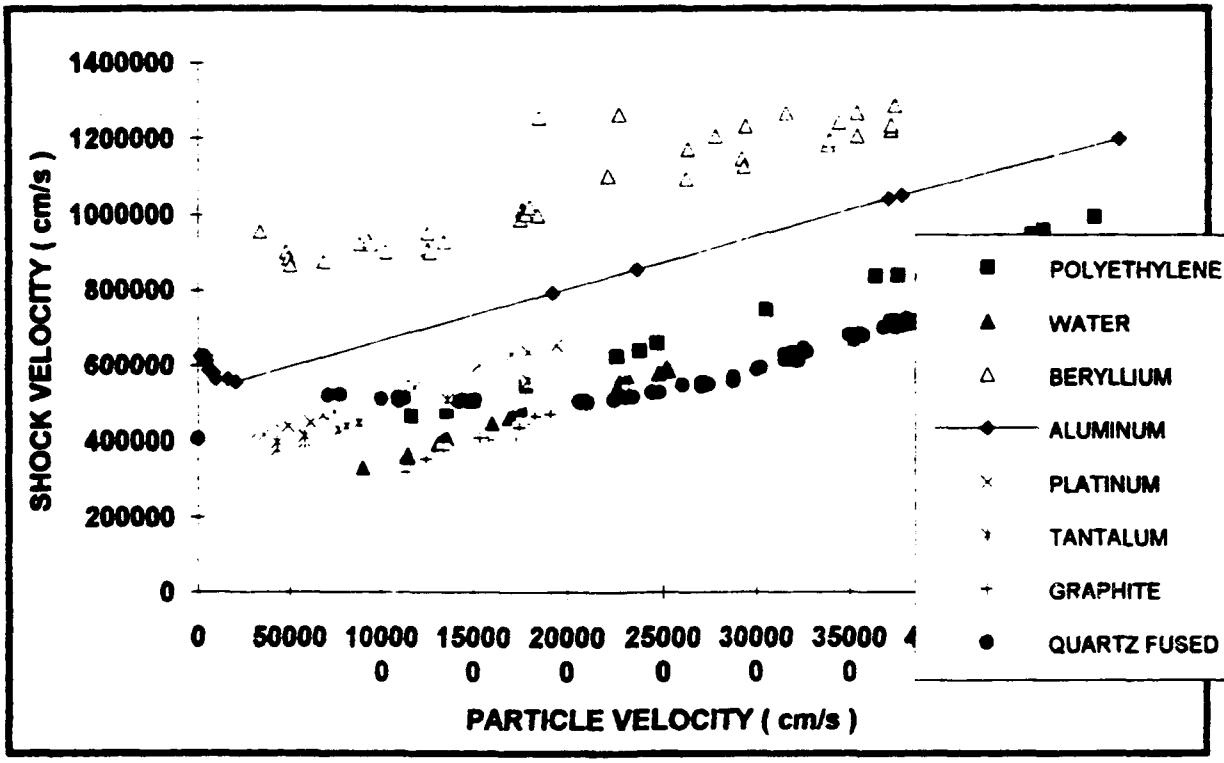


Figure 2. Shock velocity versus particle velocity for eight materials.

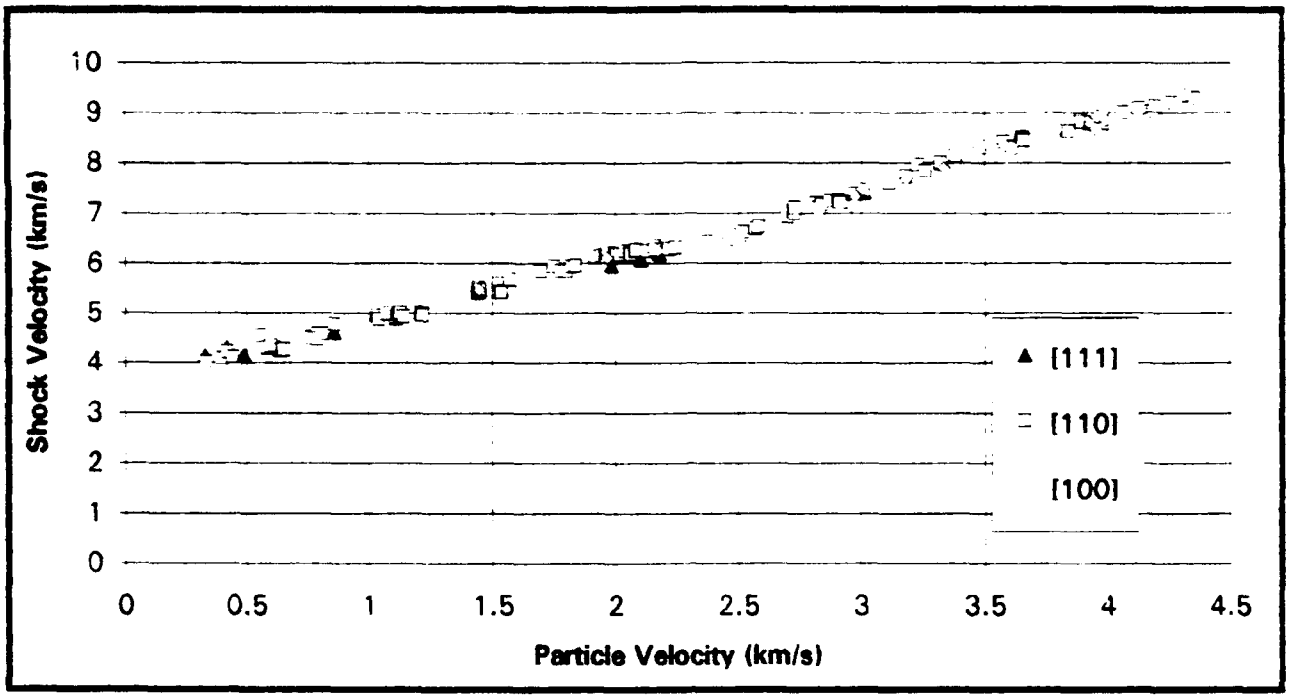


Figure 3. Shock velocity versus particle velocity for three crystal orientations in sodium chloride.

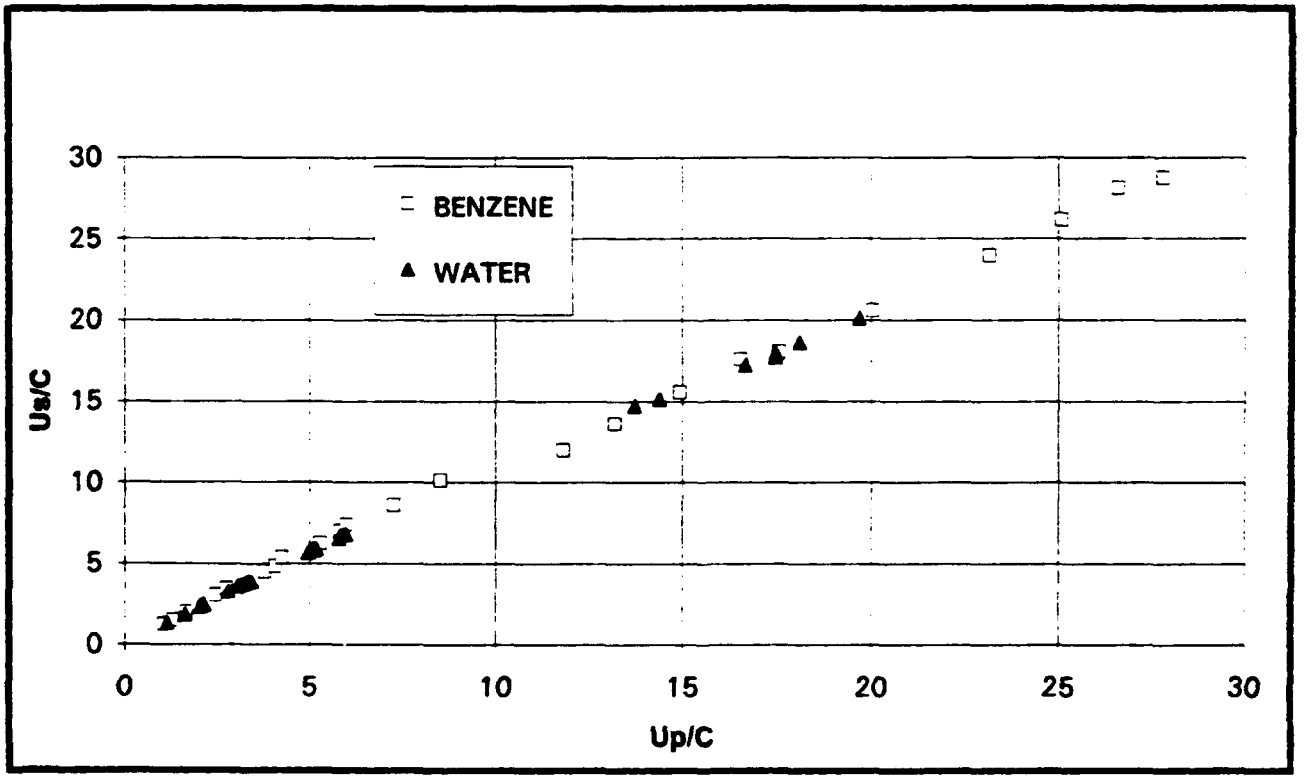


Figure 4. Normalized pressure after hypervelocity collisions as a function of normalized particle velocity for water and benzene.

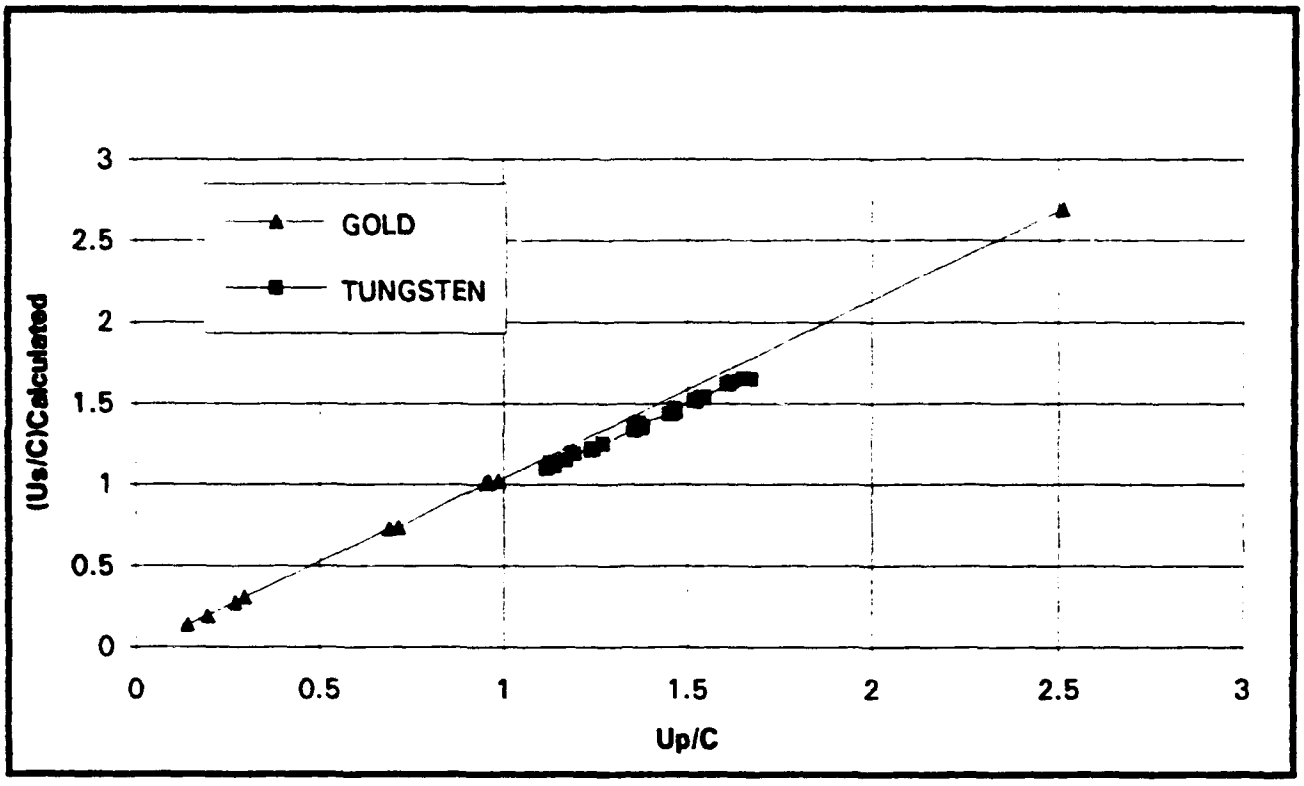


Figure 5. Calculated normalized shock velocity versus normalized particle velocity for two different metallic crystal structures.

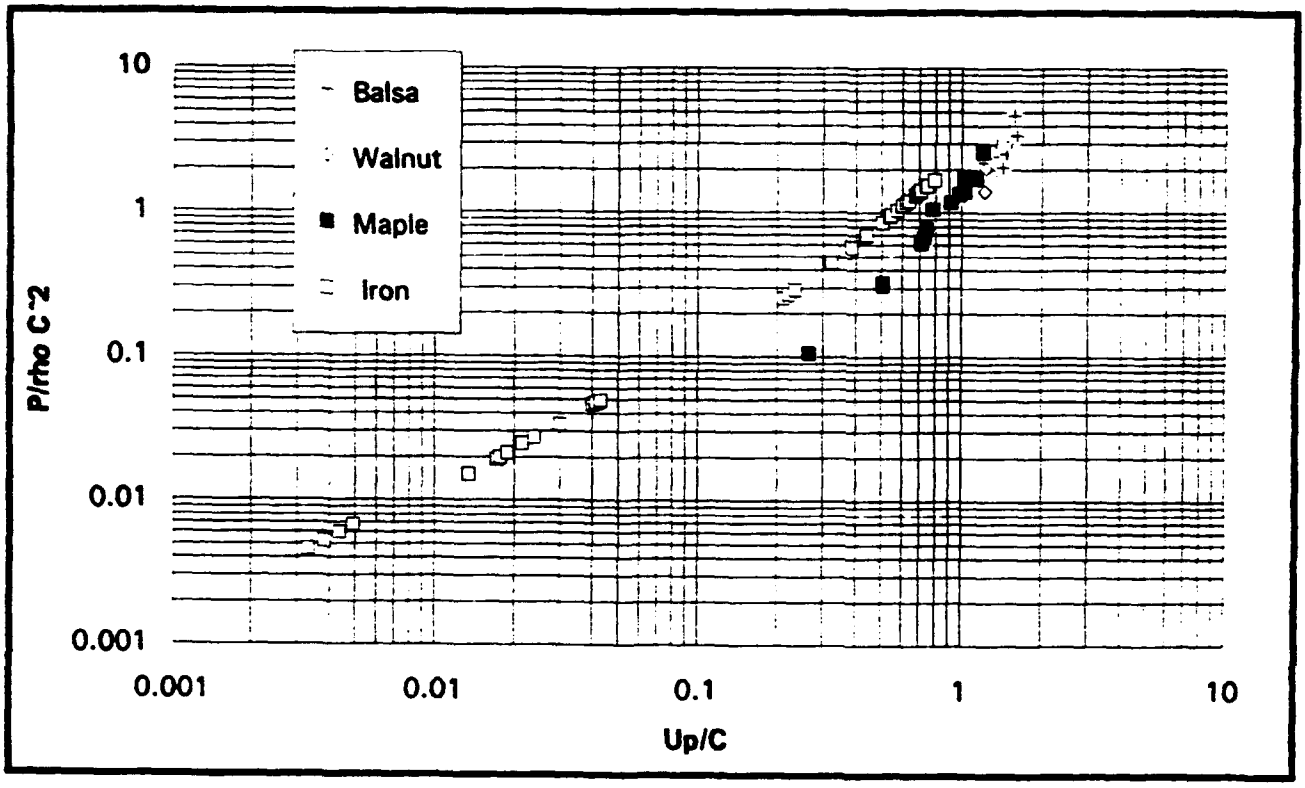


Figure 6. Normalized pressure for hypervelocity collisions as a function of normalized particle velocity for three wood materials and iron.

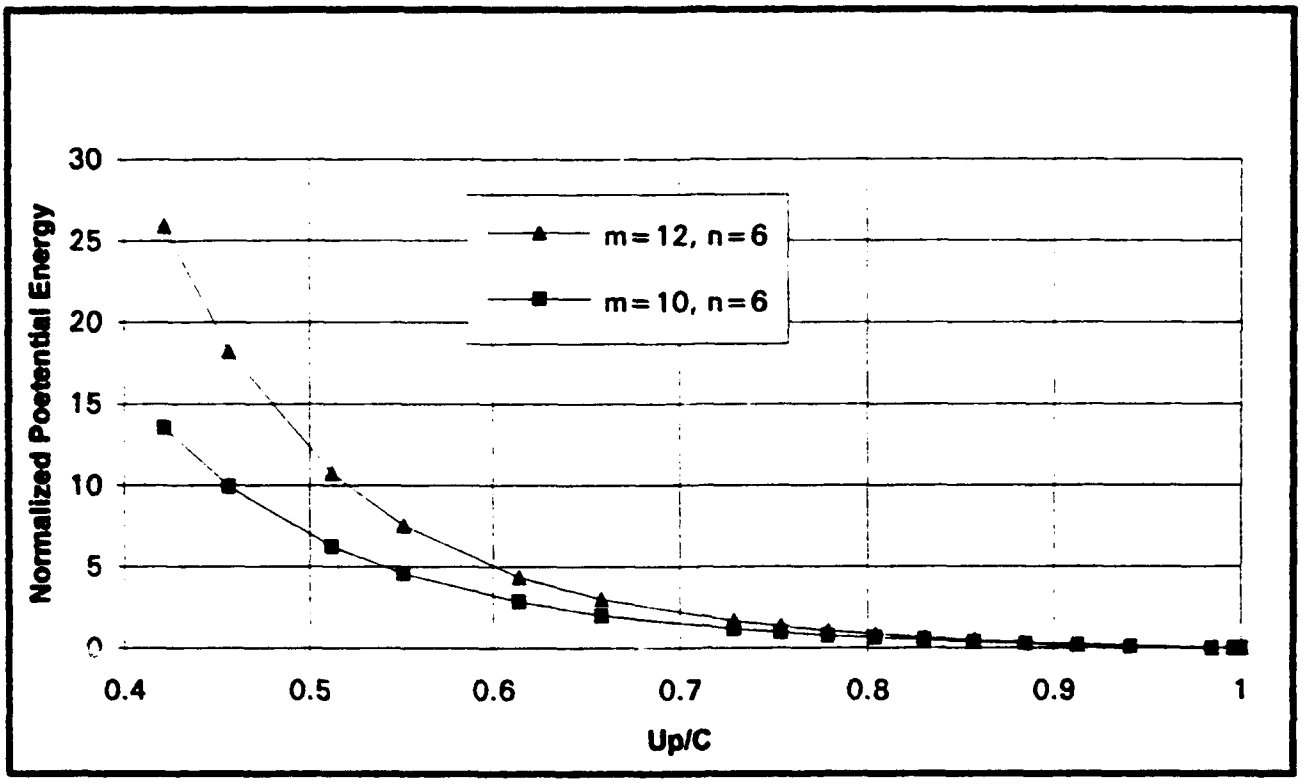


Figure 7. The normalized Lennard-Jones potential energy curve for two different exponents of the repulsion term.

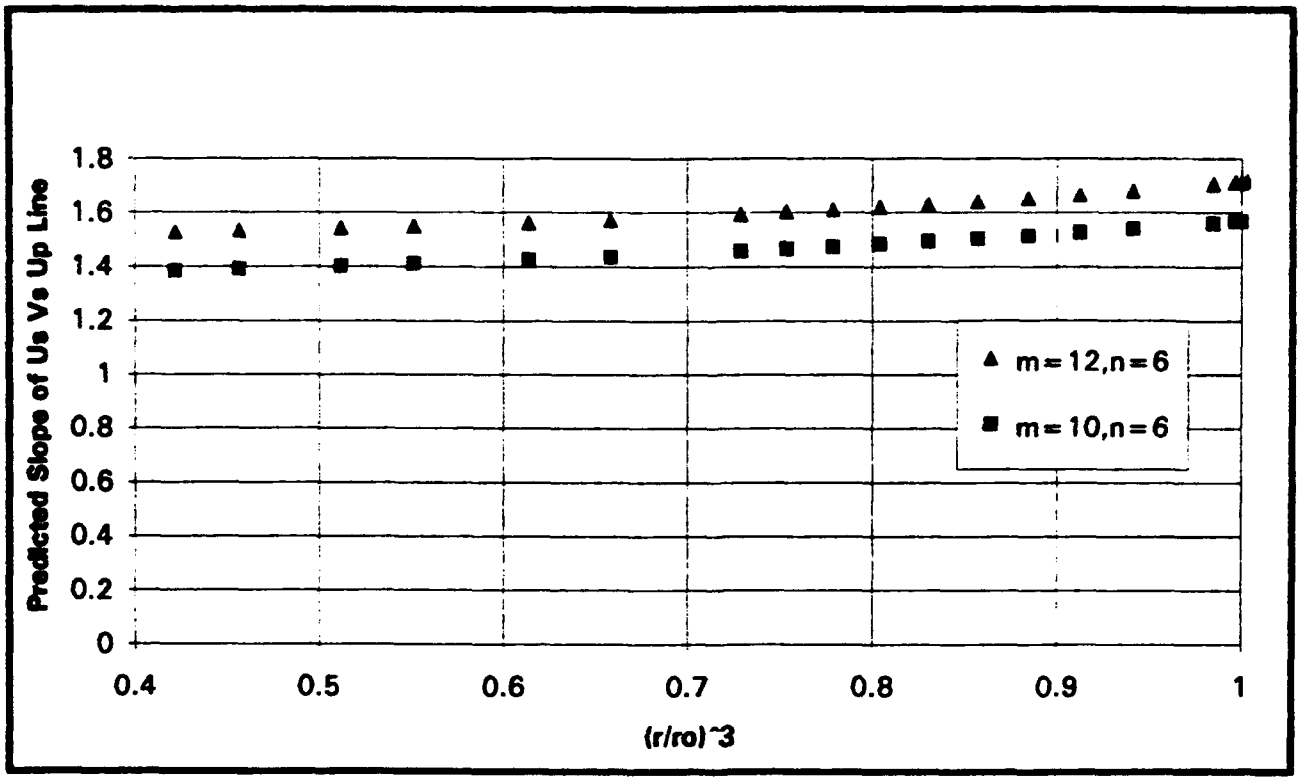


Figure 8. Predicted slope, Equation 14, of the shock velocity versus particle velocity using the Lennard-Jones equation for potential energy for two values of the exponent in the repulsion term.

SPACE SYSTEM SURVIVABILITY AND DEMONSTRATION WITH C++

Jay Bhuyan
Assistant Professor
Department of Computer Science

Tuskegee University
Tuskegee, AL 36088

Final Report for:
Summer Faculty Research Program
Phillips Laboratory
Kirtland Air Force Base, Albuquerque, NM

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1993

SPACE SYSTEM SURVIVABILITY AND DEMONSTRATION WITH C++

Jay Bhuyan
Assistant Professor
Department of Computer Science
Tuskegee University, Tuskegee, AL 36088

Abstract

This paper discusses a c++ implementation of a simulated system which consists of arbitrary offensive and defensive objects. The purpose of the simulated system is to investigate the efficacy of weapons, sensors, battle management and communication system as well as the effect of policy and strategy. The system being studied is not restricted to the space system survivability problem. It can simulate any arbitrary complex system consisting of different objects interacting with each other. An object-oriented approach to the above problem was previously developed and is called Advanced Simulation Programming Environment (ASPEN). ASPEN is already implemented in Fortran. We decided to study the c++ implementation of the ASPEN framework because c++ supports all the object-oriented characteristics like encapsulation, polymorphism and inheritance.

Introduction

The purpose of our investigation is to simulate the engagements between any arbitrary offensive and defensive system. The entire earth and its near vicinity (at least out to geosynchronous orbit) is considered in our simulation. The simulation is supposed to take into consideration thousands of objects in a limited engagement to an upward boundary of fifty thousand in a full scale exchange. The purpose of this code is to investigate the efficacy of weapons, sensors, battle management, and communication system as well as the effect of policy and strategy. The system is exposed to both natural and engagement induced stresses. The natural stress consists of weather and electromagnetic environment while the engagement induced stress consists of nuclear effect and electronic warfare.

The Defensive Technology Evaluation Code (DETEC)[5] can be assumed as a subset of Advanced Simulation Programming Environment (ASPEN)[1-4], which is an application of object-oriented approach to a complex system. ASPEN is not restricted to simulate an engagement in a war. It can simulate any system where several objects interact with each other.

ASPEN is a core of code called the framework that integrates a set of objects into a working simulation. It can be thought of as a database manager which creates, destroys, and manages different objects in a database. It provides various real world functions and an interference for the user to define the input and output to the system. It simulates an event driven process. Each event is a time tagged message stored in an event queue. The events are either generated by the user or by actions of different objects. The events are served by the framework on a first come first serve basis. The events are sent to the appropriate objects for further processing. It may result in creation of an object or in the execution of a behavior function of an object.

The support functions which are required to execute the simulation are performed by the framework. The support functions are responsible for interaction with the user, management of event queue, updating of real world environment, managing the object functions, and managing the user defined outputs.

Each object in ASPEN has an identity, behavior, and state. The object's data structure can be divided into 4 sections: identification, real world attribute, object attribute, and the associated component array section. The identity of the object is defined by the identification section, object's state is defined by the real world attribute and object attribute sections, and the object's behavior and associated objects are defined by the associated component section.

Object-oriented approach is very suitable for this kind of simulation, as it makes the incremental construction of a complex model easier. Object-oriented approach facilitates the extensibility and reuse of the software. ASPEN is implemented in Fortran. Though the resulting code is well documented and structured, it is very difficult to support the object oriented-characteristics through Fortran language. Therefore, we decided to explore the possibility of c++ implementation of the ASPEN framework.

c++ supports the object-oriented concept like objects, classes, encapsulation, function and operator overloading, inheritance and polymorphism. Encapsulation is enforced in a c++ class by declaring private and protected member variables and member functions. Inheritance is defined by defining a class in term of another class. The derived class inherits the behaviors of the base class through private, public, and protected members. Virtual function declaration enables c++ language to support polymorphism. Multiple derived functions will implement a function in different ways. Function overloading allows the programmer to implement a number of different functions with the same name. The compiler will be able to detect the required function by looking at the number and type of arguments. Operation overloading allows the programmer to redefine the meaning of standard unary and binary operation.

The rest of the paper is divided into 4 sections. Section 1 describes the working of ASPEN. Section 2 describes the simulation control table, which stores the information about the complete simulation. Section 3 describes the design of object and its implementation. Section 4 describes the design of events, event queue, and internal and global database. Section 5 concludes the paper.

1. Description of ASPEN

The ASPEN framework consists of five modules: Manager, Executive, Global processes, Engagements, and Logger. The user interacts with the system through the Manager module. The user can define a scenario, define an object, and specify the output using the Manager. Executive is responsible for the overall control of the simulation. It manages an event queue consisting of several events. The events are generated either at the time of initialization or by the processes. The executive is responsible to direct an event to an object which results in the change of the object's dynamic parameter values. Global processes are the behavior function of a mother nature object. It creates an embedded real world environment. It is responsible for updating the real world attribute of each object. The events result in the execution of a function of an object. The functions of an object are classified into following categories: Decision process, sensing, communication, action, and reaction. The logger module is responsible for creating different output files as per the user's specification. The dynamic and static values of the objects are recorded in these files.

To execute the simulation the user interacts with the system by selecting various options from a menu. The available options are selection of models, initialization of objects, initialization of event to start execution, the type of interactive display, and the information to be recorded into the post processing files. As per the specifications of the user, the objects are created in the database. The database of objects is called the real world database. Once the objects are created the embedded real world attributes are calculated and stored in the database. After the creation of the real world database the actual simulation begins by selecting an event in front of the queue. The event can either be a global process or a function to be executed by an object. Updating the velocity, location, and weather are examples of global process events. These events are passed to the global process module by the executive. The global process module executes the process corresponding to the event and the information generated by the process is placed in each object.

The engagement module processes the function type event. It also determines the

type of the function and the object responsible for executing this function. The function is executed by the object and the state of the function is changed accordingly. The execution of a function may also generate another event which is put in the event queue. Simulation can be stopped either through an interrupt or through simulation stop time event.

2. Simulation Control Table (SCT)

Simulation control table is used to store the state of the system and the locations of different blocks in the database. There are 8 different blocks in the database: database file header, description data, event queue, active events, simulation grid, global database, and the simulation objects. The SCT points to the starting locations of all blocks except the grid data block. In the case of grid data block the SCT points to the last created grid. Both the current simulation time and the simulation stop time are stored in the SCT. Time is measured in Julian day starting 1 January 4713 BC. The fraction of the day is stored as number of seconds in the universal time. We decided to represent the SCT as a structure instead of a class as we did not feel the need for hiding its members from other routines. The declarations and descriptions of different fields of SCT are given below.

```
typedef struct
{
 int julian_day;
 int universal_time
} time;

typedef struct
   {
    int state;// The current simulation state. -2 Completed, -1 Not defined,
              //0  Initialized, 1 Interrupted, 2 Executing.

    fstream database; // A random file containing the simulation database.
```

```
    fstream output; // The simulation post processing output file.

    time current; // The absolute simulation current time.
    int event_dbase; // The position of the event queue in the data base.
    event_type event_mem;// The pointer to the event queue.
    int defn_dbase;//The position of the simulation object's definition.
    defn_type *defn_mem; // pointer to the definition
    int grid_dbase;//The position of the last created grid in the database file.
    grid_type *grid_dbase;//pointer to the last created grid.
    int active_event_dbase; //The position of the active events in the database
                            //file.
    active_event_type* active_event_mem; // pointer to the first active event.
    int global_database_dbase;// The position of the global database in the
                          //database file.
    int_dbase_type *global_database_mem; // pointer to the global database.
    int communication_dbase;//The position of the communication packet in
                        //the database file.
    comm_type *communication_mem;//pointer to the communication packet
    time greenwich; //The greenwich sidereal time at the current simulation time.
    time sin_greenwich;//The sine of the greenwich sidereal time at the current
                          //simulation time.
    time cos_greenwich;//The cosine of the greenwich sidereal time at the current
simulation time.
    time start; // The absolute simulation start time.
    time stop; // The absolute simulation stop time.
  }sct_type;
```

The fields of the SCT structure are divided into two parts. Starting from the state field up to the communication field are called application independent entries and the rest are application dependent entries.

2. Description and Design of Objects

An object is defined as an entity that has three attributes: identity, behavior, and state. An ASPEN object can be assumed to be consists of 4 separate parts: identification, real world attributes, associated component array, and object attributes. The position of an object in ASPEN framework is represented either in Earth Centered Inertial coordinate system (ECI) or in Earth Centered Geographic coordinate system (ECG). In the case of earth fixed objects ECG system is being used while in the case of movable objects like satellites ECI system is used. The attributes needed to identify an object is same for all objects. Therefore, we decided to use a base class for the identity part which is defined below.

```
typedef struct
      {
      double right_ascension;
      double declination;
      double radius;
      } eci_pos;

typedef struct
      {
      double latitude;
      double longitude;
      double altitude;
      } ecg_pos;

typedef struct
      {
      eci_pos pos;
      double speed;
     } eci_vel;

typedef struct
      {
       ecg_pos pos;
       double speed;
      }ecg_vel;

class object
```

```
{
 char* side;
 char* class;
 char* type;
 char* system;
 int number; // Number assigned to the object
 int state; //Object's status : Operational, destroyed, etc.
 int template;// Object's template location in the database.
 int parent_dbase; //Object's parent location in the database.
 indi_object *parent_mem;//Pointer to object's parent.
 int loc; //Object's location in the database.
 time last_update; // Last time the object was updated.
public:
 virtual get_pos_vel(eci_pos pos, eci_vel : vel){return error;}
 virtual get_pos_vel(ecg_pos pos, ecg_vel : vel){ return error;}
 int test_parent();
 int test_active();
 set_damage(double damage);
 set_dead();
 destroy();
 test_destroy();
 int  get_parent();
 virtual change_vel(eci_vel dvel){return error;}
 virtual change_vel(ecg_vel dvel){return error;}
 virtual create_associate(sct_type sct, int loc,int error){return error;}
 virtual create_associate_dynamic(sct_type sct,int loc,int error){return
error;}
 virtual   create_associate_dynamic(sct_type   sct,int   loc,eci_pos
                dpos,eci_vel dvel int error){return error;}
 virtual   create_associate_dynamic(sct_type   sct,int   loc,ecg_pos
               dpos,ecg_vel dvel, int error) {return error;}
 };
```

The detailed explanation for some of the fields in the class object are described below.

Side: In the current simulation the objects are divided into different groups. Side determines the group of an object and is set by the user during initialization.

Class: It determines the default global processes which are applied to an object. As we have discussed earlier a super class object, called Mother Nature, consists

of a collection of glocal process functions which apply to every simulation active object.

Type: This field determines the structure of an object. Structure of an object is defined by its associated component and its attributes. Associated component consists of behavior functions of the objects and the associated objects. Object's attribute section consists of static and dynamic attributes, which are the inputs and outputs of the behavior function respectively.

System: This attribute differentiates the object of the same type. The objects of the same type are further classified into different groups depending on the values of their static attributes. The examples of static attributes are total weight, thrust, total fuel, etc. Systems are defined during initialization.

Number: The objects of a particular system are distinguished by numbers. These values are assigned by the system when an instance of an object belonging to a particular system is created.

State: This attribute determines the state of an object

Template: This attribute is the location of the defn of a system in the database.

parent: This attribute determines the position of the parent of the object in the data base;

loc: this attribute determines location of the object in the database.

get_pos_vel(sct_type sct, eci_pos pos, eci_vel, vel);

purpose: This routine retrieves the ECI position and velocity of a simulation object.

Input:

sct : Simulation control table

output:

pos: the ECI position of the object

vel : The ECI velocity of the object.

get_pos_vel(ecg_pos pos, ecg_vel vel);

purpose: This routine retrieves the ECG position and velocity of a simulation object.

output:

pos: the ECG position of the object

vel : The ECG velocity of the object.

int test_parent();

purpose: This routine determines if the parent of an object exists.

int test_active();

purpose: This routine determines if the object is active.

set_damage(double damage);

purpose: This routine sets the damage to an object.

set_dead();

Purpose: This routine marks an object as dead.

destroy();

purpose: This routine marks a simulation object as totally destroyed.

test_destroy();

purpose: This routine determines if a simulation object has been destroyed.

int get_parent();

purpose : This routine gets the location of the parent of an object.

change_vel(eci_vel dvel);

purpose: This routine changes the ECI velocity of a simulation object.

Input:

dvel : The change in the ECI velocity of the object.

change_vel(ecg_vel dvel);

purpose: This routine changes the ECG velocity of a simulation object.

Input:

dvel : The change in the ECG velocity of the object.

create_associate(sct_type sct, int loc,int error);

purpose: This routine creates active associated objects for a newly created object.

Input:

sct: Simulation control table

loc: The location of the object in the database.

Error:

0 Object created

1 invalid location

create_associate_dynamic(sct_type sct,int loc,int error);

purpose: This routine dynamically creates an active associated object.

Input:

sct : Simulation control table

loc: The location of the new object template.

Error:

0 Object created

1 invalid location

create_associate_dynamic(sct_type sct,int loc,eci_pos dpos,eci_vel dvel,int error);

purpose: Dynamically creates a simulation object using ECI offset

Input:

sct : Simulation control table

loc: The location of the new object template.

dpos: The ECI offset from the parent.

dvel: The change in velocity from the parent

Error:

0 Object created

1 invalid location

create_associate_dynamic(sct_type sct,int loc,ecg_pos dpos, ecg_vel dvel,int error);

purpose : Dynamically creates a simulation object using geographical offset.

Input:

sct: Simulation control table

loc: The location of the new object template.

dpos: The ECG offset from the parent.

dvel: The change in velocity from the parent

Error:

` Object created

1 invalid location

Class	Type
Orbital	Satellite
	Simple Interceptor
	Brilliant Pebble Satellite
	Brilliant Eyes Satellite
Aerospace	Simple Missile Booster
	Missile RV
Earth Fixed	Space Surveillance Network Center
	Mission Control Element
	Battery Control Center
	Base Escape Mission Control Element
	Base Escape Space Surveillance Network Center
	Simple Tracking Station
	Silo
	Ballistic Missile Submarine (SSBN)
	Central Command Post
	Airbase
Extended	Land Line
	Link
Aircraft	Bomber
Sensor	IR Sensor
Communication	Simple Receiver
	Simple Transmitter

Table: Defined objects by Class and Type

In the present simulation, there are 23 objects present in our database. These objects are classified into 7 different classes depending on the global processes applicable to the objects. The objects and their corresponding classes are

summarized in the above table.The definition of all these classes are mentioned below.

```
class aircraft : object {
                ecg_pos pos;    // ECG Position of the object.
                ecg_vel vel;    // ECG velocity of the object.
                int power_flag; // True means that the object is under power.
                double sunlight ; //Sunlight condition at object.
                public:
                get_pos_vel( ecg_pos pos, ecg_vel  vel)
                change_vel(ecg_vel dvel);
                };
class aerospace : object{
            eci_pos pos; // ECI position of the object.
            eci_vel vel; //ECI velocity of the object.
            double sunlight; //Sunlight condition of the object.
            int power_flag; // True means that the object is under power.
            int turn_flag;  //object's turn flag;
            double temp; // object's radiation temperature;
            double absorbability; /object's absorbability;
            public:
            get_pos_vel  eci_pos pos, eci_vel  vel);
            change_vel(eci_vel dvel);
          };
class earthfixed : object {
            ecg_pos pos; //ECG position of the object.
            double sunlight; //Sunlight condition at object.
                        };
class extended : object {
                ecg_pos pos1;// Two end points location on the earth
                ecg_pos pos2;//in geographical coordinates
                        };
class orbital : object {
                eci_pos pos;//ECI position of the object.
                eci_vel vel;// ECI velocity of the object.
                double sunlight; // Sunlight condition of the object.
                double temp;// object's radiation temperature
                double absorbability; //object's absorbability
                double emissivity; // object's emissivity
                int power; //orbital object's power flag.
                int turn; // orbital object's turn flag.
                public:
```

```
        get_pos_vel( eci_pos pos, eci_vel vel);
        change_vel(eci_vel dvel);
    };
```

The individual characteristics of each object are stored in another class called indi_object which is derived from their respective classes. The indi_object class will contain the associated objects, its functions, and the static and dynamic parameters for these functions. The following functions will be placed in the public part of the indi_object to construct the objects.

```
indi_object(sct_type sct, char* iside, char* iclass, char* isys,int
    template, int error);
indi_object(sct_type sct, char* iside, char* iclass, char* isys,
    int template, eci_pos pos, eci_vel: vel,  int error){ return error;}
indi_object(sct_type sct,char* iside, char* iclass, char* isys,
int template, ecg_pos: pos, ecg_vel vel,  int error) { return error;}
```

indi_object(sct_type sct, char* iside, char* iclass, char* isys,int template, int error);

Purpose : This routine dynamically creates a simulation object.

Input:

sct: Simulation control table.

iside : The side of the object.

iclass : The class of the object.

isys: The system of the object;

template: Location of the object's template in database.

Error:

0 object created

1 No definition for class, type, or system

indi_object(sct_type sct, char* iside, char* iclass, char* isys, int template, eci_pos pos, eci_vel vel, int error);

Purpose : This routine statically creates a simulation object using the ECI position and velocity.

Input:

sct: Simulation control table.

iside : The side of the object.

iclass : The class of the object.

isys: The system of the object;

pos : The ECI position of the object.

vel: The ECI velocity of the object.

template: The location of the object's template in the database.

Error:

0 object created

1 No definition for class, type, or system

indi_object(sct_type sct,char* iside, char* iclass, char* isys,

int template ecg_pos pos, ecg_vel vel, int IRT);

Purpose : This routine statically creates a simulation object using the ECG position and velocity.

Input:

sct: Simulation control table.

iside : The side of the object.

iclass : The class of the object.

isys: The system of the object;

pos : The ecg position of the object.

vel: The ecg velocity of the object.

template : The location of the object's template in the database.

Error:

0 object created

1 No definition for class, type, or system

3. Description and Design of Events, Simulation Grid, Internal Database, and Communication Packet

Events:

Events are messages that are transmitted between objects which result in the change of dynamic attributes of objects. Events are not directly transmitted

between objects. Created events are maintained in an event queue. They are processed on a first come first serve basis. Events can either be instantaneous or extended type. Instantaneous event is one event of negligible time duration. Extended event simulates processes of considerable time duration. It is represented by two events: one event marks the beginning of the process, while the other event marks the end of the process. Because of the presence of extended event, it is possible that more than one event may be active at a particular time. This overlapping information is stored in a data structure called active events. The design of the active events is given below.

```
typedef struct
      {
      int loc_dbase; // The position of the active event in database.
      active_event_type *next; // The pointer to the next active event
      event_type   *event_mem; // pointer to the event
      int  event_dbase // Location of the event in the database
      } active_event_type;
```

The structure of the event queue consists of time ordered events which are described below:

```
typedef struct
      {
      time event_time; // Event time
      int mode; //Event mode  instantaneous, start or extended event
      int flag; //Execute immediate flag;
      int end_dbase; // Location of the end event in database
      event_type *end; // pointer to the end event if it is in memory
      int category; //global,engagement,logger
      int type; // Event type
      int subtype; //Event subtype;
      int first_dbase; //Location of the first object in the database that
                        //performs the function asscciated with the event
      int  second_dbase;  //Location  of  the  second  object  in  the  database
```

```
                        that performs the function associated with the event
    indi_object *first_mem;//Pointer to the first object
    indi_object *second_mem;//Pointer to the second object
    int number; //Number of data in the event
    int data[size] //Array containing data
    int comm_dbase; //The location of communication packet in the database.
    comm* comm_mem; //The pointer to the communication packet
    int prev_dbase; //Location of previous event in the database
    event_type* prev_mem;//Pointer to the previous event
    int next_dbase;//Location of the next event in the database
    event_type* next;//Pointer to the next event
    } event_type;
```

Simulation Grid:

Simulation grid helps us to store the constant and derived properties of space. The space under consideration (grid) is divided into a number of cells. For each cell, we store a number of constant and derived data. Terrain data is an example of constant data, while temperature, electron density, atmospheric density, and dust density are examples of dynamic data. The structures of grids and cells are given below.

```
typedef struct
    {
    int identifier; //Name of the grid
    int prev_dbase; //Location of previous grid in the database
    grid_type* prev_mem; //pointer to previous grid
    int next-dbase; //Location of next grid in the database
    grid_type* next_mem;//Pointer to next grid
    int cell_dbase;//Location of the first cell in the database
    cell_type* cell_mem;//Pointer to the first cell
   }grid_type;
typedef struct
    {
```

```
int dimension1;//The index of the cell in dimension #1
int dimension2;//The index of the cell in dimension#2
int dimension3;//The index of the cell in dimension#3
int next_mem; //The location of the next cell in the database
cell_type* next_mem;//pointer to the next cell
int number; //Number of data
int data[max-size];//Array containing data
}cell_type;
```

Communication Packet:

Communication packets are messages that are transferred between objects directly. The pointer to the communication packet is stored in the event responsible for transferring the message. The structure of the communication packet is given below.

```
typedef struct
{
int receiver_dbase;//The location of the receiver object in the database
indi_object * receiver_mem;//The pointer to the receiver object
int trans_dbase;//The location of the transmitter object in the database
indi_object *trans_mem;//The pointer to the transmitter object
int number;//Number of data in the communication packet
int data[maxsize];//Array containing data
}comm_type;
```

Internal Database:

The purpose of the internal database is to store the data during simulation execution which is to be used by the objects in future. Internal database is classified as global database and local database. Global database is used by object functions of all sides, class, type, and system. The local database is used by the individual object and its location is determined by a pointer stored in the individual object. The structure of the internal database is given below.

```
typedef struct
{
int prev_dbase;//The location of previous internal database in database
```

```
file
      int_dbase_type *prev_mem;//The pointer to previous internal database
      int next_dbase;//The location of next internal database in database file
      int_dbase_type *next_mem;//The pointer to next internal database
      int number;//The number of data in the internal database
      int data[maxsize];//Array containing the data
      }int_dbase_type;
```

4. Conclusion

The c++ implementation of ASPEN framework will enable expansion of the system fairly easy. We will be able integrate additional models into our system with very minimal change to our previous code. In the Fortran implementation, the objects, events, internal database, etc. are used by the program by transferring the database to buffers and then to the respective data structures. In the current implementation, we are proposing to transfer the data directly from the database to the data structures. We are able to employ this method because c++ supports dynamic memory allocation and deallocation, while Fortran does not. In our implementation, we only defined the objects as classes, keeping events, communication packet, internal database, etc. as structures. However, we need to study further to find out, if there is a need for converting these structures into classes, so that the operations and the contents of these data structures can be encapsulated properly.

References

1. "ASPEN Framework Routines", by Lewis W. Gaby, CSC-ASPEN-006, 23 May 1991.
2. "ASPEN User's Guide", by Neil P. Phillibe, CSC-ASPEN-007.1, January 1993.
3. "The Advanced Simulation Programming Environment (ASPEN) philosophy", by Neil P.Phillibe, CSC-ASPEN-004.3, 26 May 1992.
4. "The Base Escape Simulation", by Neil P. Phillibe, et.al.,S3D2-C-9008, Computer Science Corporation, 25 Aug 1992.
5. "Conceptual Specification for DEfensive Technology Evaluation Code (DETEC)", by Ronald D. Christman, et. al., LA-10547-MS, Los Almos National Laboratory, January 1986.

LOCATION AND IDENTIFICATION OF UNEXPLODED ORDNANCE WITH GROUND PENETRATING IMAGING RADARS, INFRARED (IR) IMAGING, ACOUSTIC IMAGING SONARS, AND ELECTROMAGNETIC INDUCTION METHODS

Albert W. Biggs

Professor

Department of Electrical and Computer Engineering

University of Alabama in Huntsville

Huntsville, AL 35899

Final Report for:

Summer Faculty Research Program

Phillips Laboratory, Kirtland Air Force Base, NM

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D.C.

September 1993

LOCATION AND IDENTIFICATION OF UNEXPLODED ORDNANCE WITH GROUND PENETRATING IMAGING RADARS, INFRARED (IR) IMAGING, ACOUSTIC IMAGING SONARS, AND ELECTROMAGNETIC INDUCTION METHODS

Albert W. Biggs
Professor
Department of Electrical and Computer Engineering
University of Alabama in Huntsville

Abstract

At least 259 former military sites across the United States have aging munitions buried beneath them. Explosives, or unexploded ordnance (UXO), have already been found in suburban neighborhoods, public parks, and even college campuses. These UXOs range from hand grenades, antitank mines, bombs, and artillery shells to several tons of TNT.

Although toxic chemicals are also hidden at hundreds of other sites, this work is limited to cleanup of UXOs at current and abandoned military sites. The work described in this paper is a continuation of studies initiated last summer in the SFRP, and will result in proposals for development of sensors for locating and identifying buried ordnance. These sensors are ground penetrating imaging radars, acoustic imaging sonars, infrared (IR) imaging, and electromagnetic induction methods.

Military ranges have terrain or soil profiles which vary because of relative water content, types of soil, and present of other UXOs. These variations, which also change daily and seasonally, require the use of more than one sensor at a given facility or range. Descriptions of sensors provide a base for finding and identifying the buried ordnance.

LOCATION AND IDENTIFICATION OF UNEXPLODED ORDNANCE WITH GROUND PENETRATING IMAGING RADARS, INFRARED (IR) IMAGING, ACOUSTIC IMAGING SONARS, AND ELECTROMAGNETIC INDUCTION METHODS

Albert W. Biggs

INTRODUCTION

The government believes aging munitions may be buried beneath at least 259 former military sites across America. Explosives already have endangered suburban neighborhoods, public parks, and some college campuses. The hazards range from hand grenades to artillery rounds.

In addition to munitions, hundreds of sites, many dating back to World War II, hide toxic chemicals ranging from old fuel at an abandoned Arizona Air Force base to mustard gas agents near Edison, New Jersey. Other sites near Hanford, Washington, have buried drums of radioactive waste. In northwest Washington, D.C., up to 141 unexploded World War I chemical shells were dug up, about four miles from the White House. One resident was excavating his yard for a new driveway and found an artillery round from the Spanish-American war.

Although several forms of hazardous waste will be encountered at both military and other sites, only UXOs, such as artillery rounds, bombs, bazooka rounds, hand grenades, antitank/antipersonnel mines, and related ordnance, will be considered as targets for the four groups of sensors discussed in the subsequent sections. These sensors are ground penetrating

imaging radars. acoustic imaging sonars, infrared (IR) imaging, and electromagnetic induction methods.

THE US ARMY YUMA PROVING GROUND FOR SENSOR DEMONSTRATIONS

Hundreds of military and government sites are scheduled for range cleanup of buried UXOs. The US Army Yuma Proving Ground (YPG) is selected as a typical model for location and detection of buried UXOs with one or more airborne and ground based sensors. YPG has a wide collections of buried bombs, antitank/antipersonnel mines, artillery rounds, and more. These range in age from a few weeks to many years. Conditions at YPG are sufficiently variable to represent those found at other sites, so that if sensors are successful at YPG, they will respond at the other sites.

Many UXOs, mines, and similar debris at the US Army Yuma Proving Ground (YPG) in Arizona are located on or slightly below the surface, while larger UXOs like 500 pound and larger bombs are buried much deeper at Aberdeen Proving Ground (APG) in Maryland. UXOs, like mines and antitank explosives, at YPG are very sensitive to people and vehicles, exploding on contact. Areas like YPG has topographic features which are not amenable to people and conventional or unconventional vehicular traffic, and associated loss of vehicles and personnel precludes ground based sensors except at a safe and suitable distance from the area.

However, areas like APG are relatively inert to personnel and/or vehicular traffic. Even after 500 pound bombs have been detonated, resulting craters sometimes uncover UXO

bombs which were previously undetected.

Some ranges have combinations of surface, near surface, and deeply buried UXOs. The terrain is very dry sand, marshy soil, water, and mixtures of the latter. Variations in UXOs, depths of burial, and terrain indicate use of multisensor types on surface and airborne platforms.

ACOUSTIC IMAGING SONARS FOR DETECTING UXOs AND SIMILAR DEBRIS

It is highly useful to see in places where light cannot penetrate or doesn't exist. For this purpose, we can use x-rays, gamma rays, electron beams, microwaves, IR or optical, acoustic waves, and other forms of propagating energy. Acoustic waves are the most natural. For countless millennia, members of the animal kingdom have used these waves for imaging otherwise unobservable objects and inhomogeneities.

Echo-locating bats are the most interesting of nature's animals in acoustic imaging. With a unique ability to generate, hear, and process sound, they catch their meal in flight in total darkness. They make light chirping sounds with pitches too high for human detection. They process the echoes of this sound from nearby obstacles to avoid collisions or each other and to catch insects on the wing. Sound gives bats mental pictures of their surroundings. With specialized larynges, sensitive ears, and audio cortices, bats can easily and quickly find their ways through furniture in dark rooms and even through ropes and strings strung across them.

The analogy between animal sonar is made because of the similarity to man-made radar and sonar [1]. Both animal sonars and "range cleanup" sonars are continuously

looking for targets. Bat sonars are used for orientation and the detection of prey. Since the purpose of the signal is to detect a target regardless of its velocity relative to the bat, in radar literature this signal is called "Doppler tolerant" [2] - [4], where a Doppler distorted signal is still recognized by a single correlation process as a reflected version of the transmitted waveform regardless of the reflector's velocity.

The output of a correlation receiver is one version of a sonar wideband ambiguity function [5], which can be used to derive some optimally Doppler tolerant waveforms. Sonar for general target descriptions [6] - [7] will also be studied.

Surface and low altitude airborne platforms for acoustic sources will be fabricated at Phillips Laboratory, Kirtland AFB. Surface platforms for acoustic sources will be located at the surface of a fresh water tank, made by digging a rectangular hole in the sand near the Trestle facility (the other sensors will also be tested here) at Kirtland AFB, with depth 8 feet and surface area 10 X 14 ft. Walls will be concrete block, lined with rubber, and the bottom will be one inch rubber sheet. In addition to fresh water, sea water, and several mixtures of wet soil (different amounts of water from 100% mud to very dry sand) will be used as a sonar propagation medium [8]-[10]. Targets to simulate bombs, artillery projectiles, and other UXOs. less explosives) will be located on the bottom of the tank.

Airborne platforms will located on wooden towers above the preceding tanks, so that comparisons may be made on the effects on the interfaces between ground and air on acoustic wave reflection losses. Analytical models will be developed to design acoustical antennas [11], and transformers for transitions between different media [12] will be fabri-

cated to increase coupling between air, soil, and water. Models for propagation through water [13] and soil-rock [14] will be used.

GROUND PENETRATING IMAGING RADARS FOR DETECTING UXOs, MINES, BOMBS, AND SIMILAR DEBRIS

Ground penetrating radar measurements and images will be made with airborne and the Trestle platforms. The airborne platforms are similar to those afforded by slow moving aircraft like the Otters, often utilized for microwave radar imagers. As a ground based tower, the Trestle may be considered as a supplemental platform for sensing the terrain. The term multi-sensor is descriptive because IR and microwave imagers will be flown, similar to those used in the Spaceborne Imaging Radar (SIR), on flights over the Eastern Sahara with SIR-A [15] - [16] and over Al Labbah Plateau in Saudi Arabia with SIR-B [17] - [19].

The reference to the Al Labbah Plateau is made because of similarities in the terrain with that of Yuma. Similar penetration of the Yuma terrain will be evaluated in experiments at the top of the towers at the Trestle or similar facility at PL/WSR, KAFB NM. The targets will be mines (inert) from YPG, assorted ordnance debris from YPG, and aircraft fuel tanks (cigar shapes) filled with water to simulate bombs. These targets will be on the surface and buried in shallow depths of 0.5, 1.0, 1.5 and 2.0 meters. They will be undisturbed for a few days to establish thermal equilibrium with the surrounding terrain.

Newer versions of SIR, multifrequency multipolarization synthetic aperture radars (SAR) with variable image geometry, are being developed as part of the Space Station's

Earth Observation System (EOS), for the launching of the SIR-C and SIR-D planned for the early 1990's. These SARs will operate in L-band, C-band, and X-band, with quadpolarization available for the L-band and C-band. Multiincidence data will be acquired using electronic beam steering and other imaging and other imaging geometries by mechanically pitching, yawing, and rolling the antenna. The present SIR-C design includes a bandwidth of 20 MHz, to provide better resolution than that available with four looks, and a 10 MHz bandwidth for low resolution mode which will increase the swath width for a given data rate.

The reference to SIR is made because of Dr. Biggs' association with this program during his work with the Remote Sensing Laboratory at the University of Kansas. It is also made because of the subsurface imaging achieved in flights over sand buried landscape at Al Labbah Plateau in Saudi Arabia. Penetration of 1.24 meters through the sand sheet surrounding the deeper eolian deposits at Anbat and Irq at Ubaytir was easily achieved with the 1.3 GHz (23.5 cm wavelength) SIR imaging radars. Some reflections (a dark tone) were found at depths of 3.1 meters and 4.7 meters.

With metallic targets and use of low flying aircraft instead of high orbiting satellites (distance ratio of at least 1,000, and with the radar range equation varying inversely as the fourth power of distance, a power gain of 1,000,000,000,000, or 120 dB), much better resolution and penetration is achieved.

Another radar system is similar to that used by Stanford Research Institute (SRI) at the Chaco Canyon National Monument in New Mexico. The SRI radar has a center frequency of 400 MHz and a resolution of about 30 cm (one foot). In addition to locating

Indian ruins dating back to 600 AD, the SRI radar was successful in locating the docks of the Pharaohs when they made their journeys to the land of the dead.

A subcontract will be given to SRI for radar imaging flights over terrain with known UXO targets. SRI has its own airplane and ultra wideband synthetic aperture radar (SAR), known as SRI UWBR. The SRI radar is a single frequency, HH polarization impulse radar that is flown on a Beech Queen Air airplane. Four center frequencies, 200 MHz, 300 MHz, 350 MHz, or 400 MHz, are available with a single pulse (5 ns wide) for a 200 MHz bandwidth. Antennas are formed by a linear array of dipoles under each wing, one side for transmitting and the other side for receiving, and this gives a depression angle between 30 degrees and 60 degrees. A swath width of about 400 meters was recorded for flights.

Flights of the SRI UWBR radar will be made over buried UXOs in same environment as that for IR and acoustic imagers (near Trestle at Kirtland AFB). The SLAR, with SAR processing, will be obtained with a subcontract from SRI, through Dr. Roger Vickers. During the third year of imaging, demonstrations flights will be made with SRI radars at YPG, APG, and Naval offshors ranges.

Images will be made early in the morning, noon, and early evening periods. This will allow unequal thermal diffusion to occur so that maximum contrast will be achieved between the "dummy" ordnance and the New Mexico terrain to compare results with IR images.

Negotiations will be made for flights with the NASA/JPL AIR radar, which is a multiple frequency, multiple polarization instrument, operating from a NASA DC-8 airplane.

The NASA/JPL AIR radar is a SAR with narrow frequency bands, UHF-band, L-band, and C-band, with center frequencies at 439 MHz, 1.25 GHz, and 5.3 GHz, respectively. Each band allows for transmission in either horizontal or vertical polarization, and the SAR can receive in both horizontal and vertical polarization. This permits multi-polarization recording of HH, HV, VH, and VV. Each band also operates with a 38 MHz bandwidth and gives a range and azimuth resolution of approximately 4 meters by 4 meters. The radar allows look angles from 30 degrees to 70 degrees. Swath widths of about 3.5 km have been flown.

Negotiations will also be made for flights a year later with the NAWC-AD P3 radar, which is a multiple frequency, multiple polarization instrument, operating from a U.S. Navy P-3 airplane. The NAWC-AD P3 radar is a SAR with narrow frequency bands, UHF-band, L-band, and C-band, with center frequencies at 1.25 GHz, 5.25 GHz, and 9.38 GHz, respectively. Like the NASA/JPL AIR radar, the NAWC-AD P3 SAR can record multi-polarization recording of HH, HV, VH, and VV. Each band also operates with a 120 MHz bandwidth and gives a range and azimuth resolution of approximately 1.5 meters by 1.5 meters. The radar allows look angles from 30 degrees to 70 degrees. Swath widths of about 3.5 km have been flown.

Dr. Ray Smith at the U.S. Army Missile Command (MICOM), Advance Sensors Directorate, will develop adaptive deep soil models which will measure soil profiles with UXOs so that their transforms can be convoluted with those of known UXOs until the actual UXO is identified. It is adaptive use of clutter in radars. These results will be correlated with those of Tayfun Ozdemir [20] for shallow targets, Take Abe [21] - [23] for targets in snow, Martti Hallikainen [25] - [26] for the behavior of wet soil at microwave frequencies.

INFRARED (IR) IMAGING FOR DETECTING UXOs, MINES, AND SIMILAR DEBRIS

In this section, tower and airborne platforms for IR imaging are planned. When many of the UXOs, mines, and similar debris are located on or slightly below the surface, the IR imagers will be mounted on ground based towers for evaluation of their suitability for actual use in slow moving aircraft like the Otters. Both IR and microwave imagers will be flown in actual range cleanup, similar to those used in the Shuttle Imaging Radar (SIR) on flights over the Al Labbah Plateau, in Saudi Arabia [15] - [19].

The IR imager to be used at the Trestle facility at Kirtland Air Force Base (KAFB) will require a liquid nitrogen bath for cooling. It is a multi-spectral IR imager with five bands between one and five microns, on loan from the Electronic Sciences Division (ESD), RADC/EA, Hanscom AFB MA, through Dr. Paul Pelligrini. The ESD IR multispectral imager is similar to the Landsat-5 Thematic Mapper (TM) color-infrared imager with Bands 2, 3, 4, and 6 used in the Shuttle Imaging Radar (SIR) on flights over the Al Labbah Plateau, in Saudi Arabia.

The reference to the Al Labbah Plateau was made because of similarities in the terrain with that of Yuma. Similar penetration of the Yuma terrain will be evaluated in experiments at the top of the towers at the Trestle or similar facility at PL/WSR, KAFB NM. The targets will be mines (inert) from YPG, assorted ordnance debris from YPG, and aircraft fuel tanks (cigar shapes) filled with water to simulate bombs.

These targets will be on the surface and buried in shallow depths of 0.5, 1.0, 1.5 and 2.0 meters. They will be undisturbed for a few days to establish thermal equilibrium with

the surrounding terrain. Images will be made early in the morning, noon, and early evening periods. This will allow unequal thermal diffusion to occur so that maximum contrast will be achieved between the "dummy" ordnance and the New Mexico terrain (noon measurements are only made to provide a contrast).

The mention of SIR is also made because of Dr. Biggs' association with this program during his work with the Remote Sensing Laboratory at Kansas University. In addition to subsurface penetration of the multispectral TM in SIR flights, he participated [31] in using Dutch IR imagers in flights over the polders in Holland in order to detect fresh water streams (the use of excessive fertilizer had made water undrinkable for cows). These streams were on the surface two thousand years ago, but were gradually covered with silt and soil to a depth of a meter. The temperature difference between underground streams and soil was sufficient to discriminate the streams, which appeared as waving paths on the IR images.

ELECTROMAGNETIC INDUCTION METHODS FOR DETECTING UXOs, MINES, BOMBS, AND SIMILAR DEBRIS

The arguments in favor of electromagnetic induction methods for the detection of unexploded ordnance or UXOs are that the ordnance is buried to substantial depths, with random orientation, and has been in place for an extended period of time. The use of methods that do not depend on changes in surrounding soil structure and have sufficient soil penetration indicates electromagnetic induction methods with extremely low frequencies. Excellent work in measuring the conductivity of metallic cylinders [32] - [34], and both

analytical studies and experimental verifications at the Defense Research Establishment Suffield in Medicine Hat, Alberta, Canada, by Yogadhish Das, John McFee, Robert Chesney and others [35] - [38], expresses the use of electromagnetic induction methods for UXO detection. Other work by David Hill [39], Misse Nabighian [40], Lotfollah Shafai [41], Shri Krishna Singh [42], Kenneth Spies [43], and numerous papers by James Wait indicate the interest in these sensors.

Dr. H. Flathe, Federal Institute for Geosciences, Germany, has several papers [44] - [50] on determination of the earth's resistivity by surface measurements. He and Dr. Carl Baum have mutual interests in using extremely low frequency (ELF) techniques for detection of UXOs and in other research activities. Dr. Flathe has both analytical and experimental capabilities as Chief Geologist for Germany.

Froelich Rainey, director of the University Museum as well as the Museum Applied Science Center for Archeology (MASCA) at the University of Pennsylvania, is one of the pioneers in introducing scientific and technological methods into archeology. He describes [51] use of a ground penetrating radar by Stanford Research Institute (SRI) at the Chaco Canyon National Monument in New Mexico. The SRI radar has a center frequency of 400 MHz and a resolution of about 30 cm (one foot). In addition to locating Indian ruins dating back to 600 AD, the SRI radar was successful in locating the docks of the Pharaohs when they made their journeys to the land of the dead.

Dr. Rainey also describes the use of magnetometers in the search for the old Greek city of Sybaris, in what is now Calabria in southern Italy. Attempts to find this famous city

have been continuing for more than a century. The ancient surface of the plain is now deeply buried beneath silt washed down from the surrounding mountains and the courses of the rivers have altered many times during the centuries. The use of a cesium magnetometer, with the help of Sheldon Breiner of Varian Associates. detected buried walls and roof tiles at depths of 6 meters.

The same cesium magnetometer was used by the US Army to locate buried munitions in Viet Nam. The use of Varian and SRI highly sensitive magnetometers, known as a magnetic anomaly detection (MAD) system, can find perturbations in the earth's magnetic field. An alternative active system can be used where a low frequency source produces a CW electromagnetic (mostly magnetic) field that penetrates the soil and is scattered by buried ordnance. Dr. Baum, Phillips Laboratory, is actively developing analytical models in terms of Method of Singular Expansions (MSE) at ELF, to support the measurements made by SRI and Varian and Dr. Rainey. The scattered signals from the MSE resonances are detected by one or more highly sensitive magnetometers.

For either active or passive systems, outputs of the magnetometers are fed to signal processing systems that translates the array outputs into inferences regarding positions and types of ordnance scatterers. Analytical models at SRI exist to permit this translation from magnetic field array readings. The MAD system can be implemented in a ground based version or it can be mounted below a helicopter.

REFERENCES

[1] Richard A. Altes and E. L. Titlebaum, "Bat signals as optimally doppler tolerant waveforms," **J. Acoustic Soc. Am.**, vol. 48, pp. 1014-1020, Oct. 1970.

[2] P.M. Woodward, **Probability and Information Theory with Applications to Radar**, Pergamon Press, Oxford, 2nd ed., pp.68-80, 1964.

[3] J.J. Kroszczynski, "Pulse compression by means of linear-period modulation," **Proc. IEEE**, vol. 57, pp. 1260-1266, Aug. 1969.

[4] S.A. Kramer, "Doppler and acceleration tolerances of high-gain, wideband linear FM correlation sonars," **Proc. IEEE**, vol. 55, pp. 627-636, April 1967.

[5] Richard A. Altes, "Some invariance properties of the wide-band ambiguity function," **J. Acoustic Soc. Am.**, vol. 53, pp. 1154-1160, April 1973.

[6] Richard A. Altes, "Sonar for generalized target description and its similarity to animal echolocation systems," **J. Acoustic Soc. Am.**, vol. 59, pp. 97-105, Jan. 1976.

[7] A. Freedman, "A mechanism of acoustic echo formation," **Acustica**, vol. 12, pp. 10-21, 1962.

[8] James Ott and James Rice, "Sonar microwave-antenna simulator," **IEEE Trans. Antennas Propagat.**, vol. AP-34, pp. 1418-1424, Dec. 1986.

[9] Richard K. Moore and B.E. Parkins, "Omnidirectional scattering of acoustic waves from rough surfaces of known statistics," **J. Acoustic Soc. Am.**, vol. 40, pp. 170-175, July 1966.

[10] John Rouse and R.K. Moore, "Measured surface spectrum dependence of backscattering from rough surface," **IEEE Trans. Antennas Propagat.**, vol. AP-20, pp. 211-214, Mar. 1972.

[11] David I. Paul, "Acoustical radiation from a point source in the presence of two media," **J. Acoustic Soc. Am.**, vol. 53, pp. 1154-1160, Dec. 1976.

[12] Albert W. Biggs, "Slow wave transmission Line transformers," **IEEE Trans. Electromagn. Compat.**, vol. 34, pp. 242-251, Aug. 1992.

[13] Jerry L. Sutton, "Underwater acoustic imaging," **IEEE Proc.**, vol. 67, pp. 554-566, April 1979.

[14] I. A. Chaban, "Sound attenuation in sediments and rock," **Acoust. Phys.**, vol. 39, pp. 190-193, Mar.- Apr. 1993.

[15] John F. McCauley et al, "Subsurface valleys and geoarcheology of the Eastern Sahara revealed by shuttle radar," **Science**, vol. 218, pp. 1004-1020, 3 Dec. 1982.

[16] John F. McCauley et al, "Paleodrainages of the Eastern Sahara--The radar rivers revisited (SIR-A/B implications for a mid-tertiary trans-African drainage system)," **IEEE Trans. Geosci. Remote Sensing**, vol. GE-24, pp. 624-648, July 1986.

[17] Charles Elachi et al, "Spaceborn radar subsurface imaging in hyperarid regions," **IEEE Trans. Geosci. Remote Sensing**, vol. GE-22, pp. 383-388, July 1984.

[18] Gerald G. Schaber et al, "Shuttle imaging radar: Physical controls on signal penetration and subsurface scattering in the Eastern Sahara," **IEEE Trans. Geosci. Remote Sensing**, vol. GE-24, pp. 603-623, July 1986.

[19] Graydon L. Berlin et al. "SIR-B subsurface imaging of a sand-buried landscape: Al Labbah Plateau, Saudi Arabia." **IEEE Trans. Geosci. Remote Sensing**, vol. GE-24, pp. 595-602, July 1986.

[20] Tayfun Ozdemir et al. "Imaging of shallow subsurface objects: An experimental investigation," **IEEE Trans. Geosci. Remote Sensing**, vol. GE-30, pp. 472-481, May 1990.

[21] Glenn S. Smith and Waymond R. Scott, "A scale model for studying ground penetrating radars," **IEEE Trans. Geosci. Remote Sensing**, vol. GE-27, pp. 358-368, July 1989.

[22] Takeo Abe et al. "Experimental study of microwave transmission in snowpack." **IEEE Trans. Geosci. Remote Sensing**, vol. GE-28, pp. 915-921, Sep. 1990.

[23] Yoshio Yamagutchi et al, "Detection of objects buried in wet snowpack by an FM-CW radar," **IEEE Trans. Geosci. Remote Sensing**, vol. GE-29, pp. 201-207, Mar. 1990.

[24] Luen C. Chan et al, "A characterization of subsurface radar targets," **Proc. IEEE**, vol. 67, pp. 991-1000, July 1979.

[25] Martti Hallikainen et al, "Microwave behavior of wet soil--Part I: Empirical models and experimental observations." **IEEE Trans. Geosci. Remote Sensing**, vol. GE-23, pp. 25-34, Jan. 1985.

[26] Myron C. Dobson et al, "Microwave behavior of wet soil--Part II: Dielectric mixing models." **IEEE Trans. Geosci. Remote Sensing**, vol. GE-23, pp. 35-46, Jan. 1985.

[27] Merrill I. Skolnik, **An Introduction to Impulse Radar**, NRL Memorandum Report 6755, Naval Research Laboratory, Washington DC, 20 Nov. 1990.

[28] D. J. Daniels, D. J. Gunton, and H. F. Scott, "Introduction to subsurface radar," **IEE Proc.**, vol. 135, pt. F, pp. 278-319, Aug. 1988.

[29] G. K. A. Oswald, "Geophysical radar design," **IEE Proc.**, vol.135, pt. F, pp. 371-379, Aug. 1988.

[30] D. M. McCann et al, "Comparison of the seismic and ground probing radar methods in geological surveying," **IEE Proc.**, vol.135, pt. F, pp. 380-392, Aug. 1988.

[31] Albert W. Biggs, "Radar and Infrared Remote Sensing of Terain, Water, Arctic Sea Ice, and Agriculture," **32nd Symposium of EWPP/AGARD, NATO, Oberammergau, Germany**, May 24-28, 1983.

[32] R.C. Callarotti and M. Alfonzo, "Measurement of the conductivity of metallic cylinders by means of an induction method," **J. Appl. Phys.**, vol. 43, pp. 3040-3047, July 1972.

[33] C.V. Dodd and W.E. Deeds, "Analytical solutions to eddy-current probe-coil problems," **J. Appl. Phys.**, vol. 39, pp. 2829-2838, May 1968.

[34] J.W. Luquire, W.E. Deeds, and C.V.Dodd, "Axially symmetric eddy currents in a spherical conductor," **J. Appl. Phys.**, vol. 41, pp. 3976-3982, Sep. 1970.

[35] Yogadhish Das, John McFee, and Robert Chesney, " Time response of a sphere in the field of a coil: Theory and experiment," **IEEE Trans. Geosci. Remote Sensing**, vol. GE 22, pp. 360-367, July 1984.

[36] Yogadhish Das, John McFee, and Robert Chesney, " Determination of depth of shallowly buried objects by electromagnetic induction, **IEEE Trans. Geosci. Remote Sensing**, vol. GE 23, pp. 60-66, Jan. 1985.

[37] Yogadhish Das, John McFee, Jack Toews, and Gregory Stuart, "Analysis of an electromagnetic detector for real-time location of buried objects," **IEEE Trans. Geosci. Remote Sensing**, vol. GE 28, pp. 278-287, May 1990.

[38] J. McFee, H. Chesney, Y. Das, J. Toews, M. Turnbull, R. Pennell, and M. Bell, "Experimental time domain electromagnetic induction system," **Rev. Sci. Instruments**, vol. 55, pp. 968-975, June 1984.

[39] David A. Hill, "Near-field detection od buried dielectric objects," **IEEE Trans. Geosci. Remote Sensing**, vol. GE 28, pp. 278-287, May 1990.

[40] Misse N. Nabighian, "Quasi-static transient response of a conducting permeable sphere in a dipolar field," **Geophysics**, vol. 35, pp. 303-309, April 1970.

[41] A.A. Sebak and Lotfollah Shafai, "Transient response computation of spheroidal objects using impedance boundary conditions," **IEEE Trans. Antennas Propagat.**, vol. AP 32, pp. 1116-1122, Oct. 1984.

[42] Shri Krishna Singh, "Electromagnetic transient response of a conducting sphere embedded in a conductive medium," **Geophysics**, vol. 38, pp. 864-893, Oct. 1973.

[43] J. Wait and Kenneth P. Spies, "Quasi-static response of a conducting permeable sphere," **Geophysics**, vol. 34, pp. 789-792, Oct. 1969.

[44] H. Flathe, "^The role of a geological concept in geophysical research work in solving hydrogeological problems," **Geoexploration**, vol. 14, pp. 195-206, 1976.

[45] F. Bender and H. Flathe, "Groundwater conditions in the Chaco Boreal of Paraguay," **Publ. Assoc. Int. Hydrol. Scient. Gentbrugge, Germany**, vol. 57, pp. 173-185, 1961.

[46] H. Flathe, "A practical method of calculating geophysical model graphs for horizontally stratified media," **Geophys. Prospect.**. vol. 3, pp. 268-295, 1955.

[47] H. Flathe, "Geoelektrische Untersuchsung der Grundwasserver salzung im sudlicher Jordantal." **Geol. Jb., Hannover, Deutschland**, vol. 85, pp. 767-782, 1968.

[48] H. Flathe, "Interpretation of geoelectrical resistivity measurements for solving hydrogeological problems," **Minutes of Groundwater Geophysics-1967, Ottawa, Econ. Geol. Rep.**, vol. 26, pp. 580-597, 1970.

[49] H. Flathe, "The role of a geologic concept in geophysical research work for solving hydrogeological problems." **Geoexploration, Amsterdam, Netherlands**. vol. 14, pp. 195-206, 1976.

[50] H. Flathe und D. Pfeiffer, "Grundzuge der Morphologie, Geologie und Hydrogeologie im Karstgebiet Gunung Sewu/Java (Indonesien)," **Geol. Jb., Hannover, Deutschland**, vol. 83, pp. 533-562, 1965.

[51] Froelich Rainey, "New tools for an old art," **IEEE Spectrum**, pp. 39-43, Sep. 1976.

AN ANALYSIS APPROACH TO DETERMINE QUALITY FACTORS OF LARGE, COMPLEX CAVITIES

Ronald R. DeLyser
Assistant Professor of Electrical Engineering
Department of Engineering

University of Denver
2390 S. York Street
Denver, CO 80208-0177

Final Report for:
Summer Faculty Research Program
Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September, 1993

AN ANALYSIS APPROACH TO DETERMINE QUALITY FACTORS OF LARGE, COMPLEX CAVITIES

Ronald R. DeLyser
Assistant Professor of Electrical Engineering
Department of Engineering
University of Denver

Abstract

An analysis approach to determine quality factors (Q) of large, complex cavities is proposed. It consists of looking at apertures and other objects as individual sources of cavity Q. These sources of Q contribute to the total Q linearly if they are not in close proximity to each other, cavity walls, corners or edges. Analysis of the individual sources is done assuming that the energy density in the cavity is uniform and the fields away from the walls are completely random so that the fields incident on the walls can be viewed as a composite of several local plane waves of arbitrary polarization. The technique is illustrated analytically with large apertures, both loaded and unloaded, small apertures, and a large area of ferrite absorber. A discussion on using existing computer programs and how they can be modified to calculate Q for other classes of objects is also included.

AN ANALYSIS APPROACH TO DETERMINE QUALITY FACTORS OF LARGE, COMPLEX CAVITIES

Ronald R. DeLyser

Introduction

In [1] a technique to predict the statistical distribution of the electric fields in a complex cavity is described. Statistical modeling of mode stirred chambers are presented in [2] and [3]. As noted in [4], these three reference all come to the same basic conclusion, i.e., the probability density functions for one and three dimensional fields in mode-stirred chambers and large, complex cavities are chi-squared distributions. Further, the estimate of the variance of each of the field components is directly proportional to the cavity quality factor, Q. Thus the normalized accuracy for Q is the same as that for the estimate of the variance. Knowledge of chamber Q then allows prediction of the mean field strength from the input power.

How then, does one go about calculating the Q of a large, complex cavity? One way is to do it by using any of a number of computer codes using numerical schemes derived from the Method of Moments (MoM) [5], the Finite Element Method (FEM) [6], the Geometrical Theory of Diffraction (GTD) [7],or a combination of all three [8]. These can be valuable tools for specific, known systems. However, sometimes the exact nature of the system is not known, but there may be a general "generic" description. Again, we can use the numerical methods, but other approximate methods may work just as well. This is the subject of this paper.

The basis of the method presented here can be found in [9]. In that paper, the fields inside a mode-stirred chamber are analyzed by a method where the cavity wall is assumed to be locally planar. The energy density in the cavity is assumed uniform and the fields away from the walls are completely random so that the fields incident on the walls can be viewed as a composite of several local plane waves. The analysis proceeds with the classical problems of perpendicular and parallel polarized plane waves incident on a dielectric-conductor interface. The fields are derived for the dielectric region and then the magnitude squared of these fields are averaged over all possible angles of incidence. The power absorbed by the finitely conducting walls (P_{diss}) and the energy density (U) stored in the cavity is then calculated. The Q of the cavity is given by

$$Q=\frac{\omega UV}{P_{diss}} \tag{1}$$

where ω is the angular frequency and V is the cavity volume. The Q for a rectangular cavity was found in this way and agreed well (to the lowest order in k_0L, where k_0 is the wave number in free space and L is a characteristic

length of the chamber) with a theoretical expression derived using "global" modes [10].

This approach of analyzing the fields at a dielectric-conductor interface can be extended to analyses of general classes of locally planar conducting screens with apertures, loaded apertures (dielectrics, resistive screens, and meshes, etc.), slots, lossy surface materials, etc. In this manner, one can generate values for Q for each of these problems. Complex cavities with a number of these classes of problems would then have a composite Q of

$$\frac{1}{Q_{tot}} = \frac{1}{Q_1} + \frac{1}{Q_2} + \frac{1}{Q_3} + \cdots \tag{2}$$

where Q_{tot} is the composite Q, and Q_1, Q_2, Q_3 ..., are individual Q's associated with each type of problem. (*Note:* equation (2) can only be applied when the individual Q's are independent of each other, i.e. the objects are not too close to each other nor do they otherwise interact with each other in a non-linear fashion. Also, the objects should not be too close to a wall, corner, or edge of the cavity.)

Examples of calculating Q's for a large aperture, a large aperture loaded with a dielectric sheet, a ferrite material attached to the wall, and two different types of small apertures are given in the following sections. All values for Q are compared to the value of that calculated due to wall losses alone. The examples presented here are of the type where analytical methods can be employed. Numerical methods will have to be used for other classes of problems where the dimensions of the object are on the order of a wavelength. These and other types of problems are discussed in the final section of this paper.

Quality Factor Due to Wall Losses

In order to compare Q's due to the different types of problems presented, the Q due to wall losses will be calculated first. A standard size cavity where the volume to internal surface area ratio (V/S) of $L/6$ (L is a characteristic length) is chosen. This ratio is valid for a sphere with a diameter L, a cube with side length L, and a cylinder with height and diameter equal to L. L is to be large when compared to the wavelength of radiation, λ. We choose $L = 18\,\lambda$ so that V/S is $3\,\lambda$.

From [9], the Q for a large multi-mode cavity of volume V and surface area S is given by

$$Q = \frac{3V}{2S\mu_r\delta_s} \tag{3}$$

where μ_r is the relative permeability of the cavity wall. The skin depth of the cavity wall, δ_s, is given by

$$\delta_s = \sqrt{\frac{2}{\omega \mu \sigma}} \tag{4}$$

where μ and σ are the permeability and conductivity of the cavity wall, respectively. The Q of a reference cavity made with copper walls is calculated to be 6.433×10^5.

The Large Aperture

The large aperture is assumed to have a characteristic dimension which is large with respect to the wavelength but small with respect to the total surface area of the cavity so that the fields are not significantly perturbed from their distribution when the cavity is closed. Since for our standard cavity $S = 6L^2$ and $L = 18\lambda$, then $S = 1944\lambda^2$. If we choose an aperture area of $100\lambda^2$, these criteria should be satisfied. However, there is a dilemma which gives rise to two different ways to proceed. We will continue with one way and address this dilemma in sequence.

The geometry for a plane wave obliquely incident on the plane of the aperture is shown in Figure 1. The incident angles, θ and ϕ, are the usual spherical coordinate system angles, and the polarization angle, τ, is the angle between the electric field vector and the line segment OA' which is in the x-y plane and perpendicular to the line segment OA which is the projection of the propagation vector on the x-y plane. This general system is chosen so that later in this development small apertures of arbitrary polarization can be addressed. In this coordinate system the fields can be expressed as

$$\left.\begin{aligned} \mathbf{E}(x,y,z) &= \mathbf{E}_0(\theta,\phi \quad e^{jk_0 r} \\ \mathbf{H}(x,y,z) &= \frac{1}{\eta_0}\mathbf{E}_0(\quad\quad 2) e^{jk_0 r} \end{aligned}\right\} \tag{5}$$

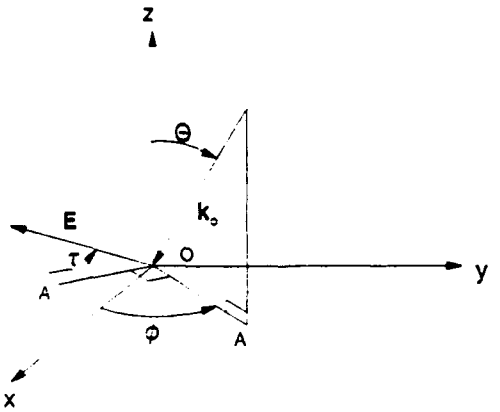


Figure 1. Coordinate system for oblique incidence of a plane wave.

where η_0 is the free space characteristic impedance and

$$\left.\begin{aligned} \mathbf{k_0} &= k_0(\sin\theta_i\cos\phi_i\,\hat{a}_x + \sin\theta_i\sin\phi_i\,\hat{a}_y + \cos\theta_i\,\hat{a}_z) \\ E_{0x}(\theta_i,\phi_i,\tau) &= E_0\,(\sin\phi_i\cos\tau - \cos\theta_i\cos\phi_i\sin\tau) \\ E_{0y}(\theta_i,\phi_i,\tau) &= -E_0\,(\cos\phi_i\cos\tau + \cos\theta_i\sin\phi_i\sin\tau) \\ E_{0z}(\theta_i,\phi_i,\tau) &= E_0\sin\theta_i\sin\tau \end{aligned}\right\} \tag{6}$$

where E_0 is the magnitude of the incident electric field.

The power exiting the aperture, S_A, is

$$P(\theta) = \frac{1}{2}\,\Re\int_{S_A}\mathbf{E}\times\mathbf{H}^*\cdot d\mathbf{S} = \frac{1}{2}\int_{S_A}\frac{E_0^2}{\eta_0}\cos\theta_i\,dS \tag{7}$$

where $\Re$ means "take the real part of". The result is independent of ϕ_i and the polarization angle τ as expected. Since the field is constant across the aperture, we have

$$P(\theta) = \frac{1}{2}\,\frac{E_0^2}{\eta_0}\cos\theta_i\,S_A\,. \tag{8}$$

The power is now averaged over all angles of incidence and polarization. In general, the average is given by

$$\langle P\rangle = \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}\int_0^{\frac{\pi}{2}} P\,(\theta_i,\phi_i,\gamma)\,\sin\theta_i\,d\theta_i\,d\phi_i\,d\tau\,. \tag{9}$$

Since the power exiting the aperture is independent of ϕ_i and τ, we have

$$\langle P\rangle = \int_0^{\frac{\pi}{2}} P(\theta_i)\,\sin\theta_i\,d\theta_i = \frac{1}{2}\,\frac{E_0^2}{\eta_0}\,S_A\int_0^{\frac{\pi}{2}}\cos\theta_i\,\sin\theta_i\,d\theta_i \tag{10}$$

so that the total averaged power exiting the aperture is

$$\langle P\rangle = \frac{1}{4}\,\frac{E_0^2}{\eta_0}\,S_A\,. \tag{11}$$

The Q of the cavity due only to the presence of the large aperture can be calculated with equation (1). Using a procedure similar to that in [9], the average energy density in the closed cavity is determined to be

$$U = 2\,\epsilon_0\,E_0^2 \tag{12}$$

where ϵ_0 is the free space permittivity. So, substitution of equations (11) and (12) into equation (1) gives, after a little manipulation

$$Q = \frac{16\pi}{\lambda}\frac{V}{S_A} \quad . \tag{13}$$

For our reference cavity, $V = L^3 = 5.832\ \lambda^3$ so that $Q = 2{,}932$. Compared to the Q due to the wall losses alone, the Q due to the large aperture will dominate the calculation for the total Q as expected.

Now we address the aforementioned dilemma. The above derivation is based on the intuitive notion that the fields at the aperture are merely due to plane waves traveling toward the aperture. However, the assumption that the fields are not significantly perturbed from those of the closed cavity would dictate that the power density at the aperture is the same as the power density in the center of the cavity. That is related to the energy density in the cavity (equation (12)) by

$$P_A = c\,U = 2\,\frac{E_0^2}{\eta_0} \tag{14}$$

where $c = 1/(\mu\epsilon_0)^{1/2}$ is the phase velocity in the cavity. Only half of this power density is propagating toward the aperture. Comparing equations (14) and (8) (the cosine term takes into acount the geometrical optics approximation of the aperture), the power density propagating toward the aperture assuming that it is equal to the power density in the cavity is twice as great as that calculated assuming only plane waves traveling toward the aperture. This second approach is taken in reference [11][1] and predicts a Q half as great as predicted above. Even though this second approach is consistent with the assumptions, it is equally inconsistent with intuition. Where does reality lie? This author believes that it lies between the two extremes.

The Large Aperture with a Dielectric Window

Now, let's cover the aperture with a dielectric of thickness d_1. The only difference in equation (10)[2] for

[1] The approach to calculating Q is also very similar to the one taken here. The author discovered this report on 26 August 1993, beforeits final publication and a number of weeks after the initial draft of the theoretical portions of this paper had been distributed internally.

[2] We continue based on the intuitive approach keeping in mind that the dilemma presented above would dictate twice the power transmitted through the aperture.

this covered aperture is a factor to account for the transmission through the dielectric. In order to use transmission coefficients previously derived [12], we look at the usual special polarizations of the electric field perpendicular and parallel to the plane of incidence. The transmission coefficient for this two layer geometry is

$$\left.\begin{aligned} T &= \frac{4\, e^{-jk_{1z}d_1}}{(1 + p_{01})(1 + p_{12})(1 + R_{01}R_{12}e^{-2jk_{1z}d_1})} \\ k_i &= \omega \sqrt{\mu_i \epsilon_i} \qquad k_{iz} = k_i \cos\theta_i \\ k_i \sin\theta_i &= k_j \sin\theta_j \end{aligned}\right\} \tag{15}$$

where

$$\left.\begin{aligned} R_{ij} &= \frac{1 - p_{ij}}{1 + p_{ij}} \\ p_{ij} &= \frac{\mu_i k_{jz}}{\mu_j k_{iz}} \text{ for TE waves, } p_{ij} = \frac{\epsilon_i k_{jz}}{\epsilon_j k_{iz}} \text{ for TM waves,} \\ i, j &= 0, 1, 2 \end{aligned}\right\} \tag{16}$$

is the reflection coefficient for electric fields for the TE case and the reflection coefficient for the magnetic fields for the TM case. Further, θ_i and θ_j are the angles of incidence and transmission for adjacent media i and j respectively; and k_i, μ_i, and ϵ_i are the wave number, permeability, and permittivity of region i respectively. We can now modify equation (10) by multiplying it by the average of the TE and TM transmission coefficients. The result is

$$\langle P \rangle = \frac{1}{4} \frac{E_0^2}{\eta_0} S_A \int_0^{\frac{\pi}{2}} (|T_\perp|^2 + |T_\parallel|^2) \cos\theta_i \sin\theta_i \, d\theta_i \tag{17}$$

Mathcad®4.0 was used to evaluate the Q due to a covered aperture the same size as the open aperture evaluated in the previous section. A dielectric slab with a dielectric constant of 4 and a thickness of .5 λ changed the cavity Q to 3.206, an increase of 9.38 % over the open aperture value. The same dielectric cover with thickness of .25 λ changed the cavity Q to 4.863, an increase of 66 %. This is reasonable as the .5 λ thickness is the optimum value for maximum transmission and the .25 λ is optimum for maximum reflection.

Other types of loading of the aperture can be analyzed. For example, if the aperture was covered with a metal mesh where the period of the mesh is small with respect to the wavelength, equivalent boundary conditions

EBCs) [13] can be used to derive the transmission coefficient [14] which would then be substituted for the parenthetical expression in equation (17). The Q of the cavity could then be studied for factors such as the fineness of the mesh, the size of the wire and, the insulation between the wires. A number of different types of periodic structures have been studied to develop EBCs (references [15], [16], [17] and [18] to name a few) which can be used to study the effects of different periodic loadings of the aperture.

Absorbing Material on the Cavity Wall

A simple extension of the analysis of dielectric covered aperture can be applied to a ferrite material backed by the cavity wall. The reflection coefficient for the two layer geometry is given by [12]

$$R = \frac{R_{01} + R_{12} e^{-j2k_1 d_1}}{1 + R_{01} R_{12} e^{-j2k_1 d_1}} \tag{18}$$

where further definitions are given in equations (16). Substitution of $1 - |R|^2$ for $|T|^2$ in equation (17) gives the average power density absorbed by the ferrite. Once more, Mathcad®4.0 is used for this evaluation and the resulting Q calculations. For a ferrite with the same area as the uncovered aperture, a thickness of 0.003 free space wavelengths, $\epsilon_r = 10$, and $\mu_r = 3 - j\,30$, the Q is calculated to be 6,795.

The Small Aperture

The effects of the aperture field for a small aperture positioned at the origin of the coordinate system shown in figure 1 can be expressed in terms of the equivalent magnetic and electric dipoles [19]. The expression for the far electric field is [20]

$$\mathbf{E}(r) = [-\omega\mu k\, \hat{a}_r \times (2\, \mathbf{p}_m) - \frac{k^2}{\epsilon}\, \hat{a}_r \times \hat{a}_r \times (2\, \mathbf{p}_e)] \frac{e^{-jkR}}{4\pi R} \tag{19}$$

where

$$\left.\begin{aligned} \mathbf{p}_m &= \bar{\alpha}_m \cdot \mathbf{H}_{sc} \\ \mathbf{p}_e &= \epsilon\, \alpha_e\, E_{scz}\, \hat{a}_z \end{aligned}\right\} \tag{20}$$

are the magnetic and electric moments, α_m and α_e are the magnetic and electric polarizabilities and H_{sc} and E_{scz} are the tangential components of the magnetic field, and the normal component of the electric field, respectively, at the aperture when the aperture is closed. For a perfectly conducting screen, these are merely twice the corresponding components of the incident fields so that

$$\mathbf{p}_m = p_{mx}\,\hat{a}_x = p_{my}\,\hat{a}_y \tag{21}$$

where

$$\begin{bmatrix} p_{mx} \\ p_{my} \end{bmatrix} = 2 \begin{bmatrix} \alpha_{mxx} & \alpha_{mxy} \\ \alpha_{myx} & \alpha_{myy} \end{bmatrix} \begin{bmatrix} H_x^i \\ H_y^i \end{bmatrix} \tag{22}$$

and

$$\mathbf{p}_e = 2\,\epsilon\,\alpha_e\,E_z^i\,\hat{a}_z \tag{23}$$

where the use of the superscript, i, represents the incident fields.

After all substitutions are made and vector products evaluated, the far electric fields are expressed as

$$\left.\begin{aligned} \mathbf{E}(r) &= [E_\theta(\theta,\phi)\,\hat{a}_\theta + E_\phi(\theta,\phi)\,\hat{a}_\phi]\,\frac{e^{-jkR}}{R} \\ E_\theta(\theta,\phi) &= -\frac{\omega\mu k}{2\pi}\,(p_{mx}\sin\phi - p_{my}\cos\phi) - \frac{k^2}{2\pi\epsilon}\,p_{ez}\sin\theta \\ E_\phi(\theta,\phi) &= -\frac{\omega\mu k}{2\pi}\,(p_{mx}\cos\theta\cos\phi - p_{my}\cos\theta\sin\phi) \end{aligned}\right\}. \tag{24}$$

The total power radiated is given by [21]

$$P_{Rad} = \frac{1}{2\eta}\int_0^{2\pi}\int_0^{\frac{\pi}{2}} \left[|E_\theta(\theta,\phi)|^2 + |E_\phi(\theta,\phi)|^2\right]\sin\theta\,d\theta\,d\phi \tag{25}$$

This expression must be evaluated for the electric and magnetic polarizations resulting from the fields at the aperture due to an incident plane wave[3]. The resulting expression in this case is a function of the incidence angle (both θ_i and ϕ_i) and the polarization angle, τ, and must be averaged over all angles of incidence and polarization. After a tedious but straightforward derivation, the total averaged power radiated by a small aperture is given by

[3] Since the square of the fields in equation (25) are proportional to the square of the incident fields, we could also do the derivation assuming that these values are the averaged values at the closed conducting screen. Doing the derivation this way gives the same result, as it should with any linear system.

$$\langle P_{Rad} \rangle = \frac{2k^3 \omega \epsilon_0 E_0^2}{9\pi} \left(\alpha_{mxx}^2 + \alpha_{mxy}^2 + \alpha_{myx}^2 + \alpha_{myy}^2 + \alpha_e^2 \right) \tag{26}$$

which is intuitively satisfying since, on the average, each polarization is weighted equally. Finally, evaluation of Q for the cavity due to a small aperture gives

$$Q = \frac{9\pi V}{k^3 \left(\alpha_{mxx}^2 + \alpha_{myx}^2 + \alpha_{mxy}^2 + \alpha_{myy}^2 + \alpha_e^2 \right)} \quad . \tag{27}$$

Some examples are in order. Consider a small circular aperture of diameter d which has polarizabilities given by

$$\left.\begin{aligned} \alpha_{mxx} &= \alpha_{myy} = \frac{d^3}{6} \\ \alpha_{myx} &= \alpha_{mxy} = 0 \\ \alpha_e &= \frac{d^3}{12} \end{aligned}\right\} . \tag{28}$$

For a diameter of 0.1 λ, the Q for the cavity with just one of these circular apertures is 1.064×10^{10}, much higher than that for the Q due to finite wall conductivity alone. If the diameter of the aperture is increased to 0.2 λ, the Q is reduced to a value of 1.662×10^8. This dramatic change is due to the fact that for the small circular aperture, the polarizabilities squared are proportional to the diameter of the aperture to the 6th power.

Next, consider a thin slot. The area of the slot is chosen to be the same as that of the circular aperture with a diameter of 0.1 λ. The slot's length to width ratio (l/w) is chosen to be 10. This gives a length of 0.28 λ. The polarizations of a narrow slot with its axis parallel to the y axis are [22]

$$\left.\begin{aligned} \alpha_{mxx} &= \alpha_e = \frac{\pi}{16} w^2 l \\ \alpha_{myy} &= \frac{\pi}{24} \frac{l^3}{\ln(4l/w) - 1} \\ \alpha_{myx} &= \alpha_{mxy} = 0 \end{aligned}\right\} . \tag{29}$$

The calculated Q for this case is 5.802×10^8.

Reference [22] gives values for many different types of small apertures as well as arrays of apertures, hatch

apertures, and apertures with impedance loading. Analytical expressions also exist in [23] for narrow slots with depth. The effects of all of these could be determined, but clearly, it will take a number of such apertures to overcome the effect on Q due to wall losses.

Other Apertures and Objects

One has to resort to the use of numerical methods for apertures and other objects with sizes not in the "small" or "large" categories. We can also extend the analysis to combinations of these object with each other, other objects, cavity walls, and cavity corners. The programs available at the Satellite Assessment Center (SAC) which are appropriate for the project are CARLOS-3D™ [24] (Code for Analysis of Radiators on Lossy Surfaces), a general purpose computer code calculating scattering from complex three-dimensional geometries using MoM; and GEMACS [8], which solves electromagnetic radiation and scattering problems using MoM and GTD for exterior analysis, and a Finite Difference (FD) formulation for solution for interior problems. These programs need modified post processing and/or excitation capabilities in order to solve problems similar to those presented above.

There are two types of problems that the computer program(s) should be able to solve: (1) the "full-up" problem, i.e., the full cavity plus aperture/object problem, and (2) the scattering problem for the planar surface-plus-object. In the first problem, the program(s) have to modified in order to calculate Q of the cavity. In the second problem, the program(s) have to be modified in order to calculate bistatic scattering for plane wave sources of arbitrary polarization and incidence angle. Further processing then has to be done to determine the scattered power averaged over all angles of incidence of the incoming wave, and the average energy density as a function of distance from the conducting screen. Assuming that the energy density is a constant not too far from the screen, it can be assumed to be constant and equal to that value throughout a large cavity. This energy density and the power dissipated or exiting the cavity can the be used to calculate the Q. In other words, the program(s) have to do numerically what has been done analytically above.

Once the capability exists to do these two types of problems, results from the second type of problem can compared to the full-up solution. An ideal alternative would be to do experiments instead of (or in addition to) the full-up numerical analysis and use those results to compare to the results of the approach presented here. Studies should be done to investigate the validity and the limits of this analysis approach. Some of the questions to be answered are:

1. What size of cavity is necessary for the "large cavity theory" to be valid?
2. What size of aperture is large enough to use the large aperture solution presented above?
3. What size of aperture is small enough to use the small aperture solution presented above?
4. How close can individual objects be and still have their contributions to Q combine linearly?

5. How far from a cavity wall, edge or corner should an object be so that it can be treated as a scatterer in unbounded space?

A description of the codes and how they should be modified to do these studies follows.

CARLOS-3D™ [24]

CARLOS-3D™ implements the MoM solution for fully arbitrary three-dimensional complex perfect electrically conducting (PEC) as well as fully or partially penetrable scatterers and is based on surface integral equations (SIE) spanning the entire external surfaces of the body and the internal boundaries between penetrable and PEC regions. The boundaries may be characterized by resistive sheets, magnetic sheets, or impedance boundary conditions (Leontovich boundary conditions.) The user has the option of selecting the following formulations: (1) the electric field integral equation (EFIE) for coated conductors, (2) the EFIE, magnetic field integral equation (MFIE), and combined EFIE/MFIE (CFIE) on closed conductors, (3) the PMCHW (after Poggio, Miller, Chu, Harrington, and Wu) on dielectric boundaries, and (4) the CFIE for combinations of open/closed geometries. The observables are mono- and bistatic scattering coefficients, far field calculations, and surface current computation.

CARLOS-3D™ can be used to model apertures and loaded apertures in an infinitely conducting screen, and material regions (of the variety described above) backed by or above an infinitely conducting screen, edge or corner. The post processing has to be modified to give the averaged scattered fields from plane waves of arbitrary incidence and polarization. A numerically efficient way of simulating the excitation to represent incident plane waves from all directions and for all polarizations needs to be investigated. (In a MoM formulation this amounts to modifying the excitation vector. The large, dense impedance matrix remains the same.) This would make computation of averaged scattered fields much simpler. Calculations of the averaged absorbed or transmitted power and the average energy density as a function of distance from the screen are then used to calculate the Q of a large cavity containing the object. This value should be compared to a full-up simulation and/or experiment of the cavity with the object. Size of the cavity should be varied in order to determine the bounds on the locally planar theory.

CARLOS-3D™ is not suitable for the full-up problem of calculating Q for large cavities with apertures or internal objects. However, a more general code from which CARLOS-3D™ was derived, CARLOS™, will soon be available to the SAC. This code can be used to model the cavity with apertures or other internal objects. The most significant difference between CARLOS-3D™ and CARLOS™ is that CARLOS™ is capable of calculating near fields, a must for internal problems. Since all of the fields can be calculated inside and outside the cavity, energy stored and power exiting or absorbed can be calculated giving the Q. Also, since the electric fields throughout the cavity are known, the statistical nature of the fields can be determined and compared to the theory presented in [1] - [4]. Studies as to how the cavity size and object size affect these quantities should also be done to bound the

statistical theory and the locally planar theory.

GEMACS [8]

The GEMACS program (version 5) supports the solution of electromagnetic radiation and scattering from complex perfectly conducting objects using MoM solution of the EFIE, MFIE, and CFIE formulation. It also uses GTD for large perfectly conducting and impedance surfaces and combines GTD and MoM (called MoM/GTD hybrid) for problems where small radiators near large scatterers are modeled. A FD solution is used for closed cavities which when combined with the MoM, GTD or MoM/GTD hybrid can be used for multiple region problems.

GEMACS can be best be used with little modification for the full-up problem of calculating cavity Q's for cavities with apertures, wires, conducting bodies, and dielectric and lossy dielectric bodies. The cavity can be excited internally with either dipole or point sources. The FD solution has as observables thin wire currents and cavity electric fields. The energy in the cavity can be calculated from the electric fields. The power into the cavity is determined by the power into the source(s) which is equal to the power dissipated by apertures, conducting bodies and lossy dielectrics. This information then can be used to calculate cavity Q. As with CARLOS™, since the electric fields throughout the cavity are known, similar studies can be done as presented above.

A further modification of GEMACS to include complex permeability for regions within the cavity will allow the modeling of ferrites in the cavity. The code can also be used to calculate scattering from apertures in a conducting screen. However, as noted above, CARLOS-3D™ can handle this problem as well as other related problems of loaded apertures and regions of layered materials backed by a conducting screen.

Test and Measurement Considerations

At this point, the statements describing the driving force behind this project should be repeated. The probability density functions for one and three dimensional fields in mode-stirred chambers and large, complex cavities have been shown to be chi-squared distributions. The estimate of the variance of each of the field components is directly proportional to the cavity quality factor, Q. Thus the normalized accuracy for Q is the same as that for the estimate of the variance. Knowledge of chamber Q then allows prediction of the mean field strength from the input power. Strictly speaking, this prediction is only valid for a cavity that is excited with a single frequency or a very narrow band of frequencies [1]. Even for a narrow band of frequencies, the Q of the cavity must be slowly varying. (This theory will henceforth be referred to as the Chi Theory.)

The analysis method outlined in this paper looks at apertures and other objects as individual sources of cavity Q. These sources of Q contribute to the total Q linearly if they are not in close proximity to each other,

cavity walls, corners or edges. Some "small" and "large" objects may be analyzed without resorting to numerical methods. Complex objects and objects that are in close proximity to each other or cavity walls and corners have to be analyzed using numerical methods. A fact taken for granted in this paper is that the calculation of Q, no matter what the method, is usually accomplished at a single frequency. In fact, Q can be a function of frequency. Equations (3) and (4) give a Q due to wall losses that is proportional to the square root of the frequency. The values of Q for small apertures is inversely proportional to the frequency cubed, and for large apertures it is proportional to frequency. The dielectric covered large aperture and the ferrite absorber have transmission and absorption that are complex functions of frequency for a given thickness of the dielectric or ferrite, thus affecting Q in a very complex way. Other objects or combinations that are analyzed using numerical methods can have rapidly varying contributions to Q due to resonances, causing the total Q to be a rapidly varying function of frequency.

All of the information in the previous paragraphs needs to be considered when measuring Q. Given the statistical nature of the power density in the cavity, a number of measurements have to be taken. The method used to get data for calculation of Q in mode stirred chamber is to take measurements at a single frequency for different positions of a tuner (paddle wheel.) Thus, in effect, the tuner is used to "stir" the modes in the cavity. In [4], 180 measurements per frequency were made using a square-law device with an output voltage proportional to the square of the electric field. Since the Q of the cavity is proportional to the mean of the power density, it is readily calculated from the mean of these measurements. The relationship used to calculate Q from the transmitted and recieved power, P_t and P_r, of two antennas inside the cavity is [25]

$$Q = \frac{16\pi^2 V}{\lambda^3} \frac{P_r}{P_t} \tag{30}$$

where V is the cavity volume.

An alternative method for measurement of Q of mode-stirred chambers takes advantage of the relationship between cavity Q and its time constant, $\tau = Q/\omega$. The measurement procedure consists of putting transmitting and receiving antennas inside the cavity and measureing the rise and decay of the power when the transmitter is turned on and off. This method was used in [11] where the average of 200 measurements (200 tuner positions) was necessary to get a smooth curve in order to do a least squares curve fit on the exponentials. The results were compared to those obtained from the power measurement method (equation (30)) and good agreement was obtained.

In [1], experimental data was presented and used in an effort to validate the Chi Theory. Succeeding reports, [26] and [27], present more experimental data and give convincing arguments for the Chi Theory. The

method used to take the large number of measurements needed for the calculations of the statistical nature of the fields and Q was to position a B-dot probe in one location in the cavity, and vary the frequency of excitation. In [26], the frequency range was 4 - 8 GHz and in [27] and [1] the ranges were 4 - 8 GHz and 1 - 12 GHz. Essentially, this technique relies on the fact that at different frequencies, the modes in the cavity are slightly different and the power density at a single location should then have the theoretical statistical distribution. This is, as noted above, only valid over a very narrow band of frequencies. It is not clear from the reports how the data from this series of experiments was reduced in order to get the positive results reported. Clearly, all of the data could not be used as the Q of the cavities must surely be varying over such a wide bandwidth.

Another method used to take measurements in large chambers is the frequency stirring method reported in [28]. Frequency stirring is accomplished by using band limited Gaussian noise as the excitation for the experiment. The high mode density found in a large cavity at a single frequency is increased substantially by using this technique. Field uniformity tests were conducted using B-dot probes to measure magnetic flux densities. The ratios of the calculated power from two such probes were then plotted as a function of frequency. Measurements at a single frequency showed variations of ± 15 dB over a frequency range of 1 - 4 GHz. A noise bandwidth of 10 MHz reduced those variations to ± 5 dB and a 50 MHz bandwidth reduced variations to ± 1.5 dB. The effect of frequency stirring is then to create a more homogeneous power density within the cavity so that only one measurement is needed to get the information needed to calculate Q, or any other quantity of interest for that matter. This method will not, of course, allow you to study the statistical nature of the fields within the cavity. One could also investigate the possibility of combining this method with the time constant method presented above in order to get a measure of Q.

A measurement technique not used in this research area, but one that could be very effective, is thermographic imaging [29][30]. This is an experimental technique which involves the infrared observation of the heat generated in a lossy material by means of microwaves. Of course, the lossy material would affect Q but careful design of the experiment would either maximize or minimize its effect depending on what the desired measurement is. The lossy material can be easily characterized and its effect on Q quantified. A single mapping of the field in the cavity can provide not only the spacial variation of the fields, but also the data necessary for calculations of average value, mean, standard deviation, and of course Q.

Conclusions

As with many studies, sometimes success is measured by the number of questions it prompts rather than the number of answers is gives. In addition to the questions posed above about effects of objects and ensembles of objects on the Q of large cavities, some of the questions to be answered concerning the Chi Theory are:

1. Is the Chi Theory invalid for all cases where the cavity is excited by a wide-band source?

2. Under what conditions is the Chi Theory valid for wide-band sources?

3. How much variation in cavity Q can be allowed and still have a reasonable outcome with the use of the Chi Theory?

4. How large must a cavity be for the Chi Theory to be valid?

5. What is the best way to measure the fields inside an overmoded cavity in order to determine their statistical distribution and thus the Q of the cavity?

All of these questions can be answered with a comprehensive study that should include experiments and numerical simulations specifically designed with these questions in mind. A proposal will soon be submitted that will include modification of GEMACS, CARLOS-3D™ and/or CARLOS™ as outlined above so that contributions to cavity Q for a wide variety of objects and ensembles of objects can be studied. Proposals would then follow specifying studies using the modified codes. The numerical studies should be used to design experiments to fully investigate and validate the theory. Well designed experiments will also validate the numerical code.

References

[1] R. H. Price *et al.*, "Determination of the statistical distribution of electromagnetic field amplitudes in complex cavities," JAYCOR rep. 880JAL129, June 1, 1988.

[2] R. R. Lentz and H. C. Anderson, "Reverberating chambers for EMC measurements," presented at the IEEE EMC Symposium, pp. 446-451, 1979.

[3] P. Corona *et al.*, "Magnification factor for mode stirred chambers," *IEEE Trans. Electromagn. Compat.*, vol. EMC-18, no. 2, pp. 54-59, May 1976.

[4] J. G. Kostas and B. Boverie, "Statistical model for a mode-stirred chamber," *IEEE Trans. Electromagn. Compat.*, vol. 33, no. 4, pp. 366-370, November 1991.

[5] R. F. Harrington, *Field Computation by Moment Methods*, Macmillan, New York, 1968.

[6] P. P. Silvester and R. L. Ferrari, *Finite Elements for Electrical Engineers*, 2nd ed., Cambridge University Press, Cambridge, 1990.

[7] J. B. Keller, "Geometrical theory of diffraction," *J. Opt. Soc. Amer.*, vol. 52, no. 2, pp. 116-130, February 1962.

[8] E. L. Coffey, "General Electromagnetic Model for the Analysis of Complex Systems (GEMACS)," RADC-TR-90-360, Rome Air Development Center, NY, December 1990.

[9] J. M. Dunn, "Local, high-frequency analysis of the fields in a mode-stirred chamber," *IEEE Trans. Electromagn. Compat.*, vol. 32, no. 1, pp. 53-58, February 1990.

[10] B. H. Liu, D. C. Chang, and M. T. Ma, "Eigenmodes and the composite quality factor of a reverberating chamber," NBS Tech. Notes, no. 1066, August 1983.

[11] D. A. Hill, J. W. Adams, M.T. Ma, A. R. Ondejka, B. F. Riddle, M. L. Crawford, and R. T. Johnk, "Aperture excitation of electrically large, lossy cavities," *NIST Technical Note 1357*, August 1993.

[12] J. A. Kong, *Electromagnetic Wave Theory*, 2nd ed., John Wiley & Sons, Inc., New York, 1990.

[13] M. I. Kontorovich, "Averaged boundary conditions on the surface of a grid with quadratic cells," *Radiotekh, Electron.*, vol. 8, no. 9, pp 1506-1515, 1963 (in Russian).

[14] M. I. Astrakhan, "Reflecting and screening properties of plane wire grids," *Radio Engineering*, vol. 23, no. 1, pp. 76-83, 1968.

[15] M. I. Kontorovich and V. P. Akimov, "Averaged boundary conditions on the surface of a plane wire grid with nonorthogonal cells," *Radiotekh, Electron.*, vol. 22, no. 6, pp 1125-1135, 1977 (in Russian).

[16] A. I. Adonina and V. V. Shcherbak, "Equivalent boundary conditions at a metal grating situated between two magnetic materials," *Soviet Physics-Technical Physics*, vol. 9, no. 2, pp. 261 - 263, August 1964.

[17] A. N. Sivov, "Electrodynamic theory of a dense plan grating of parallel conductors," *Radio Eng. Electron. Phys.*, vol. 6, pp. 429 - 440, 1961.

[18] V. I. Fedyanovich, "Weakly transparent capacitive sheet," *Sov. J. Comm. Tech. Electron.*, vol. 34, no. 7, pp 146 - 152, 1989.

[19] C. M. Butler, Y. Rahamt-Samii and R. Mittra, "Electromagnetic penetration through apertures in conducting surfaces," *IEEE Trans. Antennas Propag.*, vol. 26, pp. 82-93, 1978.

[20] A. Ishimaru, *Electromagnetic Wave Propagation, Radiation, and Scattering*, Prentice Hall, Englewood Cliffs, New Jersey, 1991.

[21] C. A. Balanis, *Antenna Theory, Analysis and Design*, John Wiley & Sons, New York, 1982.

[22] K. S. H. Lee, editor, *EMP Interaction: Principles, Techniques, and Reference Data*, Hemisphere Publishing Corporation, New York, 1986.

[23] L. K. Warne, T. E. Koontz, and K. C. Chen, "Equivalent Polarizabilities of Apertures with Depth," AFWL Interaction Note 474, March 1989.

[24] J. M. Putnam, L. N. Madgyesi-Mitschang and M. B. Gedera, "CARLOS-3D Three-Dimensional Method of Moments Code", McDonnell Douglas Aerospace - East, December 1992.

[25] P. Corona, G. Latmiral, E. Paolini, "Performance and analysis of a reverberating enclosure with variable geometry," *IEEE Trans. Electromagn. Compat.*, vol. 22, no. 1, pp. 2 - 5, February 1980.

[26] R. H. Price *et al.*, "The pepsi syndrome as it applies to high power microwave susceptibility analysis," JAYCOR report, August, 1990.

[27] R. H. Price *et al.*, "Prediction of the microwave power density probability distribution in military systems using statistical mechanics," JAYCOR report, August, 1990.

[28] T. A. Loughry, "Frequency stirring: an alternate approach to mechanical mode-stirring for the conduct of electromagnetic susceptibility testing," Phillips Laboratory, Kirtland Air Force Base, NM, PL-TR-91-1036, November, 1991.

[29] R. M. Sega, "Infrared detection of microwave scattering and diffraction," *Proceedings of the National Radio Science Meeting*," Boston, MA, June 1984.

[30] Donald W. Metzger, "Quantification of Thermographic Mapping of Microwave Fields," PhD Thesis, University of Colorado, Colorado Springs, CO. 1991.

Massively Parallel supercomputing for Large-Scale Electromagnetic Modeling

C. Y. Pan

Research Associate Professor

Department of Physics

Utah State University

Logan, Utah 84322-4415

Final Report for:

Summer Faculty Research Program

Phillips Laboratory

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D. C.

and

Utah State University

December 1993

Massively Parallel supercomputing for Large-Scale Electromagnetic Modeling

C. Y. Pan
Research Associate Professor
Department of Physics
Utah State University

Abstract

This report summarizes the research efforts of applying finite-difference time-domain (FD-TD) technique for Maxwell's equations on massively parallel supercomputer to model electromagnetic behavior of large-scale electromagnetic structure. The Air Force has real needs to model electromagnetic behavior of asserts having size larger than 10^7 grids in space. It is impossible to do so using current available modeling tool simply because of the enormous data needed in the modeling exceeds the computer storage of any supercomputer available at the time being. Based upon the serial large-scale electromagnetic FD-TD code which have been developed in the past year we have developed a primary massively parallel supercomputer version of the FD-TD code. It takes the unique advantages of recently available massively parallel supercomputer CM-5 and using a synchronous out-of-core technique. This code significantly reduces the I/O bottle net problem and speeds up the overall calculation tremendously. As a result we can simulate an object having size at 10^8 grids in space at a reasonable cost. It is 10 times bigger than the current state-of-the-art level. After some minor modification we should be able to model an object having size at 10^{9-10} grids in space. This means that modeling entire Jet Fighter model in space which is viewed as a "grant challenge", is no longer a far fetched dream at all. The technique we have developed is not limited to electromagnetic modeling, it can be applied to many other fields, such as large antenna design, ocean motion research, weather prediction, quantum field theory applications, optical switch modeling, neural network research, etc. As a by-product of this research we will also give some useful information and guidance for those who want to begin massively parallel supersomputing for their own project. Some cross over research efforts on massively parallel supercomputing can be developed to meet the interests of the Air Force.

Massively Parallel supercomputing for Large-Scale Electromagnetic Modeling

C. Y. Pan

Introduction

The large-scale methods in computational electromagentics has received tremendous interests in recent years. The motivations were from the DOD requirement for aerospace vehicles having low radar cross section (RCS) [1] to high power microwave (HPM) effect research like Phillips Laboratory (PL) involved. Therefore, a direct space-grid time domain Maxwell's solvers have been the subject of intense interest for this application, challenging previously dominant frequency-domain integral equation approaches. It is a "grand challenge" for us to be able to treat a problem with 10^{10-11} field unknowns. Although the basic science is the century-old Maxwell's equations (ca 1870), however our ability of developing engineering understanding of this basic science still grows strongly. Especially, the recently introduced massively parallel supercomputers have potential to treat Terabyte sized problem and delivery Teraflop performance, this effort really shows practical promising. Just only several months ago it was impossible that we can simulate the electromagnetic (EM) behavior of an object having size great than 10^7 grids in space. After the massively parallel supercomputer CM-5 has introduced to the market on January 1993, this becomes feasible. A lot of new problems associated with the large-scale EM modeling are born, a lot of new and exciting simulations which we ever could dream on will become available. Our understanding to the nature, to the science, to the engineering will be tremendously enhanced by the result of this venture. For example, we may simulate ultra large size object, even the entire F-16 Jet Fighter under HPM environment, this is very important for the Air Force by saving millions of dollars in field HPM test; we may be able to trace the travel of electrons in their quantum world to get a very close look to some new materials, such as high temperature superconductors; we may simulate the pico second optical switch for new generation of computers; we may predict the weather more effectively ever by highly increasing the resolution of

"monitor"; we may find the essence of some undefeated human diseases simply by enlarged our ability of "visualizing" the object of interests, etc. Professor Allen Toflove once commend on this new research efforts as "reinventing Maxwell's equations" [2], think about the giant simulation abilities comes out of this research and all the new and exciting applications came out from it I am more than agree to his comment.

We have carried out the research on large-scale EM modeling since January 1992 supported by PL of the Air Force. Two methods were developed. One is finite-difference time-domain (FDTD) approach, another one is boundary element method (BEM). This report will be focused on FDTD approach. A serial FDTD code using synchronous out-of-core technique has been developed on August 1992. In a demo we presented at PL [3] it showed that we could simulate the EM behavior under HPM environment for an object having size at 10^{6-7} grids in space. It was a significant progress in terms of the enhancement of modeling ability at PL. Before that we had a FDTD Maxwell solver, called TSAR, which was developed by the Lawrence Livermore National Laboratory. The biggest object size it could simulate was 10^6 grids in space, in another word, the biggies simulating object was $(1.5\ meter)^3$ which could not meet the EM modeling demanding of the Air Force. For example, at WSM/PL of Lt. Col. Anderson's they had a HPM test project under NATO agreement which required to model a semi-elliptic shelter at 10^9 grids in space [4]. Since the simulation tool to model such big size object was not available they have to carry out the HPM test without any simulation predictions, such as the cut-off frequency, energy flow, etc. Examples of simulating other large-scale flying bodies at the Air Force' interest can be listed in a long form, among such to simulate the entire F-16 Jet Fighter is always the dream. The serial FDTD code which we developed last year [5-7] has enhanced the EM modeling ability of 10-100 times than TSAR does. However, it still does not meet the demand of large-scale EM modeling of the Air Force yet. More seriously, the large amount of time spent in I/O process due to synchronous out-of-core technique becomes a bottle net problem. Practically it is not acceptable to obtain a simulation result after waiting several weeks. Therefore to develop the serial code to a massively parallel code becomes practically urgent. This report will

summarize the research effort on massively parallel supercomputing for the large-scale EM modeling. As a result of this Summer Research Faculty Program we have developed a primary massively parallel CM-5 code for large-scale EM modeling. The results show that we can reduce the I/O time substantially comparing with the serial code. This parallel CM-5 FD-TD code can simulate an object having 10^8 grids in space within a reasonable execution time. After some modifications we should be able to model an object having 10^{9-10} grids in space in the near future. Considering the current state-of-the-art of large-scale EM modeling is to be able to model an object having 10^7 grids in space, also considering the prediction made by leading scientist in this field, Professor Allen Taflove that it is impossible to do entire Jet Fighter modeling until 1996" [8], we can see that the present research result is indeed a significant progress toward the fulfillment of the "grand challenge" of large-scale EM modeling.

The rest of this report will be organized in four sections after this brief introduction. Section "Methodology" will introduce the FD-TD algorithm, radiation boundary condition (RBC), synchronous out-of-core technique and the CM-5 massively parallel supercomputer. Section "CM-5 version of large-scale EM modeling code" will report the detail of the massively parallel CM-5 code we have developed. Section "Large-Scale EM Modeling" will report on the primary results of the massively parallel supercomputing on EM modeling. The final section "Discussion" will identify the future research target of this program and provide some suggestion to those who want to begin the massively parallel supercomputing for their own project.

Methodology

1. FD-TD algorithm

(1) Basic theory of EM modeling

Maxwell's equations can be expressed as:

$$\nabla \cdot \vec{B} = 0 \quad , \quad \text{no magnetic monopole law} \tag{1.1}$$

$$\nabla \times \vec{E} = -\frac{1}{C}\frac{\partial \vec{B}}{\partial t} \quad , \quad \text{Faraday's law} \tag{1.2}$$

$$\nabla \cdot \vec{D} = 4\pi\rho \quad , \quad \text{Gauss' law} \tag{1.3}$$

$$\nabla \times \vec{H} = \frac{1}{C}\frac{\partial \vec{D}}{\partial t} + \frac{4\pi}{C}\vec{J} \quad , \quad \text{Ampere's law} \tag{1.4}$$

where $\vec{D}$ is the electric flux density, $\vec{E}$ is the electric field intensity, $\vec{B}$ is the magnetic flux density, $\vec{H}$ is the magnetic field intensity, $\vec{J}$ is the current density, and ρ is the charge density.

In linear media we have:

$$\vec{D} = \varepsilon\vec{E}$$

$$\vec{B} = \mu\vec{H} \tag{2}$$

where ε is the electric permittivity and μ is the magnetic permeability.

(2) Yee's algorithm

In EM modeling we are actually solving the Maxwell's equations (1) in space and time with some boundary conditions imposed, some electromagnetic source, such as incident wave or time varying current will be taken into consideration. The data collected are the time histories of the electric and magnetic fields in space. This type boundary problem generally has no exact solutions except a few solvable cases. Therefore varies computational approaches are used to do the modeling. Among them the FD-TD approach is the most commonly used approach to carry out the EM modeling. It was K.S. Yee [9] originated a set of finite difference equations for the time-dependent Maxwell's equations. Yee's algorithm show great usefulness: a. The Yee algorithm solves for both electric and magnetic fields in time and space using the coupled Maxwell's curl equations rather than solving for the electric filed alone (or the magnetic field alone) with a wave equation. b. The Yee algorithm centers its E and H-file vector components in 3-D space, so that every E-field vector component is surrounded by four circulating H-field components; and vis versa. c. The Yee algorithm centers its E and H-field vector components in time in what is commonly termed a leapfrog arrangement. All of the E-field

computations in the 3-D space of interest are completed for a particular time point using H-field data stored in the computer memory. Then, all of the H-field computations in the 3-D space are completed using E-field data just computed and stored in memory. The cycle can begin again with the re-computation of the E-field based upon the newly obtained H-fields. This process continue until time-stepping is concluded. Fig. 1 shows a typical Yee space lattice.

(3) FD-TD EM modeling implementer formula

Yee introduced the notation for space points and functions of space and time. We denote a space point in a uniform, rectangular lattice as:

$$(i,\ j,\ k) = (i\Delta x,\ j\Delta y,\ k\Delta z) \tag{3.1}$$

Here, Δx, Δy, Δz are the lattice increments in the x, y, and z coordinate directions, respectively; and i, j, and k are integers. Further we denote any function, f, of space and time evaluated at a discrete point in the space lattice and at a discrete point in time as:

$$f\ (i\Delta x,\ j\Delta y,\ k\Delta z,\ n\Delta t) = f^{n}_{i,\ j,\ k} \tag{3.2}$$

Here, Δt is the time increment, assumed uniform over the observation interval, and n is an integer. With these notations, the first order space and time derivatives are as:

$$\frac{\partial f}{\partial x}\Big|_{(i\Delta x,\ j\Delta y,\ k\Delta z,\ n\Delta t)} = \frac{f^{n}_{i+1/2,\ j,\ k} - f_{i-1/2,\ j,\ k}}{\Delta x} + O[(\Delta x)^2] \tag{4.1}$$

$$\frac{\partial f}{\partial t}\Big|_{(i\Delta x,\ j\Delta y,\ k\Delta z,\ n\Delta t)} = \frac{f_{i,\ j,\ k}^{\,n+1/2} - f_{i,\ j,\ k}^{\,n-1/2}}{\Delta t} + O[(\Delta t)^2] \tag{4.2}$$

We can derive FD-TD expressions based on Yee's algorithm for the E and H-fields components as the following:

$$\begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix}^{n+1/2}_{i,\ j,\ k} = \alpha \cdot \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix}^{n-1/2}_{i,\ j,\ k} + \beta \cdot \begin{bmatrix} \dfrac{E_y|^{n}_{i,\ j,\ k+1/2} - E_y|^{n}_{i,\ j,\ k-1/2}}{\Delta z} - \dfrac{E_z|^{n}_{i,\ j+1/2,\ k} - E_z|^{n}_{i,\ j-1/2,\ k}}{\Delta y} \\ \dfrac{E_z|^{n}_{i+1/2,\ j,\ k} - E_x|^{n}_{i-1/2,\ j,\ k}}{\Delta x} - \dfrac{E_x|^{n}_{i,\ j,\ k+1/2} - E_x|^{n}_{i,\ j,\ k-1/2}}{\Delta z} \\ \dfrac{E_x|^{n}_{i,\ j+1/2,\ k} - E_x|^{n}_{i,\ j-1/2,\ k}}{\Delta y} - \dfrac{E_y|^{n}_{i+1/2,\ j,\ k} - E_y|^{n}_{i-1/2,\ j,\ k}}{\Delta x} \end{bmatrix}$$

$$\begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}^{n+1}_{i,j,k} = \alpha' \cdot \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}^{n}_{i,j,k} + \beta' \cdot \begin{bmatrix} \dfrac{H_z|^{n+1/2}_{i,j-1/2,k} - H_z|^{n+1/2}_{i,j-1/2,k}}{\Delta y} - \dfrac{H_y|^{n+1/2}_{i,j,k+1/2} - H_y|^{n+1/2}_{i,j,k-1/2}}{\Delta z} \\ \dfrac{H_x|^{n+1/2}_{i,j,k+1/2} - H_x|^{n+1/2}_{i,j,k-1/2}}{\Delta z} - \dfrac{H_z|^{n+1/2}_{i-1/2,j,k} - H_z|^{n+1/2}_{i-1/2,j,k}}{\Delta x} \\ \dfrac{H_y|^{n+1/2}_{i+1/2,j,k} - H_y|^{n+1/2}_{i-1/2,j,k}}{\Delta x} - \dfrac{H_x|^{n+1/2}_{i,j+1/2,k} - H_x|^{n+1/2}_{i,j-1/2,k}}{\Delta y} \end{bmatrix}$$

Where

$$\alpha = \frac{1 - \dfrac{\rho'_{i,j,k}\,\Delta t}{2\mu_{i,j,k}}}{1 + \dfrac{\rho'_{i,j,k}\,\Delta t}{2\mu_{i,j,k}}} \tag{6.1}$$

$$\beta = \frac{\dfrac{\Delta t}{\mu_{i,j,k}}}{1 + \dfrac{\rho'_{i,j,k}\,\Delta t}{2\mu_{i,j,k}}} \tag{6.2}$$

$$\alpha' = \frac{1 - \dfrac{\sigma_{i,j,k}\,\Delta t}{2\varepsilon_{i,j,k}}}{1 + \dfrac{\sigma_{i,j,k}\,\Delta t}{2\varepsilon_{i,j,k}}} \tag{6.3}$$

$$\beta' = \frac{\dfrac{\Delta t}{\varepsilon_{i,j,k}}}{1 + \dfrac{\sigma_{i,j,k}\,\Delta t}{2\varepsilon_{i,j,k}}} \tag{6.4}$$

and ρ' , σ are defined as

$$\vec{J}_m = \rho'\vec{H}, \quad \vec{J}_e = \sigma\vec{E} \tag{6.5}$$

2. Radiation Boundary Conditions

A basic consideration with the FD-TD approach to solve electromagnetic field problems is that most such problems are usually considered to be "open" problems where the domain of the computed field is ideally unbounded. Clearly, no computer can store an unlimited amount of data, and therefore, the field computation zone must be limited in size. The computation zone must be large enough to enclose the structure of interest, and a suitable boundary condition on the outer perimeter of the computation zone must be used to simulate the extension of the computation zone to infinity. This boundary condition suppresses spurious reflections of outward-propagating numerical wave analogs to some acceptable level, permitting the FD-TD solution to remain valid for all time steps (especially after spurious reflected wave analogs return to the vicinity of the modeled structure). Outer lattice boundary conditions of this type have been called either radiation boundary condition (RBS), absorbing boundary conditions (ABS), or lattice truncation condition. A simple and successful FD scheme for the RBC was introduced by Mur [10]. For simplicity, using 2nd-order Mur RBC we give the following time-stepping algorithm for components of W along x = 0 lattice boundary: (see Fig. 2)

$$W|_{0,j,k}^{n+1} = -W|_{1,j,k}^{n-1} + \frac{c\Delta t-\Delta x}{c\Delta t+\Delta x}\left[W|_{1,j,k}^{n+1} + W|_{0,j,k}^{n-1} \right] + \frac{2\Delta x}{c\Delta t+\Delta x}\left[W|_{0,j,k}^{n} + W|_{1,j,k}^{n} \right]$$

$$+ \frac{(c\Delta t)^2\Delta x}{2\Delta y^2(c\Delta t+\Delta x)}\left[W|_{0,j+1,k}^{n} - 2W|_{0,j,k}^{n} + W|_{0,j-1,k}^{n} + W|_{1,j+1,k}^{n} - 2W|_{1,j,k}^{n} + W|_{1,j-1,k}^{n} \right]$$

$$+ \frac{(c\Delta t)^2\Delta x}{2\Delta z^2(c\Delta t+\Delta x)}\left[W|_{0,j,k+1}^{n} - 2W|_{0,j,k}^{n} + W|_{0,j,k-1}^{n} - W|_{1,j,k+1}^{n} - 2W|_{1,j,k}^{n} + W|_{1,j,k-1}^{n} \right] \tag{7}$$

3. Synchronous out-of-core technique

Since the size of the object of interest is too big to be simulated in-core. For example, in order to simulate an object having size at 10^9 grids in space it required in-core RAM at 100 Gbytes which exceeds the limit of any current available supercomputers. A natural alternative is to use out-of-core

technique. We can divide the object along its z-direction into n_z 2-D slices where n_z is an integer. Because the nature of FD-TD we need retain at least three slices in-core while we updating the data. Actually what we did was to update FD-TD calculation in the 2nd-slice which is at z point by using the information in the 1st-slice and the 3rd-slice which are at $(z-\Delta z)$ -point and $(z+\Delta z)$ -point, respectively. After the 2nd-slice has been updated writing the updating information and the 1st-slice information (which has been used) back to disk, in the meantime read in a new-slice information from the disk which is at $(z+2\Delta z)$ point. In the meantime relabel the number of the slices in-core, i.e. old 2nd-slice becomes new 1st-slice, old 3rd-slice becomes new 2nd-slice, read in slice becomes new 3rd-slice (see Fig. 3). This process of read/write continues until running all z-direction. The programming for this out-of-core technique is very nasty, however it does make it possible to simulate a large-scale problem. For example, for a problem having 10^9 grids in space it only requires 10 - 100 Mbyte RAM. It is obviously that the I/O process then becomes enormous, induced so called I/O bottle net problem. By taking advantage of the unique feature of parallel read/write function of the CM-5 and using other I/O speed up algorithm this trade off I/O problem can be reduced to an acceptable level. In fact we have achieved to speed up the I/O by a factor of 10 using the CM-5 version of FD-TD code. It can be reduced even more if we add some modifications in the code which will be explained in the Discussion section.

4. The CM-5 massively parallel supercomputer

In this research we are taking advantages of massively parallel supercomputer. CM-5 is one of them, other machines are Paragon (Intel), KSR (University of Cornell), T-3D (Cray Research Co.), etc. CM-5 is made by Thinking Machine Co. (TMC) who has several years in making massively parallel supercomputers. They have marketed CM-2, CM-200 before CM-5 appears on January 1993. After comparisons I find out that the CM-5 is the best candidate for the time being to do massively parallel supercomputing in terms of its hardware stability and profound software support. The main features of the CM-5 are listed as the following [11]:

a. Massively data parallel function. Current CM-5 has 32, 64,, 128, 512 up to 1,024 nodes.

Each node has 32 Mbyte of memory and four vector units, delivers up to 128 Mflops.

b. It can treat Terabyte-sized problem (current CM-5 has 32 Gbyte RAM) and it can have Teraflops performance (current CM-5 can delivery 0.1 - 0.5 Teraflop)

c. A MIMD machine which can performance control parallel computation through the CM Message Passing Library (CMMD).

d. It has good I/O capacity. The current CM-5 has data vaults of 30 Gbyte disk with a band width 25 Mbyte/sec. Its scalable disk array (SDA) can perform parallel read/write which has a 130 Gbyte disk with a band width 100 Mbyte/sec (it can be increased more).

e. High efficiency compiler support CM Fortran, C* and *List program languages.

f. It has high performance programming environment, called "Prism" which can provide line by line debugging, analyzing parallel performance, visualizing data, etc.

g. It has a strong library support which called CM Scientific Software Library (CMSSL).

h. The program execution environment includes checkpoints, timers, timesharing and NQS.

i. It has a data visualization software, called AVS.

The main hardware of the CM-5 can be seen in Fig. 4.

CM-5 version of large-scale EM modeling code

As a result of this research a primary CM-5 version of large-scale EM modeling code has been developed. This program is written in CM Fortran. Since it is quite length I will not print the whole program in this report. A program flow chart can be seen in Fig. 5, and the program can be found in enclosed disk. The validity of this program has been checked by comparing the result obtained by serial code and this parallel code for the same EM problem (see Fig. 6). The primary modeling results will be reported in next section. The main features of this CM-5 code are listed as the following:

1. It takes advantages of massively parallel feature of CM-5, therefore it is about 10 times faster than serial code on CRAY-2.

2. It uses the out-of-core technique, therefore it can treat an object having size 10^8 grids

in space. After some modification it should be able to treat an object having size at 10^{9-10} grids in space in the near future (see the following text).

3. It implements a 2nd-order Mur RBC. It should be pointed out that this primary version is attempting to use the least computer memory in order to be able to simulate the biggest object as possible. The trade off is that it slows the execution time. Efforts of writing a more fast version is on going.

4. It uses SDA to do parallel read/write. The program implementer is successful, the result is encouraging (see the following section).

Massively Parallel Supercomputing

After checking the validity of this CM-5 code I have run several tests. The relative results are summarized as the following:

1. Test the effective of SDA. I used the CM-5 at the National Center for Supercomputing Application at University of Illinois. Test is done on a 64 x 64 x 64 grids in space using 256 nodes partition. The EM model is a 3-D cubic box having a time varying current sitting at the center of the box. Comparison is made between using data vaults and SDA (Table 1).

Table 1

I O hardware	CPU/per time step
Data Vaults	14.44 sec
SDA	2.33 sec

2. Comparison between serial code and parallel code is obtained by the simulation on a same EM model indicated in 1. Its size is 10^6 grids in space with fixed BC (Table 2).

Table 2

Type of computer	Type of code	CPU/per time step
Sparc-2	Seral	2,670 sec
CRAY-2	Seral	825 sec
CM-2	Parallel	330 sec
CM-5	Parallel	78.5 sec

3. Comparison between seral code and parallel code for the same EM model as 2. but having a 2nd-order Mur RBC (Table 3).

Table 3

Type of computer	Type of code	CPU/per time step
Sparc-2	Seral	3,471 sec
CRAY-2	Seral	1,319 sec
CM-2	Parallel	992 sec
CM-5	Parallel	96 sec

4. Same model as used in 2 and 3, the size is 10^8 grids in space

CM-5: CPU/per time step = 377 sec

Discussions

As a summary, this research results a massively parallel version of large-scale EM modeling code. It has achieved important progress towards meeting the EM modeling demands of the Air Force and the goal of "grand challenge". However, it needs to be further developed. Several further research tasks are urgent in order to delivery a satisfactory large-scale Maxwell's solver to the Air Force in the near future. They are as the following:

1. It is important to develop another fast version to further reduce the RBC execution time. The present version is the most memory saving version but it is slow for the RBC subroutine. Right

now the RBC subroutine uses about 70% of all FD-TD updating time. It can be resolved by using another type RBC, namely the Lian RBC [12] which can speed up about 3 times than Mur RBC.

2. It is important to further reduce the I/O time for the out-of-core technique. We will fully take advantage of parallel read/write function of the CM-5. Right now the machine we have used at Army Research Center has 100 Mbyte/sec I/O band width. We are going to use the CM-5 at Los Alamos National Laboratory whose SDA has I/O band width at 250 Mbyte/sec. Also we are going to implement the advanced time step algorithm which was proposed by WSR/PL [13] into this CM-5 code. By doing all these modifications it is estimated that we can further reduce the I/O time by a factor of 5-10.

3. It is also important to implement the MIMD program to fully taking advantage of the CM-5. Right now all the nodes are doing same instruction. By using CMMD we should let different node doing different job at entire FD-TD calculation. This will be definitely increase the execution speed, especially for the RBC part. CMMD was only available last May which is not stable yet, we believe that by the end of this year it should be more stable.

4. In order to delivery a practical useful software package we also need to add pre-process, such as mesh generator and pro-process, such as visualization software.

Acknowledgment

I would like to express my deep appreciation for the support from the Air Force Office of Scientific Research to give me the opportunity to do this summer research. The constant support from Mr. William D Prather, Lt. Col. Anderson, Capt. Tom Timmerman and Capt. Raley marek at PL are highly appreciated. I am grateful for the help from Dr. H.Q. Lin, Dr. John G. Kennedy and Dr. Dan Fraser. My appreciation also extends to my university colleges W. John Raitt and Wilford Hansen. The support of CM-5 time provided by the Army Research Center and the National Center for Supercomputer Application is highly appreciated.

Reference

[1], [2] Allen Taflove, "Basic and Application of Finite-Difference Time-Domain (FD-TD) Technique for Modeling Electromagnetic Wave Interactions", Workshop, University of Chicago (June 1992).

[3] C.Y. Pan *et* al. "Large-Scale Methods in Computational Electromagnetics", Seminar, Phillips Laboratory (PL) (August 1992).

[4] C.Y. Pan, "Work Report(31): Electromagnetic Waves in Semi-Elliptic Waveguides", WSR/PL (June 1993).

[5] C.Y. Pan, "Work Report (15): A New Method of Electromagnetic Modeling", WSR/PL (Jan. 1992).

[6] Thomas J. Timmerman, J. Raley Marek, C.Y. Pan, Jeff T. MacGillivray, and Sydney A. Blocher, "Progress in the State-of-the-Art of the Numerical Method of HPM Coupling", Proceeding of the 6th National Conference on HPM Technology (1993).

[7] Michael C. Governal and C.Y. Pan, "The Electromagnetic Behaviors of Large Object under High Power Microwave Environment", *J. Appl. Phys.* **73** (10), 6803 (1993).

[8] Private talk with Allen Taflove (1992).

[9] K. S. Yee, "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media", *IEEE Trans. Antennas Propagat.*, vol. AP-14, 302 (1966)

[10] G. Mur, "Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic field equations", *IEEE Trans. Electromagn. Compat.* **23**, 377 (1981).

[11] C.Y. Pan, "Massively Parallel Supercomputing on Electromagnetic Modeling", Seminar, PL (September 1993).

[12] Z. Liao, H. L. Wong, B. Yang and Y. Yuan, "A Transmitting Boundary for Transient Wave Analysis", *Scientia Sinica (Series A)*, vol. XXVII, no. 10, 1061 (1984).

[13] Jeff T. MacGillivray and J. Raley Marek, "Time Advanced Technique for FD-TD Calculation", WAR/PL Report (1993).

Fig. 1. Position of the EM field components about a cubic unit cell of the Yee space lattice.

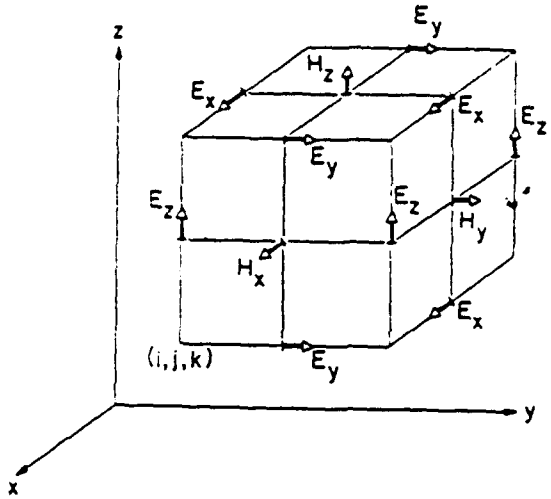


Fig. 2. Field components near the x = 0 grid boundary used in the Mur differencing scheme.

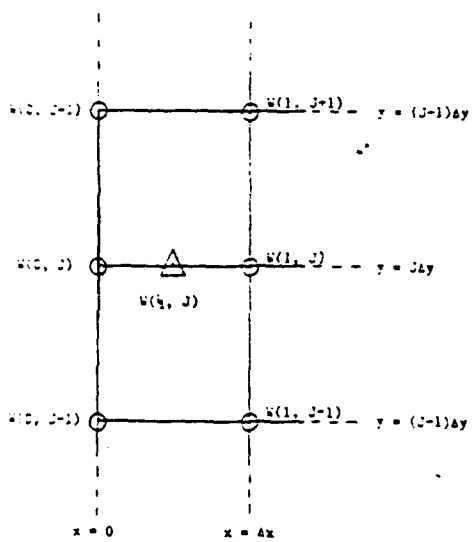


Fig. 3. The synchronous out-of-core technique.

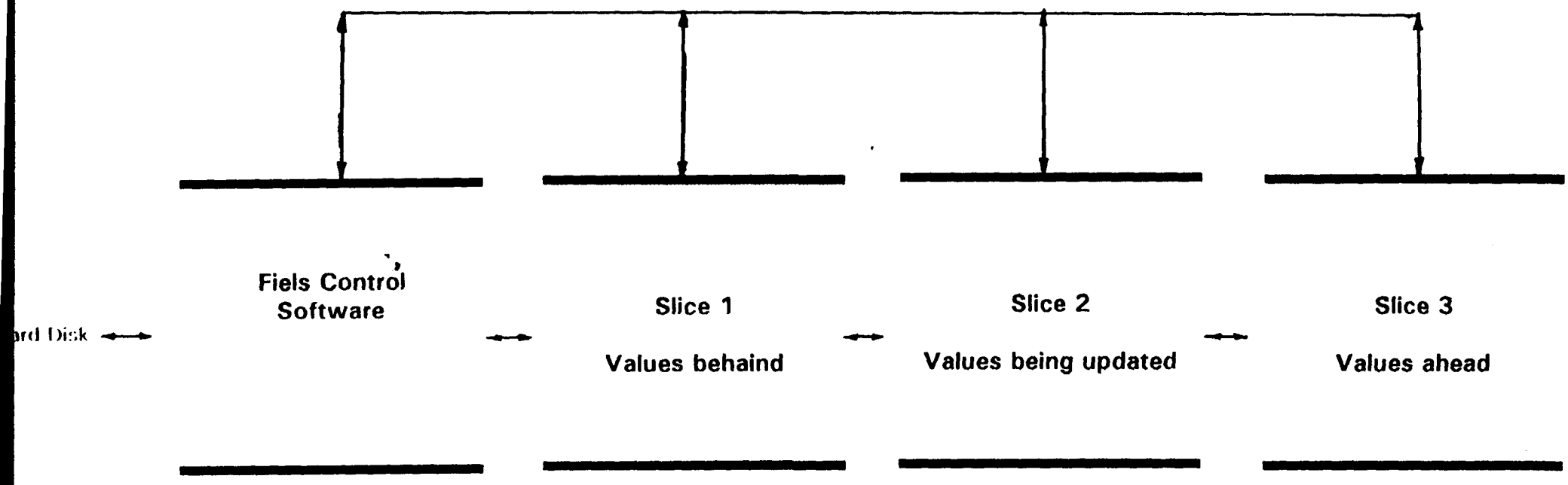


Fig. 4. The CM-5 massively parallel supercomputer

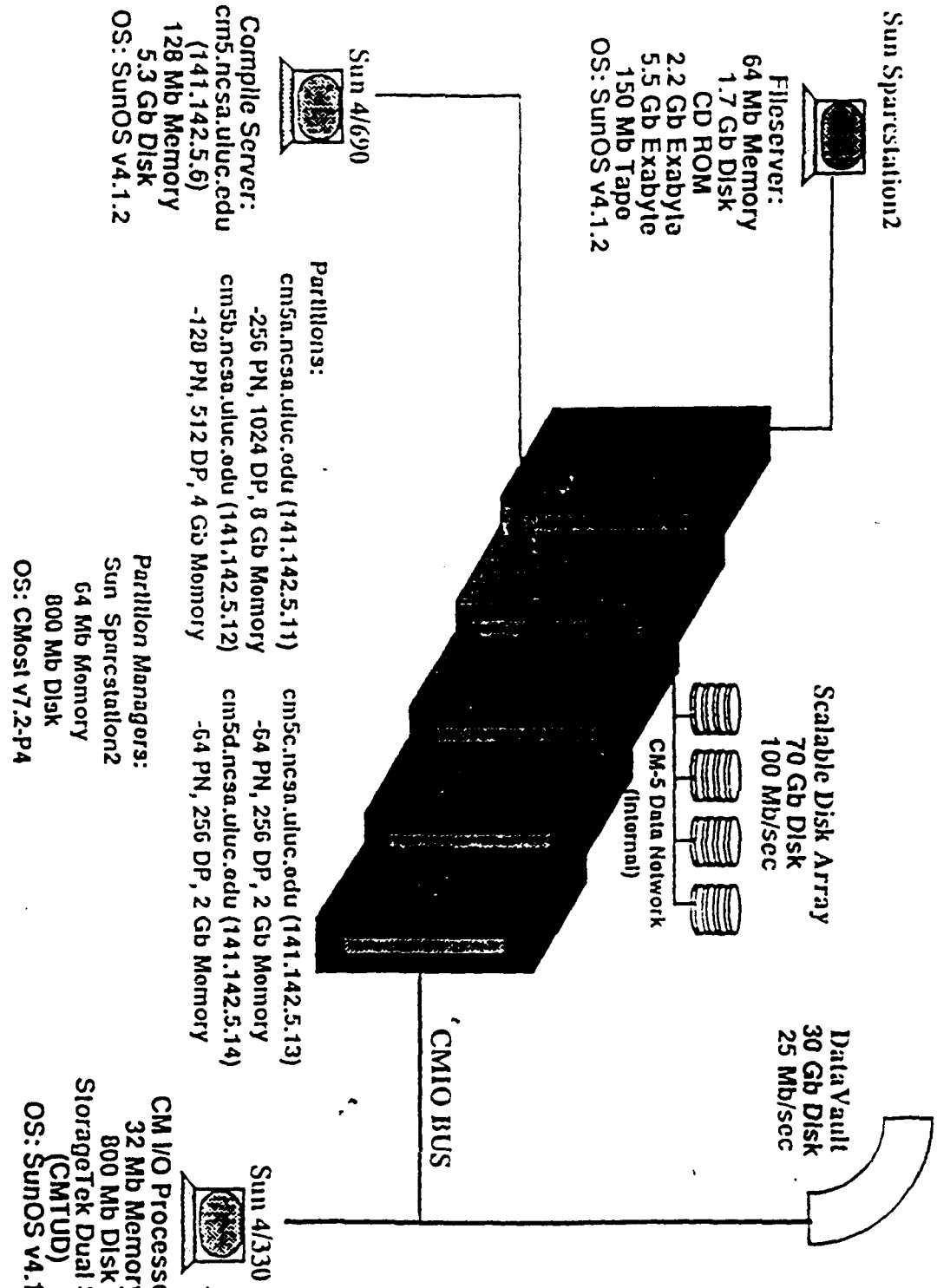


Fig. 5. The flow chart of CM-5 version EM modeling program.

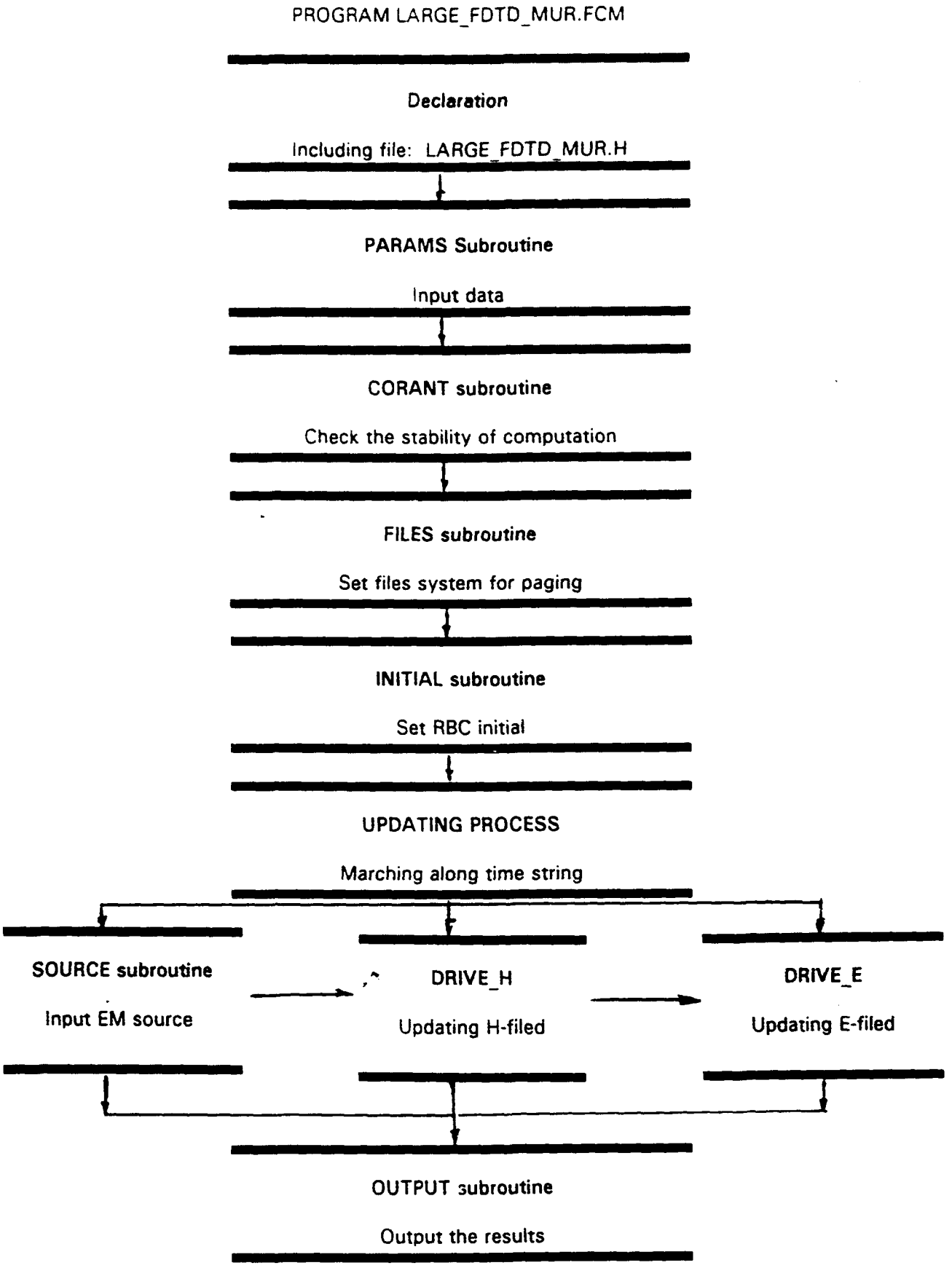
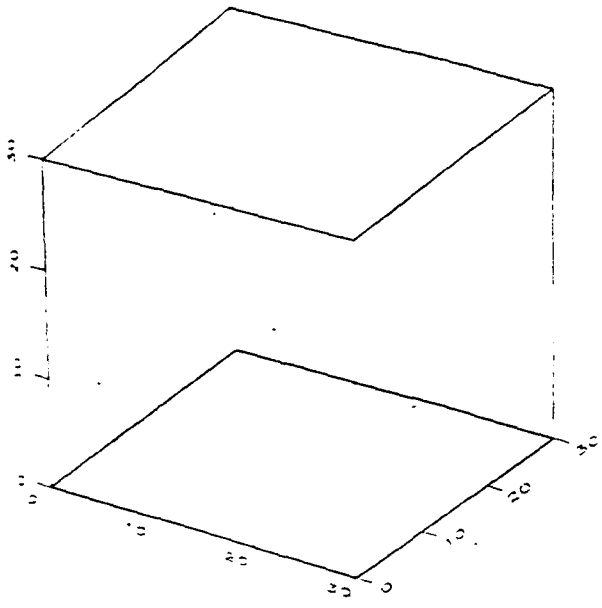
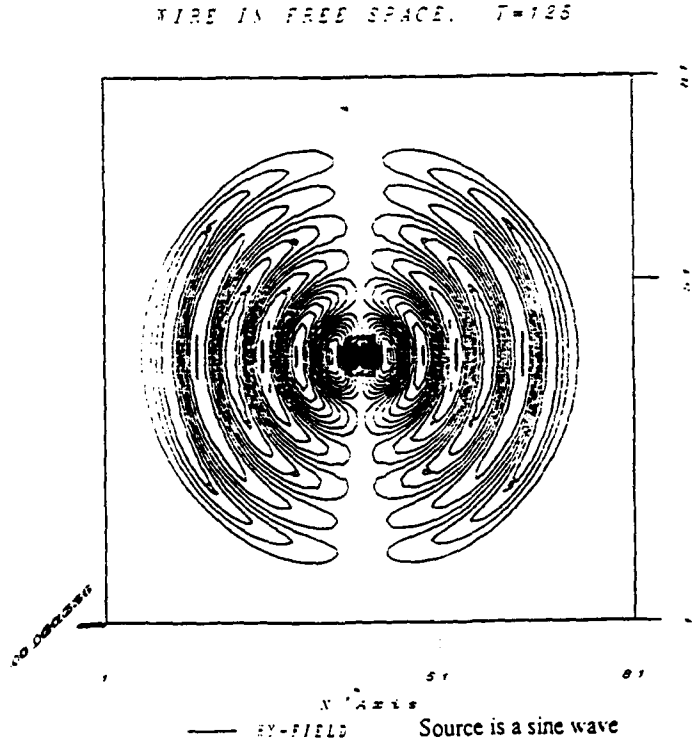


Fig. 6. A test EM model and comparison by using serial and parallel code, respectively.





Source is a sine wave

A UNIFORM DIELECTRIC LENS FOR LAUNCHING A SPHEROIDAL WAVE INTO A PARABOLOIDAL REFLECTOR

Alexander P. Stone
Professor
Department of Mathematics and Statistics

University of New Mexico
Albuquerque, New Mexico 87131

Final Report for:
Summer Faculty Research Program

Phillips Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D. C.

August 1993

A UNIFORM DIELECTRIC LENS FOR LAUNCHING A SPHEROIDAL WAVE INTO A PARABOLOIDAL REFLECTOR

Alexander P. Stone
Professor
Department of Mathematics and Statistics
University of New Mexico

Abstract

In this paper we consider dielectric-lens designs for the specific case of launching an approximate spherical TEM wave onto an impulse radiating antenna (IRA). Restrictions on launch angles are derived yielding a range of acceptable lens parameters. An equal transit-time condition on ray paths is imposed to ensure the correct spherical wavefront. Some reflections, ideally small, at the lens boundary are allowed. Illustrations and numerical tables are presented from which examples of these lenses may be constructed.

A UNIFORM DIELECTRIC LENS FOR LAUNCHING
A SPHEROIDAL WAVE INTO A PARABOLOIDAL REFLECTOR

Alexander P. Stone

1 Introduction

Consider an impulse radiating antenna (IRA) in the form of a paraboloidal reflector fed by a conical transmission line suitable for nguiding a spherical TEM wave [1] as indicated in Fig. 1.1. The paraboloidal reflector is assumed to have a circular edge of radius a with

$$D = 2a \equiv \text{ diameter} \tag{1.1}$$

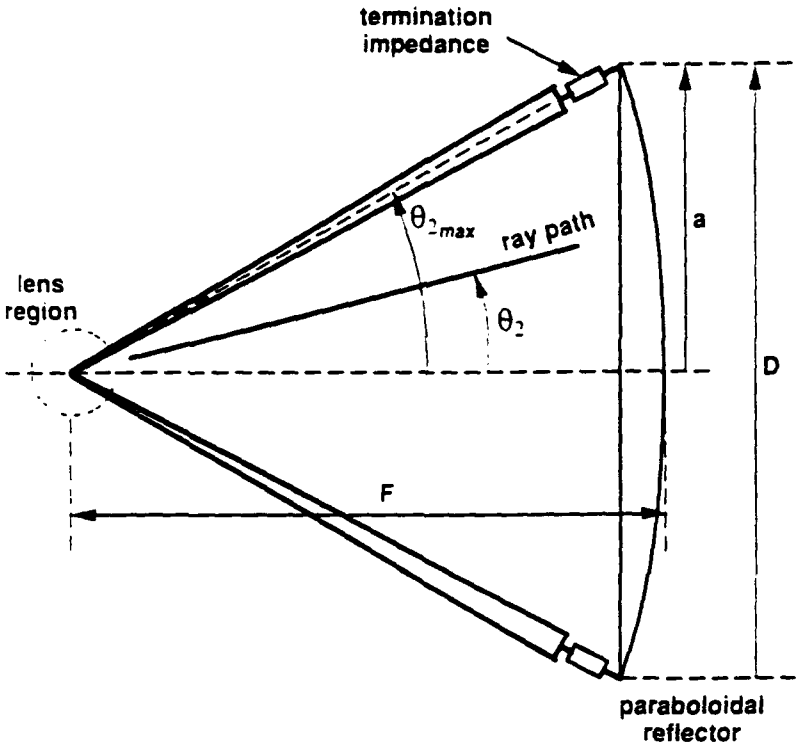


Fig. 1.1: Reflector IRA

The apex of the conical feed is located a distance

$$F \equiv \text{ focal distance} \tag{1.2}$$

from the center of the reflector. As discussed in [2] the angle from the apex of the conical transmission line (focal point) to the edge of the reflector is

$$\theta_{2_{max}} = \text{arccot}\left(\frac{2F}{D} - \frac{D}{8F}\right) = 2\arctan\left(\frac{D}{4F}\right)$$

$$= 2\arctan\left(\frac{a}{2F}\right) \tag{1.3}$$

If, as an example of commonly used reflector parameters $F/D = 0.4$, then $\theta_{2_{max}} \simeq 64.01°$. Centering our coordinate system on the conical apex, then $0 \leq \theta_2 \leq \theta_{2_{max}}$ represents the range of interest of angles for launching an electromagnetic wave toward the reflector, the axis of rotation symmetry of this reflector being taken as the z axis in the usual spherical coordinates.

As discussed in [2], as one extrapolates the desired wave on the TEM launch back toward the apex the electric field is larger and larger, until at some position before reaching the apex electrical breakdown conditions are exceeded. This is especially important in transmission where high voltages (and corresponding high powers) are desired. If the required spacing of the conical conductors at this cross section is larger than radian wavelengths at the highest frequencies of interest, or larger than some small rise time (times the speed of light) of interest, then particular care needs to be taken in synthesizing the fields at this cross section (on some aperture spherical surface). One way to achieve the increased dielectric strength, allowing one to extrapolate the desired wave back to smaller cross sections, where a switch or some other appropriate electrical source is located, is by the use of a dielectric lens. Various kinds of lenses can be considered, including those which in an ideal sense can launch the exact form of spherical TEM wave desired [7]. Here we consider a simple uniform dielectric lens which meets the equal-time requirement for the desired spherical wave, but has some (preferably small) reflections at the lens boundary which distort somewhat the desired spatial distribution (TEM) of the fields on the aperture sphere.

The lens region in Fig. 1.1 is shown on an expanded scale in Fig. 1.2. In the notation of [5] the apex, or focal point for the spherical wave (outside the lens) launched toward the reflector, is the origin ($\vec{r}_2 = \vec{0}$) of the $\vec{r}_2$ coordinate system. Here we illustrate some cut at a constant o, the lens being a body of revolution. Defining

$$\epsilon_r \equiv \frac{\epsilon_1}{\epsilon_2} \equiv \text{ relative permeability of lens} \tag{1.4}$$

we let the permeability both inside and outside the lens be μ_0. The outer permittivity ϵ_2 will often be taken as ϵ_0 in practical cases, and the lens permittivity ϵ_1 will be taken as that for dielectrics of interest such as for polyethylene or transformer oil (i.e. 2.26).

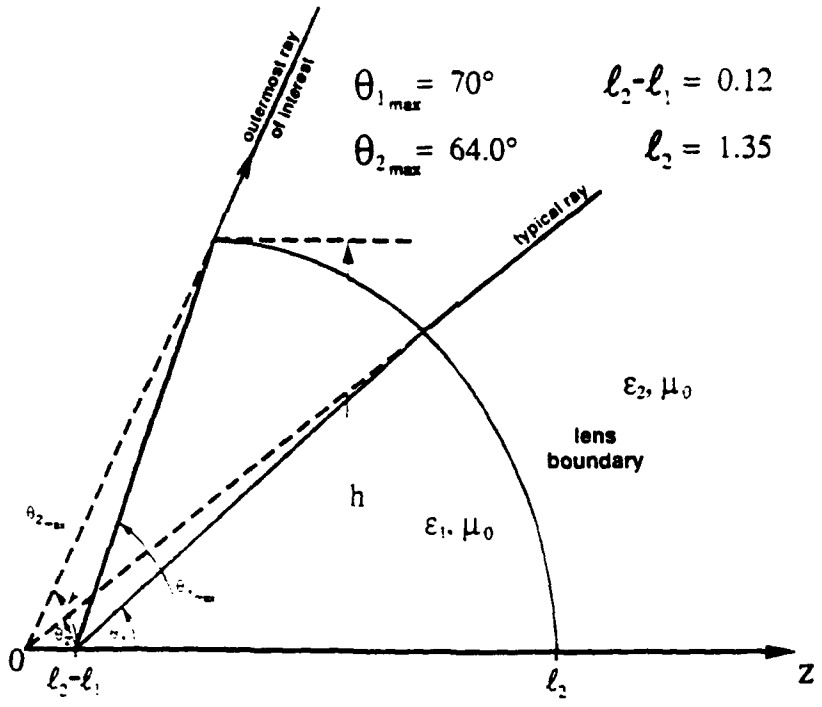


Fig. 1.2: Lens For Launching Spherical Wave

The $\theta_{2_{max}}$ previously introduced is now the maximum of θ_2, describing the rays leaving the lens toward the reflector. Inside the lens there are rays emanating from $(x, y, z) = (0, 0, \ell_2 - \ell_1)$ with the angle θ_1 with respect to the z axis. With the inside and outside rays meeting at the lens boundary the various angles are related. Corresponding to $\theta_{2_{max}}$ there is also a $\theta_{1_{max}}$ with $0 \leq \theta_1 \leq \theta_{1_{max}}$. For normalization purposes the position on the lens boundary for this outermost ray of interest is defined as having a cylindrical radius h. For later use this position will remain fixed for a given $\theta_{2_{max}}$ for various shapes of the lens boundary given by varying $\theta_{1_{max}}$. Note that the scaling lengths in [5] are related by the focal-length formula

$$\ell_0^{-1} = \ell_1^{-1} + \ell_2^{-1} \tag{1.5}$$

So given ℓ_1 and ℓ_2 one finds ℓ_0 for use in the formulas of [5], and ℓ_0 is scaled in units of h.

2 Restrictions on Launch Angles

As indicated in Fig. 2.1 there is a potential problem with the lens concerning the fatness (extent of cylindrical radius Ψ) and the maximum angle $\theta_{2_{max}}$ for launching the spherical wave outside the lens. In particular as θ_2 approaches $\theta_{2_{max}}$ from below, the lens boundary should not cross over the outermost ray of interest defined by $\theta_2 = \theta_{2_{max}}$. Referring to Fig. 2.1 then we require that the slope of the lens boundary, where the boundary meets this ray, should satisfy

$$\theta_{\ell_{max}} \leq \theta_{2_{max}} \tag{2.1}$$

Note that the radius of the lens boundary Ψ_b can be allowed to exceed h, still meeting the restriction of (2.1).

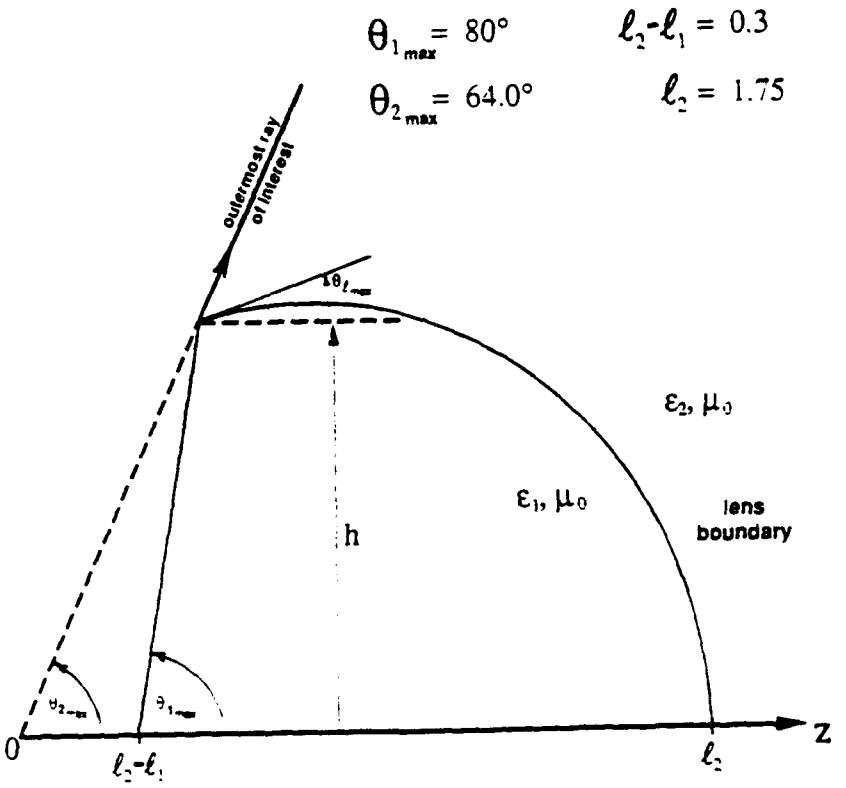


Fig. 2.1: Restriction of Lens Boundary to Inside Outermost Ray of Interest

Take the limit of equality in (2.1) to define critical angles (subscript "c"). This case is illustrated in Fig. 2.2. where the region where the critical ray meets the boundary is

expanded. Appealing to Snell's law in which the phase velocities of the waves in the two media are matched along the boundary gives

$$\sqrt{\epsilon_1}\sin(\psi_i) = \sqrt{\epsilon_2}\sin(\psi_t) \tag{2.2}$$

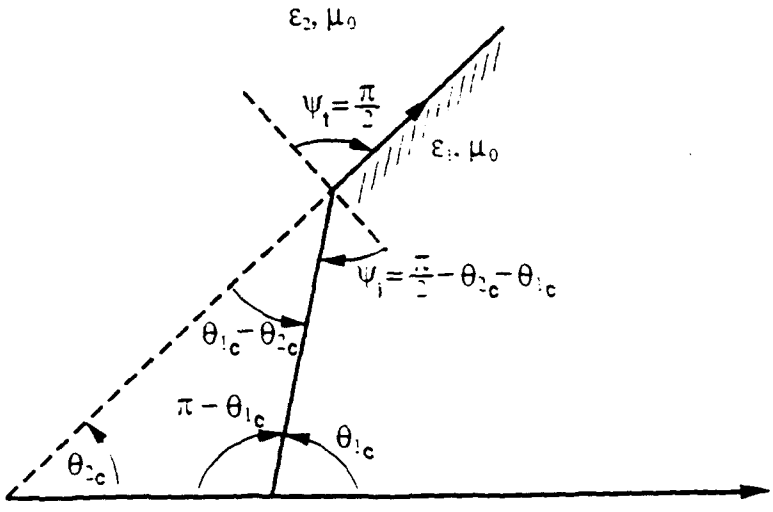


Fig. 2.2: Critical Angle of Lens Boundary

Setting

$$\theta_{\ell_c} = \theta_{2_c} \tag{2.3}$$

then we have

$$\psi_t = \frac{\pi}{2} \tag{2.4}$$

i.e., the wave in medium 2 is propagating parallel to the lens boundary. By geometric construction we have

$$\psi_i = \frac{\pi}{2} + \theta_{2_c} - \theta_{1_c} \tag{2.5}$$

which gives

$$\begin{aligned} \sin\left(\frac{\pi}{2} + \theta_{2_c} - \theta_{1_c}\right) &= \cos(\theta_{2_c} - \theta_{1_c}) = \cos(\theta_{1_c} - \theta_{2_c}) \\ &= \epsilon_r^{-\frac{1}{2}} \\ \theta_{1_c} &= \theta_{2_c} + \arccos(\epsilon_r^{-\frac{1}{2}}) \end{aligned} \tag{2.6}$$

where one needs the principle value for the arccos noting that $\theta_{1_c} > \theta_{2_c}$ in the construction of Fig. 2.2.

Noting in (2.2) that a transmission angle $\psi_t \leq \pi/2$ gives $\theta_{\ell_{max}} \leq \theta_{2_{max}}$ as an acceptable lens boundary in Fig. 2.1, then we have

$$\begin{aligned} \sin(\psi_i) &= \sin\left(\frac{\pi}{2} + \theta_{2_{max}} - \theta_{1_{max}}\right) = \cos(\theta_{1_{max}} - \theta_{2_{max}}) \\ &\geq \epsilon_r^{-\frac{1}{2}} \\ \theta_{1_{max}} &- \theta_{2_{max}} \leq \arccos(\epsilon_r^{-\frac{1}{2}}) \\ \theta_{1_{max}} &\leq \theta_{2_{max}} + \arccos(\epsilon_r^{-\frac{1}{2}}) \end{aligned} \tag{2.7}$$

Now we can also allow $\theta_{1_{max}} < \theta_{2_{max}}$, but then the geometrical construction in Figs. 2.1 and 2.2 do not apply and the lens boundary becomes concave to the right [5]. For present considerations, since $\theta_{1_{max}}$ describes the path of the conical-transmission-line conductors in the lens region, a region which lowers the characteristic impedance from that of the conical transmission line outside the lens, our interest centers on $\theta_{1_{max}}$ near $\pi/2$ which maximizes the transmission -line characteristic impedance in the lens. There are other considerations as well, such as high voltages (breakdown) in the lens near the conical apex at $(x, y, z) = (0, 0, \ell_2 - \ell_1)$, which push in the same direction. So for this paper our attention concerning $\theta_{1_{max}}$ is limited to

$$\theta_{2_{max}} \leq \theta_{1_{max}} \leq \min\left[\frac{\pi}{2}, \theta_{2_{max}} + \arccos(\epsilon_r^{-\frac{1}{2}})\right] \tag{2.8}$$

For our example cases

$$\begin{aligned} \epsilon_r &= 2.26 \\ \arccos\ (\epsilon_r^{-\frac{1}{2}}) &\simeq 48.3^\circ \end{aligned} \tag{2.9}$$

Note that for large ϵ_r the allowable range of $\theta_{1_{max}}$ is constrained close to $\theta_{2_{max}}$.

3 Special Case of $\theta_1 = \theta_2$: Spherical Lens

A very simple lens is that of a sphere of radius b centered on the origin with

$$\begin{aligned} \ell_2 &= \ell_1 = 2\ell_0 = b \\ \theta_2 &= \theta_1,\ \theta_{2_{max}} = \theta_{1_{max}} \end{aligned} \tag{3.1}$$

In this case, if the conical transmission line in medium 2 has a characteristic impedance Z_{c_2}, then continuing the conical conductors back into the lens gives a characteristic impedance there of

$$Z_{c_1} = \epsilon_r^{-\frac{1}{2}} Z_{c_2} \tag{3.2}$$

While the TEM modal distribution is the same on both sides of the lens boundary there is a reflection at the boundary with reflection coefficient

$$R = \frac{Z_{c_2} - Z_{c_1}}{Z_{c_2} + Z_{c_1}} = \frac{\epsilon_r^{\frac{1}{2}} - 1}{\epsilon_r^{\frac{1}{2}} + 1} \tag{3.3}$$

and transmission coefficient

$$T = 1 + R = \frac{2\epsilon_r^{\frac{1}{2}}}{\epsilon_r^{\frac{1}{2}} + 1} \tag{3.4}$$

For our example case we have

$$\begin{aligned} \epsilon_r &= 2.26 \\ R &\simeq 0.20, \quad T \simeq 1.20 \end{aligned} \tag{3.5}$$

The reflected wave in turn reflects off the source point (apex) with an amplitude dependent on the source impedance, say -1 reflection for a short circuit. This reflection in turn passes through the lens boundary as another spherical TEM wave.

Note that in principle the lens should be a complete sphere (4π steradians, volume $4\pi b^3/3$) for the above analysis to exactly apply. Otherwise the missing portions of the lens can introduce other modes which affect the fields at the observer with $\theta \leq \theta_2 \leq \theta_{2_{max}}$, complicating the waveform during the times of significance for the reflections.

This points to a possible disadvantage for this kind of spherical lens. Other lens shapes, while meeting the equal-time requirement for the first wave through the lens going into a spherical wave outside the lens, can break up the wavefront for successive waves by sending non-spherical waves back from the lens boundary which need not (in large part) converge on the source point.

4 Brewster-Angle Considerations

One can reduce reflections at the lens boundary by changing the direction of incidence for appropriate polarization (E wave) by use of Brewster angle considerations [3, 4, 6]. Referring to Fig. 4.1, and using a subscript "B" for this case we have

$$\begin{aligned} \cos(\psi_{iB}) &= \left[\frac{\epsilon_r}{\epsilon_r + 1}\right]^{\frac{1}{2}} = \sin(\psi_{tB}) \\ \sin(\psi_{iB}) &= [\epsilon_r + 1]^{-\frac{1}{2}} = \cos(\psi_{tB}) \\ \cot(\psi_{iB}) &= \epsilon_r^{\frac{1}{2}} = \tan(\psi_{tB}) \\ \psi_{iB} + \psi_{tB} &= \frac{\pi}{2} = 90^\circ \end{aligned} \tag{4.1}$$

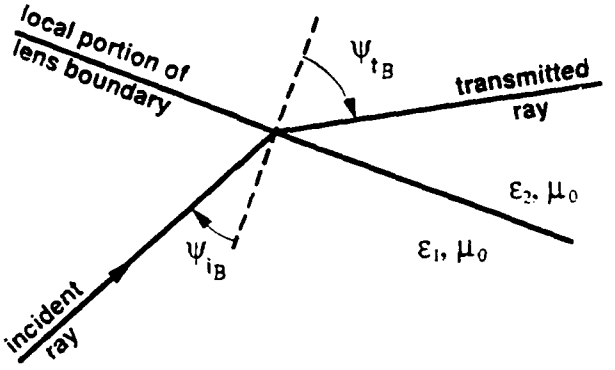


Fig. 4.1: Total Transmission of E Wave at Brewster Angle

For our example case we have

$$\begin{aligned} \epsilon_r &= 2.26 \\ \upsilon_{iB} &\simeq 33.6^\circ \\ \upsilon_{tB} &\simeq 56.4^\circ \end{aligned}$$

Noting that the angle υ_{tB} of the transmitted ray is greater than 0° (transmission of normally incident wave) but less than 90° (corresponding to transmission parallel to the lens boundary as in Fig. 2.2), then for $\theta_{1_{max}}$ chosen near the critical case there are $\theta_1 < \theta_{1\,max}$ and $\theta_2 < \theta_{2_{max}}$ which satisfy this Brewster-angle condition. So increasing $\theta_{1_{max}}$ above $\theta_{2_{max}}$ as in (2.8) can make some of the rays have a better transmission through the lens boundary.

5 Lens Shapes

The equal-time condition for a diverging spherical wave in a medium with permittivity ϵ_1 going into another diverging spherical wave in a second medium with permittivity ϵ_2 is given by the equation

$$\sqrt{\epsilon_1}\left\{\left[[z_b - \ell_2 + \ell_1]^2 + \Psi_b^2\right]^{\frac{1}{2}} - \ell_1\right\} = \sqrt{\epsilon_2}\left\{\left[z_b^2 + \Psi_b^2\right]^{\frac{1}{2}} - \ell_2\right\} \tag{5.1}$$

where ℓ_1, ℓ_2, z_b and Ψ_b are as in Fig. 1.2. This expression is derived in [5] where it appears as (4.9). If we use the coordinate relations on the lens boundary, given by

$$z_b - \ell_2 = \Psi_b \cot(\theta_{1_b}) - \ell_1 = \Psi_b \cot(\theta_{2_b}) - \ell_2 \tag{5.2}$$

the equal-time condition can then be re-expressed in the form

$$\left(\frac{\ell_2}{\ell_1} - \epsilon_r^{\frac{1}{2}}\right)\left[\sin(\theta_{1_b})\cos(\theta_{2_b}) - \cos(\theta_{1_b})\sin(\theta_{2_b})\right] = \left(\frac{\ell_2}{\ell_1} - 1\right)\left[\sin(\theta_{1_b}) - \epsilon_r^{\frac{1}{2}}\sin(\theta_{2_b})\right] \tag{5.3}$$

where $\epsilon_r = \epsilon_1/\epsilon_2$. There are also constraints on θ_1 and θ_2, derived in the earlier sections of this paper, and these are

$$\theta_{1_{max}} \leq \theta_{2_{max}} + \arccos(\epsilon_r^{-\frac{1}{2}})$$

$$\theta_{2_{max}} \leq \theta_{1_{max}} \leq min\left[\frac{\pi}{2}, \theta_{2_{max}} + \arccos(\epsilon_r^{-\frac{1}{2}})\right] \tag{5.4}$$

Since $\theta_{2_{max}} = 2\arctan(\frac{D}{4F})$ from (1.3), we may choose various F/D ratios to determine lens shapes. Thus, for a given F/D value, $\theta_{2_{max}}$ is determined. If we then choose h (Fig. 1.2) and ϵ_r and $\theta_{1_{max}}$ consistent with our constraints, we obtain the lens parameters ℓ_1/h and ℓ_2/h. Since θ_2 is a function of θ_1 (or vice versa) a lens boundary curve is generated for given $\epsilon_r, \ell_1/h$, and ℓ_2/h with $\theta \leq \theta_1 \leq \theta_{1_{max}}$, $0 \leq \theta_2 \leq \theta_{2_{max}}$. If we now select a new $\theta_{1_{max}}$, another lens can be specified for our choice of F/D. Thus by varying F/D a large collection of lens designs are obtainable.

The process described above will now be analyzed in detail. Let us fix $\theta_{2_{max}}$ (via choice of F/D), choose h and select $\theta_{1_{max}}$ in accordance with the constraints (5.4). The parameters ℓ_1/h and ℓ_2/h are determined by the transit-time relation and we find that

$$\frac{\ell_2}{\ell_1} = \frac{\epsilon_r^{\frac{1}{2}}\left[\sin\left(\theta_{1_{max}} - \theta_{2_{max}}\right) + \sin\left(\theta_{2_{max}}\right)\right] - \sin\left(\theta_{1_{max}}\right)}{\sin\left(\theta_{1_{max}} - \theta_{2_{max}}\right) + \epsilon_r^{\frac{1}{2}}\sin\left(\theta_{2_{max}}\right) - \sin\left(\theta_{1_{max}}\right)} \tag{5.5}$$

Since the geometry indicated in Fig. 1.2 gives the relation

$$\frac{\ell_2 - \ell_1}{h} = \cot(\theta_{2_{max}}) - \cot(\theta_{1_{max}}) \tag{5.6}$$

we can find explicit formulas for ℓ_1/h and ℓ_2/h. The results are

$$\begin{aligned} \frac{\ell_1}{h} &= \frac{\sin(\theta_{1_{max}} - \theta_{2_{max}}) + \epsilon_r^{\frac{1}{2}}\sin(\theta_{2_{max}}) - \sin(\theta_{1_{max}})}{(\epsilon_r^{\frac{1}{2}} - 1)\sin(\theta_{1_{max}})\sin(\theta_{2_{max}})} \\ \frac{\ell_2}{h} &= \frac{\epsilon_r^{\frac{1}{2}}\left[\sin(\theta_{1_{max}} - \theta_{2_{max}}) + \sin(\theta_{2_{max}})\right] - \sin(\theta_{1_{max}})}{(\epsilon_r^{\frac{1}{2}} - 1)\sin(\theta_{1_{max}})\sin(\theta_{2_{max}})} \end{aligned} \tag{5.7}$$

Thus ℓ_1 and ℓ_2 are determined once $\theta_{1_{max}}$, $\theta_{2_{max}}, \epsilon_r$ and h are known. We can then use, once more, the transit time condition to obtain θ_2 as a function of θ_1, (regarding $\epsilon_r, \ell_2/\ell_1$ as known), and thereby determine the lens boundary curve, which in general is a quartic curve.

To find θ_2 as a function of θ_1 we obtain a quadratic equation in either $\cos(\theta_2)$ or $\sin(\theta_2)$ from (5.3), which we rewrite in the form

$$B\Big[\sin(\theta_1)\cos(\theta_2) - \cos(\theta_1)\sin(\theta_2)\Big] = A[\sin(\theta_1) - \epsilon_r^{\frac{1}{2}}\sin(\theta_2)]$$
$$A = (\ell_2/\ell_1) - 1, \qquad B = (\ell_2/\ell_1) - \epsilon_r^{\frac{1}{2}}. \tag{5.8}$$

Algebraic manipulations then yield a quadratic in $\cos(\theta_2)$, and the quadratic formula then yields

$$\cos(\theta_2) = \frac{AB\sin^2(\theta_1) + |B\cos(\theta_1) - A\epsilon_r^{\frac{1}{2}}|\sqrt{\Big[B^2 - 2AB\epsilon_r^{\frac{1}{2}}\cos(\theta_1) + A^2\epsilon_r\Big] - A^2\sin^2(\theta_1)}}{B^2 - 2AB\epsilon_r^{\frac{1}{2}}\cos(\theta_1) + A^2\epsilon_r} \tag{5.9}$$

Similarly, from a quadratic in $\sin(\theta_2)$ we may obtain

$$\sin(\theta_2) = \frac{A(A\epsilon_r^{\frac{1}{2}} - B\cos(\theta_1)) + |B|\sin(\theta_1)\sqrt{\Big[B^2 - 2AB\epsilon_r^{\frac{1}{2}}\cos(\theta_1) + A^2\epsilon_r\Big] - A^2\sin^2(\theta_1)}}{B^2 - 2AB\epsilon_r^{\frac{1}{2}}\cos(\theta_1) + A^2\epsilon_r} \tag{5.10}$$

Finally, to obtain the lens boundary curve we need to compute the coordinates z and Ψ as a function of θ_2 (and θ_1). The geometry in Fig. 1.2 then yields the results

$$z = \frac{(\ell_2 - \ell_1)\tan(\theta_1)}{\tan(\theta_1) - \tan(\theta_2)} \tag{5.11}$$

$$\Psi = z\tan(\theta_2) = \frac{(\ell_2 - \ell_1)\tan(\theta_1)\tan(\theta_2)}{\tan(\theta_1) - \tan(\theta_2)} \tag{5.12}$$

The case $\theta_{1_{max}} = \pi/2$ leads to some simplification in the above formulas. We find, for example, that when $\theta_{1_{max}} = \pi/2$,

$$\frac{\ell_2}{\ell_1} = \frac{\epsilon_r^{\frac{1}{2}}[\cos(\theta_{2_{max}}) + \sin(\theta_{2_{max}})] - 1}{\cos(\theta_{2_{max}}) + \epsilon_r^{\frac{1}{2}}\sin(\theta_{2_{max}}) - 1} \tag{5.13}$$

$$\frac{\ell_1}{h} = \frac{\cos(\theta_{2_{max}}) + \epsilon_r^{\frac{1}{2}}\sin(\theta_{2_{max}}) - 1}{(\epsilon_r^{\frac{1}{2}} - 1)\sin(\theta_{2max})} \tag{5.14}$$

$$\frac{\ell_2}{h} = \frac{\epsilon_r^{\frac{1}{2}}[\cos(\theta_{2_{max}}) + \sin(\theta_{2max})] - 1}{(\epsilon_r^{\frac{1}{2}} - 1)\sin(\theta_{2_{max}})} \tag{5.15}$$

and the coordinates (z, Ψ) are as given in (5.11) and (5.12). When $\theta_1 \to \theta_2 \to 0$ we find

$$\begin{aligned} \frac{\sin(\theta_2)}{\sin(\theta_1)} &\to \frac{\ell_1}{\ell_2} \\ \frac{\tan(\theta_2)}{\tan(\theta_1)} &\to \frac{\ell_1}{\ell_2} \\ z &\to \ell_2 \\ \Psi &\to 0 \end{aligned} \tag{5.16}$$

as expected.

Numerical results are obtainable from the preceding analysis. In Figure 5.1 we show various lens boundaries corresponding to a value of F/D corresponding to 0.4. For this choice of F/D, $\theta_{2_{max}} = 64.01°$, and lens boundary curves are obtained for choices of $\theta_{1_{max}}$

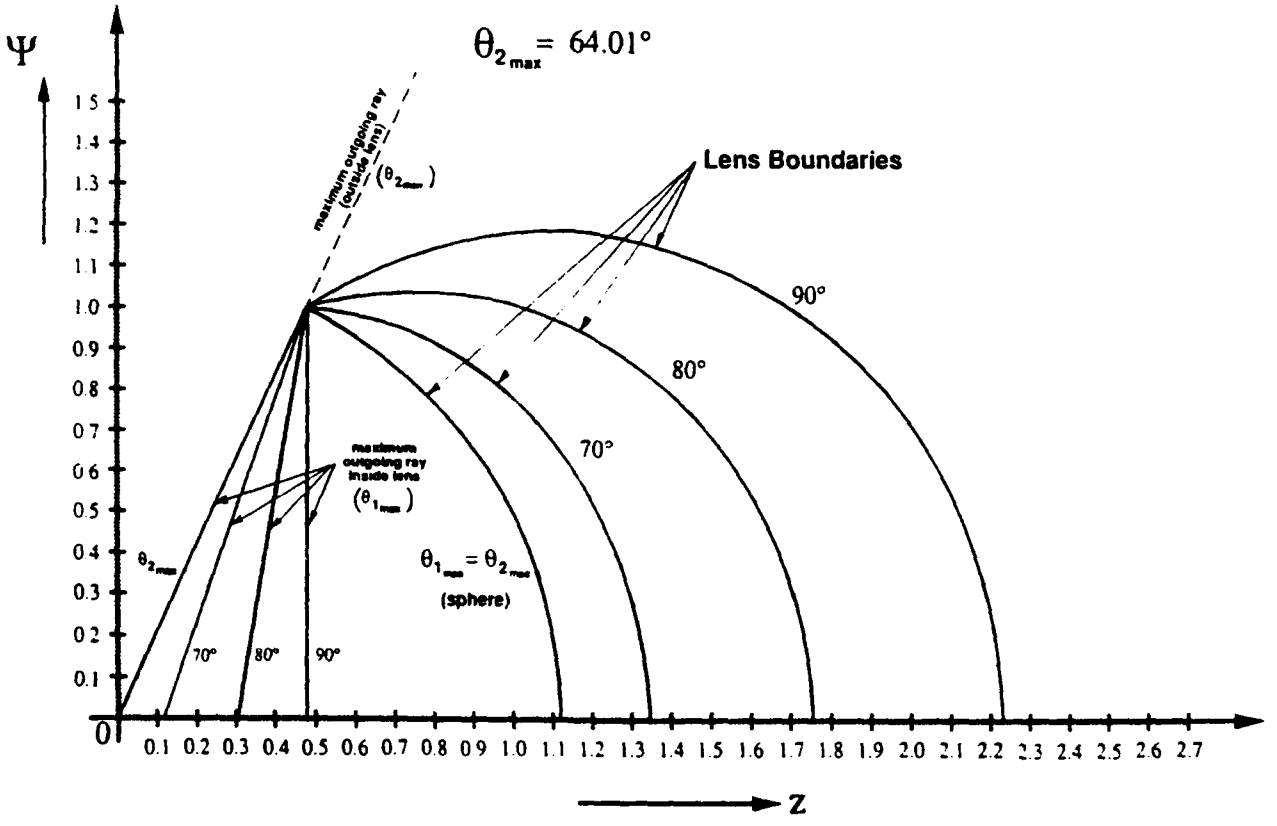


Fig. 5.1: Lens Shape for $F/D = 0.4$

equal to $70°, 80°, 90°$ as well as $\theta_{2_{max}}$ itself. In Table 5.1 numerical data is presented for the case $\theta_{1_{max}} = 90°$ with F/D = 0.4 by allowing θ_1 and θ_2 to vary up to their maximum values and calculating the coordinates z/h and Ψ/h.

θ_1	θ_2	z	Ψ
.000	.000	2.236	.000
3.000	2.345	2.232	.091
6.000	4.690	2.222	.182
9.000	7.032	2.205	.272
12.000	9.371	2.181	.360
15.000	11.706	2.151	.446
18.000	14.036	2.115	.529
21.000	16.359	2.073	.608
24.000	18.675	2.025	.684
27.000	20.981	1.972	.756
30.000	23.278	1.914	.823
33.000	25.563	1.851	.886
36.000	27.836	1.785	.943
39.000	30.093	1.715	.994
42.000	32.335	1.642	1.040
45.000	34.560	1.567	1.080
48.000	36.764	1.490	1.113
51.000	38.947	1.412	1.141
54.000	41.107	1.332	1.163
57.000	43.240	1.253	1.178
60.000	45.345	1.173	1.188
63.000	47.418	1.095	1.191
66.000	49.457	1.017	1.189
69.000	51.458	.941	1.182
72.000	53.417	.867	1.169
75.000	55.332	.796	1.151
78.000	57.196	.728	1.129
81.000	59.005	.662	1.102
84.000	60.753	.600	1.072
87.000	62.435	.542	1.039
90.000	64.044	.488	1.002

Table 5.1: Lens Shape Data for F/D = 0.4, with $0 \leq \theta_1 \leq \theta_{1_{max}}$ and $0 \leq \theta_2 \leq \theta_{2_{max}}$ and $\theta_{1_{max}} = 90°$, $\theta_{2_{max}} = 64.0°$

6 Concluding Remarks

It has been noted that one significant application of the type of dielectric lens discussed here is its possible use for launching an approximation to a spherical TEM wave onto an IRA. As one goes to higher and higher voltage pulses launched onto the TEM feed, the possibility of electrical breakdown is a consideration. The desire for faster and faster pulse rise times means that one needs to establish the TEM-mode field distribution over the cross section of the feed. In this case one would require that such pulses be treated as waves, and one in general wants to match these waves from one region to another with a minimum of distortion and reflection.

The lenses considered here have been treated from an "equal-time" point of view. Thus some reflections are introduced at boundaries. However, these reflections can be small in certain cases, and the wave passing into the second medium can approximate a dispersionless TEM wave. Future investigations will take up the issue of impedance matching.

References

[1] C. E. Baum, Radiation of Impulse-Like Transient Fields, Sensor and Simulation Note 321, November 1989.

[2] C. E. Baum, Configurations of TEM Feed for an IRA, Sensor and Simulation Note 327, April 1991.

[3] C. E. Baum, Wedge Dielectric Lenses for TEM Waves Between Parallel Plates, Sensor and Simulation Note 332, September 1991.

[4] C. E. Baum, J. J. Sadler, and A. P. Stone, A Prolate Spheroidal Uniform Isotropic Dielectric Lens Feeding a Circular Coax, Sensor and Simulation Note 335, December 1991.

[5] C. E. Baum, J. J. Sadler and A. P. Stone, Uniform Isotropic Dielectric Equal-Time Lenses for Matching Combinations of Plane and Spherical Waves, Sensor and Simulations Note 352, December 1992.

[6] C. E. Baum, J. J. Sadler, and A. P. Stone, Uniform Wedge Dielectric Lenses for Bends in Circular Coaxial Transmission Lines, Sensor and Simulations Note 356, December 1992.

[7] C. E. Baum and A. P. Stone, *Transient Lens Synthesis: Differential Geometry in Electromagnetic Theory*, Hemisphere Publishing Corp., 1991.

THE PRINCIPLE OF INCREASE OF CROSS-SECTIONAL AREA AND ITS APPLICATION TO THE FRAGMENTATION OF SATELLITES

Arjun Tan

Professor

Department of Physics

Alabama A & M University

Meridian Street

Normal, AL 35762

Final Report for

Summer Faculty Research Program

Phillips Laboratory

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D.C.

August 1993

THE PRINCIPLE OF INCREASE OF CROSS-SECTIONAL AREA AND ITS APPLICATION TO THE FRAGMENTATION OF SATELLITES

Arjun Tan
Professor
Department of Physics
Alabama A & M University

Abstract

The changes in the average cross-sectional areas of common regular solids upon fragmentation are calculated. It is shown that the average cross-sectional area of an idealized flat plate remained constant on fragmenting but those of all other types of solids increased. Two factors responsible for the increase in the average cross-sectional areas of solids are identified. For sheet materials, it is the curvature which causes the increase. For solid objects, the increase is due to the creation and exposure of new surfaces. For fragmentaions of real solids, a principle of increase of the average cross-sectional area is arrived at. Its similarity with the entropy principle is discussed. The fragmentation of solids is identified as an irreversible process and the average cross-sectional area is recognized as a measure of mechanical disorder. Finally, the increase of the average cross-sectional area is calculated for documented satellite fragmentation events, both in space and on ground. The results are in general agreement with our analysis.

THE PRINCIPLE OF INCREASE OF CROSS-SECTIONAL AREA AND ITS APPLICATION TO THE FRAGMENTATION OF SATELLITES

Arjun Tan

INTRODUCTION

The cross-sectional area (CSA) is an important parameter in the study of collisions, scattering, ultracentrifuge and ballistics among others. In space research, the CSA determines the collision probability of a satellite in orbit (cf. Kessler, 1981). In the fragmentation of the Solwind P78-1 satellite, it was observed that the total CSA of the fragments was several times greater than the CSA of the satellite itself (Remillard, 1990). The same phenomenon is observed in almost all recorded satellite fragmentation events. It is this increase in CSA which is directly responsible for the increased potential hazard to large space structures, future space missions and useful orbiting satellites. This study analyzes the changes in CSA in the fragmentations of regular solids and leads to a fundamental principle of nature that the CSA of any solid must increase upon fragmentation. The results are applied to known cases of satellite breakups both on ground and in space.

THE CSAs OF COMMON REGULAR SOLIDS

We first calculate the CSAs of some common regular solids (S), the average value of the CSA over all possible orientations (<S>) and the ratio of <S> to the surface area of the solid (A).

The CSA of a Sphere. The CSA of a sphere of radius r in any direction is its projection on a plane perpendicular to that direction and is obviously equal to πr^2.

Formal Proof. We shall use spherical coordinates with the center of the sphere as origin (Fig. 1). An elementary surface area on the sphere is given by

$$ds = r^2 \sin\theta \, d\theta \, d\phi \quad . \tag{1}$$

The projection of ds on a plane perpendicular to any direction is equal to ds cosα, where α is the angle between the normal to the surface element and the given direction. Now the angle α between two points (θ_1, ϕ_1) and (θ_2, ϕ_2) on the sphere is given by (*cf.* Arfken, 1970)

$$\cos\alpha = \cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 \cos(\phi_2 - \phi_1) \tag{2}$$

Choose the x-axis as the line of sight. Then $\theta_1 = \pi/2$, $\phi = 0$; $\theta_2 = \theta$, $\phi_2 = \phi$; and

$$\cos\alpha = \sin\theta \cos\phi \quad . \tag{3}$$

From Eqs. (1) and (3),

$$ds \cos\alpha = r^2 \sin^2\theta \, d\theta \cos\phi \, d\phi \quad . \tag{4}$$

Integrating over the entire hemisphere facing the positive x-direction, we get

$$S = r^2 \int_0^{\pi} \sin^2\theta \, d\theta \int_{-\pi/2}^{\pi/2} \cos\phi \, d\phi = \pi r^2 \quad . \tag{5}$$

Since the sphere is symmetrical about any axis passing through the origin, $\langle S \rangle = S$ and $\langle S \rangle / A = 1/4 = 25\%$.

The CSA of a Flat Plate of Arbitrary Shape. Let the center-of-mass of the plate be the origin and let the x-axis be the direction of the line of sight (Fig. 2). The S = A cosα, where A is the surface area of the plate and α is the angle between the outward normal and the x-axis and is given by Eq. (3). Averaging over all possible orientations of the plate, we get

$$\langle S \rangle = A \, 2 \int_0^{\pi} \sin\theta \, d\theta \int_{-\pi/2}^{\pi/2} \cos\phi \, d\phi \; / \; 2 \int_0^{\pi} d\theta \int_{-\pi/2}^{\pi/2} d\phi = \frac{4}{\pi^2} A \quad . \tag{6}$$

Thus $\langle S \rangle / A = 4/\pi^2 = 40.5\%$.

The CSA of a Rectangular Parallelopiped. Let the dimensions of the rectangular parallelopiped be a, b and c. Then the total surface area is A = 2 (ab + bc + ca). In general three adjacent surfaces (out of a total six) are visible. The average CSA of the solid is equal to the sum of the averages of the

three adjacent surfaces. Since each surface can have any arbitrary orientation, from Eq. (6),

$$\langle S \rangle = \frac{4}{\pi^2} (ab + bc + ca) \quad . \tag{7}$$

Thus $\langle S \rangle / A = 2/\pi^2 = 20.26\%$.

The CSA of a Rectangular Cylinder. This is a special case of the above with b = a and c = h (the length of the cylinder). Here, A = 2a (a + 2h) and $\langle S \rangle = (4/\pi^2)$ a (a + 2h); whence $\langle S \rangle / A = 2/\pi^2 = 20.26\%$.

The CSA of a Cube. In this case, b = c = a, $A = 6a^2$, $\langle S \rangle = 12a^2/\pi^2$ and $\langle S \rangle / A = 2/\pi^2 = 20.26\%$

The CSA of a long Cylinder. Consider a long cylinder, the areas of whose end surfaces are negligible compared with that of the curved surface. Let r be the radius and h the length of the cylinder. Then the area of its curved surface $A = 2\pi rh$ and the area of an axial section of the cylinder is $A' = 2rh$. It is this latter area which will be responsible for the resulting CSA. Let the center-of-mass of the cylinder be the origin of coordinates and the x-axis be the line of sight. Then the orientation of the cylinder in ϕ does not alter the CSA, the orientation in θ only does. Consequently, we can choose the cylinder to lie in the z-x plane. Then θ is the angle between the normal to the cylinder in the z-x plane and the line of sight (x-axis), whence $S = A' \cos\theta$. Averaging over θ, we get

$$\langle S \rangle = 2A' \int_0^{\pi/2} \cos\theta \, d\theta \; / \; 2 \int_0^{\pi/2} d\pi = \frac{2}{\pi} A' \quad . \tag{8}$$

Again, $\langle S \rangle / A = 2/\pi^2 = 20.26\%$.

The CSA of a Regular Cylinder. In the case of a general cylinder, only one of the two (flat) end surfaces contributes to the CSA at any time. Let S be the CSA of the cylinder, S_1 be that of the curved surface and S_2 that of one end surface. Then $\langle S \rangle = \langle S_1 \rangle + \langle S_2 \rangle$. From the results of the previous sections, we obtain

$$\langle S \rangle = \frac{4}{\pi} r (r + h) \quad . \tag{9}$$

Since $A = 2\pi r (r + h)$, we have once again $\langle S \rangle / A = 2/\pi^2 = 20.26\%$. Note that the same result is also obtained by generalizing the rectangular cylinder into an n-sided polygonal cylinder and letting n tend to infinity.

The CSA of a Frustrum of a Cone. Let r_1 and r_2 be the radii of the end surfaces ($r_1 < r_2$), h the height of the frustrum and l the length of the edge. Then $l^2 = h^2 + (r_2 - r_1)^2$ and $A = \pi r_1^2 + \pi r_2^2 + \pi(r_1 + r_2)l$. Divide the frustrum into thin slabs of varying radii. The each slab may be approximated by a flat cylinder. Let S = CSA of the frustrum, S_l = CSA of the curved surface and S_t = CSA of the end surface. The $\langle S \rangle = \langle S_l \rangle + \langle S_t \rangle$. Adding contributions from the thin slabs, one gets

$$\langle S_l \rangle = \frac{2}{\pi} (r_1 + r_2) h \quad . \tag{10}$$

Since only one end surface contributes to the CSA at any time, we take the average CSA of the end surfaces:

$$\langle S_t \rangle = \frac{2}{\pi} (r_1^2 + r_2^2) \quad . \tag{11}$$

Adding (10) and (11):

$$\langle S \rangle = \frac{2}{\pi} [r_1^2 + r_2^2 + (r_1 + r_2) h] \quad . \tag{12}$$

The CSA of a Cone. In the previous result, put $r_1 = 0$ and $r_2 = r$. Then $A = \pi r(r + h)$ and

$$\langle S \rangle = \frac{2}{\pi} r (r + h) \quad . \tag{13}$$

THE CHANGE OF AVERAGE CSA IN THE FRAGMENTATION OF SOLIDS

We shall next study specific cases of fragmentation of solids and examine the change in the average CSA upon fragmentation. In the follwing, S' is the sum of the CSAs of all the fragments produced and $\langle S' \rangle$ is the average of S' over all possible orientations.

Change in the average CSA in the breakup of a Flat Plate. Consider a flat plate of arbitrary shape and area A. Let the plate break up into n flat fragments having areas A_1, A_2, A_3, ..., A_n. Then $A_1 + A_2 + A_3 + \ldots A_n = A$. Since each fragment itself is a flat plate, we have

$$\langle S' \rangle = \frac{4}{\pi^2} (A_1 + A_2 + A_3 + \ldots + A_n) = \frac{4}{\pi^2} A \quad . \tag{14}$$

This is equal to the average CSA of the original plate <S>. Thus the average CSA remains unchanged when a flat plate breaks up into various fragments.

Change in the Average CSA when a Rectangular Solid fragments into smaller solids. Let the rectangular solid having dimensions a, b and c break up into n identical solids having dimensions proportional to those of the parent solid. For this to be effective, n has to be the cube of a whole number. Then each fragment has the dimensions $a/n^{1/3}$, $b/n^{1/3}$ and $c/n^{1/3}$. Thus, we have

$$A' = n\, 2 \left(\frac{ab}{n^{2/3}} + \frac{bc}{n^{2/3}} + \frac{ca}{n^{2/3}} \right) = n^{1/3} A \quad . \tag{15}$$

Also, by Eq. (7)

$$\langle S' \rangle = n \frac{4}{\pi^2} \left(\frac{ab}{n^{2/3}} + \frac{bc}{n^{2/3}} + \frac{ca}{n^{2/3}} \right) = n^{1/3} \langle S \rangle \quad . \tag{16}$$

Thus, both the surface area and the average CSA increase as the cube root of the number of fragments. As an example, for $n = 125$, $\langle S' \rangle = 5\langle S \rangle$.

Change in the Average CSA when a Solid Sphere melts and disintegrates into smaller spheres. Let R be the radius of the sphere. Then its volume $V = (4/3)\pi R^3$ and $A = 4\pi R^2$ and $\langle S \rangle = S = \pi R^2$. Let the solid turn into n spheres of equal volume. From the conservation of volume, the radius of each small sphere is determined to be $r = R/n^{1/3}$. Thus

$$A' = n\, 4\pi \left(\frac{R}{n^{1/3}} \right)^2 = n^{1/3} A \quad . \tag{17}$$

and

$$\langle S' \rangle = n\pi \left(\frac{R}{n^{1/3}} \right)^2 = n^{1/3} \langle S \rangle \quad . \tag{18}$$

The dependence of the CSA increase on the number of fragments is identical to that of the preceeding case.

Change in the Average CSA when a Hollow Sphere breaks up into a large number of small Fragments. For a hollow sphere of radius R, <S> = S = πR^2. If the sphere breaks into a large number (n) of fragments, each of which can be treated as "flat", then S' = $4\pi R^2$, and

$$\langle S' \rangle = \frac{4}{\pi^2} 4\pi R^2 = \frac{16}{\pi^2} \langle S \rangle = 1.62 \langle S \rangle \quad . \tag{19}$$

Determination of n for flat fragment approximation. Let us assume that each fragment is circular having a radius r. Next apply the simple pendulum criterion: r/R = 5° = π/36 rad. Then the area of the fragment is ds = πr^2 = (π^2/36^2)R^2 and the number of fragments n = S/ds = 5184/π^2 = 525.

Change in the Average CSA when a Hollow Cylinder breaks up into a large number of small Fragments. As before, consider a hollow cylinder of radius r and lengtrh h fragment into a large number of small fragments, each of which may be treated as flat. We then have S' = $2\pi rh + 2\pi r^2$, and

$$\langle S' \rangle = \frac{4}{\pi^2} S' = \frac{8}{\pi} r (r + h) \quad . \tag{20}$$

By Eq. (9), <S'> = 2 <S>. In other words, the increase in the average CSA is twofold.

Change in the Average CSA when a Hollow Rectangular Cylinder breaks up into flat Fragments. Consider the rectangular cylinder of the previous section having dimensions a, a and h. Here, S' = A = 2a (a + 2h), and

$$\langle S' \rangle = \frac{4}{\pi^2} S' = \frac{8}{\pi^2} a (a + 2h) \quad . \tag{21}$$

From the results of the previous section, <S'> = 2 <S>. Again, the increase in the average CSA is twofold.

Change in the average CSA when a Hollow Cone breaks into flat fragments. Consider a hollw cone, the radius of whose base is r and the height of which is h. If the cone breaks up into a large number of fragments, each of which may be considered flat, the $S' = A = \pi r^2 + \pi r (h^2 + r^2)^{1/2}$, whence

$$\langle S' \rangle = \frac{4}{\pi^2} S' = \frac{4}{\pi} r [r + (h^2 + r^2)^{1/2}] \quad . \tag{22}$$

and

$$\langle S' \rangle / \langle S \rangle = 2 [r + (r^2 + h^2)^{1/2}] / (r + h) > 2 \quad . \tag{23}$$

The increase in the average CSA in this case is greater than two-fold.

PRINCIPLES OF CHANGE OF AVERAGE CSA I2N THE FRAGMENT2ATION OF SOLIDS

In the examples discussed above, one can notice that the average CSA increased in all fragmentation cases except one (the flat plate), where it remained constant. A closer scrutiny reveals the fact that there are two factors which are responsible for the increase. For thin sheet objects (flat plate, hollow sphere, hollow cylinder, etc.), it is the curvature which is responsible for the increase of CSA upon fragmentation. The average CSA remains constant only if the object sheet is flat (i.e., it has zero curvature). Any curvature (convexity or concavity) in the parent object results in an increase in the average CSA.

For solid objects (e.g., rectangular solids, solid spheres, solid cylinders, etc.), the surface area must necessarily increase upon fragmentation due to the creation of new surfaces. In consequence, the CSA and the average CSA must also increase. Fig. 3 illustrates this phenomenon for the breakup of a long cylinder into two and four fragments. Obviously, more newer surfaces (shaded) are created for each additional fragment. These new surfaces now contribute to the total CSA of the fragments, which consequently must increase, the increase being a function of the number of fragments produced.

On generalizing our results, we arrive at two fundamental principles of nature, which can be stated as follows:

Principle 1. The average CSA remains constant in the fragmentation of a flat plate.

Principle 2. The average CSA increases in the fragmentation of all other solids.

NB. There appears to be an exact parallel in the principles of change of CSA and entropy. For, the entropy principle states that the entropy of the universe may remain constant or increase in a thermal process, but cannot decrease.

We shall next re-examine Principle 1 for the real situation. The flat plate approximation is a mathematical abstraction. In reality, all plates must have finite thicknesses. And it is this extension in the third dimension which will contribute to the CSA and actually increase it. As an example, consider a square plate fragment into four small square plates (Fig. 4). Four new surfaces (shaded) are now exposed to the field of view, which will add to and increase the CSA. Thus, we can now modify our statements into a single general *law* which states: *The average CSA always increases in actual fragmentations of solids.*

NB. There is now absolute parallel between the principles of increase of average CSA and entropy, since the entropy of the universe increases in all real (i.e., irreversible) processes.

This parallelism between the average CSA and entropy is not surprising. It is well-known that the entropy is a measure of disorder and that natural events proceed in the direction of maximum disorder. The average CSA, similarly can be viewed as measure of mechanical disorder. Since the fragmentation of a solid must be an irreversible mechanical process, the average CSA is bound to increase upon it.

It is interesting to note that for solid objects in 1, 2, 3 and 4 dimensions, normally two geometrical measures are defined, of which only one (having the same dimension as the object) is conserved (Table I). For sheet materials, the total surface area is still conserved in a breakup, even though the CSA is not.

TABLE I. CONSERVATION OF GEOMETRICAL MEASURES

	Length	Surface Area	Volume	Hypervolume

1-d String	Yes			
2-d Sheet	No	Yes		
3-d Solid		No	Yes	
4-d Object			No	Yes

Increase of CSA due to Exposure of Internal Structures. If a hollow fragmenting object has internal structures, these may break off from the parent structure to form additional fragments. They may further fragment themselves to add to the fragments. Inside the parent structure, they are not contributory to the CSA. But once exposed, they have their own CSAs which contribute to the increase of CSA on fragmentation. As an example, consider a hollow sphere containing several spherical structures inside, fragment, exposing and releasing these structures (Fig. 5). The increase in CSA is then due to fragmentation plus the exposure of the internal structures.

It must be remembered that the internal structures may themselves fragment. Also, the internal structures may possess sub-structures, thus furthering the increase of the total CSA. In principle, *there is no upper limit to the increase of CSA in a fragmentation event.*

APPLICATIONS OF THE PRINCIPLES OF CHANGE OF AVERAGE CSA TO SATELLITE BREAKUPS ON GROUND AND IN SPACE

As of now, over one hundred satellites and rocket bodies have fragmented in orbit (cf. Nauer, 1992). Also some satellites and satellite mock-ups have been fragemnted in laboratory tests by hypervelocity impact (cf. Maethner, 1993). This study examines the CSA change in some of these recorded events in the light of the principles of change of CSA arrived at earlier. It must be stated that all satellites have complicated structures which may not be described easily. Nevertheless, most have some overall general shapes which can be usefully approximated.

Solar Panels. Solar panels are mainly planar and the flat plate approximation may be applied reliably. According to Principle 1, the breakup of solar panels should not increase the average CSA.

Cylindrical Rocket Bodies. The final stage rocket in normally left in space and functions like an independent satellite. Several such rocket bodies have exploded in orbit (cf. Nauer, 1992). Rocket bodies are by and large cylindrical in shape and also contain relatively fewer internal structures as compared with payload satellites. Even if the entire mass were assumed to be in a cylindrical drum, the thickness of a typical Delta 2nd stage rocket turns out to be merely 1 cm (Tan, 1992). Thus the hollow cylinder approximation may be applicable. According to our analysis, the average CSA doubles in a thorough fragmentation of a hollow cylinder.

Cylindrical Satellites. Payload satellites usually have various shapes, but the circular cylindrical shape predominates (cf. Caprara, 1986). Some, like the meteorological ITOS satellites are square cylindrical in shape (Caprara, 1986).

Spherical Satellites. Some payload satellites, including the first ever launched (Sputnik I) were perfect spheres. According to our analysis, the increase in average CSA in the breakup of a hollow sphere is 62%. Payload satellites are generally more densely packed than rocket bodies and the internal structures will provide additional increases in the average CSA.

Conical Shrouds. The nose cone shielding the payload during a launch is normally conical with a rounded tip and is also left in orbit together with the payload and the final stage rocket. The cone has very few internal structures and the frustrum approximation may be justifiable. However, no breakup of a nose cone has been reported so far.

RCS vs. CSA. The actual CSA of an object is not exactly the same as the cross-sectional area measured by RADAR (RCS). The RCS, like the CSA, of course varies with the orientation of the object. Sometimes, there is an order of magnitude variation in the measured RCS of the same object. The RCS of orbiting satellites are therefore averaged and entered in the catalog (cf. USSPACECOM RCS Catalog). The functional relationship between RCS and CSA has been obtained by

Badhwar and Anz-Meador (1989) after analyzing data from hundreds of orbiting satellites, rocket bodies and debris. If <R> is the average value of the RCS, then

$$\langle S \rangle = 0.5712 \langle R \rangle^{0.7666} \quad . \tag{24}$$

<R> must be converted to <S> before interpreting the data.

RESULTS AND DISCUSSIONS

As stated earlier, the results obtained are used to analyze several recorded satellite fragmentation events.

Solwind P78-1 Fragmentation. Solwind P78-1, a solar observation satellite was destroyed by a projectile in the only confirmed US ASAT test against a live target in space. The main body of the satellite was roughly cylindrical having a diameter of 2.1 m and length of 1.3 m (cf. Nauer, 1992). Attached to the main body was one nearly square solar panel of area 47 ft^2 (cf. Remillard, 1990), which translates to 4.366 m^2. The average CSAs on the main body and the panel are, according to Eqs. (6) and (9): $\langle S \rangle_{panel} = 1.77$ m^2 and $\langle S \rangle_{body} = 3.14$ m^2.

In order to analyze the Solwind breakup, the RCSs of the 267 largest fragments (out of a total 285 cataloged) were converted to their average CSAs by means of Eq. (24) and the aggregate CSA of the fragments summed out to be 16.56 m^2. Since the largest fragment had a CSA of only 0.57 m^2, or smaller than that of the main body or solar panel, both the main body and the solar panel must have been fragmented in the event. If the ASAT had hit the solar panel only, the main body (a relatively more robust structure) could have survived, but in that case, there would have been no increase in the average CSA of the fragments (flat plate). Thus the main body itself must have been struck by the ASAT. Subtracting $\langle S \rangle_{panel}$ (which does not change in the fragmentation) from the aggreagate CSA of the fragments, we obtain the average CSA of the body fragments $\langle S' \rangle_{body\ fragments} = 14.79$ m^2. As $\langle S \rangle_{body} = 3.14$ m^2, there was at least a 4.71-fold increase in the average CSA in the breakup of the main body. Since the fragmentation of a hollow cylinder merely doubles the average CSA, the balance must have been contributed

by the internal contents of the main body. This dramatic increase in the average CSA illustrates the fact that the satellite was densely packed with instruments and suffered a thorough fragmentation as a result of the impact. Recall that if a solid (with no cavities) broke up into 285 equal fragments having shapes identical to that of the parent solid, the increase in the average CSA would have been $285^{1/3}$, or 6.58-fold.

Titan 3C-4 Rocket Fragmentation. A Titan 3C-4 transtage rocket fragmented over the Indian Ocean, due presumably, to the accidental ignition of hypergolic fuel on board. It was circular cylindrical in shape, having a diameter of 3 m and length 6 m (cf. Nauer, 1992). According to Eq. (9), $\langle S \rangle_{rocket}$ = 14.3 m^2. The fragmentation was prolific with 469 trackable objects, the larget having an average CSA of only 0.82 m^2. Here the largest 83 fragments had an aggregate CSA of $2\langle S \rangle_{rocket}$. The aggregate CSA of all fragments exceeds this figure but not by much (the contributions from the smaller fragments decreased rapidly). Since the fragmentation of a hollow cylinder exactly doubles the average CSA, the rocket body was relatively hollow inside (as compared with payload satellites).

Ground-based Satellite Fragmentation Experiments. Satellite fragmentation tests using hypervelocity impact have been carried out recently in the laboratories to specifically characterize the fragment debris and to compare with satellite fragmentation models (cf. Maethner, 1993). The Satellite Orbital Debris Characterization Impact Test series was carried out at USAF Arnold Engineering Development Center during December 1991 - January 1992. It was sponsored by the Defense Nuclear Agency (DNA) and conducted by the USAF Wright Laboratory's Munitions Division (WL/MNSA). The resulting debris was characterized by General Research Corporation (GRC) and follow-up analysis was performed by Kaman Sciences Corporation. The targets included an Oscar satellite (a member of the Transit series of US Navy's Navigation Satellites) and a mock-up of Oscar. The targets were impacted by 150 gm projectiles at speeds of around 6 km/s.

The Federal Republic of Germany Impact Test (FRGIT) series was carried out in November 1991 at the GmbH test facility in Germany. It was also sponsored by DNA and conducted by WL/MNSA and the debris was again characterized by GRC. In

these tests, 3 mock satellites were impacted by 10 gm projectiles at speeds of around 10 km/s.

The fragment dimensions were individually measured, the average CSAs were calculated using Eqs. (7) and (9) and the results entered in Table II. The 5- to 7-fold increase in the average CSAs are quite consistent with the theoretical results obtained in this study.

Table II. Change in Average CSA in Ground-based Satellite Fragmentations.

	A, cm^2	<S>, cm^2	<S'>, cm^2	<S'>/<S>
FRGIT 1	7,998.49	1,620.83	11,288.27	6.96
FRGIT 2	7,998.49	1,620.83	9,492.47	5.86
FRGIT 3	7,998.49	1,620.83	8,466.29	5.22
Mock 2	8,685.74	1,742.87	11,307.81	6.49
Oscar	8,685.74	1,742.87	11,936.95	6.85

REFERENCES

G. Arfken, *Mathematical Methods for Physicists*, Academic Press, New York, 1970.

G. D. Badhwar and P. D. Anz-Meador, *Earth, Moon and Planets, 45*, 29, 1989.

G. Caprara, *The Complete Encyclopedia of Space Satellites*, Portland House, New York, 1986.

D. J. Kessler, *Icarus, 48*, 39, 1981.

S. R. Maethner, *Proc. SPIE/OE Conf.*, Orlando, 1993.

D. J. Nauer, *Teledyne Brown Engg. Rept. CS92-TR-JSC-007*, 1992.

S. K. Remillard, *Master's Thesis, Air Force Inst. Tech.*, 1990.

A. Tan, *USAF Phillips Lab. Int. Rept. WSSD-92-003*, 1992.

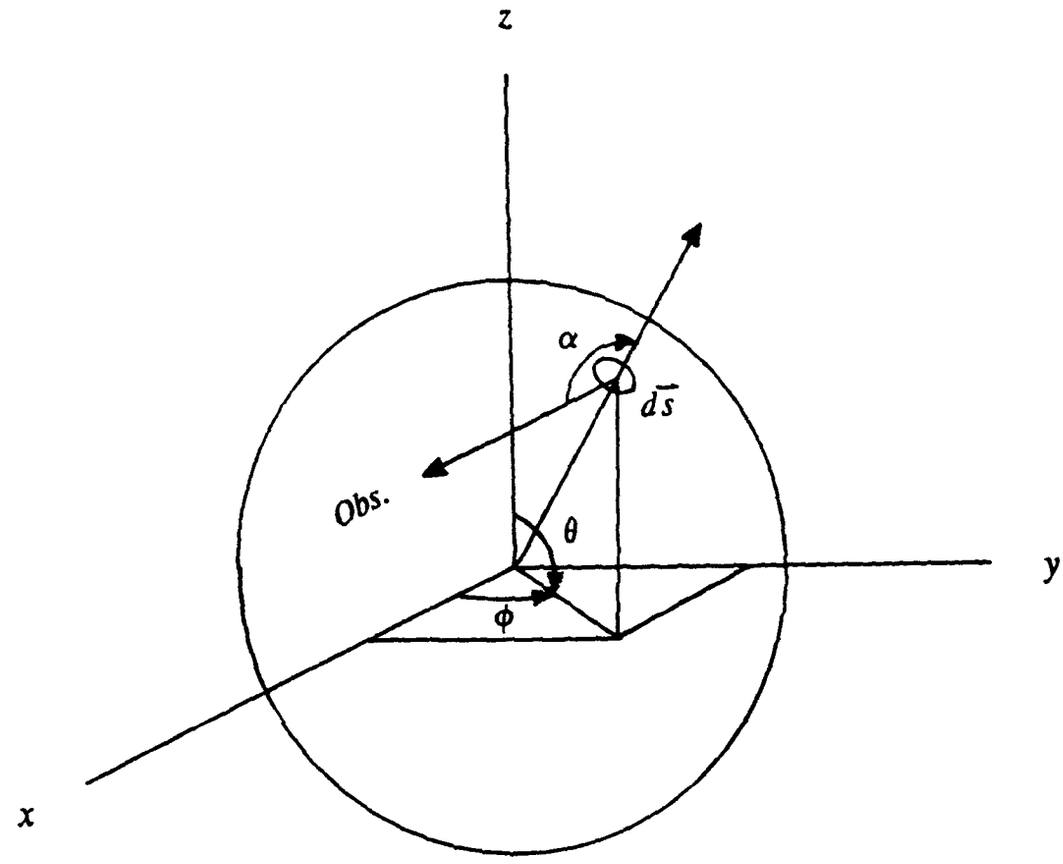


Fig. 1.

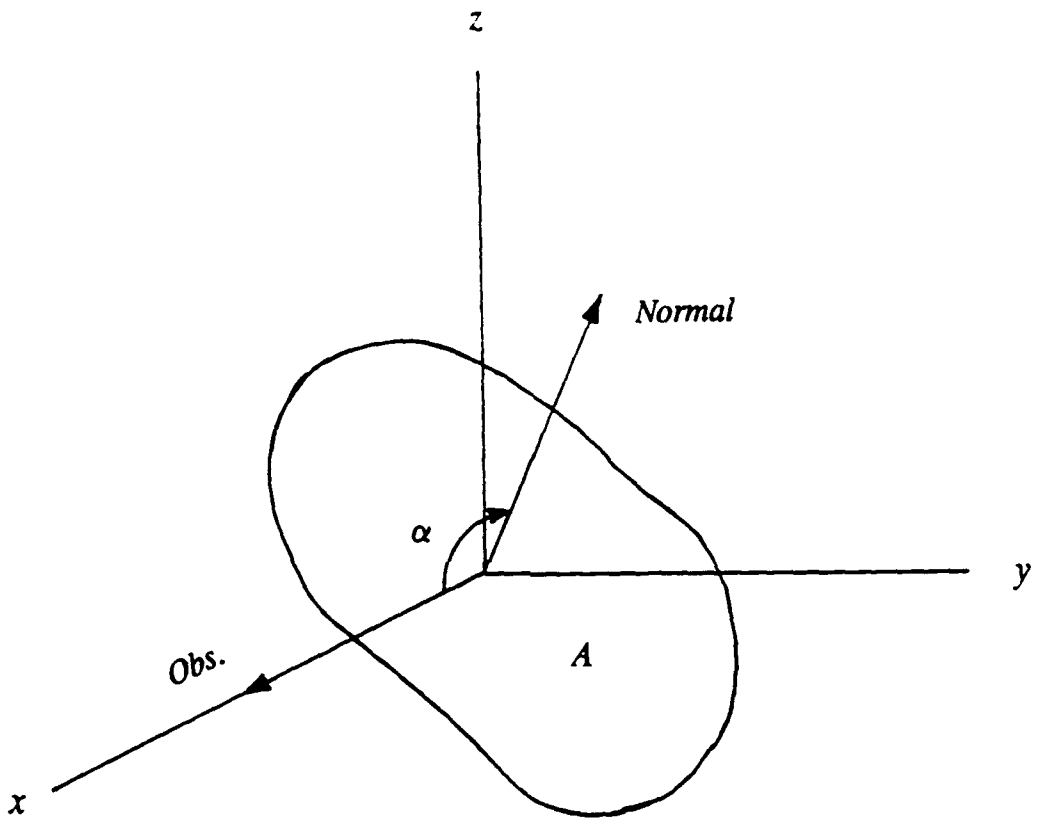


Fig. 2.

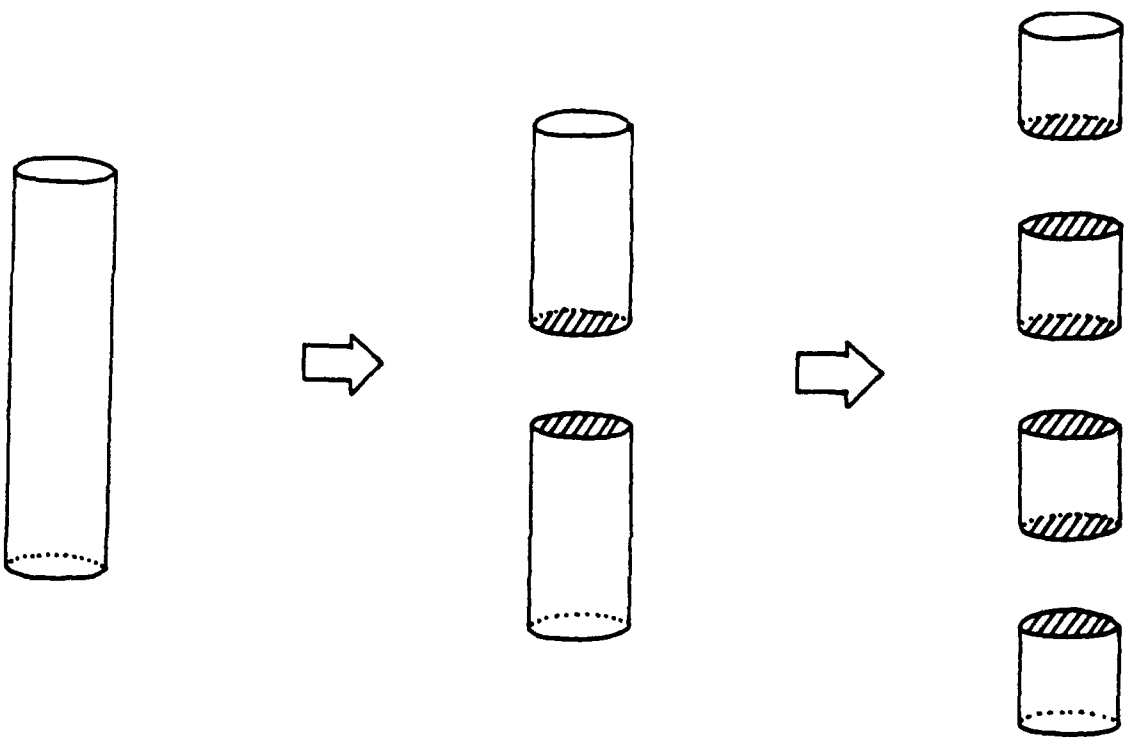


Fig. 3.

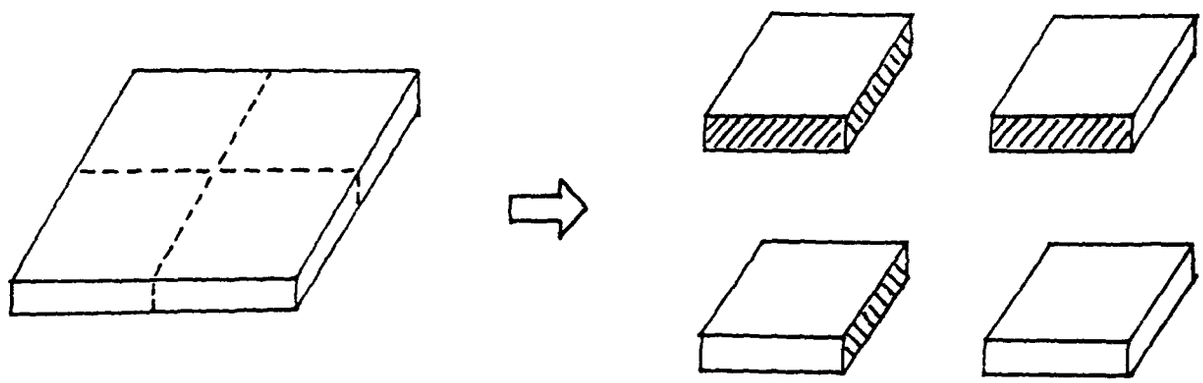


Fig. 4.

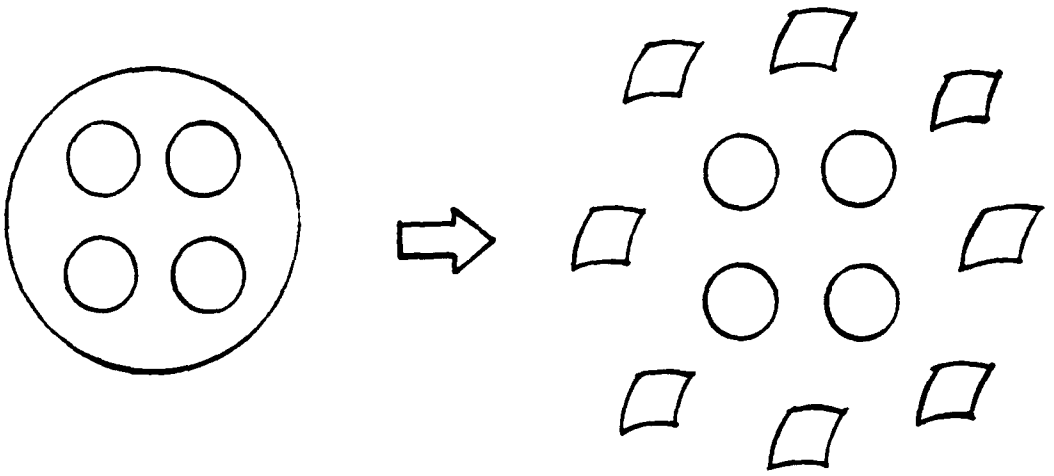


Fig. 5.

Chia Thio's report was unavailable at time of publication.

A PROBABILITY MODEL
FOR PREDICTION OF IMPACT WITH SPACE DEBRIS

Wayne J. Zimmermann

Professor

Department of Mathematics & Computer Science

Texas Woman's University

P.O.Box 22865

Denton, TX 76204

Final Report for:

Summer Faculty Research Program

Phillips Laboratory

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D.C.

August 1993

A PROBABILITY MODEL
FOR PREDICTION OF IMPACT WITH SPACE DEBRIS

Wayne J. Zimmermann

Introduction

As a result of our extensive excursions into the region referred to as low earth orbit (LEO) there is an ever increasing amount of man-made space debris which, in turn, increases the threat to the survivability of all space platforms, particularly those of designed for extended operation. Only recently, the last ten or fifteen years, have we realized that the low earth orbit is a finite resource. If we wish to continue to utilize this resource we must limit the amount of debris generated while developing schemes to remove some of the existing debris.

As to what constitutes man-made debris? Currently much of the existing debris includes old boosters, pieces of exploded bolts, parts of old dysfunctional satellites resulting from impacts and pieces of booster system resulting from an explosion or impact.

The present debris population is significant in that its associated threat is real. Hence, the space community must address a large number of problems related to this threat. Related questions include: Which trajectory should a flight engineer select if one is to minimize the probability of impact?

What measurements are needed to determine the probability of impact? What is the distribution of the debris mass in LEO? Of size? Of shape? What techniques can be used to measure such properties of a distribution? Is the distribution static? Is there a tendency for the debris to move to a state in which the distribution is uniform? Does the distribution vary so greatly that it is chaotic, unpredictable? If debris is not static is it growing? If so, how? Would it be possible to measure the growth rate? Is so how? Is space debris cascading towards uncontrolled growth? Equivalently, if an impact between two objects occurs is there enough new debris particles to cause another impact within a shorter period of time period [0,T] which in turn causes the process to repeat itself? If so, what is the cascade rate? If an object breaks up what effect does the breakup have on the debris distribution? How quickly can the effects of a breakup be ascertained? Is a breakup oriented relative to the satellite? Does orientation of the breakup alter the distribution of the debris? Can a breakup be simulated with confidence? Is is possible to determine the confidence in the models? If so how? What type of simulation should be used in modeling a breakup? Hydrocode, statistics or fractals? What type of object should be used in modeling a breakup due to explosion, a solid sphere or a hollow sphere filled with small steel shot, or a cylinder with electronic

boards? Should a breakup due to an explosion be preformed in a fluid and thereby allow improved data gathering? Would the environment, fluid, used in a breakup test cause the experiment to be void?

Clearly it is impossible to answer all of these questions in this paper. Hence we will restrict our discussion to a single well defined problem concerned with determination of probability of impact.

Overview of Space Debris Problem

To better understand the problem of determining the effect that space debris has on the life span of an operational system this section provides an overview of the phenomenon and indicates how the problem might be approached.

Consider the following situation: A flight is scheduled for the near future with its objective being that of placing a satellite into a selected orbit. Now it is known that LEO currently contains a significant amount of debris, both natural and man-made and that of the many problems facing the flight engineer there is the problem of selecting a trajectory which will not contain a life threating object capable of impacting the satellite for a period of T years. If all foreign objects are known, that is, their mass, size, shape and orbit parameters are known then the problem, though difficult, is solvable. If

there is a lack of knowledge, particularly the orbit parameters, or if there is a breakup of one or more foreign objects and the breakup occurs after launch time or if the size of the object are to small to identify but large enough to be real threat then the problem cannot be solved exactly. Further, there is the problem of defining threat, just how large must the piece of debris be if it is to pose a threat? If the debris particle is significantly small then there is the problem of observing the object. Clearly problems of detection introduce complexities in orbit determination. Currently, debris particles of one centimeter and larger are of concern since their velocity and mass are sufficiently large to cause considerable damage to the object.

A summary of the space debris problem can be found in greater detail in the text: Space Debris, [1].

Formal Definition of Problem

Let us begin with two well defined statments concerned with the overall space debris problem.

Version 1. Given a distribution of debris which is a function of time, solar pressure, height, number of explosions, impacts and gravitational forces, and also given an object in an orbit θ with a velocity V, determine the probability that a functional device does not impact a foreign object in a time

period [0,T].

Definition. Flux is defined as the number of impacts an object encounters in one year over a unit area while traveling at a velocity V. Dimensionally F is number of impacts per square meter per second.

Version 2. Given an object in an orbit θ with a velocity V determine the flux, that is, the number of impacts occurring per unit area per year.

Clearly a small flux correlates to a small probability of impact for a given time period [0,T]. In both versions the solution requires that the distribution of the debris population is known.

A Related Problem. Of considerable importance is the problem: how can a measure of confidence in the modeling of the flux and the measurement analysis be ascertained? The answer to this last question is importance since probability is a function of the value assigned to flux.

Determination of Debris Density

Clearly the determination of the environment is necessary if a solution to the degree of threat posed by the debris. To determine this threat it is necessary to perform some measurements in the region of interest. This can be done with radar or optically with either coupled with some form of computer

enhancement and sensitive sensor system. The manner in which the data by such devices can vary greatly. One such method, at least for the optical systems, is the stare mode. This approach is to simply count and record the number of object sited in a short period of time just before sunset. In [2] Seniw discusses the use of radar based measurements for determining flux. He also presents an algorithm which determines a total population estimate. This segment will be restricted to a discussion to that algorithm as modified for the processing of optical data.

The technique presented by Seniw, [2], begins with the assumptions that the objects are uniformily distributed in the right ascension and latitude. A question associated with this assumption is: Does this mean that the average mean distance between any two object is the same?

Seniw also assumes that the process is characterized by independent times of arrival and a constant rate of arrival. Hence, the process was taken to be Poisson which implies that the mean = standard deviation = τT.

Another question is: If a set of measurements has the property that the mean equals the standard deviation can it be said that the associated process is Poisson?

Definition. If the rate of arrival is proportional to the population then the population can be approached as a Poisson Process.

The process given here satisfies these assumption when viewed on short durations. Based on these remarks the following procedure is used to determine the distribution of the debris in the region defined by LEO.

Method: To determine the population the following steps are executed.

1. Select a star location at random for the telescope to maintain. For a period of n (say thirty) minutes count the number of objects moving through the field. The scope is to remain fixed on the selected star background.
2. Based on the time it takes for an object to pass through the sindow of observation and assuming circular oribits and estimating the direction of the object calculate the approximate distance between every pair of debris particles.
3. Repeat Step 1 for 15 days; selected at random.
4. Perform measurements. Determine the mean, μ, the mean distance between two particles.
5. Store measured data in a file containing records of the following form:

   IDENTIFICATION -- assigned.

   BETA (ß) -- angle of field of view.

   ECCENTRICITY -- computed for tracked objects.

TIME	-- time for object to cross field.
HEIGHT	-- computer by tracking or assuming circular trajectory.
LAT	-- latitude (measured).
LONG	-- longitude.
INCLINATION	-- inclination (measured).
ALBEDO	-- albedo (measured).
SIZE	-- computed based on albedo.
PERIOD	-- period of object.

6. Partition data into three bins for height--low, medium, and high.

7. Partition each partition by size, say, >=0.1, >=1.0, >=5.0, >=10.0, >=15.0, >=20.0, >=25.0, >=50.0

8. For each subpartition determine the total population using

$$N(h,d) = 2\pi \tau T_i r_i[\cos^2\theta - \cos^2 I_i]^{1/2} / (h_i \beta)$$

where the measured parameters are: τ is the average of the measurements, θ is the latitude, β is the angle of the field of view, and the computed parameters are: T_i the average period of the object, h_i is the average altitude of the objects, I_i is the average inclination of the objects, and $r_i = r_e + h_i$.

9. Use the population to determine the density and flux.

Density = population in bin/volume of bin.

9. Continued.

Define

flux-volume = area * velocity * time.

Flux = density * (flux-volume)

One of the weakness in the formula N(h,d) is due to the assumption that the process of debris detection is Poisson. This should be verified.

Flux / Probability Relation

This section derives a relation between FLUX and probability of no impact.

Question: How can the probability of no impact be determined? How does it relate to flux? To answer this question we consider a related problem for modeling atomic collision.

Assumption: The debris is relatively motionless as the object moves through segment 1. See Figure 1. The debris is assumed to move instanteously and then remain relatively motionless with respect to the object during the period for which the object occupies segment 2. This assumption is made for each successive segment 3 to n.

Notation: Let A_{ED} denote the effective area of the debris. Thus

$$A_{ED} = \Sigma \pi (r_i + r_s)^2$$

where r_i is the radius of the ith particle and r_s is the radius of the satellite. Let A_r denote the cross-sectional area of the toriod, the orbit. Then

$$P_r[\text{Impact occurs in seg 1}] = A^{cd} / A_r = P_I.$$

Thus

$$P_r[\text{Impact occurs in seg 2}] = (A_{cd}/A_r)*(A_r - A_{cd}) / A_r$$
$$= (1 - P_I)\ P_I.$$

In general

$$P_r[\text{Impact occurs in seg n}] = (1 - P_I)^{n-1}\ P_I.$$
$$= P_{I,n}$$

Since impact occurs in only one segment (assume destruction on impact) then

$$P_r[\text{Impact within segments 1 to n-1}] =$$
$$= P_{I,1} + P_{I,2} + \ldots + P_{I,n-1}$$
$$= P_I\ [1+(1 - p_I)+(1 - p_I)^2+\ldots+ (1 - p_I)^{n-1}]$$
$$= P_I\ (1 - p_I^n)/(1 - P_I)$$

If $\hat{a}_i$ denotes the average cross-sectional area of a debris particle then the probability

$$P_I = \hat{a}_i f/A_r$$

where f denotes the number of particle in the ith segment. Here

$$\hat{a}_i = \pi\ (\Sigma(r_i + r_s)^2)\ /\ f$$

If r_i = 0.01 m, r_s = 1 m and A_r = 10 m^2 then $P_I = 3.2047 \times 10^{-4}\ f_i$ Further if we assume the distribution of debris is uniform then $f_i = f$ for all i where f denote the flux. Hence

P_r[Impact in n segments] = $\alpha f[1 - (\alpha f)^n] / [1 - \alpha f]$

where $\alpha = 3.2047 \times 10^{-4}$. Thus

P_r[No Impace in n segments] = 1 - P_r[Impact in n segs]

$$= 1 - P_I(1 - P_I^n) / (1 - P_I)$$

$$= 1 - P_I(1 - 2P_I + P_I^{n+1})/(1 - P_I)$$

Since $\alpha f << 1$ then the term P_{In+1} is almost zero. Thus

$$P_r[\text{No Impact}] = (1 - 2 P_I)/(1 - P_I).$$

In one year at a altitude of 600 km a satellite will circle the earth approximately 5400 times. If a single revolution is divided into 72 segments the n is very large. Hence $P_I^n = 0$. Returning

$$P_r = P_r[\text{No Impact}] = (1 - 2 \alpha f) / (1 - \alpha f)$$

Solving for f, the flux in a single segment, we have

$$f = (1 - P_r)/(\alpha(2 - P_r))$$

Using this relationship we can generate the following table:

P_r	f
.9000	283.61
.9900	30.89
.9990	3.12
.9999	.31

Interval Analysis

We are now in a position to address the related problem: how can some measure of confidence of the flux model be

ascertained? Initially, it was suggested that we obtained a measure similar to that used in reliability analysis. Due to the complexity of the problem the approach taken here is a modified sensitivity analysis.

This approach is based on the use of confidence intervals derived from the best fit to the acquired data. Once this is done we then reformulate the selected model using interval analysis. This provides us with an upper bound and a lower bound which are based on the associated statistical confidence intervals.

Review of Interval Analysis. We include a few concepts on interval analysis. By an interval we mean:

$$[a,b] = \{x : a \leq x \leq b \}.$$

We begin by noting that intervals can be treated as numbers: Hence, if X, Y and Z are intervals the arithmetic operations are defined as follows:

* $X + Y = \{ x + y : x \in X, y \in Y \}$

* $- X = - [X_{Lb}, X_{Ub}] = [-X_{Ub}, -X_{Lb}]$

* $1 / X = \{ 1/x : x \in X \}$. If X contains 0 then 1 / X is not defined.

* $XY = \{ xy : x \in X, y \in Y \}$

XY is an interval and its endpoints can be computed from $XY_{Lb} = \min\{X_{Lb}Y_{Ub}, X_{Lb}Y_{Ub}, X_{Ub}Y_{Lb}, X_{Ub}Y_{Ub}\}$

$XY_{Ub} = \max\{X_{Lb}Y_{Ub}, X_{Lb}Y_{Ub}, X_{Ub}Y_{Lb}, X_{Ub}Y_{Ub}\}$

* An interval matrix is a matrix whose elements are intervals

Two special intervals are: [0,0], denoted by 0 and [1,1], denoted by 1. Hence $X + 0 = X$ and $X * 1 = X$.

Algebraic properties are an immediate property of the previous definitions. They include:

$$X + (Y + Z) = (X + Y) + Z$$

$$X(YZ) = (XY)Z$$

$$X + Y = Y + X$$

$$XY = YX$$

Note, the distributive law does not always hold, that is, it is not always true that:

$$X\ (Y + Z) = XY + XZ$$

What is true is that $X\ (Y + Z)$ is contained in $XY + XZ$.

For a real x and an interval X, xX is an interval.

Review of Engineering Flux Model. To see how the interval analysis can be applied we review the various parameters appearing in the Engineering Flux Models. Kessler, Reynolds, Anz-Meador, [3], presents the model as:

$$F(d,h,i,t,S) = k\ \Phi(h,S)\ PHI(i)\ [F_1(d)g_1(t) + F_2(d)g_2(t)]$$

where

d: denotes the size of the debris,

h: altitude,

i: inclination,

t: time in years,

S: solar activity for the previous year.

k: a number associated with the tumbling and value is 1 for a randomly tumbling surface.

PHI(i): Debris correction factor (tabular) as a function of inclination

$F_1(d)$: size distribution for small fragments--

$:= 1.05 \times 10^{-5}\ d^{-2.5}$

$g_1(t)$: growth rate for small particles --

$:= (1 + 2 \cdot p)^{(t-1985)}$

$F_2(d)$: size distribution for large fragments--

$:= 7.0 \times 10^{10}(d+700)^{-6}$

$g_2(t)$: growth rate for large particles -- $:= (1 + p)^{(t-1985)}$

Figure 2 [Kessler] is a comparison of the flux arising from Jan '87 track population plus the Catalog vs the catalog.

The model

$$F(d,h,i,t,S) = k\ \Phi(h,S)\ PHI(i)\ [F_1(d)g_1(t)+F_2(d)g_2(t)]$$

indicates that the flux is a function of the diameter of the particle, the altitude, the inclination, time and solar activity. What the Kessler model fails to do is provide the probability of impact. Having already addressed this point we now consider the problem of determining confidence intervals.

To do this we simply apply a nonlinear regression fit to the data used to generate $\Phi(h,S)$. In doing so we found that if

the data, which is generally represented in log-log form, is transformed to rectangular coordinates the resulting graphical display indicates that the flux is a mixture of two normal distributions. Each distribution could be attributed to a specific altitude. Hence, the log-log representation could be approximated by a mixture of second degree equations. Based on this argument we determined that the quadratic regression and its associated 90% confidence intervals could be used to provide boundaries in the models by using the first the lower value in the model for calculating flux, then use the larger value in the same model. See Figure 3.

Conclusions

The problem is at best one of estimating flux by use of a model. It appears that the best procedure for determining the flux or probability would be to simulate the phenomenon on a computer, maintaining the trajectories of a large number of debris particles, say twenty thousand. It is known that the smaller particles are dynamically interesting, that is, they possess strange trajectories and their locations are difficult to predict. Further, small particles are simply difficult to observe. As a result, estimating parameters defining their population is almost impossible. Yet, it is known that for the relative velocities involved in the situation described here

these small particles do present a real threat to functional satellites. Further, the creation of new debris particles as a result of explosions or impacts alter the environment. Many of these new particles are not known even though they are large enough to track. Most are too small to observe. In modeling, it would be necessary to predict the velocities and the mass of such particles. This is a current area of research referred as Breakup Modeling. Once all segments of the problem are modeled the probability of impact could be estimated by the Monti Carlo method.

References

[1] Moore, R.E., Interval Analysis, Prentice Hall, Englewood Cliffs, 1966

[2] Seniw, W.P., Space Population Study--Final Report, MIT--Lincoln Laboratory, March 1992, Rome AFB, Lincoln Laboratory (Point of Contact: Gerald P. Banner)

[3] Kessler, Reynolds, Anz-Meador, Orbital Debris Environment for Spacecraft Designed to Operate in Low Earth Orbit, NASA Technical Memo. 100471, April 1989

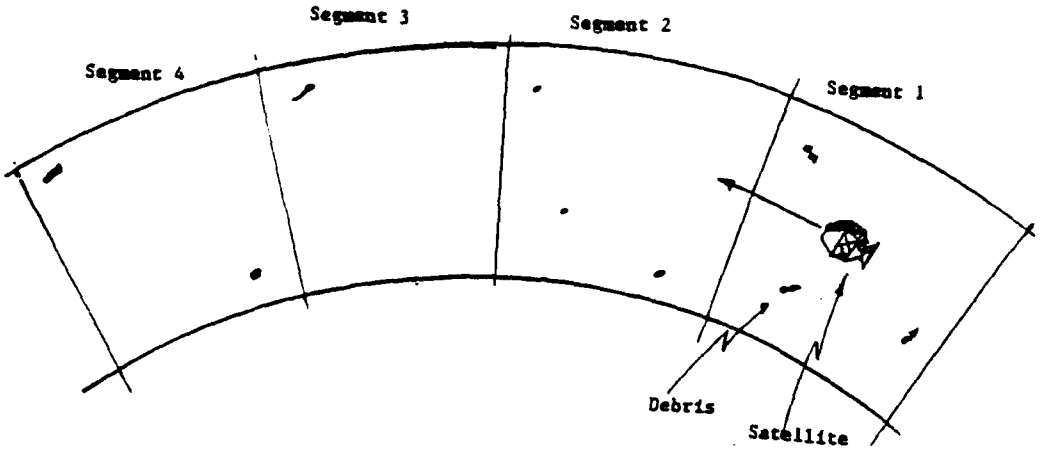


Figure 1. A diagram illustrating a 2-D model of debris environment possessing small relative velocities.

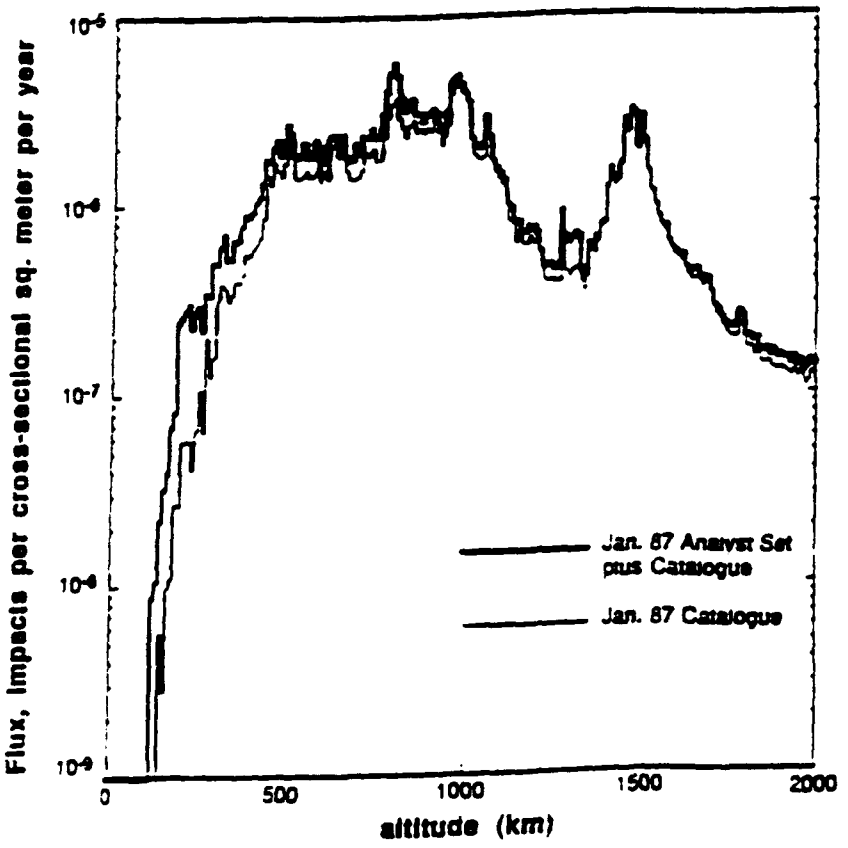


Figure 2. A comparison of the flux arising from the January 1987 tracked population (Analyst Set plus Catalog) and the January 1987 Catalog.

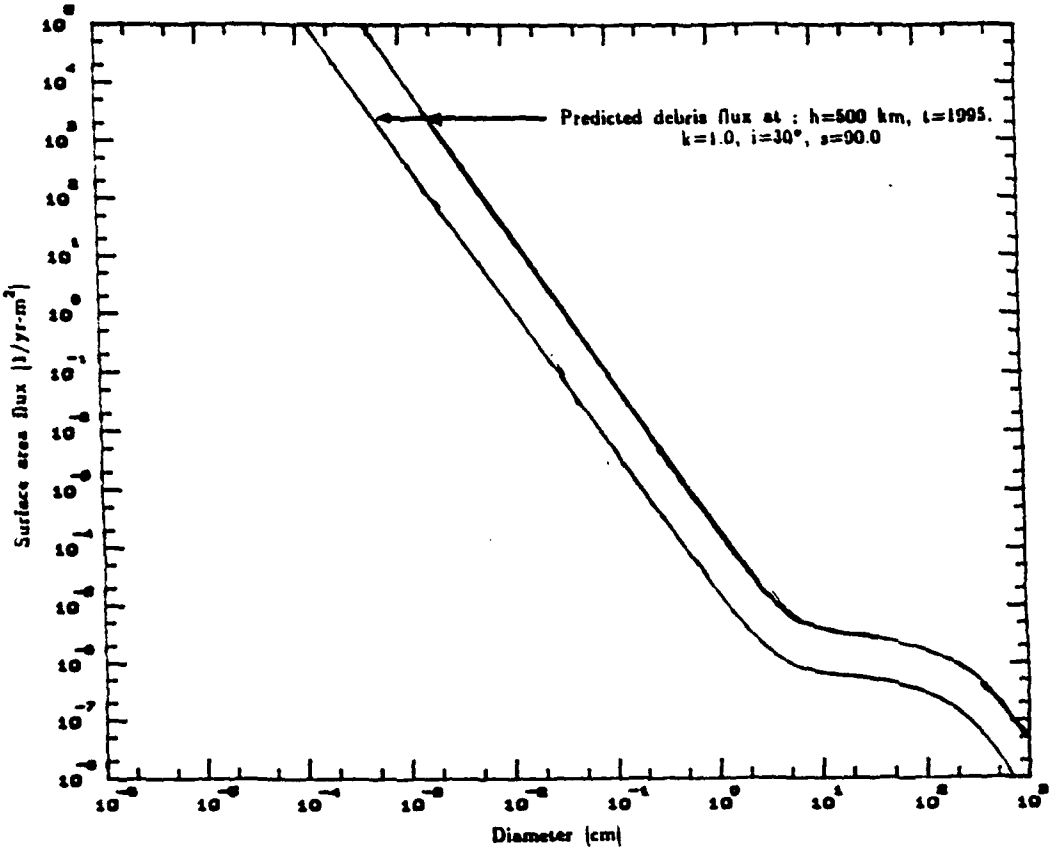


Figure 3. Predicted flux levels using the Engineering Model with interval analysis.