Copy 1.

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1992

SUMMER RESEARCH EXTENSION PROGRAM FINAL REPORTS

VOLUME 4B

WRIGHT LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 UPLANDER WAY
CULVER CITY, CA 90230-6608

SUBMITTED TO:

LT. COL. CLAUDE CAVENDER
PROGRAM MANAGER

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

BOLLING AIR FORCE BASE
WASHINGTON, D.C.

MAY 1993

REPOR, DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(leave blank)*

2. REPORT DATE
28 Dec 92

3. REPORT TYPE AND DATES COVERED
Annual 1 Sep 91 - 31 Aug 92

4. TITLE AND SUBTITLE

1992 Summer Faculty Research Program (SFRP)
Volume: 4B (SREP)

5. FUNDING NUMBERS

F49620-90-C-0076

6. AUTHOR(S)

Mr Gary Moore

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Research & Development Laboratoreis (RDL)
5800 Uplander Way
Culver City CA 90230-6600

8. PERF

AFOSR-TR-95
0725

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR/NI
110 Duncan Ave., Suite B115
Bldg 410
Bolling AFB DC 20332-0001
Lt Col Claude Cavender

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The purpose of this program is to develop the basis for cintinuing research of interest to the Air Force at the institution of the faculty member; to stiumlate co continuing relations among faculty members and professional peers in the Air Force to enhance the research interests and capabilities of scientific and engineering educators; and to provide follow-on funding for research of particular promise that was started at an Air Force laboratory under the Summer Faculty Research Program.

During the summer of 1992 185 university faculty conducted research at Air Force laboratories for a period of 10 weeks. Each participant provided a report of their research, and these reports are consolidated into this annual report.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1992

SUMMER RESEARCH EXTENSION PROGRAM FINAL REPORTS

VOLUME 4B

WRIGHT LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES
5800 Uplander Way
Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Lt. Col. Claude Cavender

Program Manager, RDL
Scott Licoscos

Program Administrator, RDL
Gwendolyn Smith

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Bolling Air Force Base
Washington, D.C.
May 1993

PREFACE

This volume is part of a five-volume set that summarizes the research of participants in the 1992 AFOSR Summer Research Extension Program (SREP). The current volume, Volume 4B of 5, presents the final reports of SREP participants at Wright Laboratory.

Reports presented in this volume are arranged alphabetically by author and are numbered consecutively -- e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3, with each series of reports preceded by a 22-page management summary. Reports in the five-volume set are organized as follows:

VOLUME	TITLE
1A	Armstrong Laboratory (part one)
1B	Armstrong Laboratory (part two)
2	Phillips Laboratory
3	Rome Laboratory
4A	Wright Laboratory (part one)
4B	Wright Laboratory (part two)
5	Air Force Civil Engineering Laboratory, Arnold Engineering Development Center, Frank J. Seiler Research Laboratory, Wilford Hall Medical Center

1992 SUMMER RESEARCH EXTENSION PROGRAM FINAL REPORTS

Wright Laboratory

Wright Laboratory (cont'd)

VOLUME 4B (cont'd)

1992 SUMMER RESEARCH EXTENSION PROGRAM (SREP) MANAGEMENT REPORT

1.0 BACKGROUND

Under the provisions of Air Force Office of Scientific Research (AFOSR) contract F49620-90-C-0076, September 1990, Research & Development Laboratories (RDL), an 8(a) contractor in Culver City, CA, manages AFOSR's Summer Research Program. This report is issued in partial fulfillment of that contract (CLIN 0003AC).

The name of this program was changed during this year's period of performance. For that reason, participants' cover sheets are captioned "Research Initiation Program" (RIP), while the covers of the comprehensive volumes are titled "Summer Research Extension Program" (SREP). The program's sponsor, the Air Force Office of Scientific Research (AFOSR), changed the name to differentiate this program from another which also bore its original name.

Apart from this name change, however, the program remained as it has been since its initiation as the Mini-Grant Program in 1983. The SREP is one of four programs AFOSR manages under the Summer Research Program. The Summer Faculty Research Program (SFRP) and the Graduate Student Research Program (GSRP) place college-level research associates in Air Force research laboratories around the United States for 8 to 12 weeks of research with Air Force scientists. The High School Apprenticeship Program (HSAP) is the fourth element of the Summer Research Program, allowing promising mathematics and science students to spend two months of their summer vacations at Air Force laboratories within commuting distance from their homes.

SFRP associates and exceptional GSRP associates are encouraged, at the end of their summer tours, to write proposals to extend their summer research during the following calendar year at their home institutions. AFOSR provides funds adequate to pay for 75 SREP subcontracts. In addition, AFOSR has traditionally provided further funding, when available, to pay for additional SREP proposals, including those submitted by associates from Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs). Finally, laboratories may transfer internal funds to AFOSR to fund additional SREPs. Ultimately the laboratories inform RDL of their SREP choices, RDL gets AFOSR approval, and RDL forwards a subcontract to the institution where the SREP associate is employed. The subcontract (see Attachment 1 for a sample) cites the SREP associate as the principal investigator and requires submission of a report at the end of the subcontract period.

Institutions are encouraged to share costs of the SREP research, and many do so. The most common cost-sharing arrangement is reduction in the overhead, fringes, or administrative changes institutions would normally add on to the principal investigator's or research associate's labor. Some institutions also provide other support (e.g., computer run time, administrative assistance, facilities and equipment or research assistants) at reduced or no cost.

When RDL receives the signed subcontract, we fund the effort initially by providing 90% of the subcontract amount to the institution (normally $18,000 for a $20,000 SREP). When we receive the end-of-research report, we evaluate it administratively and send a copy to the laboratory for a technical evaluation. When the laboratory notifies us the SREP report is acceptable, we release the remaining funds to the institution.

2.0 THE 1992 SREP PROGRAM

SELECTION DATA: In the summer of 1991, 170 faculty members (SFRP associates) and 142 graduate students (GSRP associates) participated in the summer program. Of those, 147 SFRPs and 10 GSRPs submitted SREP proposals; 88 SFRP SREPs and 7 GSRP SREPs were selected for funding (total: 95).

	Summer 1991 Participants	Submitted SREP Proposals	SREPs Funded
SFRP	170	147	88
GSRP	142	10	7

The funding was provided as follows:

Contractual slots funded by AFOSR	75
Laboratory-funded	13
Additional funding from AFOSR	7
Total	95

Seven HBCU/MI associates from the 1991 summer program submitted SREP proposals; five were selected (one was lab-funded; four were funded by additional AFOSR funds).

By laboratory, the applications submitted and selected show in the following table:

	Applied	Selected
Air Force Civil Engineering Laboratory	6	4
Armstrong Laboratory	34	20
Arnold Engineering Development Center	12	2
Frank J. Seiler Research Laboratory	5	3
Phillips Laboratory	30	18
Rome Laboratory	16	11
Wilford Hall Medical Center	1	1
Wright Laboratory	53	36
TOTAL	157	95

Note: Phillips Laboratory funded 2 SREPs; Wright Laboratory funded 11; and AFOSR funded 7 beyond its contractual 75.

ADMINISTRATIVE EVALUATION: The administrative quality of the SREP associates' final reports was satisfactory. Most complied with the formatting and other instructions RDL provided to them. In the final days of December 1992 and in the first two months of 1993, several associates called and requested no-cost extensions of up to six months. After consultation with our AFOSR Contracting Officer's Representative, RDL approved the requests but asked that all such associates provide an interim report to be included in this volume. That caused an AFOSR-approved delay beyond the 1 April 1993 submission of this report. The subcontracts were funded by $1,893,616 of Air Force money. Institutions' cost sharing amounted to $948,686.

TECHNICAL EVALUATION: The form we used to gather data for technical evaluation and the technical evaluations of the SREP reports are provided as Attachment 2. This summary evaluation is shown by SREP number. The average rating range was from 3.1 to 5.0. The overall average for those evaluated was 4.6 out of 5.00. The three rating factors with the highest average scores were:

- o The USAF should continue to pursue the research in this RIP report.
- o The money spent on this RIP report was well worth it.
- o I'll be eager to be a focal point for summer and RIP associates in the future.

Thus it is clear that the laboratories place a high value on AFOSR's Summer Research Program: SFRP, GSRP, and SREP.

3.0 SUBCONTRACTS SUMMARY

Table 1 lists contractually required information on each SREP subcontract. The individual reports are published in volumes as follows:

Laboratory	Volume
Air Force Civil Engineering Laboratory	5
Armstrong Laboratory	1
Arnold Engineering Development Center	5
Frank J. Seiler Research Laboratory	5
Phillips Laboratory	2
Rome Laboratory	3
Wilford Hall Medical Center	5
Wright Laboratory	4

TABLE 1: SUBCONTRACTS SUMMARY

Researcher's name	Institution	Location	Highest Degree	Subcontract Number	Duration	Department	Amount	Sharing
Abbott, Ben A	Vanderbilt University	Nashville, TN 37235	MS	135	01/01/92-12/31/92	Dept of Electrical Engineering	19966.00	0.00
Acharya, Raj	State University of New York, Buffalo	Buffalo, NY 14260	PhD	151	01/01/92-12/31/92	Dept of Electrical & Comp Engrg	20000.00	0.00
Adams, Christopher M	Oklahoma State University	Stillwater, OK 74078	PhD	68	01/01/92-12/31/92	Dept of Chemistry	20000.00	0.00
Anderson, Richard A	University of Missouri, Rolla	Rolla, MO 65401	PhD	50	01/01/92-12/31/92	Dept of Physics	20000.00	5000.00
Arora, Vijay K	Wilkes University	Wilkes-Barre, PA 18766	PhD	3	10/01/91-09/30/92	Dept of Electrical & Comp Engrg	19996.00	36208.00
Ball, William P	Duke University	Durham, NC 27706	PhD	71	01/01/92-12/31/92	Dept of Civil & Environmental Eng	20000.00	26747.00
Battles, Frank P	Massachusetts Maritime Academy	Buzzard's Bay, MA 025321803	PhD	152	01/01/92-12/31/92	Dept of Basic Sciences	20000.00	22000.00
Bieniek, Ronald J	University of Missouri, Rolla	Rolla, MO 65401	PhD	147	01/01/92-12/31/92	Dept of Physics	19945.00	4000.00
Blystone, Robert V	Trinity University	San Antonio, TX 78212	PhD	127	01/01/92-12/31/92	Dept of Biology	20000.00	14783.00
Cha, Soyoung S	University of Illinois, Chicago	Chicago, IL 60680	PhD	011	01/01/92-12/31/92	Dept of Mechanical Engineering	20000.00	3842.00
Chandra, D. V. Satish	Kansas State University	Manhattan, KS 66506	PhD	89	01/18/92-10/17/92	Dept of Electrical Engineering	20000.00	1117[illegible].00
Chenette, Eugene R	University of Florida	Gainesville, FL 32611	PhD	106	01/01/92-12/31/92	Dept of Electrical Engineering	20000.00	0.00
Christensen, Douglas A	University of Utah	Salt Lake City, UT 84112	PhD	83	01/01/92-12/31/92	Dept of Electrical Engineering	19999.00	5000.00

Chubb, Gerald P Ohio State University Columbus, OH 43235	PhD	26	01/01/92-12/31/92	Dept of Aviation	20000.00	7600.00
Courter, Robert W Louisiana State University Baton Rouge, LA 70803	PhD	8	10/01/91-09/30/92	Dept of Mechanical Engineering	20000.00	445.00
Dey, Pradip P Hampton University Hampton, VA 23668	PhD	120	01/01/92-12/31/92	Computer Science Department	19921.00	0.00
Draut, Arthur W Embry Riddle Aeronautical University Prescott, AZ 86301	PhD	133	01/06/92-05/08/92	Computer Science Dept	19431.00	0.00
Dreisbach, Joseph University of Scranton Scranton, PA 185104626	PhD	108	12/01/91-12/01/92	Dept of Chemistry	20000.00	4000.00
Dror, Itiel Harvard University Cambridge, MA 02138	BS	76	01/01/92-12/31/92	Dept of Psychology	20000.00	0.00
Drost-Hansen, W. University of Miami Coral Gables, FL 33124	PhD	124	12/01/91-12/01/92	Dept of Chemistry	20000.00	12000.00
Dunleavy, Lawrence P University of South Florida Tampa, FL 33620	PhD	41	01/01/92-12/31/92	Dept of Electrical Engineering	20000.00	6463.00
Evans, Joseph B University of Kansas Lawrence, KS 66045	PhD	96	01/01/92-12/31/92	Dept of Electrical & Comp Engrg	20000.00	0.00
Flowers, George T Auburn University Auburn, AL 368495341	PhD	73	01/01/92-12/30/92	Dept of Mechanical Engineering	19986.00	12121.00
Gantenbein, Rex E University of Wyoming Laramie, WY 82071	PhD	22	01/01/91-12/31/92	Dept of Computer Science	20000.00	26643.00
Garcia, Ephrarim Vanderbilt University Nashville, TN 37235	PhD	32	12/01/91-11/30/92	Dept of Mechanical Engineering	20000.00	9659.00
German, Fred J Auburn University Auburn University, AL 36830	PhD	49	01/01/92-12/31/92	Dept of Electrical Engineering	20000.00	0.00
Gould, Richard D North Carolina State University Raleigh, NC 276957910	PhD	87	01/01/92-12/31/92	Dept of Mech and Aerospace Engrg	20000.00	14424.00
Gove, Randy L University of Alabama, Huntsville Huntsville, AL 35899	MS	122	01/01/92-12/31/92	Dept of Physics	20000.00	3469.00
Grabowski, Marek University of Colorado, Colorado Springs Colorado Springs, CO 809337150	PhD	92	01/01/92-12/31/92	Dept of Physics	19700.00	0.00

Gunaratne, Manjriker University of South Florida Tampa, FL 33620	PhD 90 01/01/92-12/31/92 Dept of Civil Engrg & Mechanics 19994.00 10062.00
Hall, Ernest L University of Cincinnati Cincinnati, OH 452210072	PhD 134 01/01/92-12/31/92 Dept of Robotics Research 19975.00 0.00
Hamilton, William L Salem State College Salem, MA 01970	PhD 47 01/01/92-12/31/92 Dept of Geography 20000.00 32000.00
Hamilton, Kirk L Xavier University of Louisiana New Orleans, LA 70125	PhD 57 01/01/92-12/31/92 Dept of Biology 20000.00 16100.00
Harris, Harold H University of Missouri, St.Louis St. Louis, MO 63121	PhD 94 01/01/92-12/31/92 Dept of Chemistry 19300.00 8600.00
Hartung, George H University of Hawaii Honolulu, HI 96822	PhD 46 01/01/92-12/31/92 Dept of Physiology 20000.00 7530.00
Hatfield, Steven L University of Kentucky Lexington, KY 40506	BS 23 01/01/92-12/31/92 Dept of Materials Science & Engrg 20000.00 28625.00
Hedman, Paul O'Dell Brigham Young University Provo, UT 84602	PhD 17 01/01/92-12/31/92 Dept of Chemical Engineering 19999.00 6928.00
Heister, Stephen D Purdue University West Lafayette, IN 47907	PhD 5 01/01/92-12/31/92 School of Aero & Astronautics 20000.00 4419.00
Hess, David J University of Texas, Austin Austin, TX 78713	BA 149 01/01/92-12/31/92 Dept of Psychology 19914.00 8784.00
Hoffman, R. W Case Western Reserve University Cleveland, OH 44106	PhD 99 01/01/92-12/31/92 Dept of Physics 19770.00 0.00
Huerta, Manuel A University of Miami Coral Gables, FL 33124	PhD 62 01/01/92-12/31/92 Dept of Physics 20000.00 1207.00
Hui, David University of New Orleans New Orleans, LA 70148	PhD 116 01/01/92-12/31/92 Dept of Mechanical Engineering 20000.00 0.00
Iyer, Ashok University of Nevada, Las Vegas Las Vegas, NV 89154	PhD 74 01/01/92-12/31/92 Dept of Electrical & Comp Engrg 20000.00 18549.00
Khonsari, Michael M University of Pittsburgh Pittsburgh, PA 15260	PhD 53 01/01/92-12/31/92 Dept of Mechanical Engineering 20000.00 32958.00
Kibert, Charles J University of Florida Gainesville, FL 32611	PhD 2 01/01/92-12/31/92 Dept of Fire Testing & Research 20000.00 6928.00

Klarup, Douglas G	PhD	84	01/01/92-12/31/92
University of Montana	Dept of Chemistry		
Missoula, MT 59812	20000.00		0.00
Koblasz, Arthur J	PhD	145	01/01/92-09/30/92
Georgia Institute of Technology	Dept of Civil Engineering		
Atlanta, GA 30332	19956.00		0.00
Kornreich, Philipp	PhD	35	10/01/91-09/30/92
Syracuse University	Dept of Electrical & Comp Engrg		
Syracuse, NY 13244	20000.00		0.00
Kuo, Spencer P	PhD	59	01/01/92-12/31/92
Polytechnic University	Dept of Electrical Engineering		
Farmingdale, NY 11735	20000.00		9916.00
Langhoff, Peter W	PhD	115	01/01/92-12/31/92
Indiana University	Dept of Chemistry		
Bloomington, IN 47402	20000.00		35407.00
Lee, Byung-Lip	PhD	93	01/01/92-12/31/92
Pennsylvania State University	Dept of Engrg Science & Mechanics		
University Park, PA 16802	20000.00		8173.00
Leigh, Wallace B	PhD	118	01/01/92-12/31/92
Alfred University	Dept of Electrical Engineering		
Alfred, NY 14802	19767.00		18770.00
Liddy, Elizabeth	PhD	104	01/01/92-12/31/92
Syracuse University	Dept of Information Studies		
Syracuse, NY 132444100	20000.00		0.00
Liu, Cheng	PhD	6	11/01/99-12/31/92
University of North Carolina, Charlotte	Dept of Engineering Technology		
Charlotte, NC 28270	20000.00		0.00
Main, Robert G	PhD	28	01/01/92-06/30/92
California State University, Chico	Dept of Communication Design		
Chico, CA 959290504	20000.00		7672.00
Mains, Gilbert J	PhD	52	01/01/92-12/31/92
Oklahoma State University	Dept of Chemistry		
Stillwater, OK 74078	19071.00		8746.00
Marathay, Arvind S	PhD	51	01/01/92-12/31/92
University of Arizona	Dept of Optical Sciences		
Tucson, AZ 85721	20000.00		0.00
Martin, Charlesworth R	PhD	125	01/01/92-12/31/92
Norfolk State University	Dept of Physics & Engineering		
Norfolk, VA 23504	20000.00		0.00
Mayes, Jessica L	BS	16	01/01/92-12/31/92
University of Kentucky	Dept of Material Science & Engrng		
Lexington, KY 405034203	20000.00		28625.00
Mulligan, Benjamin E	PhD	54	01/01/92-12/31/92
University of Georgia	Dept of Psychology		
Athens, GA 30602	19895.00		13677.00
Munday, Edgar G	PhD	38	10/01/91-10/30/92
University of North Carolina, Charlotte	Dept of Mechanical Engineering		
Charlotte, NC 28223	20000.00		11638.00

Nurre, Joseph H Ohio University Athens, OH 45701	PhD	56	01/01/92-12/31/92	Dept of Electrical & Comp Engrg	19842.00	15135.00
Orkwis, Paul D University of Cincinnati Cincinnati, OH 452210070	PhD	14	10/01/91-10/30/92	Dept of Engineering Mechanics	19966.00	23017.00
Patra, Amit L University of Puerto Rico Mayaquez, PR 00681	PhD	69	01/01/92-12/31/92	Dept of General Engineering	20000.00	2750.00
Peters II, Richard A Vanderbilt University Nashville, TN 37235	PhD	160	01/01/92-12/31/92	Dept of Electrical Engineering	20000.00	0.00
Pollack, Steven K University of Cincinnati Cincinnati, OH 452200012	PhD	31	01/01/92-12/31/92	Dept of Materials Sci & Engrg	20000.00	14877.00
Prescott, Glenn E University of Kansas Lawrence, KS 66045	PhD	72	01/01/92-12/31/92	Dept of Electrical Engineering	20000.00	8000.00
Price, James L University of Iowa Iowa City, IA 52242	PhD	48	01/01/92-12/30/92	Dept of Sociology	20000.00	8600.00
Qazi, Salahuddin SUNY, Utica Utica, NY 13504	PhD	129	01/01/92-12/31/92	Dept of Electrical Engineering	20000.00	25000.00
Rappaport, Carey M Northeastern University Boston, MA 02115	PhD	58	01/01/92-06/30/92	Dept of Electrical & Comp Engrng	19999.00	0.00
Rawson, Jenny L North Dakota State University Fargo, ND 58105	PhD	144	01/01/92-12/31/92	Dept of Electrical Engineering	19997.00	19826.00
Riccio, Gary E University of Illinois, Urbana Urbana, IL 61821	PhD	80	01/01/92-12/31/92	Dept of Human Perception	20000.00	0.00
Rotz, Christopher A Brigham Young University Provo, UT 84602	PhD	136	12/01/91-12/31/92	Dept of Manufacturing Engineering	20000.00	11814.00
Schwartz, Martin University of North Texas Denton, TX 762035068	PhD	55	01/01/92-12/31/92	Dept of Chemistry	20000.00	18918.00
Senseman, David M University of Texas, San Antonio San Antonio, TX 78285	PhD	77	12/01/91-11/30/92	Dept of Information	20000.00	19935.00
Sensiper, Martin University of Central Florida Orlando, FL 32816	BS	15	11/01/91-05/31/92	Dept of Electrical Engineering	20000.00	0.00
Shamma, Jeff S University of Texas, Austin Austin, TX 78713	PhD	70	01/01/92-12/31/92	Dept of Electrical Engineering	20000.00	0.00

Shively, Jon H California State University, Northridge Northridge, CA 91330	PhD 140 01/01/92-12/31/92 Dept of CIAM 20000.00 14553.00
Singh, Sahjendra N University of Nevada, Las Vegas Las Vegas, NV 89014	PhD 79 01/01/92-12/31/92 Dept of Electrical Engineering 20000.00 20595.00
Smith, Gerald A Pennsylvania State University University Park, PA 16802	PhD 63 07/01/92-07/01/93 Dept of Physics 20000.00 0.00
Stephens, Benjamin R Clemson University Clemson, SC 29634	PhD 114 01/01/92-12/31/92 Dept of Psycology 19988.00 4250.00
Sudkamp, Thomas Wright State University Dayton, OH 45435	PhD 97 01/01/92-08/31/92 Dept of Computer Science 20000.00 18739.00
Sydor, Michael University of Minnesota, Duluth Duluth, MN 55804	PhD 11 01/01/92-12/31/92 Dept of Physics 20000.00 0.00
Tankin, Richard S Northwestern University Evanston, IL 60208	PhD 44 01/01/92-12/31/92 Dept of Mechanical Engineering 20000.00 29103.00
Taylor, Michael D University of Central Florida Orlando, FL 32816	PhD 141 05/01/92-07/31/92 Dept of Mathematics 20000.00 1587.00
Teegarden, Kenneth J University of Rochester Rochester, NY 14627	PhD 98 01/01/92-12/31/92 Dept of Optics 20250.00 60600.00
Tew, Jeffrey D Virginia Polytech Instit and State Univ Blacksburg, VA 24061	PhD 137 03/01/92-09/30/92 Dept of Industrial Engineering 17008.00 4564.00
Tipping, Richard H University of Alabama Tuscaloosa, AL 35487	PhD 81 01/01/92-05/31/92 Dept of Physics & Astronomy 20000.00 15000.00
Tripathi, Ram C University of Texas, San Antonio San Antonio, TX 78249	PhD 105 01/01/92-12/31/92 Dept of Mathematics 20000.00 2274.00
Wells, Fred V Idaho State University Pocatello, ID 83209	PhD 155 01/01/92-12/31/92 Dept of Chemistry 20000.00 8000.00
Whitefield, Phillip D University of Missouri, Rolla Rolla, MO 65401	PhD 25 01/01/92-12/31/92 Dept of Chemistry 19991.00 25448.00
Wolfenstine, Jeffrey B University California, Irvine Irvine, CA 92717	PhD 18 01/01/92-12/31/92 Dept of Mechanical Engineering 20000.00 11485.00
Wolper, James S Idaho State University Pocatello, ID 83209	PhD 138 01/15/92-09/30/92 Dept of Mathematics 20000.00 4828.00

Zavodney, Lawrence D	PhD	148	01/01/92-12/31/92
Ohio State University	Dept of Engineering Mechanics		
Columbus, OH 43210	20000.00		0.00
Zimmerman, Wayne J	PhD	111	01/01/92-12/31/92
Texas Women University	Dept of Mathematics		
Denton, TX 76204	19990.00		8900.00

ATTACHMENT 1:

SAMPLE SREP SUBCONTRACT

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
1993 SUMMER RESEARCH EXTENSION PROGRAM SUBCONTRACT 93-36

BETWEEN

Research & Development Laboratories
5800 Uplander Way
Culver City, CA 90230-6608

AND

University of Delaware
Sponsored Programs Admin.
Newark, DE 19716

REFERENCE: Summer Research Extension Program Proposal 93-36
Start Date: 01/01/93 End Date: 12/31/93
Proposal amount: $20000.00

(1) PRINCIPAL INVESTIGATOR: Dr. Ian W. Hall
Materials Science
University of Delaware
Newark, DE 19716

(2) UNITED STATES AFOSR CONTRACT NUMBER: F49620-90-C-09076

(3) CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER (CFDA): 12.800
PROJECT TITLE: AIR FORCE DEFENSE RESEARCH SOURCES PROGRAM

(4) ATTACHMENTS 1 AND 2: SREP REPORT INSTRUCTIONS

*** SIGN SREP SUBCONTRACT AND RETURN TO RDL ***

1. BACKGROUND: Research & Development Laboratories (RDL) is under contract (F49620-90-C-0076) to the United States Air Force to administer the Summer Research Programs (SRP), sponsored by the Air Force Office of Scientific Research (AFOSR), Bolling Air Force Base, D.C. Under the SRP, a selected number of college faculty members and graduate students spend part of the summer conducting research in Air Force laboratories. After completion of the summer tour participants may submit, through their home institutions, proposals for follow-on research. The follow-on research is known as the Research Initiation Program (RIP). Approximately 75 RIP proposals annually will be selected by the Air Force for funding of up to $20,000; shared funding by the academic institution is encouraged. RIP efforts selected for funding are administered by RDL through subcontracts with the institutions. This subcontract represents such an agreement between RDL and the institution designated in Section 5 below.

2. RDL PAYMENTS: RDL will provide the following payments to RIP institutions:

- 90 percent of the negotiated RIP dollar amount at the start of the RIP Research period.
- the remainder of the funds within 30 days after receipt at RDL of the acceptable written final report for the RIP research.

3. INSTITUTION'S RESPONSIBILITIES: As a subcontractor to RDL, the institution designated on the title page will:

a. Assure that the research performed and the resources utilized adhere to those defined in the RIP proposal.

b. Provide the level and amounts of institutional support specified in the RIP proposal.

c. Notify RDL as soon as possible, but not later than 30 days, of any changes in 3a or 3b above, or any change to the assignment or amount of participation of the Principal Investigator designated on the title page.

d. Assure that the research is completed and the final report is delivered to RDL not later than twelve months from the effective date of this subcontract. The effective date of the subcontract is one week after the date that the institution's contracting representative signs this subcontract, but no later than January 15, 1992.

e. Assure that the final report is submitted in the format shown in Attachment 1.

f. Agree that any release of information relating to this subcontract (news releases, articles, manuscripts, brochures, advertisements, still and motion pictures, speeches, trade association meetings, symposia, etc.) will include a statement that the project or effort depicted was or is sponsored by: Air Force Office of Scientific Research, Bolling AFB, D.C.

g. Notify RDL of inventions or patents claimed as the result of this research in a format specified in Attachment 1.

h. RDL is required by the prime contract to flow down patent rights and technical data requirements in this subcontract. Attachment 2 to this subcontract contains a list of contract clauses incorporated by reference in the prime contract.

4. All notices to RDL shall be addressed to:

RDL Summer Research Program Office
5800 Uplander Way
Culver City, CA 90230-6608

5. By their signatures below, the parties agree to the provisions of this subcontract.

____
Abe S. Sopher
RDL Contracts Manager

____
Date

____
Signature of Institution Contracting Official

____
Typed/Printed Name

____
Title

____
Institution

____
Date/Phone

Attachment 1

Final Report Format

1. All RIP Principal Investigators will submit a final report of the research conducted.

2. One copy of the report is due to RDL no later than twelve months after the effective date of the RIP subcontract. At the same time, submit one copy to the Air Force laboratory focal point.

3. The title page should contain the title of the research, the Principal Investigator and or other co-investigators, the month and year of issue, the university with department and address, and acknowledgement of sponsorship by AFOSR (see clause 3f of this subcontract).

4. For text, use a font that is 12 characters per inch (elite) and as close to letter quality as possible. Start with the title in all caps one and one-half inches from the top of the first page; if the title requires two or more lines, single space it. Double space below the title, and then center and type the researcher's title and name. Then space twice and begin the double-spaced text.

   Use a one-and-one-half-inch left margin and a one-inch right margin for the body of the text. Center page numbers at the foot of each page, one inch from the bottom. Each page should have a one-inch margin at the top. The format should be that of a standard research paper: it should begin with a one-paragraph abstract (on its own page) summarizing your work and should be followed by an introduction, a discussion of the problem, a results section, and a conclusion. Since multiple copies of your report may be required, assure that all pages can be readily copied to a black-and-white 8 1/2" by 11" page. (No colors, such as blue or green, that don't photocopy well, and no foldouts, please.)

5. The report must be accompanied by a separate statement on whether or not any inventions or patents have resulted from this research. If yes, use a DD Form 882 (supplied by RDL on request) to indicate the patent filing date, serial number, title, and a copy of the patent application, and patent number and issue date for any subject invention in any country in which the subcontractor has applied for patents.

Attachment 2

Contract Clauses

This contract incorporates by reference the following clauses of the Federal Acquisition Regulations (FAR), with the same force and effect as if they were given in full text. Upon request, the Contracting Officer or RDL will make their full text available (FAR 52.252-2).

FAR CLAUSES	TITLE AND DATE
52.202-1	DEFINITIONS (APR 1984)
52.203-1	OFFICIALS NOT TO BENEFIT (APR 1984)
52.203-3	GRATUITIES (APR 1984)
52.203-5	COVENANT AGAINST CONTINGENT FEES (APR 1984)
52.304-6	RESTRICTIONS ON SUBCONTRACTOR SALES TO THE GOVERNMENT (JUL 1985)
52.203-7	ANTI-KICKBACK PROCEDURES (OCT 1988)
52.203-12	LIMITATION ON PAYMENTS TO INFLUENCE CERTAIN FEDERAL TRANSACTIONS (JAN 1990)
52.204-2	SECURITY REQUIREMENTS (APR 1984)
52.209-6	PROTECTING THE GOVERNMENT'S INTEREST WHEN SUBCONTRACTING WITH CONTRACTORS DEBARRED, SUSPENDED, OR PROPOSED FOR DEBARMENT (MAY 1989)
52.212-8	DEFENSE PRIORITY AND ALLOCATION REQUIREMENTS (MAY 1986)
52.215-1	EXAMINATION OF RECORDS BY COMPTROLLER GENERAL (APR 1984)
52.215-2	AUDIT - NEGOTIATION (DEC 1989)
52.222-26	EQUAL OPPORTUNITY (APR 1984)
52.222-28	EQUAL OPPORTUNITY PREAWARD CLEARANCE OF SUBCONTRACTS (APR 1984)
52.222-35	AFFIRMATIVE ACTION FOR SPECIAL DISABLED AND VIETNAM ERA VETERANS (APR 1984)
52.222-36	AFFIRMATIVE ACTION FOR HANDICAPPED WORKERS (APR 1984)

52.222-37	EMPLOYMENT REPORTS ON SPECIAL DISABLED VETERANS AND VETERANS OF THE VIETNAM ERA (JAN 1988)
52.223-2	CLEAN AIR AND WATER (APR 1984)
52.232-6	DRUG-FREE WORKPLACE (MAR 1989)
52.224-1	PRIVACY ACT NOTIFICATION (APR 1984)
52.224-2	PRIVACY ACT (APR 1984)
52.225-13	RESTRICTIONS ON CONTRACTING WITH SANCTIONED PERSONS (MAY 1989)
52.227-1	AUTHORIZATION AND CONSENT (APR 1984)
52.227-2	NOTICE AND ASSISTANCE REGARDING PATENT AND COPYRIGHT INFRINGEMENT (APR 1984)
52.227-10	FILING OF PATENT APPLICATIONS - CLASSIFIED SUBJECT MATTER (APR 1984)
52.227-11	PATENT RIGHTS - RETENTION BY THE CONTRACTOR (SHORT FORM) (JUN 1989)
52.228-6	INSURANCE - IMMUNITY FROM TORT LIABILITY (APR 1984)
52.228-7	INSURANCE - LIABILITY TO THIRD PERSONS (APR 1984)
52.230-5	DISCLOSURE AND CONSISTENCY OF COST ACCOUNTING PRACTICES (SEP 1987)
52.232-23	ASSIGNMENT OF CLAIMS (JAN 1986)
52.237-3	CONTINUITY OF SERVICES (APR 1984)
52.246-25	LIMITATION OF LIABILITY - SERVICES (APR 1984)
52.249-6	TERMINATION (COST-REIMBURSEMENT) (MAY 1986)
52.249-14	EXCUSABLE DELAYS (APR 1984)
52.251-1	GOVERNMENT SUPPLY SOURCES (APR 1984)

DoD FAR CLAUSES	TITLE AND DATE
252.203-7001	SPECIAL PROHIBITION ON EMPLOYMENT (MAR 1989)
252.203-7002	STATUTORY COMPENSATION PROHIBITIONS AND REPORTING REQUIREMENTS RELATING TO CERTAIN FORMER DEPARTMENT OF DEFENSE (DoD) EMPLOYEES (APR 1988)
252.223-7500	DRUG-FREE WORK FORCE (SEP 1988)
252.225-7001	BUY AMERICAN ACT AND BALANCE OF PAYMENTS PROGRAM (APR 1985)
252-225-7023	RESTRICTION ON ACQUISITION OF FOREIGN MACHINE TOOLS (JAN 1989)
252.227-7013	RIGHTS IN TECHNICAL DATA AND COMPUTER SOFTWARE (OCT 1988)
252.227-7018	RESTRICTIVE MARKINGS ON TECHNICAL DATA (OCT 1988)
252.227-7029	IDENTIFICATION OF TECHNICAL DATA (APR 1988)
252.227-7034	PATENTS - SUBCONTRACTS (APR 1984)
252.227-7037	VALIDATION OF RESTRICTIVE MARKINGS ON TECHNICAL DATA (APR 1988)
252.231-7000	SUPPLEMENTAL COST PRINCIPLES (APR 1984)
252.231-7001	PENALTIES FOR UNALLOWABLE COSTS (APR 1988)
252.231-7003	CERTIFICATION OF INDIRECT COSTS (APR 1986)
252.251-7000	ORDERING FROM GOVERNMENT SUPPLY SOURCES (APR 1984)
252.271-7001	RECOVERY OF NONRECURRING COSTS ON COMMERCIAL SALES OF DEFENSE PRODUCTS AND TECHNOLOGY AND OF ROYALTY FEES FOR USE OF DoD TECHNICAL DATA (FEB 1989)

7 November 1991

AFOSR/PKO
Bldg. 410, Room C-124
Bolling AFB, DC 20332-6448

Attn: Ms. Kathleen Wetherell

Dear Ms. Wetherell:

Enclosed for your approval is the model subcontract for the Research Initiation Program under the Summer Research Programs (Contract F9620-90-C-0076). The blanks will be filled by merging information from our dBase IV database.

Sincerely,

Abe S. Sopher
Contracts Manager

cc: AFOSR/NI (Lt. Col. Cavendar)

ATTACHMENT 2:

SAMPLE TECHNICAL EVALUATION FORM AND TECHNICAL EVALUATION SUMMARY

1992 RESEARCH INITIATION PROGRAM TECHNICAL EVALUATION

RIP NO: 92-2
RIP ASSOCIATE: Dr. Charles Kibert

Provided are several evaluation statements followed by ratings of (1) through (5). A rating of (1) is the lowest and (5) is the highest. Circle the rating level number you best feel rates the statement. Document additional comments on the back of this evaluation form.

Mail or fax the completed form to:

RDL
Attn: 1992 RIP TECH EVALS
5800 Uplander Way
Culver City, CA 90230-6608
(Fax: 310 216-5940)

1. This RIP report has a high level of technical merit 1 2 3 4 5

2. The RIP program is important to accomplishing the lab's mission 1 2 3 4 5

3. This RIP report accomplished what the associate's proposal promised 1 2 3 4 5

4. This RIP report addresses area(s) important to the USAF 1 2 3 4 5

5. The USAF should continue to pursue the research in this RIP report 1 2 3 4 5

6. The USAF should maintain research relationships with this RIP associate 1 2 3 4 5

7. The money spent on this RIP effort was well worth it 1 2 3 4 5

8. This RIP report is well organized and well written 1 2 3 4 5

9. I'll be eager to be a focal point for summer and RIP associates in the future 1 2 3 4 5

10. The one-year period for complete RIP research is about right 1 2 3 4 5

****USE THE BACK OF THIS FORM FOR ADDITIONAL COMMENTS*

LAB FOCAL POINT'S NAME (PRINT): ____________________

OFFICE SYMBOL: __________ PHONE: ______________

TECHNICAL EVALUATION SUMMARY

	Technical Evaluation Questionnaire Rating Factors										
Subcontract no.	1	2	3	4	5	6	7	8	9	10	Average
135	5	4	5	4	4	4	4	4	5	5	4.4
50	4	4	5	4	4	4	4	3	5	5	4.2
3	4	3	3	3	3	3	3	3	3	4	3.2
71	4	4	4	4	3	5	5	4	5	5	4.3
152	3	4	3	4	4	3	4	3	4	5	3.7
147	5	5	5	5	5	5	5	5	5	4	4.9
011	4	4	5	4	5	5	5	4	5	4	4.5
106	5	5	4	5	5	5	5	5	5	5	4.9
83	5	4	5	5	5	5	5	5	5	4	4.8
26	5	4	4	5	5	5	5	5	4	4	4.6
8	5	3	4	4	5	5	5	3	5	5	4.4
120	1	5	2	4	5	3	2	1	4	4	3.1
133	3	2	4	5	5	4	3	4	3	5	3.8
108	5	4	4	5	5	5	5	5	5	5	4.8
76	5	5	5	5	5	5	5	5	5	3	4.8
122	5	5	5	5	5	4	5	5	5	5	4.9
92	4	5	5	5	5	5	5	5	5	5	4.9
47	5	5	5	5	5	4	4	5	5	5	4.8
57	4	4	4	5	5	4	4	4	4	2	4.0
17	5	5	5	5	5	5	5	5	5	5	5.0
5	5	3	4	4	4	5	5	5	4	3	4.2
62	5	4	5	4	4	5	5	5	5	5	4.7
74	4	3	4	4	4	4	5	4	4	5	4.1
53	4	3	4	4	3	4	3	5	3	4	3.7
84	5	4	4	5	5	5	5	5	5	4	4.7
145	4	4	5	4	5	5	5	5	5	4	4.6
35	5	5	5	5	5	5	5	5	5	5	5.0

	Technical Evaluation Questionnaire Rating Factors										
Subcontract no.	1	2	3	4	5	6	7	8	9	10	Average
59	5	4	5	5	5	5	5	5	5	5	4.9
115	5	5	5	5	5	5	5	5	5	5	5.0
118	4	5	5	5	5	5	5	4	5	4	4.7
104	5	3	4	3	5	4	5	5	4	5	4.3
6	3	5	5	5	3	5	5	4	5	3	4.3
28	5	4	5	5	5	4	5	4	4	4	4.5
51	5	5	4	5	5	5	5	5	5	4	4.8
16	5	5	5	5	5	4	5	5	5	5	4.9
54	5	4	5	4	5	4	5	5	5	5	4.7
56	3	3	5	4	5	3	4	5	5	5	4.2
69	4	5	4	5	5	4	5	5	5	5	4.7
72	5	5	5	5	5	5	5	5	5	5	5.0
129	5	5	5	5	5	5	5	5	5	5	5.0
58	3	4	5	4	3	4	5	4	4	4	4.0
144	5	5	5	5	5	5	5	5	5	5	5.0
80	5	5	5	5	5	5	5	5	4	4	4.8
136	5	4	5	5	5	5	5	5	5	4	4.8
55	5	5	5	5	5	5	5	5	5	4	4.9
77	5	4	3	4	3	4	4	4	5	4	4.0
15	5	4	5	5	5	5	5	4	5	5	4.8
70	5	4	4	5	5	5	5	5	5	4	4.7
140	5	5	5	5	5	5	5	5	5	5	5.0
79	4	3	5	4	5	4	5	5	4	5	4.4
63	5	5	5	5	5	5	5	5	5	5	5.0
97	5	4	4	5	5	5	5	5	5	5	4.8
11	5	4	4	4	4	5	4	4	5	3	4.2
44	5	5	5	5	5	5	5	5	5	5	5.0
141	5	4	5	4	4	5	5	5	5	4	4.6
98	5	5	5	5	5	5	5	5	5	5	5.0

	Technical Evaluation Questionnaire Rating Factors										
Subcontract no.	1	2	3	4	5	6	7	8	9	10	Average
81	4	4	3	4	4	4	4	5	5	4	4.1
105	5	5	5	5	5	5	5	5	5	5	5.0
25	4	4	4	5	5	5	4	5	4	2	4.2
18	5	3	5	5	5	3	5	5	5	4	4.5
138	5	4	5	5	5	5	5	3	5	3	4.5
111	5	5	5	5	5	5	5	5	5	5	5.0
Avg by factor:	4.5	4.2	4.5	4.6	4.7	4.6	4.7	4.6	4.7	4.4	4.6

EFFECTS OF DAMAGE ON VIBRATION FREQUENCY AND DAMPING BEHAVIOR OF CANTILEVERED LAMINATED PLATES

David Hui
Professor

Department of Mechanical Engineering

University of New Orleans
New Orleans, LA 70148

Final Report for:
Research Initiation Program
Wright Laboratory

sponsored by

Air Force Office of Scientific Research
Bolling Air Force Base, Washington, DC

December 1992

EFFECTS OF DAMAGE ON VIBRATION FREQUENCY AND DAMPING BEHAVIOR OF CANTILEVERED LAMINATED PLATES

David Hui
Professor
Dept. of Mechanical Engineering
University of New Orleans

ABSTRACT

The frequency and damping behavior of cantilevered plates are examined theoretically and experimentally. The laminated plates were previously damaged in a series of Hopkinson Bar impact experiments performed at the Cold Regions Research and Engineering Laboratory, and the extent of internal delamination damage was known from C-Scan technique taken before and after the Hopkinson Bar impact. It appears that the natural frequency of the undamaged and damaged plates remains relatively unchanged and thus, agree with the theoretically calculated frequency of an undamaged plate. However, the damping capacity of a damaged plate increases detectably with the extent of damage, depending on the layup and other factors.

1. INTRODUCTION

The use vibration technique for non-destructive detection of damage in fibre composite structure is an important topic in the aerospace industries since it is a very inexpensive practical method (Adams et al. 1978 and Cawley and Adams 1979). Such vibration tests can also be used to detect fatigue crack damage in composites (Schultz and Warwick 1971 and DiBenedetto 1972). Highsmith and Reifsnider (1982) found that the crack density from the shear lag analysis is related to the reduction of stiffness; Allen et al. (1987) reported the prediction and experimental investigation on damage dependent damping in laminated beams. The assessment of damage in GRP Laminates by stress wave emission and dynamic mechanical measurement was presented by Sims et al (1977). Further examination on the effects of damping on vibration of composites were reported by Singh et al (1991). The damping capacity is used as an indication of the extent of damage by Torvik and Bourne (1980). Despite the numerous attempts, the direct correlation between vibration frequency and damping with the extent of damage has not been firmly established and most of the reported work are preliminary in nature.

Vibration of cantilevered plates is a subject which has been thoroughly investigated by many authors since closed form solution exists (Weaver et al. 1990). In particular, vibration of cantilevered laminated plates was investigated by Crawley (1979a,b), Crawley and Dugundji (1980) and Chu and Ochoa (1989), among others. Much less is known about the effects of barely visible damage of laminated plate on its vibration frequency and the damping behavior. Intuitively, it appears that a damaged structure would be more flexible, that is, less stiff, and thus, the frequency should decrease. Such a decrease in the frequency is well known in a cantilevered isotropic homogeneous beam which contains a single slit (Shen and Grady 1992). It can be conjectured that a damage plate due to a previous non-projectile impact would vibrate at a lower frequency than if the plate had not been damaged. Correspondingly, the plate should possess higher damping capacity.

It is hoped that the vibration and damping behavior of damaged cantilevered plate would enable one to predict the level of damage, and thus, serves as a non-destructive test of the extent of damage in the plate. The present work aims to examine the effects of previosly known impact induced delamination damage on the frequency and damping behavior of a cantilevered laminated plate. Both theoretical and experimental investigations are reported.

The well known vibration solution of cantilevered isotropic-homogeneous beam is extended to laminated beams with damping. In the experimental work, laminated "wide" beams are set in free cantilever beam vibration motion by a light non-damaging mpact in a laboratory pendulum apparatus. The natural frequency and damping behavior was obtained by analysis of the transient response of the plate by attaching strain gauges.

The work is part of an overall effort in the understanding of the energy absorption of laminated plates subjected to spherical projectile impact. There exist a need to estimate the energy absorption of these plates undergoing vibration motion, after the projectile has either perforated the plate or bounce back from the plate in the case of non-perforation impact.

2. Theoretical Vibration Behavior of Cantilevered Beams Undamped and With Distributed Damping

The purpose of this chapter is to present derivations for the eigen functions, natural frequencies and free vibration response for cantilever beams without and with distributed viscous damping. The general equation for transverse free vibration of a beam, in the particular case of a prismatic beam in which the flexural rigidity **E I** does not vary with x, is given by(Weaver[19]

$$(\partial^4 v(x,t)/\partial x^4) + (1/a^2)\ (\partial^2 v(x,t)/\partial t^2) = 0 \qquad (1)$$

where,

$$a^2 = (E\ I)/(\rho\ A)$$

When a beam vibrates transversely in one of its natural modes, the deflection at any location varies harmonically with time, as

follows:

$$v(x,t) = X(x)\ [\ B_1 \cos(\omega t) + B_2 \sin(\ \omega t)] \qquad (2)$$

Substitution of equation (1) into equation (2) results in

$$d^4X(x)/dx^4 - (\omega^2/a^2)\ X(x) \qquad = 0 \qquad (3)$$

In solving this fourth order ordinary differential equation, we introduce the notation

$$(\omega^2/a^2) = k^2$$

or

$$d^4X(x)/dx^4 - k^2X(x) = 0 \qquad (4)$$

The general form of the solutions for equation (4) becomes

$$X(x) = C_1 \sin(kx) + C_2 \cos(kx) + C_3 \sinh(kx) + C_4 \cosh(kx) \qquad (5)$$

Taking the derivatives, one obtains:

$$(dX(x)/dx) = k_1 [C_1 \cos(kx)\text{-}C_2 \sin(kx) + C_3 \cosh(kx)+C_4 \sinh(kx)]$$

$$(d^2X(x)/dx^2) = k_2 [\text{-} C_1 \sin (kx)\text{-}C_2 \cos(kx) + C_3 \sinh(kx)+C_4 \cosh(kx)]$$

$$(d^3X(x)/dx^3) = k_3 [\text{-} C_1 \sin(kx)+C_2 \cos(kx) + C_3 \sinh(kx)+ C_4\cosh(kx)]$$

$$(d^4X(x)/dx^4) = k_4 [C_1 \sin(kx)+C_2 \cos(kx) + C_3 \sinh(kx)+C_4 \cosh(kx)]$$

By inspection , the O.D.E in x is identically satisfied. We can write equation(5) in the following equivalent form. Note that the new constants are not same as that in equation(6) are not the same as those in equation(5).

$$X(x) = C_1[\cos(kx)+\cosh(kx)] + C_2[\cos(kx)\text{-}\cosh(kx)]$$
$$+ C_3[\sin(kx)+\sinh(kx)] + C_4[\sin(kx)\text{-}\sinh(kx)] \qquad (6)$$

Doing a similar procedure as was done for equation(5), clearly indicates that equation(6) identically satisfies the equation(4).

We now solve the equation for the cantilever beam boundary conditions:

(i) $(X)_{x=0} = 0$, (ii) $(dX/dx)_{x=0} = 0$,

(iii) $(d^2X/dx^2)_{x=L} = 0$ & (iv) $(d^3X/dx^3)_{x=L} = 0$

Substituting the boundary conditions in equation(6):

$C_1 = 0$; and $C_3 = 0$;

so that;

$$X(x) = C_2\{ \cos(kx)\text{-}\cosh(kx) + (C_4/C_2)[\sin(kx)\text{-}\sinh(kx)]\} \qquad (7)$$

Taking the derivatives of equation(7) results is

$$(dX(x)/dx) = k\ C_2\{\sin(kx)-\sinh(kx) + (C_4/C_2)[\cos(kx)-\cosh(kx)]\}$$

$$(d^2X(x)/dx^2) = k^2\ C_2\{-\cos(kx)-\cosh(kx) + (C_4/C_2)[-\sin(kx)-\sinh(kx)]\}$$

$$(d^3X(x)/dx^3) = k^3\ C_2\{\sin(kx)-\sinh(kx) + (C_4/C_2)[-\cos(kx)-\cosh(kx)]\}$$

$$(d^4X(x)/dx^4) = k^4\ C_2\{\cos(kx)-\cosh(kx) + (C_4/C_2)[\sin(kx)-\sinh(kx)]\}$$

We now substitute the boundary conditions (iii) and (iv) in equation (7).

The third boundary condition for vanishing moment at the free end x = L, yields.

$$\cos(kL) + \cosh(kL) + (C_4/C_2)[\sin(kL) + \sinh(kL)] = 0$$

and the fourth boundary condition zero shear at free end yields.

$$-\sin(kL) + \sinh(kL) + (C_4/C_2)[\cos(kL) + \cosh(kL)] = 0$$

Setting the determinant of the above systems of two homogeneous system of equations to zero, which is the condition for the existence of a solution for the ratio C_4/C_2, gives

$$[\cos(kL)+\cosh(kL)]^2 - [\sin)kL)+\sinh(kL)]\ [-\sin(kL)+\sinh(kL)] = 0$$

After simplifying the above equation becomes,

$$\mathbf{\cos(kL)\ \cosh(kL) = -1} \qquad \mathbf{(8)}$$

The roots of the equation(8) give the eigen values for this vibration problem. The first few are given by (Weaver et al. 1990),

$$(kL)_1 = 1.875$$

$$(kL)_2 = 4.694$$

$$(kL)_3 = 7.855$$

The natural frequencies are then given by (E is Young's modulus, I is moment of inertia, ρ is density, and A is the cross sectional area of the wide beam),

$$\omega_i = (k_i L)^2\ (a/L^2), \qquad a = [EI/(\rho A)]^{1/2}$$

therefore, the ratio of the coefficients is given by:

$$(C_4/C_2)= [\sin(kL) - \sinh(kL)]/[\cos(kL) + \cosh(kL) \qquad (9)$$

In order to find C_2, it is necessary to satisfy the orthonormality condition, for the eigen functions, equation(7) with the values of C_4/C_2 from the equation(9) as follows

$$(1/L^3) \int_{x=0}^{x=L} X^2(x)\, dx = 1$$

After substituting the eigen function in the above equation we get

$$(C^2{}_2/L^3) \int_{x=0}^{x=L} \{\cos^2(kx) + \cosh^2(kx) - 2 \cos(kx) \cosh(kx)$$

$$+(C_4/C_2)^2 [\sin^2(kx) + \sinh^2(kx) - 2 \sin(kx) \sinh(kx)] + 2[\cos(kx) - \cosh(kx)] [\sin(kx) - \sinh(kx)](C_4/C_2)\} dx = 1$$

Defining a non-dimensional coordinate x' such that x' = kx, one obtains

$$(C_2/L)^2(1/kL) \int_{x'=0}^{x'=L} \{\cos^2(x') + \cosh^2(x') - 2 \cos(x') \cosh(x')$$

$$+(C_4/C_2)^2 \quad [\sin^2(x') \quad + \quad \sinh^2(x') \qquad 2\sin(x') \sinh(x')] + 2[\cos(x') -\cosh(x')] [\sin(x') -\sinh(x')](C_4/C_2)\} dx' = 1$$

The integral above can be expressed in terms of the following elementary integral:

$$J_{cS} = \int_{0}^{kL} \cos(x)\sinh(x)\,dx$$

$$J_{cS} = (1/2)\,[\sin(kL)\sinh(kL) + \cos(kL)\cosh(kL)] - 1$$

Similarly,

$$J_{sC} = \int_{0}^{kL} \sin(x)\cosh(x)\,dx$$

$$J_{sC} = (1/2)\,[\sin(kL)\sinh(kL) - \cos(kL)\cosh(kL)] - 1$$

$$J_{cC} = \int_{0}^{kL} \cos(x)\cosh(x)\,dx$$

$$J_{cC} = (1/2)\,[\sin(kL)\cosh(kL) - \cos(kL)\sinh(kL)]$$

$x=kL$

$$J_{sS} = \int_{0}^{kL} \sin(x)\sinh(x)\,dx$$

$$J_{sS} = (1/2)\,[\sin(kL)\cosh(kL) - \cos(kL)\sinh(kL)]$$

$$I_{cc} = \int^{kL} \cos^2(x)\,dx$$

0

$$I_{cc} = (1/2)\,[kL + (1/2)\sin(kL)]$$

$$I_{ss} = \int_{0}^{kL} \sin^2(x)\,dx$$

$$I_{ss} = (1/2)\,[kL - (1/2)\sin(kL)]$$

$$I_{sc} = \int_{0}^{kL} \sin(x)\cos(x)\,dx$$

$$I_{sc} = (1/4)\,[\cos(2kL) - 1]$$

Similarly for the hyperbolic functions

$$I_{CC} = \int_{0}^{kL} \cosh^2(x)\,dx$$

$$I_{CC} = (1/2)\ [kL + (1/2)\ \sinh(kL)]$$

$$I_{SS} = \int_{0}^{kL} \sinh^2(x)\ dx$$

$$I_{SS} = (1/2)\ [-kL + (1/2)\ \sinh(kL)]$$

Finally,

$$I_{SC} = \int_{0}^{kL} \sinh(x)\ \cosh(x)\ dx$$

$$I_{cc} = (1/4)\ [\cosh(2kL) - 1]$$

In terms of the above elementary integral, the orthonormality condition yields,

$$\mathbf{kL = (C_2/l)^2/\{I_{cc} + I_{cc} + (C_4/C_2)^2\ (I_{ss} + I_{SS} - 2J_{sS}) + 2(C_4/C_2)\ (I_{cs} + J_{cS} - J_{sC} + I_{Sc})\}} \tag{10}$$

Now C_2 can be solved for from equation(10). Notice that there is a distinct value of C_2 and a distinct eigen function for every value of eigen frequency given by equation(8).

The purpose of the preceding was to derive the eigen functions X(x) and the eigen frequencies (kL_1, kL_2, kL_3, kL_4, etc.) for the undamped free vibration of a cantilever beam.

We now, introduce damping distributed into the beam by assuming

(i) Damping is of the viscous type and proportional to the normal strain rate.

(ii) Damping is small a under damped.

(iii) The damped structure will vibrate with same the eigen functions as the undamped structure.

From elementary strength of materials theory, M is the bending moment of the internal normal stress about the neutral axis of the beam, I is the momentum of inertia and E is the Young's modulus of the plate. The bending moment without damping is given by

$$\mathbf{M = (EI)\ (\partial^2 v/\partial x^2)} \qquad \mathbf{(11a)}$$

If instead of a purely elastic beam in which the fibre stresses on a cross-section are proportional to disturbance from the neutral axis, y, and are equal to

$$\mathbf{x = M/I; \qquad y = E(\partial^2/\partial x^2); \qquad y = E\ e_x} \qquad \mathbf{(11b)}$$

we are dealing with a viscoelastic material, the stress will include a contribution that is proportional to the strain rate. Then the total stress will be given by

$$\sigma_x = \mathbf{E\ e_x + C_A\ e'_x = E\ (\partial^2/\partial x^2)\ y + C_A\ (\partial^3/\partial x^2 \partial t)\ y} \qquad \mathbf{(11c)}$$

Hence, by integrating the stress of equation(1c) over the cross-section, the combined bending moment due to elastic and viscous stresses will be given by

$$\mathbf{M = E\ I\ (\partial^2/\partial x^2)\ + C_A\ I\ (\partial^3/\partial x^2 \partial t)} \qquad \mathbf{(11d)}$$

For the complete generality, with regard to damping we also treat the beam as if it were resting on a viscous foundation, in which case the deflection of the beam is resisted by a viscous distributed pressure proportional to the rate of deflection, or

$$\mathbf{P = -C_B\ \partial v/\partial t} \qquad \mathbf{(11e)}$$

The modification of the partial differential equation for the motion of an undamped beam given in equation(1) that the results from introduction of distributed damping due to viscous bending stresses and viscous foundation load is given by equation(12).

$$\mathbf{EI\ (d^4v(x,t)/dx^4) + C_A\ I\ (d^4v(x,t)/[dx^3dt]) + C_B\ (dv(x,t)/dt)}$$
$$\mathbf{= -\rho A\ (d^2v(x,t)/dt^2} \qquad \mathbf{(12)}$$

Assume that the solution is,

$$v(x,t) = \sum X_i(x)\ \phi_i(t) \qquad (13)$$

Where $X_i(x)$ are the undamped eigen functions

$$v(x,t) = X_1(x)\ \phi_1(t) + X_2(x)\ \phi_2(t) + X_3(x)\ \phi_3(t) + ...$$

The governing partial differential equation (P.D.E)is,

$$\mathbf{(d^4v(x,t)/dx^4) + (C_A/E)\ (d^5v(x,t)/[dt\ d^4x]) + (C_B/E)\ (dv(x,t)/dt)}$$
$$\mathbf{= (1/a^2)\ (d^2v(x,t)/dt^2)} \qquad \mathbf{(14)}$$

let $C_A' = C_A/E$ and $C_B' = C_B/EI$

The linear P.D.E. becomes, (i is the mode number 1,2,3,....) after substitution of equation(13),

$$\mathbf{\Sigma\ (d^4X_i/dx^4)\phi_i(t) + [C_A'(d^4X_i/dx^4) + C_B'X_i(x)](d\phi_i(t)/dt) + (1/a_i)^2X_i(x)d^2\phi_i(t)/dt^2) = 0}$$

Note that

$$\mathbf{d^4v/dx^4 = k^4\ X(x)}$$

$$\sum_{i=1}^{\infty} k^4{}_iX_i(x)\,\phi_i(t)+[C_A'k^4{}_i + C_B']X_i(x)](d\phi_i(t)/dt) + (1/a_i)^2X_i(x)(d^2\phi_i(t)/dt^2) = 0$$

To get rid of the dependence on the axial coordinate, we x by use the Orthonormality properties of the eigen functions and the equation by the eigen function $X_j(x)$ and then integrate from x=0 to x=L so that,

$$\int_0^L X_i(x)\,X_j(x)\,dx = \delta_{ij}\ \{\ 0 \text{ if } i \neq j;\ 1 \text{ if } i = j$$

This results in the ordinary differential equation (O.D.E)

$$k^4\phi_i(t)+[C_Ak^4{}_i + C_B]\,(d\phi_i(t)/dt) + (1/a)^2\,(d^2\phi_i(t)/dt^2) \quad = \quad 0$$

for i = 1,2,3,....∞

since, $(1/a^2) = (k^2{}_i/\omega^2{}_i)$

$$\phi_i(t)+[C_A' + C_B'/k^4{}_i]\,(d\phi_i(t)/dt) + (1/\omega_i)^2\,(d^2\phi_i(t)/dt^2) = 0$$

Now assume that the form of the solution to this O.D.E is,

$$\phi_i(t) = B_i\,\exp(p_i\,t)$$

Inserting this into the O.D.E. yields,

$$\{(p_i/\omega_i + [C_A' + C_B'/k^4{}_i)]\,p_i + (\omega_i)^2\}\,\phi_i = 0$$

Solving this quadratic equation for p_i, one obtains,

$$p_i = (1/2)\{-(\omega_i)^2\,[C_A'+(C_B'/k^4{}_i)] \pm ([(\omega_i)^4\,[C_A'+(C_B'/k^4{}_i)^2]]-4(\omega_i)^2)^{1/2}\}$$

Assume C_B' is negligible. Then;

$$p_i = (1/2)\{-(\omega_i)^2\,C_A' \pm ([(\omega_i)^4\,C_A' - 4(\omega_i)^2)^{1/2}\}$$

Denoting the symbol $j = (-1)^{1/2}$

$$p_i = (1/2)\{-(\omega_i)^2 C_A' \pm j[4(\omega_i)^2-(\omega_i^4 C_A']^{1/2}\}$$

For small damping $(\omega_i^4 C_A'/2)^2 << 1$.

Therefore, dropping this small term,

$$p_i = (\omega_i)^2 (-C_A'/2) + i(\omega_i) \tag{15}$$

Thus, the solution is of the form, with damping

$$\phi_i(t) = [G_i \sin(\omega_i t) + H_i \cos(\omega_i t] \exp(-\omega^2{}_i C_A t/2) \tag{16}$$

Therefore, the deflection is,

$$v(x,t) = \sum_{i=1}^{i=\infty} X_i(x)\, \phi_i(t) \tag{16b}$$

The initial conditions due to an impulse load are,

(i) $v(x, t=0) = 0$ for all x that is, zero deflection; and

(ii) $(\partial v/\partial t)\ (x, t=0) = f(x)$ for all x, that is, an initial velocity distribution is specified due to the impulse load.

Now write the P.D.E expressing the impulsive forcing function in terms of a product of a delta function in space and a delta function in time as

$$C_A' v(x,t)_{,txxxx} + v(x,t)_{,xxxx} + (1/a_i^2) v(x,t)_{,tt}$$
$$= (\tau p/(EI))\, \delta(x - L)\, \delta(t) \tag{17}$$

By integrating equation(17) with respect to time over a small time interval containing the instant at which the impact is applied, we obtain a expression for the initial velocity distribution as follows:

$$(1/a^2)\ \partial v(x,0)/\partial t) = (\tau p/(EI))\delta(x\ -L) \qquad (17b)$$

Since the initial displacement distribution vanishes, the solution has the form of equation(16) with $H_i = 0$ and with $G_i = (a^2/\omega_i)\ (\tau p/(EI))\ X_i(x=L)$

or as shown in equation(18).

$$v(x,t) = \sum_{i=1}^{i=\infty} X_i(x)[(a^2/\omega_i)(\tau p/(EI)X_{i(x=L)}\ \mathrm{Sin}(\omega_i t)\ \exp(-\omega^2{}_i C_A t/2)] \qquad (18)$$

3 EXPERIMENTAL APPROACH

In the experimental work, graphite-epoxy laminated plates are used in the cantilever "wide" beam vibration tests. These plates were previously damaged in the low-velocity non-perforation Hopkinson Bar impact where the energy absorption of the plates was analysed using the wave propagation phenomenon (Dutta et al. 1991). The laminated plates are set in free cantilever beam vibration motion by a light non-damaging impact in a laboratory pendulum apparatus (see the schematic diagram of Figure 1). The natural frequency and damping behavior of the laminated plates are obtained by analysing the transient response recorded by the strain gauges attached to the plates. The natural frequency and the damping of the damaged plates are then compared to that of similar undamaged plate. Details of the experimental setup can be found in the MS thesis (Tamilvanan 1992).

The damaged plates are selected from those impacted in previous Hopkinson Bar experiments. The extent of damage was known from C-Scan. The plates are made of AS4/3502 graphite-epoxy material. A typical lamination sequence for a 32 layer laminate is,

(45, 0, 0,45, -45, 0,0, 90, 45, -45, 0,0,45,-45, -45, 45,
0, 0, -45, 45, 90, 0, 0, -45, 45, 0, 0, -45, 45, 0, 0, -45, 45, 0,0)

The laminated plate has an outermost protective layer made of a slightly different woven material.

The cantilevered laminated plate is positioned in a vertical plane by a vise clamped along the horizontal edge. A pendulum apparatus consisting of an aluminum rod swinging in a vertical plane perpendicular to the laminated plate is employed to strike the plate centrally near its upper edge. Strain gauges are mounted on both sides of the plate.

The pendulum is manually drawn to a constant initial angle as indicated by a protractor, rigidly mounted on the apparatus and then releasing it to strike the plate. When the pendulum rod hits near the top edge of the plate, the strain gauges are used to detect the surface strains at the both (upper and lower) surfaces of the plate. The strain data are analysed on line using a Tektronics Fast Fourier Analyzer. The position of the spikes in the power spectral density of the response enables one to determine the first several natural frequencies of the cantilevered plate.

Before starting the experiments, some precautionary measures are taken. The bridge circuit was balanced correctly for both strain gauges and accelerometers (not used). The trigggering is set in the Fast Fourier Analyzer for the required time and amplitude. The triggering was set to the manual mode. Releasing the pendulum was consistently done from the same angle to get the repeatability of the plate vibration response. The strain gauge mounting technique is an important procedure to ensure accurate reading of the strains. The surface of the plate was carefully smooth and clean using a cleaning solvent. The strain gauge was covered by Scotch Tape. The Perama Bond 910 Catalyst was coated on one side of the strain gauge and allowed to dry for a sufficient time. A drop of Alpha Cynaoacrylate Ester was applied to the clean surface of the plate. Then the strain gauge was placed on the plate and pressed down for 2 minutes. Finally the gauge is soldered while assuring that there is no short circuit.

Wave forms of the resulting strain gauge readings are recorded and analysed so that one can obtain the damping constant and the logarithmic decrement of the transient vibration oscillations, and the spectrum of the frequencies. The natural logarithmic of the amplitude of the transient oscillations are plotted against time and the slope of the best least square fit line is determined.

4 DISCUSSIONS OF RESULTS

The natural frequencies of the cantilevered laminated plates are determined from the spikes of the power spectral density. A typical strain gauge response curve and the corresponding power spectral density are shown in Figure 2 and 3 respectively (Figure A16,A17 in Tamilvanan 1992). It can be seen that there exists a strong dominant peak at a frequency value that closely agrees with the computed fundamenal frequency for the first mode. As expected, the predominant frequency found corresponds to the first bending mode.

In order to make the results applicable to other plate materials and dimensions, the non-dimensional frequency defined by Weaver (1990) is used. Figure 4 (A34 in Tamilvanan 1992) shows that the fundamental frequency is not sensitive to the extent of damage as measured by the energy absorption in the Hopkinson Bar experiments. In addition, there is no noticeable difference between the natural frequencies of the damaged and undamaged plates.

Based on the time domain free vibration response waveforms, the cantilevered plates respond primarily in under-damped decaying oscillations at the first fundamental frequency. A small degree of sub-harmonic modulation of the oscillation amplitude is also evident. This behavior along with the distributed random frequency content in the power spectral density suggest that the plates may display a certain amount of non-linear dynamic behavior. Figure 5 (A36 Tamilvanan 1992) shows a graph of the damping coefficient (assuming viscous damping in the logarithmic decrement calculation) versus energy absorption of the plates. A considerable amount of scatter is evident and thus, an average linear curve fit is used along with a pair of parallel lines bounding the region in which results fall within plus or minus one standard deviation. There is a noticable positive slope to these lines, indicating that the damping coefficient increases linearly with the energy absorped. Naturally, the assumption is made that the energy absorption is a good measure of the extent of damage and this turns out to be a reasonable one in the deply analysis of some of these plates (Maryala 1992).

The scattering of the experimental data is partially attributed to the different layups as indicated by points A, B and C in Figure 5 (see Tamilvanan 1992). Further, the orientation of the fibers in the outer plies lying underneath the strain gauges are not identical on all the plates. Despite the above scattering, it appears that for layup A with a woven outer ply, damping coefficient increases linearly with the energy absorped as seen in Figure 6. The slope is considerably greater for layup C than for layup A.

5. CONCLUSIONS

This work attempts to correlate the extent of damage with the vibration frequency and damping coefficient of cantilevered laminated plates. For the limited amount of velocity range and plate layups, one can conclude that,

(i) The natural fundamental frequencies of damaged and undamaged plates do not vary very much and they agree with the theoretical calculation for the undamaged plate.

(ii) The natural fundamental frequencies of damaged plates are not affected by the impact energy absorped or by implication, the extent of damage

(ii) The damping capacity of damaged laminated plates increases linearly with the amount of internal delamination incurred during a non-perforation impact at moderate velocities.

REFERENCES

Adams, R.D., Cawley, C.J. and Stone, B.J. (1978), "A Vibration Testing for Non-Destructively Assessing the Integrity of the Structures", J. of Mechanical Engineering Sciences, Vol. 20, pp. 93-100.

Allen, D.H., Harris, C.E. and Highsmith, A.L. (1987), "Prediction and Experimental Observation of Damage Dependent Damping in Laminated Composite Beams", The Role of Damping in Vibration and Noise Control, ASME DE-Vol. 5, September, pp. 253-263.

Cawley, P. and Adams, R.D. (1979), "A Vibration Technique for Non-Destructive Testing of Fiber Composite Structures", J. of Composite Materials, Vol. 13, pp. 161-175.

Chu, P. and Ochoa,O.O. (1989), "Free-Vibration and Damping Characterization of Composites", Composite Material Technology 1989, edited by D. Hui and T.J. Kozik, pp. 127-129.

Crawley, E.F. (1979a), "A Vibration Technique for Non-Destructive Testing of Fibre Composite Structures", J. of Composite Materials, Vol. 13, April, pp. 161-175.

Crawley, E.F. (1979b), "The Natural Modes of Graphite/Epoxy Cantilever Plates and Shells", J. of Composite Materials, Vol. 13, July, pp. 195-205.

Crawley, E.F. and Dugundji, J. (1980), "Frequency Determination and Non-Dimensionalization for Composite Cantilever Plates", J. of Sound and Vibration, Vol. 72, No. 1, pp. 1-10.

DiBenedetto, A.T., Gauchel, J.V., Thomas, R.L. and Barlow, J.W. (1972), J. of Materials, Vol. 7, pp. 211-215.

Dutta, P.K., Hui, D. and Altamirano, M. (1991), "Energy Absorption of Graphite/Epoxy Plats using Hopkinson Bar Impact", CRREL Report 91-20, October.

Highsmith, A. and Reifsnider, K.L. (1982), "Stiffness Reduction Mechanisms in Composite Laminates", ASTM Symp. on Damage in Composite Materials, Basic Mechanisms, Accumulation, Toleranec and Characterization, ASTM STP 775.

Maryala, S. (1992), "Damaged Morphology of Laminated Plates under Hopkinson Bar Impact", MS thesis, University of New Orleans, September.

Schultz, A.B. and Warwick, D.N. (1971), "Vibration Response: A Non-Destructive Test for Fatigue Crack Damage in Filament-Reinforced Composites", J. of Composite Materials, Vol. 5, pp. 394-404.

Shen, M.H.H. and Grady, J.E. (1992), "Free Vibrations of Delaminated Beam", AIAA Journal, Vol. 30, No. 5, May, pp. 1361-1370.

Sims, G.D., Dean, G.D., Read, B.E. and Wester, B.C. (1977), "Assessment of Damage in GRP Laminates by Stress Wave Emission and Dynamic Mechanical Measurement", J. of Material Science, Vol. 12, pp. 2329-2342.

Singh, S.S., Rohatgi, P.K. and Keshavram, B.N. (1991), "Vibrational Damping Behavior of Composite Materials", Composite Material Technology 1991, ASME PD-Vol. 37, edited by D. Hui and T.J. Kozik, pp. 1-5.

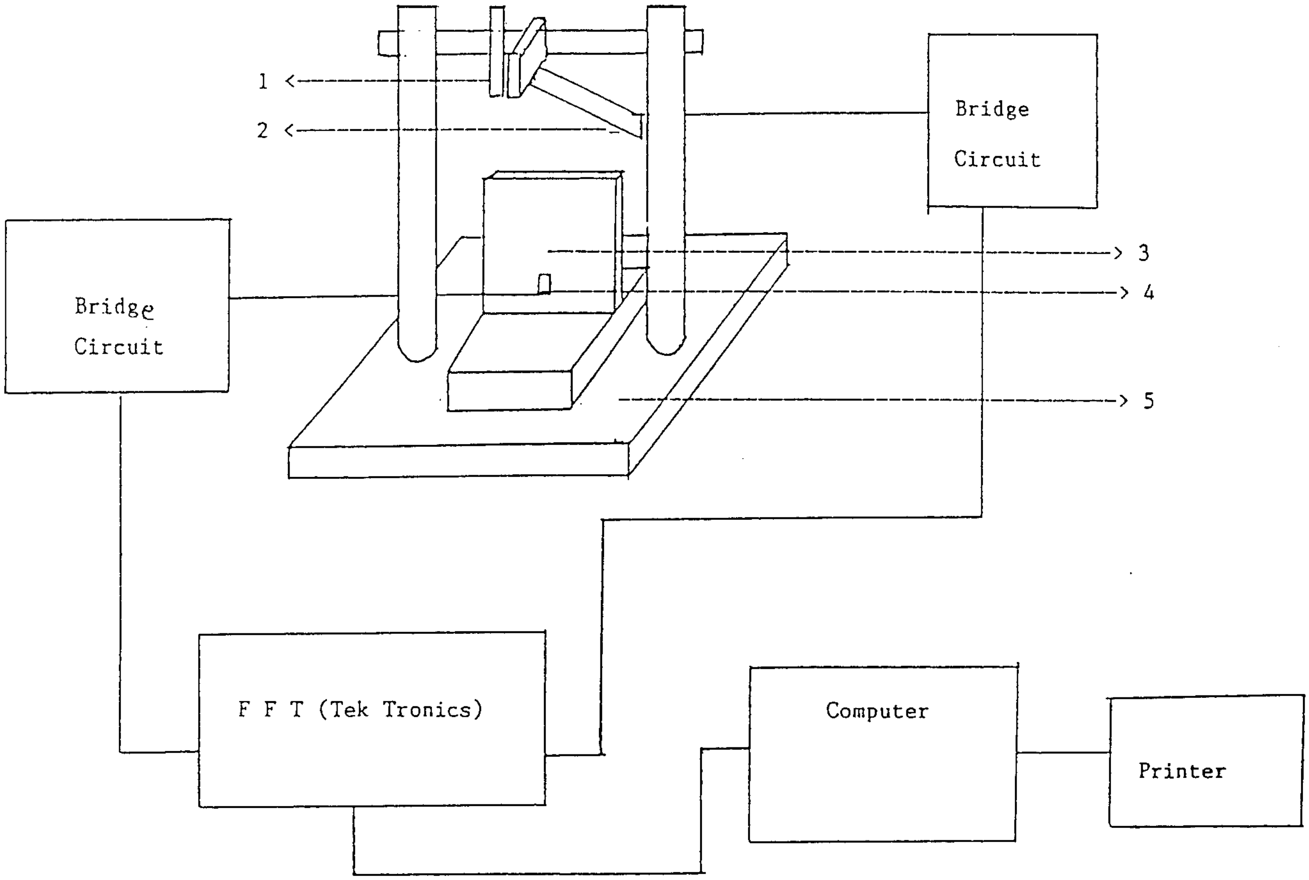
Tamilvanan, A. (1992), "Effects of Damage on Vibration Frequency and Damping Behavior of Cantilevered Laminatd Carbon Fibre Composite Plates", University of New Orleans, MS Thesis, September.

Torvik, P.C. and Bourne, C. (1980), "Material Damping as a Means of Quantifying Fatigue Damage in Composites", Shock and Vibration Bulletin, US Naval Research Laboratory, Vol. 50, pp. 1-11.

Weever, W.Jr., Timoshenko, S. and Young, D. H. (1990), "Vibration Problems in Engineering", John Wiley and Sons, New York, 5th edition.

Figure 1

Schematic Diagram of The Experimental Setup



1 => Angle Marker
2 => Pendulum with aceelerometer
3 => Testing Specimen
4 => Strain gages (Both Sides)
5 => Fixture

Figure 2

Plate # DP3C

Strain Gage Response In The Time Domain to Pendulum Impact

Data: Temp:70°F, Velocity: 159.0 ft/s Energy Absorbed: 14.2 lb-in

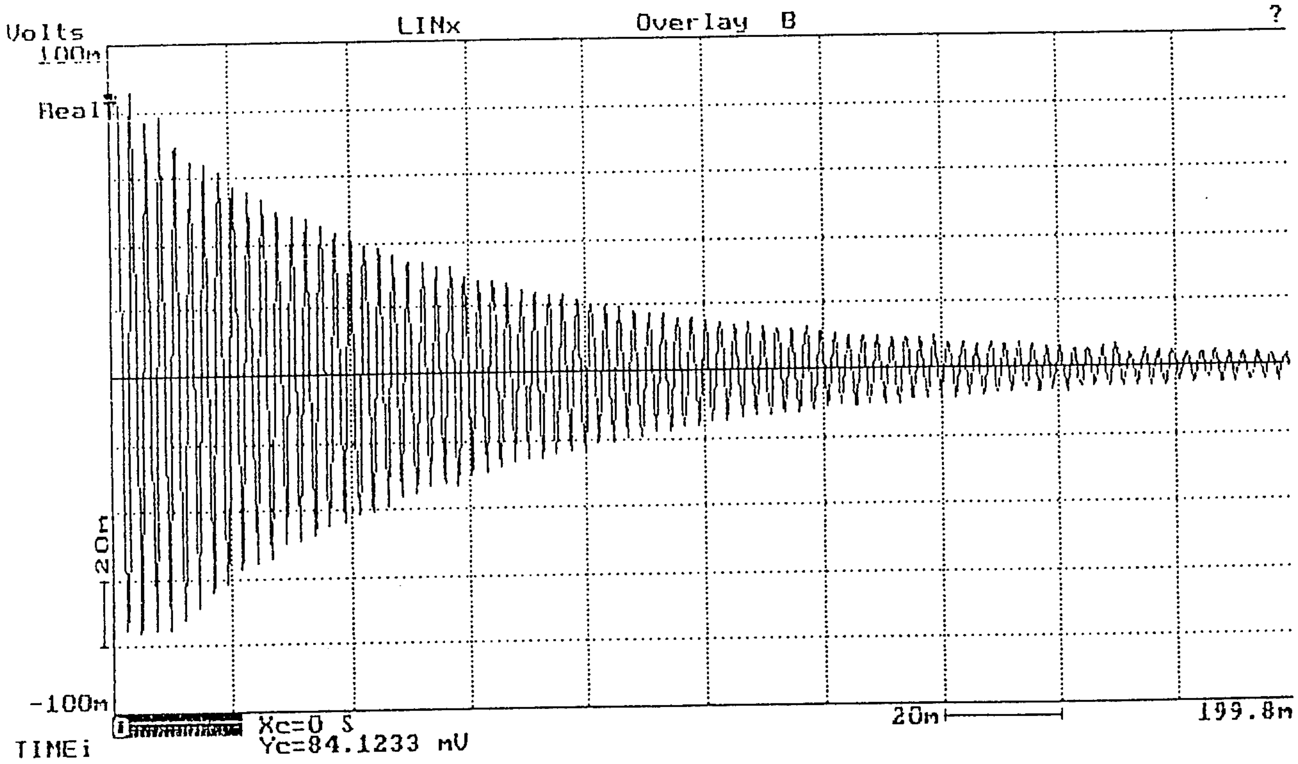


Figure 3

Plate # DP3C

Power Spectral Density

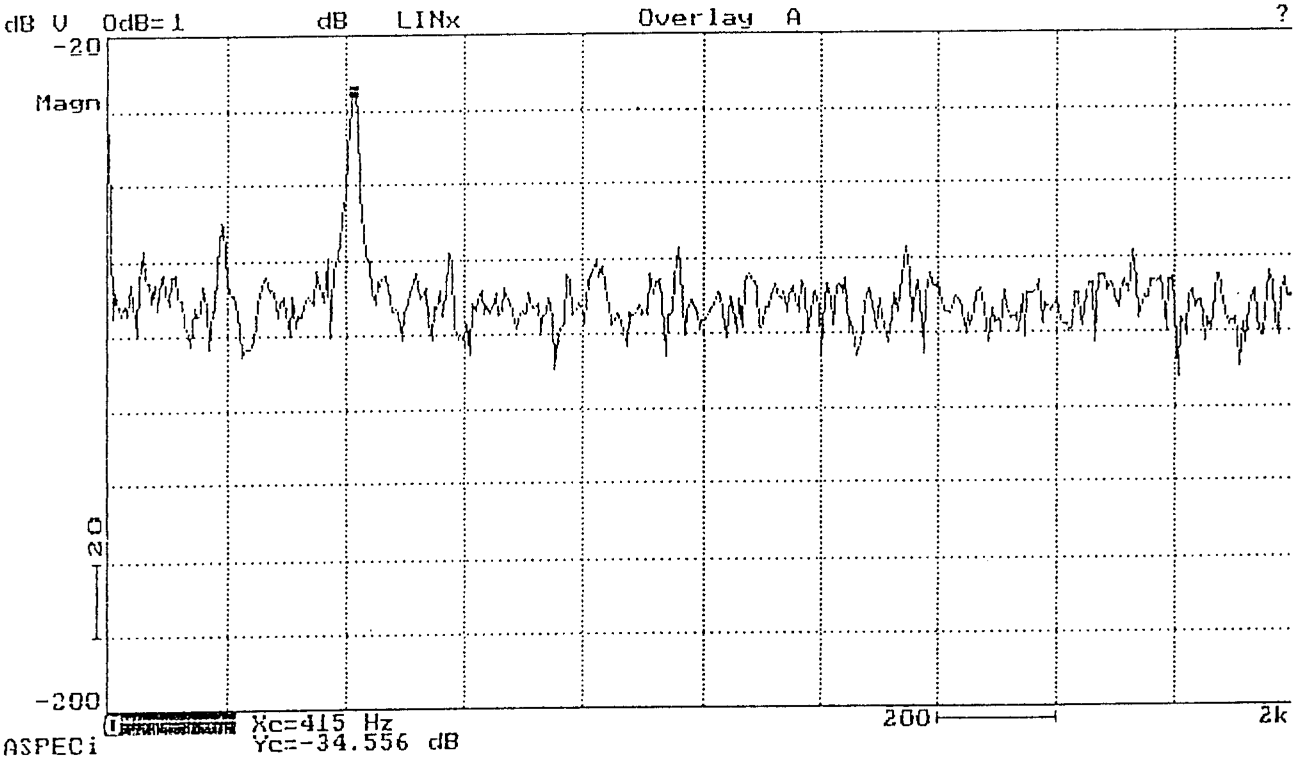


Figure 4

Non-Dimensional Frequency (K L) vs Energy Absorbed

Temp: 70°F; Best Fit: Y = 9.4381 E-5 + 1.7291

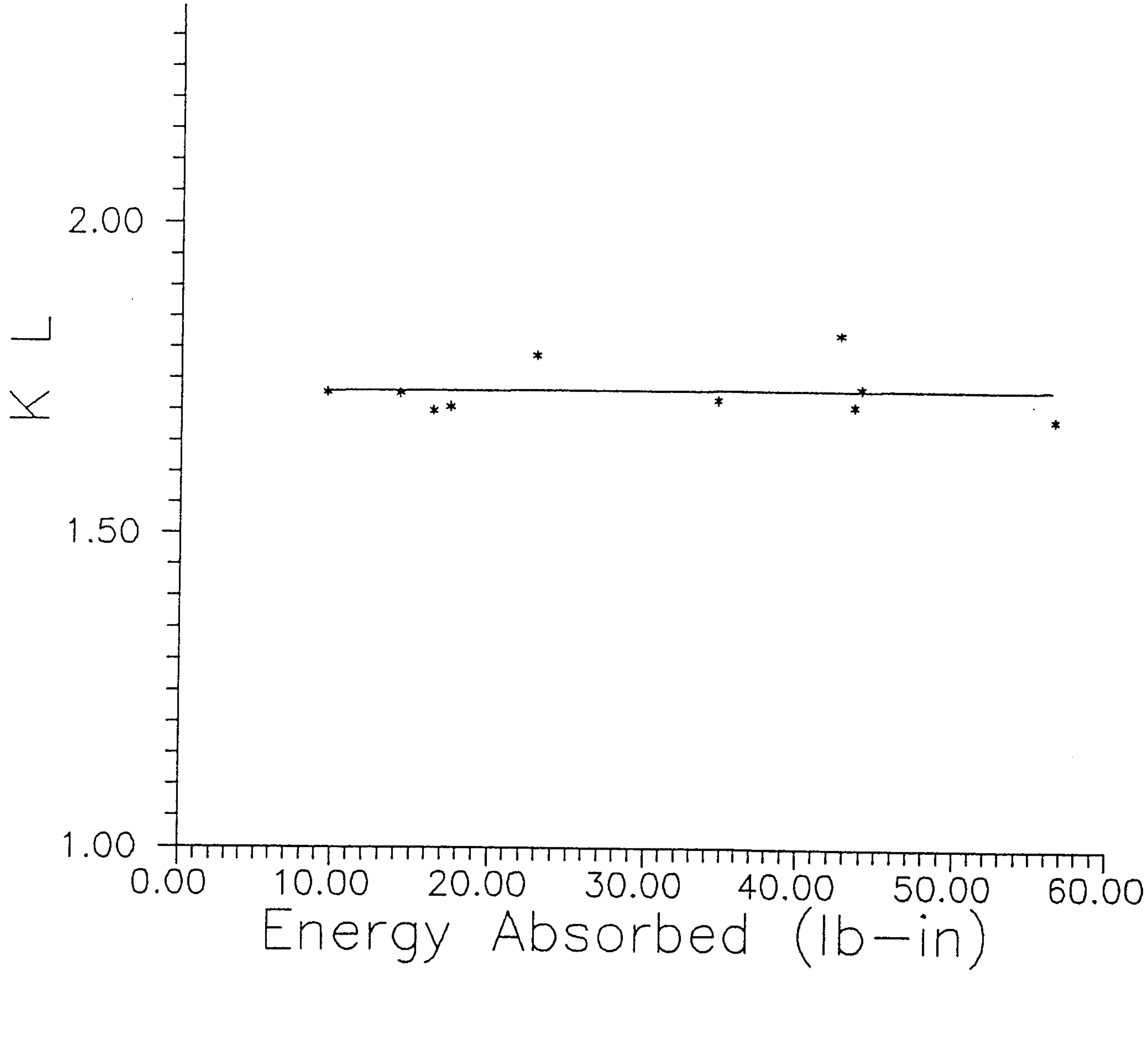


Figure 5

Long Term Response

Energy Absorbed vs Damping Coeff.

Temp: 70°F; Best Fit: Y = 0.0136 X + 0.2308

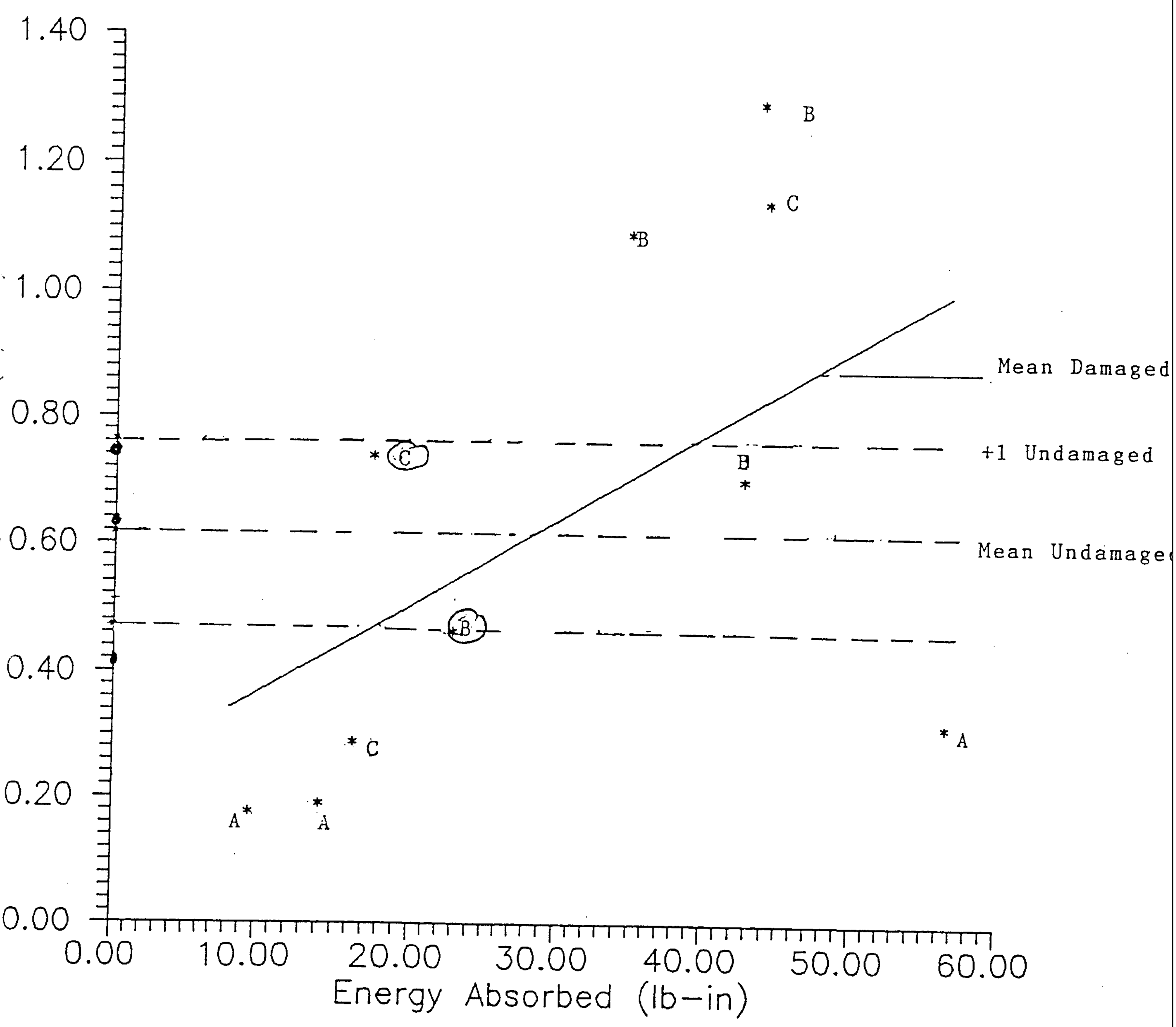


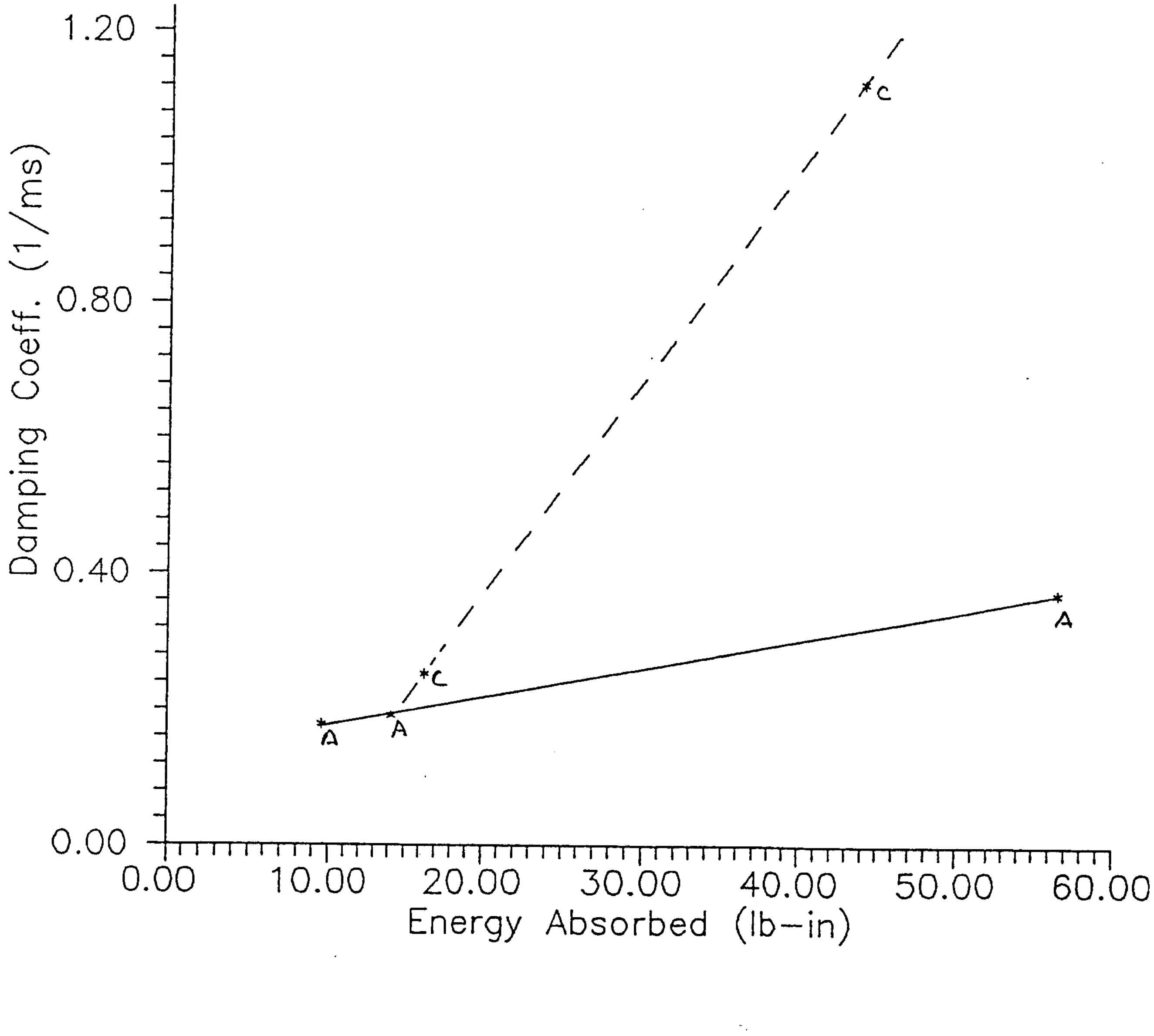
Figure 6 Long Term Response

Energy Absorbed vs Damping Coeff.

Outside Ply Angle of The Plate

Type A: (Upper 45° ply angle & Lower Woven Ply) ——

Type C: (Upper -45° ply angle & Lower -45° ply angle) - -



HIGH TEMPERATURE ELASTOHYDRODYNAMIC LUBRICATION

Michael M. Khonsari
Associate Professor
Department of Mechanical Engineering

University of Pittsburgh
Pittsburgh, PA 15261

FINAL REPORT (Phase I) for:

Wright-Patterson Air Force Laboratories

Sponsored by:

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

and

University of Pittsburgh

December 1992

ABSTRACT

The stringent requirements of Integrated High Performance Turbine Technology (IHPTET) Program calls for significant rise in the operating speed as well as temperature. To meet these requirements alternative lubricants must be explored. One possibility is to utilize the mixture of a solid lubricant such as MOS_2 or graphite particles in a carrier fluid such as ethylene glycol. The objective of this research is to derive the appropriate governing equations for an elastohydrodynamic (EHL) contact that is consistent with experimental observations and predict realistic bearing performance parameters.

A thorough review of the literature revealed that the majority of published papers in EHL have utilized constitutive equations for the lubricant that do not obey experimental observations. This is found to be the case even when the lubricant contains only one phase, i.e. liquid lubricant. Therefore, there exists a serious need for a consistent theoretical development that can be used with realistic constitutive equation.

This report contains a general formulation and solution methodology that can treat an elastohydrodynamic lubrication problem that can accept any simple, non-Newtonian constitutive equation. To demonstrate the applicability of the formulation, we shall restrict this report to the presentation of the results for the well-known rheological equation developed by Bair and Winer. To our knowledge, this is the first time that such results have been made possible.

During the course of this research, we have made a significant accomplishment in extending the analysis to include thermal effects as well as formulation of the appropriate equations for the mixture problem. These results are quite extensive and require a significant amount of discussion. We have recently made an oral presentation to our sponsors at the Wright-Patterson Laboratories which included possible extension work for future research.

The details of these extensions are now being written and will be submitted directly to the Wright Patterson Laboratories in the near future.

1 Introduction

The lubrication regime of many vital machine elements is classified as elastohydrodynamic. Although extremely thin, the height of the lubricating film in elastohydrodynamic lubrication (EHL) is normally larger than the combined surface roughness so that metal-to-metal contact does not occur. Thus, in addition to the elastic deformation of the bounding surfaces, the mechanics of fluid motion plays an important role and requires careful consideration. Characteristically in EHL applications, surfaces are nonconformal, pressures are exceedingly high (~ 1 GPa), and the entrained fluid undergoes a rapid transitory motion (~ 10^{-3} s) through the EHL "contact" region.

Original contributions made by Crook (1961) showed that a theoretical analysis which treats the fluid as linearly viscous (Newtonian) and takes the variation of viscosity with temperature into account can predict the traction coefficient within reasonable accuracy as compared with experimental measurements. Nevertheless, the range of operating speeds and contact pressures in many applications exceed those of Crook's experiments and it is now well known that at high pressures the lubricant shear stress displays marked non-linearity with increasing shear rate. Therefore, proper treatment of the problem requires consideration of non-Newtonian fluid mechanics with a realistic constitutive equation.

Johnson & Tevaarwerk (1977), in a notable contribution, proposed a non-linear Maxwell model of the form:

$$\dot{\gamma} = \frac{1}{G}\frac{d\tau}{dt} + \frac{\tau_0}{\mu} sin h^{-1}\left(\frac{\tau}{\tau_0}\right), \qquad (1)$$

where $\dot{\gamma}$ and τ represent the shear rate and the shear stress, respectively. Parameter G denotes the elastic modulus and τ_0 is what Evans & Johnson (1986a) refer to as the Eyring stress. Eyring stress represents the stress at which the fluid first begins to exhibit non-linearity when plotted against shear rate.

The first term of the right member in equation (1) represents the fluid viscoelasticity which is small when the shear rate is large. This term is often negligible in many EHL applications on the grounds that the Deborah number is small. The second term characterizes a shear thinning behaviour in the lubricant and is referred to as the Ree-Eyring constitutive equation (Ree & Eyring 1955), or is sometimes simply called the "hyperbolic sine" law.

The validity of the Ree-Eyring rheological model in an EHL application was put to experimental test by Conry et al. (1980) with successful results, particularly for relatively high shear rates. This constitutive equation has since been used in a number of recent theoretical analyses with logical extensions to include thermal effects; see for example Sui & Sadeghi (1991) and Wang et al. (1991, 1992).

The hyperbolic-sine law predicts that the shear stress in the lubricant rises monotonically when increasing the rate of shear. This particular behaviour, however, is contrary to experimental observations that attest to the existence of a limiting shear stress beyond which the lubricant deforms plastically such that increasing the shear rate does not alter the shear stress. One of the first reports on this subject appears to have been made by Bair & Winer (1979). These authors maintain that the limiting shear stress is reached at high shear rates even when the pressure in the EHL contact is sufficiently small such that the lubricant remains in liquid form (cf. Bair & Winer 1990).

It is now generally agreed upon that a limiting shear stress can essentially be considered as a property of the fluid and that it is representative of the material shear strength. In fact, experimental devices have been developed that are intended to operate at high shear rates and high pressures when the limiting shear stress is reached (cf. Jacobson 1985).

Neglecting the fluid viscoelastic contribution, the proposed rheological model [illegible] Bair & Winer (1979) takes on the form:

$$\dot{\gamma} = -\frac{\tau_L}{\mu}\ell n\left(1-\frac{\tau}{\tau_L}\right). \tag{2}$$

In contrast to the Ree-Eyring model, which has been the subject of many analyses, it appears that no attempt has been made to theoretically investigate the behavior of EHL contact with the original form given in equation (2). The difficulty appears to be in the functional form of equation (2), for it does not easily lend itself to the derivation of a Reynolds-type equation from which one can determine the pressure distribution. To this end, various approximation schemes have been reported that circumvent this difficulty. Gecim & Winer (1980), for example, replaced the natural logarithm in equation (2) with an inverse hyperbolic tangent and proceeded to perform a Grubin-type analysis to examine the lubricant's non-Newtonian characteristics. A more recent formulation (cf. Wang & Zhang 1987; Khonsari et al. 1990) approximated equation (2) with a hyperbolic function and sought a first-order perturbation solution based upon a method proposed by Dien & Elrod (1983). Utilizing this method, one can derive a form of the Reynolds equation for a fluid that exhibits non-Newtonian behavior with the restrictive assumption that the pressure-driven, or Poiseuille, component of the velocity is negligible such that the flow is regarded as strongly Couette dominated. It has been recently suggested in the literature that a first-order perturbation analysis may be inaccurate when the slide/roll ratio is small and particularly if the surfaces are in pure rolling motion (cf. Wang & Zhang 1992). Therefore, an alternative approach is desirable.

The purpose of the present paper is to offer a theoretical formulation and solution methodology that allows one to analyze an EHL problem with <u>any</u> simple non-Newtonian constitutive equation, including that of the Bair-Winer or Ree-Eyring model. Indeed, the proper equations that can accurately predict the behaviour of traction force in the elastohydrodynamic lubricant may be governed by various rheological models, depending on the operating conditions (cf. Evans & Johnson 1986b). In the formulation of the governing equations to be presented, we shall refrain from altering the functional form of the constitutive equation and avoid the restrictive assumptions and approximations caused by perturbation analyses. To illustrate the procedure, we shall focus our attention on isothermal EHL line-contact problems, such as those encountered in rollers and gear teeth applications, and examine situations where a combination of rolling and sliding may be present. Theoretical development (§2), solution methodology (§3), and presentation of numerical simulations (§4) comprise the content of this paper.

2 Theoretical Development

Consider the flow of a simple, inelastic non-Newtonian fluid sheared between two surfaces with a relatively thin film gap. The axial direction in a line-contact problem is much greater than that of all the other dimensions involved, hence the flow in the axial direction is assumed to be negligible. An order-of-magnitude analysis on the equation governing the conservation of momentum shows that in the absence of inertia effects, the terms involving cross-film velocity gradients are the dominant terms (cf. Tanner 1963). Therefore, the momentum equation reduces to:

$$\frac{\partial}{\partial y}\left[\mu *(I_2)\frac{\partial u}{\partial y}\right]=\frac{dp}{dx}, \qquad (3)$$

where x and y are the coordinates in the direction of motion and across the film, respectively. Parameter p denotes the hydrodynamic pressure and u is the component of the velocity in the x direction. Parameter μ^* represents, what we shall refer to as, the equivalent viscosity which is a function of the second invariant of the strain rate tensor, I_2. For the problem under consideration we have:

$$I_2=\left(\frac{\partial u}{\partial y}\right)^2. \qquad (4)$$

We shall now proceed to derive the governing equation for the pressure distribution for a line-contact EHL utilizing a similar development as described by Paranjpe (1992) for hydrodynamic journal bearings. Integrating equation (3) twice and applying the no-slip boundary condition at the surfaces yields the component of the velocity in the direction of motion. The result is:

$$u = U_1 + \frac{U_2 - U_1}{h}\mu_{e0}\int_0^y \frac{1}{\mu^*}dy + \frac{dp}{dx}\left[\int_0^y \frac{y}{\mu^*}dy - h\frac{\mu_{e0}}{\mu_{e1}}\int_0^y \frac{1}{\mu^*}dy\right], \tag{5}$$

where

$$\frac{1}{\mu_{e0}} = \frac{1}{h}\int_0^h \frac{1}{\mu^*}dy$$

$$\frac{1}{\mu_{e1}} = \frac{1}{h^2}\int_0^h \frac{y}{\mu^*}dy$$

$$\frac{1}{\mu_{e2}} = \frac{1}{h^3}\int_0^h \frac{y^2}{\mu^*}dy\,.$$

In equation (5), parameter h represents the film thickness and U_1 & U_2 denote the velocity of the surface. The shear rate is, therefore, given as:

$$\dot{\gamma} = \frac{\partial u}{\partial y} = \frac{U_2 - U_1}{h}\frac{\mu_{e0}}{\mu^*} + \frac{1}{\mu^*}\left(y - h\frac{\mu_{e0}}{\mu_{e1}}\right)\frac{dp}{dx}\,. \tag{6}$$

Substitution of the velocity distribution into the equation of continuity integrated across the film, viz.,

$$\int_0^h \frac{\partial(\rho u)}{\partial x}dy + \int_0^h \frac{\partial(\rho v)}{\partial y}dy = 0, \tag{7}$$

yields the generalized Reynolds equation for simple non-Newtonian fluid as given below:

$$\frac{d}{dx}\left[\rho h^3\left(\frac{1}{\mu_{e2}}-\frac{\mu_{e0}}{\mu_{e1}^2}\right)\frac{dp}{dx}\right]=\frac{1}{2}(U_2+U_1)\frac{d}{dx}(\rho h)+\frac{1}{2}(U_2-U_1)\frac{d}{dx}\left[\rho h\left(1-2\frac{\mu_{e0}}{\mu_{e1}}\right)\right]. \tag{8}$$

The above generalized Reynolds equation with the appropriate boundary conditions describes the pressure distribution. It is recalled that the generalized form of the Reynolds equation is applicable to any simple non-Newtonian constitutive equation, including that of Bair-Winer. We note that the equivalent viscosity $\mu^*(I_2)$ appears as the integrand in the μ_{e0}, μ_{e1} and μ_{e2} terms, with the limits of integration extended over the film thickness. The generalized Reynolds equation has the proper form for incorporating thermal effects into the analysis since it is well-established that the variation of viscosity across the film thickness plays an important role in hydrodynamic lubrication. The interested reader is referred to Paranjpe (1992) and Khonsari (1987) for further discussion on this subject.

It is worthwhile to note in passing that the generalized Reynolds equation (8) can be easily reduced to the classical Reynolds equation for Newtonian fluids. If one assumes that the viscosity remains constant across the film and that the fluid is Newtonian, with $\mu^*(I_2) = \mu$, equation (8) reduces to the following equation:

$$\frac{d}{dx}\left(\frac{\rho h^3}{\mu}\frac{dp}{dx}\right)=6(U_1+U_2)\frac{d}{dx}(\rho h). \tag{9}$$

Let us now introduce the following dimensionless parameters:

$$\begin{gathered}
X=\frac{x}{b};\quad Y=\frac{y}{h};\quad P=\frac{p}{E}\frac{4R}{b};\quad H=\frac{h}{R}\left(\frac{R}{b}\right)^2 \\
\bar{\xi}=\frac{\xi}{b};\quad \bar{\rho}=\frac{\rho}{\rho_0};\quad \bar{\mu}^*=\frac{\mu^*}{\mu_0};\quad \bar{\mu}=\frac{\mu}{\mu_0} \\
\frac{1}{\bar{\mu}_{e0}}=\int_0^1\frac{1}{\bar{\mu}^*}dY;\quad \frac{1}{\bar{\mu}_{e1}}=\int_0^1\frac{Y}{\bar{\mu}^*}dY;\quad \frac{1}{\bar{\mu}_{e2}}=\int_0^1\frac{Y^2}{\bar{\mu}^*}dY.
\end{gathered} \tag{10}$$

In above equations, the parameter b is the semi-width of the Hertzian contact; parameters E & R represent the equivalent modulus of elasticity and the equivalent radius of the bounding surfaces, respectively.

Integrating equation (8) once and putting the result in dimensionless form yields:

$$H^3 Q_2 \frac{dP}{dX} = NH\left(1 + \frac{S}{2} Q_1\right) + \frac{1}{\bar{\rho}} C \qquad (11)$$

where

$$S = \frac{U_2 - U_1}{u_r}, \quad u_r = \frac{1}{2}(U_2 + U_1)$$

$$N = 4\bar{U}\left(\frac{\pi}{8\bar{W}}\right)^2, \quad \bar{U} = \frac{\mu_0 u_r}{ER}, \quad \bar{W} = \frac{w}{ER}$$

$$Q_1 = 1 - 2\frac{\bar{\mu}_{e0}}{\bar{\mu}_{e1}}, \quad Q_2 = \frac{1}{\bar{\mu}_{e2}} - \frac{\bar{\mu}_{e0}}{\bar{\mu}_{e1}^2}\ .$$

Parameters u_r and S represent the rolling speed and slide/roll ratio, respectively. Parameter C is the integration constant determined by using the boundary conditions given below:

$$P = 0 \quad at \quad X = -\infty,$$

$$P = \frac{dP}{dX} = 0 \quad at \quad X = X_b\,, \qquad (12)$$

where X_b is the dimensionless position where the film ruptures.

Consideration of the elastic deformation of the bounding surfaces leads to the following dimensionless equation for the film thickness, H,

$$H = H_{00} + \frac{X^2}{2} - \frac{1}{2\pi}\int_{-\infty}^{X_b} P(\bar{\xi})\,\ell n\left(X - \bar{\xi}\right)^2 d\bar{\xi}\ . \qquad (13)$$

The traction coefficient is the ratio of the traction force to the load-carrying capacity, both per unit length, viz.,

$$f = \frac{\int_{-\infty}^{x_b} \mu * \dot{\gamma}\, dx}{\int_{-\infty}^{x_b} p\, dx}. \tag{14}$$

To illustrate the generality of the equations and the solution procedure, we shall employ, as an example, the Bair-Winer's constitutive equation as given in (2). Introducing a dimensionless shear rate parameter $\lambda = \frac{\mu|\dot{\gamma}|}{\tau_L}$ and rearranging equation (2) yields $\tau/\tau_L = 1 - e^{-\lambda}$. From the latter equation, an expression for the equivalent viscosity in terms of the shear rate immediately follows:

$$\frac{\mu *}{\mu} = \frac{1 - e^{-\lambda}}{\lambda}. \tag{15}$$

It is to be noted that the absolute value sign is inserted in the definition of λ in order to keep the proper sign.

Finally, we must address the variation of properties with pressure. Roeland's expression for viscosity (Roelands et al. 1963) and the expression given by Dowson & Higginson (1966) for density variation are appropriate for most EHL applications. The relationships are:

$$\mu = \mu_0 exp\left\{\left(\ell n\, \mu_0 + 9.668\right)\left[-1 + \left(1 + 5.1 \times 10^{-9} p\right)^z\right]\right\}$$

$$\rho = \rho_0\left(1 + \frac{d_1 p}{1 + d_2 p}\right), \tag{16}$$

where μ_0 and ρ_0 are the reference viscosity and density, respectively. Constants d_1, d_2 and z depend upon the type of the lubricant.

The limiting shear stress varies with pressure in accordance to the following equation (cf. Bair & Winer 1990):

$$\tau_L = \tau_{L0} + \beta p, \tag{17}$$

where τ_{L0} and β are constants unique to a specified oil.

3 Numerical Formulation

The system of the governing equations presented in § 2 are treated numerically using the finite difference method for a specified constitutive equation for the lubricant; since, in general $\mu^* = \mu^*(\dot{\gamma})$ and $\dot{\gamma} = \dot{\gamma}(\mu^*)$, a trial-and-error routine is needed to determine the equivalent viscosity and the shear rate $\dot{\gamma}$. For the first iteration, the Hertzian pressure distribution is assumed and once μ^* and $\dot{\gamma}$ are determined, the program proceeds to solve the Reynolds equation. The method of Newton-Raphson was chosen for this purpose (cf. Houpert & Hamrock 1986). A description of the formulation follows.

The dimensionless Reynolds equation (10) is first written in the following form:

$$f_i = H_i^3(P_{i+1} - P_i) - \Delta X \cdot \frac{1}{Q_{2i}}\left[NH_i\left(1 + \frac{S}{2}Q_{1i}\right) + \frac{1}{\bar{\rho}_i}C\right] \quad i = 1, n\,. \tag{18}$$

Using the Newton-Raphson method, the solution is obtained by letting f = 0 and solving for the unknown parameters P_2, P_3, ..., P_n, H_{00} and C from the matrix:

$$[A]\{\chi\} = \{B\}, \tag{19}$$

where

$$[A]=\begin{bmatrix} \frac{\partial f_1}{\partial C} & \frac{\partial f_1}{\partial P_2} & \cdots & \frac{\partial f_1}{\partial P_n} & \frac{\partial f_1}{\partial H_{00}} \\ \frac{\partial f_2}{\partial C} & \frac{\partial f_2}{\partial P_2} & \cdots & \frac{\partial f_2}{\partial P_n} & \frac{\partial f_2}{\partial H_{00}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial f_n}{\partial C} & \frac{\partial f_n}{\partial P_2} & \cdots & \frac{\partial f_n}{\partial P_n} & \frac{\partial f_n}{\partial H_{00}} \\ 0 & \frac{4\Delta X}{3} & \cdots & \frac{2\Delta X}{3} & 0 \end{bmatrix}^{(k)},$$

$$\{\chi\}=\left(C \quad P_2 \quad P_3 \quad .... \quad P_n \quad H_{00}\right)^{T\,(k+1)},$$

$$\{B\}=\begin{bmatrix} \frac{\partial f_1}{\partial C}C+\sum_{\ell=2}^{n}\frac{\partial f_1}{\partial P_\ell}P_\ell+\frac{\partial f_1}{\partial H_{00}}H_{00}-f_1 \\ \frac{\partial f_2}{\partial C}C+\sum_{\ell=2}^{n}\frac{\partial f_2}{\partial P_\ell}P_\ell+\frac{\partial f_2}{\partial H_{00}}H_{00}-f_2 \\ \vdots \\ \frac{\partial f_n}{\partial C}C+\sum_{\ell=2}^{n}\frac{\partial f_n}{\partial P_\ell}P_\ell+\frac{\partial f_n}{\partial H_{00}}H_{00}-f_n \\ \frac{\pi}{2} \end{bmatrix}^{(k)},$$

where we have made use of $\int_{-\infty}^{X_b} P dX = \pi/2$ in the (n + 1)th row of equation (19).

The partial differential coefficients of Jacobian factors in [**A**] are given as follows:

$$\frac{\partial f_i}{\partial C}=-\Delta X\frac{1}{Q_{2i}}\frac{1}{\bar{\rho}_i} \qquad (20)$$

$$\frac{\partial f_i}{\partial H_{00}}=3H_i^2(P_{i+1}-P_i)-\Delta X\frac{1}{Q_{2i}}N\left(1+\frac{S}{2}Q_{1i}\right)+\frac{\Delta X}{Q_{2i}^2}\left[NH_i\left(1+\frac{S}{2}Q_{1i}\right)+\frac{1}{\bar{\rho}_i}C\right]\frac{\partial Q_{2i}}{\partial H_i}$$

$$-\frac{\Delta X}{Q_{2i}} NH_i\left(\frac{S}{2}\frac{\partial Q_{1i}}{\partial H_i}\right) \tag{21}$$

$$\frac{\partial f_i}{\partial P_\ell} = H_i^3\left(\delta_{i+1,\ell}-\delta_{i,\ell}\right) + \frac{\partial f_i}{\partial H_{00}}\cdot\frac{\partial H_i}{\partial P_\ell} + \frac{\Delta X}{Q_{2i}^2}\left[NH_i\left(1+\frac{S}{2}Q_{1i}\right)+\frac{1}{\bar{\rho}_i}C\right]\cdot\frac{\partial Q_{2i}}{\partial P_\ell}$$
$$-\frac{\Delta X}{Q_{2i}}NH_i\left(\frac{S}{2}\cdot\frac{\partial Q_{1i}}{\partial P_\ell}\right) - \frac{\Delta X}{Q_{2i}}C\cdot\frac{\partial}{\partial P_\ell}\left(\frac{1}{\bar{\rho}_i}\right) \tag{22}$$

where $\delta_{i,\ell}$ the Kronecker delta defined as:

$$\delta_{i,\ell} = \begin{cases} 1, & when \quad i=\ell \\ 0, & when \quad i\neq\ell \end{cases}$$

and

$$\frac{\partial Q_{1i}}{\partial H_i} = -2\bar{\mu}_{e0}\int_0^1 \frac{\partial}{\partial H_i}\left(\frac{1}{\bar{\mu}^*}\right)Y\,dY + \frac{2\bar{\mu}_{e0}^2}{\bar{\mu}_{e1}}\int_0^1 \frac{\partial}{\partial H_i}\left(\frac{1}{\bar{\mu}^*}\right)dY \tag{23}$$

$$\frac{\partial Q_{2i}}{\partial H_i} = \int_0^1 \frac{\partial}{\partial H_i}\left(\frac{1}{\bar{\mu}^*}\right)Y^2\,dY - \frac{2\bar{\mu}_{e0}}{\bar{\mu}_{e1}}\int_0^1 \frac{\partial}{\partial H_i}\left(\frac{1}{\bar{\mu}^*}\right)Y\,dY + \frac{\bar{\mu}_{e0}^2}{\bar{\mu}_{e1}^2}\int_0^1 \frac{\partial}{\partial H_i}\left(\frac{1}{\bar{\mu}^*}\right)dY \tag{24}$$

$$\frac{\partial Q_{1i}}{\partial P_\ell} = -2\bar{\mu}_{e0}\int_0^1 \frac{\partial}{\partial P_\ell}\left(\frac{1}{\bar{\mu}^*}\right)Y\,dY + \frac{2\bar{\mu}_{e0}^2}{\bar{\mu}_{e1}}\int_0^1 \frac{\partial}{\partial P_\ell}\left(\frac{1}{\bar{\mu}^*}\right)dY \tag{25}$$

$$\frac{\partial Q_{2i}}{\partial P_\ell} = \int_0^1 \frac{\partial}{\partial P_\ell}\left(\frac{1}{\bar{\mu}^*}\right)Y^2\,dY - \frac{2\bar{\mu}_{e0}}{\bar{\mu}_{e1}}\int_0^1 \frac{\partial}{\partial P_\ell}\left(\frac{1}{\bar{\mu}^*}\right)Y\,dY + \frac{\bar{\mu}_{e0}^2}{\bar{\mu}_{e1}^2}\int_0^1 \frac{\partial}{\partial P_\ell}\left(\frac{1}{\bar{\mu}^*}\right)dY\,. \tag{26}$$

Equations (23)-(26) are left in the general form in terms of the equivalent viscosity $\bar{\mu}^*$. Now for the Bair-Winer's constitutive equation, it can be shown that:

$$\frac{\partial}{\partial H_i}\left(\frac{1}{\bar{\mu}^*}\right) = D\frac{\partial\left|\bar{\dot{\gamma}}\right|}{\partial H_i} \tag{27}$$

and

$$\frac{\partial}{\partial P_\ell}\left(\frac{1}{\bar{\mu}\,*}\right)=D\frac{\partial\left|\bar{\dot{\gamma}}\right|}{\partial P_\ell}-\frac{1}{\bar{\mu}\,*^2}e^{-\lambda}\frac{\partial\bar{\mu}}{\partial P_\ell}+\frac{\bar{\mu}}{\bar{\mu}\,*^2}\frac{1}{\lambda\bar{\tau}_L}\left[(1+\lambda)e^{-\lambda}-1\right]\frac{\partial\bar{\tau}_L}{\partial P_\ell} \tag{28}$$

where

$$D=\frac{\bar{\mu}}{\bar{\mu}\,*^2\lambda}\left[\frac{1}{\left|\bar{\dot{\gamma}}\right|}\left(1-e^{-\lambda}\right)-\frac{\bar{\mu}}{\bar{\tau}_L}e^{-\lambda}\right].$$

The dimensionless shear rate and limiting shear stress are defined as:

$$\bar{\dot{\gamma}}=\frac{\mu_0\dot{\gamma}}{\tau_{L0}}=\frac{B_2}{\bar{\mu}\,*}\left(\frac{B_1}{H}\bar{\mu}_{e0}+\frac{H}{4}\left(Y-\frac{\bar{\mu}_{e0}}{\bar{\mu}_{e1}}\right)\frac{dP}{dX}\right)\quad and\quad \bar{\tau}_L=\frac{\tau_L}{\tau_{L0}}, \tag{29}$$

where

$$B_1=S\bar{U}\left(\frac{\pi}{8\bar{W}}\right)^2\quad B_2=\frac{E}{\tau_{L0}}\left(\frac{8\bar{W}}{\pi}\right)^2.$$

Furthermore,

$$\frac{\partial\bar{\dot{\gamma}}}{\partial H_i}=\frac{B_2}{\bar{\mu}^*}D_2+B_2D_1\frac{\partial}{\partial H_i}\left(\frac{1}{\bar{\mu}^*}\right)-\frac{B_2}{\bar{\mu}^*}\left[D_3\bar{\mu}_{e0}^2\int_0^1\frac{\partial}{\partial H_i}\left(\frac{1}{\bar{\mu}^*}\right)dY+D_4\bar{\mu}_{e1}^2\int_0^1\frac{\partial}{\partial H_i}\left(\frac{1}{\bar{\mu}^*}\right)Y\,dY\right] \tag{30}$$

$$\frac{\partial\bar{\dot{\gamma}}}{\partial P_\ell}=\frac{B_2}{\bar{\mu}\,*}\cdot\frac{H_i}{4}\left(Y-\frac{\bar{\mu}_{e0}}{\bar{\mu}_{e1}}\right)\frac{1}{\Delta X}\left(\delta_{i+1,\ell}-\delta_{i,\ell}\right)+B_2D_1\frac{\partial}{\partial P_\ell}\left(\frac{1}{\bar{\mu}\,*}\right)$$
$$-\frac{B_2}{\bar{\mu}\,*}\left[D_3\bar{\mu}_{e0}^2\int_0^1\frac{\partial}{\partial P_\ell}\left(\frac{1}{\bar{\mu}\,*}\right)dY+D_4\bar{\mu}_{e1}^2\int_0^1\frac{\partial}{\partial P_\ell}\left(\frac{1}{\bar{\mu}\,*}\right)Y\,dY\right] \tag{31}$$

and

$$\frac{\partial\left|\dot{\bar{\gamma}}\right|}{\partial\varphi}=\begin{cases}\dfrac{\partial\dot{\bar{\gamma}}}{\partial\varphi} & \dot{\bar{\gamma}}\geq 0\\ \\ -\dfrac{\partial\dot{\bar{\gamma}}}{\partial\varphi} & \dot{\bar{\gamma}}<0\end{cases}\quad .$$

Parameters D_i (i = 1, 4) in equations (30) and (31) are expressed as follows:

$$D_1=\frac{B_1}{H}\bar{\mu}_{e0}+\frac{H}{4}\left(Y-\frac{\bar{\mu}_{e0}}{\bar{\mu}_{e1}}\right)\frac{dP}{dX} \tag{32}$$

$$D_2=-\frac{B_1}{H^2}\bar{\mu}_{e0}+\frac{1}{4}\left(Y-\frac{\bar{\mu}_{e0}}{\bar{\mu}_{e1}}\right)\frac{dP}{dX} \tag{33}$$

$$D_3=\frac{B_1}{H}-\frac{H}{4}\frac{dP}{dX}\frac{1}{\bar{\mu}_{e1}} \tag{34}$$

$$D_4=\frac{H}{4}\frac{dP}{dX}\frac{\bar{\mu}_{e0}}{\bar{\mu}_{e1}^2}\,. \tag{35}$$

To numerically evaluate expressions (27) and (28), we proceed by setting:

$$\zeta_k=\frac{\partial}{\partial H_i}\left(\frac{1}{\bar{\mu}*}\right)$$

and then substituting equation (30) in (27). The resulting equations take on the form:

$$a_k\zeta_k+b_k\sum_{j=1}^{m}\zeta_j\Delta Y+c_k\sum_{j=1}^{m}\zeta_j Y\Delta Y=d_k,\quad k=1,m \tag{36}$$

where

$$a_k=1-B_2D_1D;\quad b_k=\frac{\bar{\mu}_{e0}^2}{\bar{\mu}*}B_2D_3D$$

$$c_k=\frac{\bar{\mu}_{e1}^2}{\bar{\mu}*}B_2D_4D;\quad d_k=\frac{1}{\bar{\mu}*}B_2D_2D\qquad when\ \ \dot{\bar{\gamma}}\geq 0$$

and

$$a_k = 1 + B_2 D_1 D; \quad b_k = -\frac{\bar{\mu}_{e0}^2}{\bar{\mu}\,*} B_2 D_3 D$$

$$c_k = -\frac{\bar{\mu}_{e1}^2}{\bar{\mu}\,*} B_2 D_4 D; \quad d_k = -\frac{1}{\bar{\mu}\,*} B_2 D_2 D \qquad when \quad \bar{\dot{\gamma}} < 0 . \tag{37}$$

Similarly, setting:

$$\vartheta_k = \frac{\partial}{\partial P_\ell}\left(\frac{1}{\bar{\mu}\,*}\right),$$

equation (28) becomes:

$$a_k \vartheta_k + b_k \sum_{j=1}^{m} \vartheta_j \Delta Y + c_k \sum_{j=1}^{m} \vartheta_j Y \Delta Y = e_k, \quad k = 1, m \tag{38}$$

where a_k, b_k and c_k are given in (37), and

$$e_k = \pm \frac{1}{\bar{\mu}\,*} B_2 D \frac{H_i}{4}\left(Y_k - \frac{\bar{\mu}_{e0}}{\bar{\mu}_{e1}}\right)\frac{1}{\Delta X}\left(\delta_{i+1,\ell} - \delta_{i,\ell}\right) - \frac{1}{\bar{\mu}\,*^2} e^{-\lambda} \frac{\partial \bar{\mu}}{\partial P_\ell}$$

$$+ \frac{\bar{\mu}}{\bar{\mu}\,*^2} \frac{1}{\lambda \bar{\tau}_L}\left[(1+\lambda) e^{-\lambda} - 1\right] \frac{\partial \bar{\tau}_L}{\partial P_\ell}, \tag{39}$$

where the positive sign is chosen when $\overline{\dot{\gamma}} > 0$. Equations (36) and (38) each define a system of algebraic equations of the form [**A**] {ζ} = {D} and [**A**] {ϑ} = {E} which we solved numerically using the Gauss elimination method at every fixed X position.

4 Discussion

The formulation presented in §3 provides an efficient procedure for predicting the realistic behaviour of lubricating oil in the elastohydrodynamic line contact and therefore should serve as a powerful analytical tool. It is not the intention of this paper to present extensive performance parameters; nevertheless, it is worthwhile to examine some typical results based on the numerical procedure described in §3. The simulated results for the film thickness and the pressure profile with the Bair-Winer's rheological model are illustrated in Figure 1. For comparison purposes, the same operating conditions and material properties were also programmed with the assumption that the lubricant displays purely Newtonian behavior throughout the entire EHL contact zone (Figure 2).

The input values which characterized the viscosity and density variation in accordance to relationships given in equation (16) were $z = 0.59$, $d_1 = 0.6 \times 10^{-9}$ Pa^{-1}, and $d_2 = 1.7 \times 10^{-9}$ Pa^{-1}. For the Bair-Winer's model, a limiting shear stress of $\tau_{L0} = 1.385 \times 10^{7}$ Pa with $\beta = 0.05$ were assumed; see equation (17).

Comparison of Figures 1 and 2 reveals that the assumption of Newtonian fluids tends to predict a thicker film than that of the non-Newtonian fluid model which takes into account the shear-thinning property of the lubricating oil in the EHL contact. From the pressure profiles in Figure 1, it is immediately apparent that in the absence of thermal effects, sliding tends to suppress the magnitude of the pressure spike which occurs in the vicinity of the minimum film thickness. In contrast, sliding does not influence the pressure profile when the fluid is assumed to be Newtonian since the term $Q_1 = 0$ in equation (11), hence the contribution of sliding on the pressure distribution vanishes identically.

Figure 1
near here

Figure 2
near here

The inadequacy of the linearly viscous rheological model quite clearly manifests itself in an unrealistic prediction of the traction coefficient particularly at high sliding/rolling ratios (Figure 3). In contrast, the non-Newtonian simulations exhibit a marked non-linearity in the way the traction coefficient varies when plotted against the rolling/sliding ratio. This trend is consistent with experimental observations.

Figure 3
near here

Interestingly, however, at very small sliding ratios both the generalized non-Newtonian simulations and those for the linearly viscous model are in fair agreement. The extent of this conformity may indeed vary depending on the operating conditions, in particular the magnitude of the load.

5 Concluding Remarks

Realistic prediction of the behaviour of a lubricant in elastohydrodynamic contact requires consideration of proper rheological model(s) that take the non-Newtonian characteristics of the fluid into account. In this paper a general formulation and solution methodology is presented that can treat an elastohydrodynamic lubrication problem with any simple, non-Newtonian constitutive equation. Extension of this work to include thermal effects with comparison with experimental measurement and presentation of the results for the mixture problems has also been accomplished based on the formulation presented here. A report which contains this information will be submitted directly to the Wright-Patterson Air Force Laboratories in the near future.

References

Bair, S. and Winer, W.O. 1979 A rheological model for EHL contact based on primary laboratory data, *J. Lub. Tech.*, *Trans. ASME*, **101**, pp. 258-265.

Bair, S. and Winer, W.O. 1990 The high shear stress rheology of liquid lubricants at pressures of 2 to 200 MPa, *J. Trib.*, *Trans. ASME*, **112**, pp. 246-252.

Conry, T.F., Johnson, K.L. and Owen, S. 1980 Viscosity in thermal regime of EHD traction, *Thermal Effects in Tribology*, Proc. 6th Leeds-Lyon Symp. on Trib., Dowson, D., Taylor, C.M., Godet, M. and Berthe, D. eds., Mech. Eng. Pub. Ltd., London, pp. 219-227.

Crook, A.W. 1961 Lubrication of rollers Part III, *Philos. Trans. R. Soc.*, London, Ser. **A 254**, pp. 237-258.

Dien, I.K. and Elrod, H.G. 1983 A generalized Reynolds equation for non-Newtonian fluids, with application to journal bearings, *J. Lub. Tech.*, *Trans. ASME*, **105**, pp. 385-390.

Dowson, D. and Higginson, G.R. 1966 *Elastohydrodynamic Lubrication*, Pergamon Press.

Evans, C.R. and Johnson, K.L. 1986a The rheological properties of elastohydrodynamic lubricants, *Proc. Inst. Mech. Engrs.*, Ser. C **200**, pp. 303-312.

Evans, C.R. and Johnson, K.L. 1986b Regimes of traction in elastohydrodynamic lubrication, *Proc. Inst. Mech. Engrs.*, Ser. C **200**, pp. 313-324.

Gecim, B. and Winer, W.O. 1980 Lubricant limiting shear stress effect on EHD contact, *J. Lub., Trans. ASME*, **102**, pp. 213-221.

Houpert, L.G. & Hamrock, B.J. 1986 A fast approach for calculating film thickness and pressure in elastohydrodynamic lubricated contacts at high loads, *J. Trib., Trans. ASME*, **108**, pp. 411-420.

Jacobson, B.O. 1985 A high pressure-short time shear strength analyser for lubricants, *J. Trib., Trans. ASME*, **107**, pp. 220-223.

Johnson, K.L. and Tevaarwerk, J.L. 1977 Shear behaviour of elastohydrodynamic oil films, *Proc. R. Soc. London*, Ser.**A 356**, pp. 215-236.

Khonsari, M.M. 1987 A review of thermal effects in hydrodynamic bearings, part I: slider and thrust bearings, *STLE Trans.*, **30**, pp. 19-25.

Khonsari, M.M., Wang, S.H. and Qi, Y. 1990 A theory of thermo-elastohydrodynamic lubrication liquid-solid lubricated cylinders, *J. Trib., Trans. ASME*, **112**, pp. 259-265.

Paranjpe, R. 1992 Analysis of non-Newtonian effects in dynamically loaded finite journal bearing including mass conserving cavitation, *J. Trib., Trans. ASME*, to appear.

Ree, T. and Eyring, H. 1955 Theory of non-Newtonian flow, part I: solid plastic system; II: solution system of high polymers, *J. Appl. Phys.*, **26**, pp. 793-809.

Roelands, C.J.A., Vlugter, J.C. and Waterman, H.I. 1963 The viscosity-temperature pressure relationship of lubricating oils and its correlation with chemical constitution, *J. Basic Engr., Trans. ASME*, **11**, pp. 601-610.

Sui, P.C. and Sadeghi, F. 1991 Non-Newtonian thermal elastohydrodynamic lubrication, *J. Trib., Trans. ASME*, **113**, pp. 390-396.

Tanner, R.I. 1963 Non-Newtonian lubrication theory and its application to the short journal bearing, *Aust. J. Appl. Sci.*, **14**, pp. 129-136.

Wang, J. and Zhang, H.H. 1992 A higher order perturbational approach in the lubricated EHL contacts with non-Newtonian lubricant, *J. Trib., Trans. ASME*, **114**, pp. 95-99.

Wang, S., Cusano, C. and Conry, T.F. 1991 Thermal analysis of elastohydrodynamic lubrication of line contact using the Ree-Eyring fluid model, *J. Trib., Trans. ASME*, **113**, pp. 232-243.

Wang, S., Cusano, C. and Conry, T.F. 1992 Thermal non-Newtonian elastohydrodynamic lubrication of line contact under simple sliding conditions, *J. Trib., Trans. ASME*, **114**, pp. 317-327.

Wang, S.H. and Zhang, H.H. 1987 Combined effects of thermal and non-Newtonian character of lubricant on pressure, film profile, temperature rise, and shear stress in EHL, *J. Trib., Trans. ASME*, **109**, pp. 660-670.

Figure 1 Pressure distribution and film thickness (Bair-Winer model)

______, S = 0; ...., S = 0.1

W = 5.52 x 10^{-5}; G = 5152; U = 2.28 x 10^{-11}

Figure 2 Pressure distribution and film thickness (Newtonian model)

W = 5.52 x 10^{-5}; G = 5152; U = 2.28 x 10^{-11}

Figure 3 Traction coefficient as a function of sliding/rolling ratio, S

_____ Bair-Winer model; ___ . ___ . ___ Newtonian model

W = 5.52 x 10^{-5}; G = 5152; U = 2.28 x 10^{-11}

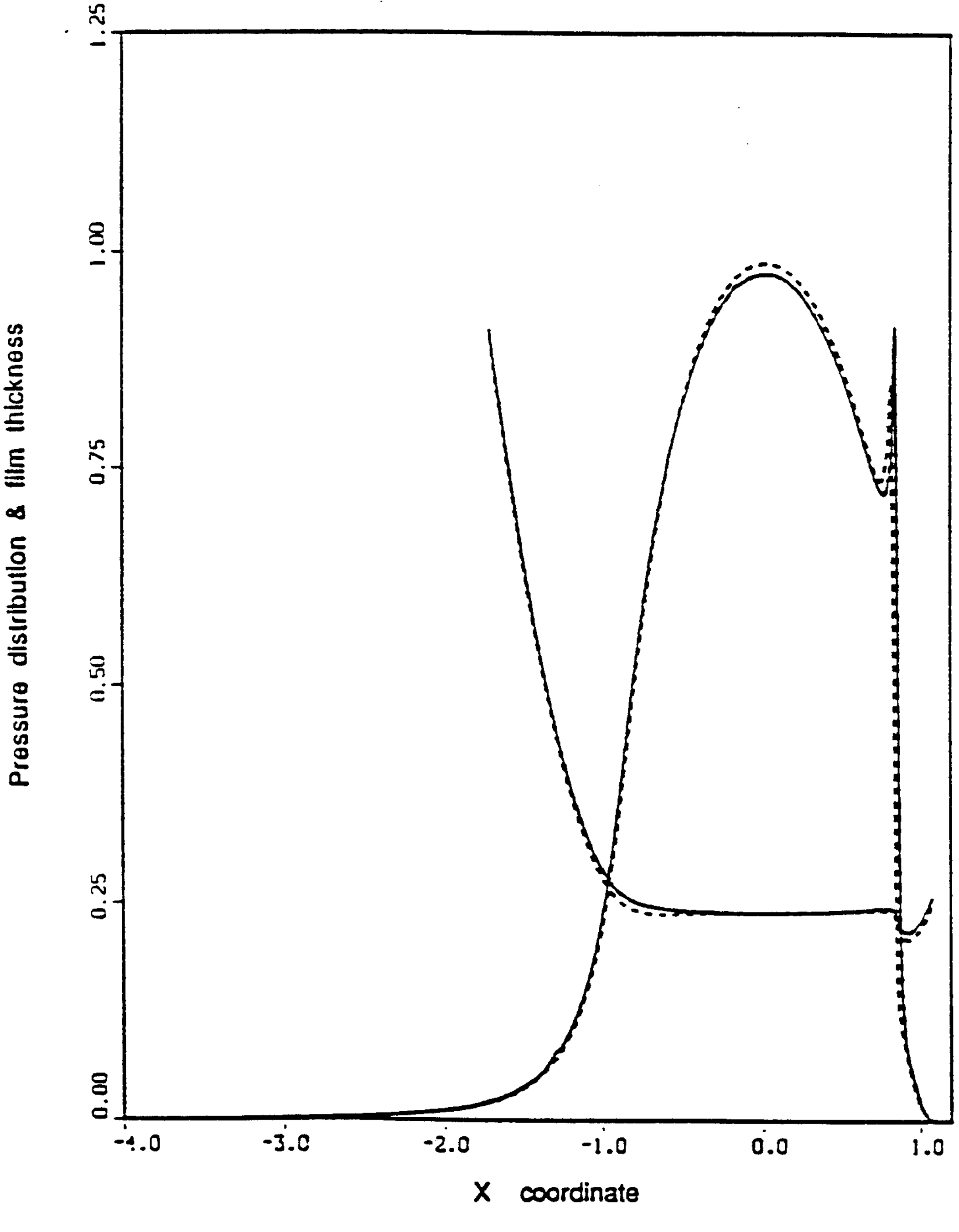


Figure 1 Pressure distribution and film thickness (Bair-Winer model)

______, S = 0; ...., S = 0.1

$W = 5.52 \times 10^{-5}$; $G = 5152$; $U = 2.28 \times 10^{-11}$

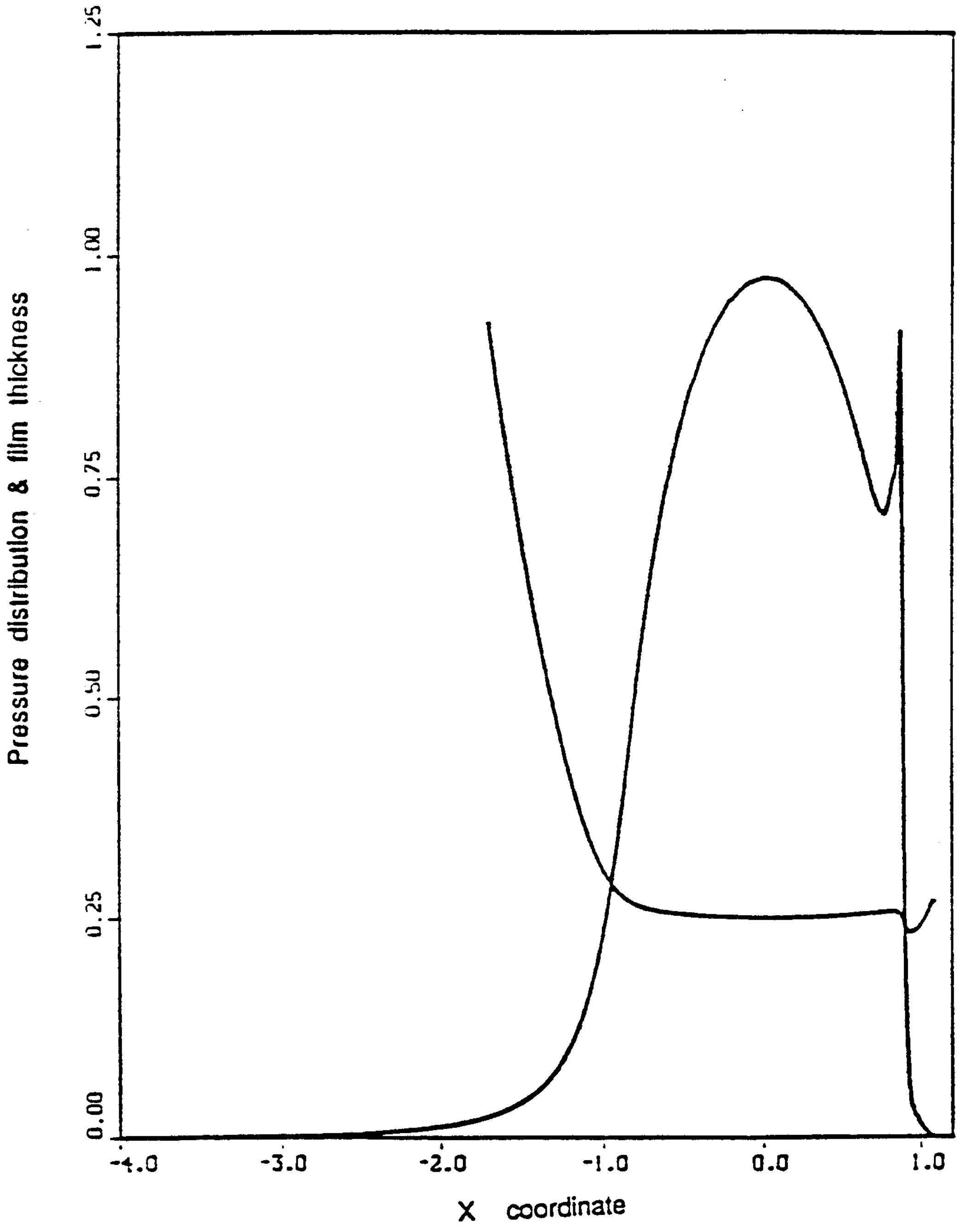


Figure 2 Pressure distribution and film thickness (Newtonian model)

W = 5.52 x 10^{-5}; G = 5152; U = 2.28 x 10^{-11}

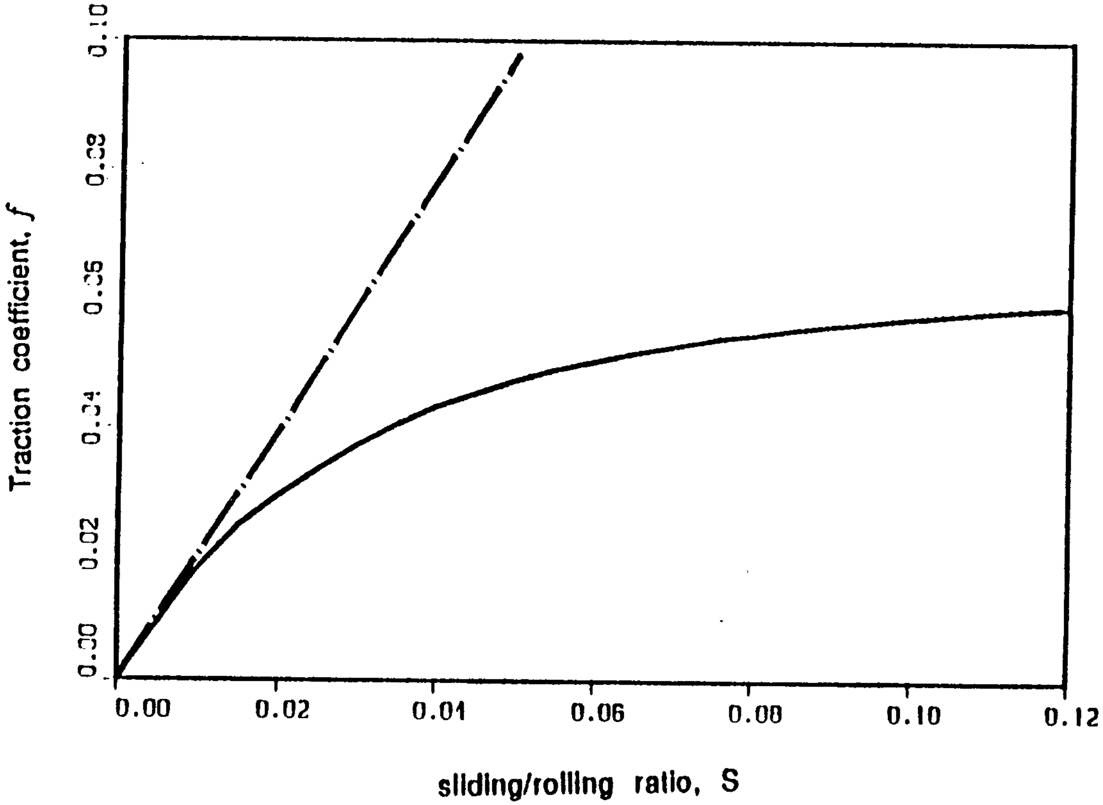


Figure 3 Traction coefficient as a function of sliding/rolling ratio, S

____ Bair-Winer model; ___ . ___ . ___ Newtonian model

$W = 5.52 \times 10^{-5}$; $G = 5152$; $U = 2.28 \times 10^{-11}$

STUDY OF DAMAGE ACCUMULATION AND FATIGUE FRACTURE MECHANISMS OF CORD-RUBBER COMPOSITES

B. L. Lee and D. S. Liu
Associate Professor and Graduate Student

The Pennsylvania State University
Department of Engineering Science and Mechanics
227 Hammond Building
University Park, PA 16802

Final Report for:

1992 Research Initiation Program
Air Force Wright Laboratory
Flight Dynamics Laboratory, Vehicle Subsystems Division

Sponsored by:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
(F49620-90-C-0076 R&D Laboratories)

01/01/92 - 12/31/92

December, 1992

TABLE OF CONTENTS

17-3

STUDY OF DAMAGE ACCUMULATION AND FATIGUE FRACTURE MECHANISMS OF CORD-RUBBER COMPOSITES

B. L. Lee and D. S. Liu
Associate Professor and Graduate Student

The Pennsylvania State University
Department of Engineering Science and Mechanics
227 Hammond Building
University Park, PA 16802

ABSTRACT

Fatigue fracture mechanisms and their dependence on cyclic loading frequency were assessed in the case of nylon fiber-reinforced elastomer matrix composite which represents the actual carcass of bias aircraft tires. Under uniaxial tension, the angle-plied carcass composite specimens were subjected to a considerably large interply shear strain before failure. The composite specimens exhibited semi-infinite fatigue life when stress amplitude was below a threshold level, i.e.,*fatigue endurance limit.*. Under cyclic stresses exceeding the endurance limit, localized damage in the form of fiber-matrix debonding and matrix cracking was formed and developed into the delamination eventually leading to gross failure of the composite. The process of damage accumulation was accompanied by a continuous increase of *cyclic strain* as well as *temperature*. Fatigue lifetime and the resistance to damage accumulation of aircraft tire carcass composite were strongly influenced by cyclic frequency. The use of higher cyclic frequency resulted in shorter fatigue lifetime at a given stress amplitude and lower endurance limit. The extent of *dynamic creep at gross failure*, which is defined as the increase of cyclic strain beyond initial elastic deformation, was roughly independent of stress amplitude under the frequency of 1 Hz, but decreased with higher stress amplitude when the frequency was raised to 10 Hz. Obviously a critical level of dynamic creep exists for gross failure of the composite and this level appears to be independent of the stress amplitude at low frequency. When the frequency is high enough, heat generation due to hysteretic loss is expected to degrade the materials. In this situation, the critical level of dynamic creep for gross failure is reduced by the loss of matrix flexibility as well as fiber-matrix bonding strength, and the degree of reduction becomes greater under higher stress amplitude.

STUDY OF DAMAGE ACCUMULATION AND FATIGUE FRACTURE MECHANISMS OF CORD-RUBBER COMPOSITES

B. L. Lee and D. S. Liu

I. INTRODUCTION

Our research effort (*1,2*) aims at the establishment of laboratory test methods and analytical models for the prediction of *structural durability* of aircraft tires. The effort for the prediction of tire durability will be based on the understanding of deformation and fracture behavior of material elements in critical regions. As reviewed in our previous report (*1*), aircraft tires are subjected to unusual combinations of speed and load compared with other types of pneumatic tires. Commonly-used aircraft tires of 49X17/26PR bias construction are rated at a speed of 358 km/hr (224 mph) and a load of 173 500 N (39 000 lbf). Such extreme combinations of speed and load result in large deflections and significant heat generation in the tires which are composed of fiber-reinforced elastomer matrix composites with various types of filled elastomers. As confirmed by field experience, these conditions cause damage in critical regions of aircraft tires such as shoulder, bead, lower sidewall or tread (*3,4*). The accumulation of damage eventually leads to catastrophic failures of whole tires. The nature and origin of cumulative damage are varied and quite complex.

For example, the failure of tire carcass (angle-plied composite plies comprising a pressure vessel part of each tire) in the shoulder area, often called a ply separation, occurs in the form of *delamination* which involves crack propagation mainly in the elastomer matrix and fiber-matrix interface. So-called bead area cracking and lower sidewall break also involve crack propagation in the elastomer matrix of carcass ply with some indications of *fiber fracture* as well. The processes of damage accumulation and structural failure of tire carcass in the shoulder, bead and lower sidewall areas are attributed to a combination of mechanical overloading and heat generation along with the resultant deterioration of constituent materials (*4,5*). In contrast, crack initiation in the filled elastomer of tread groove and subsequent propagation of cracks into the carcass of aircraft tires are attributed to the presence of strong centrifugal force resulting from unusually high speed particularly when so-called *standing wave* is present (*1,6-9*). In the case of *tread* failure, its major cause, the occurrence of standing waves can be avoided by a proper design of tires (*6*).

However, as far as the *carcass* plies in the shoulder, bead and lower sidewall areas are concerned, the occurrence of cumulative damage is difficult to avoid. Moreover it is not a simple task to identify exact failure modes and determine specific operating conditions or structural parameters which control the process of damage accumulation. This information is necessary to predict the useful life expectancy of an aircraft tire carcass. The operational life of aircraft tires is currently certified by costly dynamometer testing in which the tires are subjected to various combinations of speed and footprint load representing typical operating conditions in the field (*3,4*). The dynamometer testing clearly provides an accelerated means of evaluating the structural durability and integrity of aircraft tires. However, the usefulness of dynamometer tests in present form is limited by their empirical nature. The test results reflect merely the sensitivity of each particular tire design and construction to a given set of loading conditions, unless underlying mechanisms of property degradation, damage accumulation and structural failure of tires are identified.

In view of the limitations confronting a current means of predicting tire durability, our study plans to investigate the deformation and fracture mechanisms of angle-plied fiber-reinforced elastomer composite specimens which simulate the material elements of bias aircraft tire carcass under the cyclic loading. In order to identify the stress, strain or temperature parameters that control the process of fatigue damage accumulation, these tire carcass composite specimens will be subjected to more clearly defined condition of *uniaxial tension*, *biaxial tension*, *out-of-plane bending* or *in-plane shear* which in turn represents an individual component of complex loading in the actual tires. Eventually, it is hoped a fatigue law can be formulated that will serve to predict the life of the tire carcass on a more analytical basis even under highly complex loading history. When fully established, the results of our study are expected to complement empirical nature of dynamometer tests. In implementing the proposed plans, our initial research effort has been concentrated in examining the fatigue behavior of angle-plied carcass composite specimens under *uniaxial tension* which represents circumferential loading in the *shoulder* area of aircraft tires.

For a better determination of the failure modes, our study (*1,2,10*) utilized the *model* composites reinforced by steel wire cables. Subsequently the research program included nylon fiber cord-reinforced composite which represents the actual *aircraft tire carcass.* Under uniaxial cyclic loading which simulates fluctuating circumferential tension in the footprint region of tires, these angle-plied composite specimens were found to exhibit high levels of *interply shear* deformation. As reviewed in our previous reports, the deformation behavior of angle-plied composite laminates was analyzed by a number of investigators in the past (*11-17*). Interply shear strain develops in angle-plied laminates when the constituent plies exhibit in-plane shear deformation of opposite direction but the action is prevented by mutual constraint due to interply bonding. Compared with the case of fiber-reinforced plastic composites (*11*), fiber-reinforced elastomer composites exhibit unusually high level of interply shear strain which results from the load-induced change of reinforcement angle allowed by extreme compliance of elastomer matrix.

Our previous study confirmed that, at an axial tensile strain of 10 percent, an interply shear strain of around 30 percent develops in the nylon fiber-reinforced composite specimen representing the *bias aircraft tire carcass* with an initial reinforcement angle of +/-38 degree. Experimental study of the load-displacement response and interply shear strain variations was accompanied by detailed stress analysis based on finite element method. Linear elastic orthotropic or isotropic material elements were used for modeling. Our predictions were in reasonably good agreement with the experimental results. In examining fatigue fracture mechanisms, our previous study utilized mainly the *model* composite systems reinforced by steel wire cables at a reinforcement angle of +/-23 or +/-19 degree which results in a higher level of interply shear strain than the carcass composites. Above a critical value of interply shear strain, these angle-plied composites were found to exhibit localized failure initiated in the form of *fiber-matrix debonding*. Fiber-matrix debonding was first observed by Breidenbach and Lake (referred as 'socketing') in their pioneering works (*18,19*). Debonding is started around the cut ends of fibrous cord reinforcements at the edge of the finite width coupons. This phenomenon is justified since the maximum interply shear strain occurs at the edge of the specimen.

A new finding of our study (*1,2,10*) was that the critical load for the onset of fiber-matrix debonding constitutes a threshold level for semi-infinite fatigue life, i.e. *fatigue endurance limit*, of the model composites. The physical meaning of an endurance limit is that, with cyclic stresses lower than this threshold, fiber-matrix debonding is never initiated nor developed. Under cyclic tensile stresses exceeding the endurance limit, fiber-matrix debonding was found to be worsened progressively and developed into *matrix cracking* and *delamination* eventually leading to gross failure of the composites. The propagation of delamination and matrix damage in the elastomer composites under fatigue loading was analyzed in detail by Breidenbach and Lake (*18,19*), Huang and Yeoh (*20*), and Mandell et al (*21*). Related to this subject, a wealth of information exists in the area of fatigue fracture of filled elastomers (*22-29*). However, the correlation of the damage growth rate of composites with crack propagation characteristics of elastomers was not straightforward. The discrepancy was attributed to the departures of the real cracks in the composite laminates from the ideal shape assumed in fracture mechanics theory.

Our study showed that the damage accumulation in the forms of debonding, matrix cracking and delamination is accompanied by local strain increase (referred as *dynamic creep* hereafter), heat generation and acoustic emission (AE). The dynamic creep rate and the rate of temperature increase were inversely proportional to the fatigue life according to a power law (*2*). In monitoring of AE, distinctly different rates of signal accumulation could be assigned to the debonding and delamination failure modes (*2*). The same study also examined the effects of different stress parameters, including stress amplitude and mean stress, on the fatigue resistance of the model composites to a limited extent. Stress amplitude was found to play a dominant role in determining fatigue lifetime. Reviewing the progress of past investigations, it is clear that the discussed key findings of fatigue behavior study need to be confirmed in the case of nylon fiber-reinforced composite which represents the actual aircraft tire carcass.

In the present study of fatigue behavior of aircraft tire carcass composite, the following three topics are of immediate interest: (a) the effects of stress, strain and temperature *history* on the fatigue fracture mechanisms particularly under *high-frequency* cyclic loading; (b) the effects of *load sequence* and *frequency sequence* on the level of cumulative damage; (c) *monitoring* of damage accumulation process and attendant degradation of material properties. Under high-frequency cyclic loading, heat build-up due to hysteretic nature of constituent materials may play more prominent roles in controlling the degree of materials degradation and therefore the fatigue life of elastomer matrix composite. The question of how different load sequence or frequency sequence influences the damage accumulation process needs to be answered for the prediction of residual fatigue life of the composites under a complex loading history.

Finally our study of fatigue fracture mechanisms of tire carcass composites must include the development of reliable damage monitoring techniques as its integral part. When fully established, these experimental techniques for damage monitoring are expected to be *in-situ* or *real-time* means of predicting the residual life of tire carcass under fatigue loading. Within this context, our study has examined the above-mentioned three areas of interest for the angle-plied nylon fiber-reinforced composite which represents the actual aircraft tire carcass. Preliminary findings of these efforts will be reported in a series of technical papers. This paper, which is the first one of the series, will discuss the fatigue fracture mechanisms of aircraft tire carcass composite and their dependence on cyclic loading frequency.

II. OBJECTIVES

The present study was undertaken to assess the fatigue fracture mechanisms and their dependence on cyclic loading frequency in the case of fiber-reinforced elastomer matrix composite system which represents the actual aircraft tire carcass.

III. EXPERIMENTS

Angle-plied composite laminate specimens were prepared from the calendered plies used for construction of actual standard-grade carcass of *KC-135 aircraft tire*. The composite system was made of carbon black-filled proprietary elastomer compound matrix and 1260/2 nylon cord reinforcement laid at an angle of +/-38 degree (Table 1). The materials were supplied by The Goodyear Tire & Rubber Company (Akron, OH). To avoid tension-bending coupling, the laminates were prepared with a symmetric ply lay-up. The end tabs were added to the specimens to prevent grip failures that could be experienced during mechanical testing. Coupon specimens of 19mm width were machined from these panels and edges were polished on a grinding wheel.

The following four different *length-to-width ratios* were employed in testing of flat coupons of aircraft tire carcass composite: 5.3, 4.0, 3.3 and 0.5. In simulation of circumferential loading of a tire in the footprint region, these composite coupon specimens were subjected to uniaxial tension in *load-controlled cyclic mode*. Cyclic testing was performed to define S-N curves under the frequencies of 1, 5 and 10 Hz. The specimens were run until gross failure occurs. Fatigue testing was performed with continuous monitoring of displacement and specimen temperature. Temperature build-up as a function of fatigue life was recorded with a thermocouple attached to the specimen surface. Local strain was estimated by measuring the displacement of line markings drawn on the specimen edge.

IV. RESULTS AND DISCUSSIONS

Deformation and Failure Modes As discussed earlier, angle-plied laminate structure of fiber-reinforced elastomer composite exhibits unusually high level of interply shear strain under tension. Our previous study (*1*) showed that interply shear strain of up to 130 percent can be induced by static tensile strain of 10 percent in the case of model composite specimens (reinforced by wire cables at an angle of +/-19 degree). Compared with the case of model composites, the aircraft tire carcass composite system with nylon fiber reinforcement exhibited a smaller magnitude of interply shear deformation at a given axial strain because of its larger reinforcement angle of +/-38 degree. Static tensile strain of 10 percent resulted in an interply shear strain of around 30 percent in nylon fiber-reinforced carcass composite specimens (*1*). However, it should be noted that the axial strain at gross failure under static tension was more than 40 percent for nylon fiber composite, while the model composite failed at an axial strain of about 15 percent. Consequently, aircraft tire carcass composite specimens are subjected to a considerably large interply shear strain during fatigue loading before failure.

As in the case of model composites (*1,2,10*), the aircraft tire carcass composite system with nylon fiber reinforcement exhibited semi-infinite fatigue life (up to 10^7 cycles) when cyclic stress amplitude was below a threshold level, i.e. *fatigue endurance limit* (Figure 1). Under cyclic stresses exceeding the endurance limit, localized damage in the form of fiber-matrix *debonding* was readily observed around the cut ends of reinforcing fibers at the edge of the finite width coupons (Figure 2). A critical level of interply shear strain for fiber-matrix debonding corresponded to the axial stress of approximately 20% of static strength of the composites. Because of much shorter distances between the fibrous cords than those of model composites, the development of fiber-matrix debonding into *matrix cracking* could not be observed clearly in this material system. Debonding and matrix cracking were widened with increasing resultant strain and developed into the *delamination* eventually leading to gross failure of the composites (Figure 2). This sequence of failure modes was observed under cyclic loading as long as minimum stress remains tensile. One unique feature of the aircraft tire carcass composite was a relatively long period of time sustained after the onset of partial delamination. This tendency was more pronounced at lower stress amplitudes.

Specimen Size Effect The observed modes of damage accumulation, fiber-matrix debonding and matrix cracking leading to the delamination, are basically matrix-dependent failure processes. Since reinforcing fibers are laid off-axis in the angle-plied composite laminate specimens, fiber fracture *never* occurs in axial tension unless most fibers are gripped at both ends. From a geometric consideration, it is clear that the *length-to-width (L/W) ratio* of the specimen should be less than 1.28 (Cot 38 deg) to have this artificial situation. Our experiment confirmed that, in the case of composite specimens with a L/W ratio of 0.5, gripping of most reinforcing fibers at both ends leads to *fiber fracture* instead of delamination.

The effect of specimen width on the fatigue lifetime profile was further examined in this study by testing the specimens of three different L/W ratios (3.3, 4.0, 5.3). All of these specimens failed in *delamination*. As shown in Figures 1 and 3, the stress range vs fatigue life (S-N) curves of the aircraft tire carcass composites were found to be slightly dependent on L/W ratio. The specimens of higher L/W ratio tended to have shorter fatigue lives at a given stress range (i.e. 2 times amplitude). This trend probably resulted from diminishing effect of end constraint (grips) on Poisson's contraction of specimen width, which leads to a larger interply shear strain at mid-section in the case of longer specimen. However, the observed difference was close to the range of data scatterness.

Frequency Effects on S-N Data The effects of cyclic loading frequency on the fatigue resistance of nylon fiber-reinforced composites were examined with a *constant minimum stress* of 1.4 MPa (0.2 ksi). A series of S-N curves were generated under the frequencies of 1, 5 and 10 Hz. At a selected stress range, the frequency range was broadened to include 20 and 30 Hz. The result of fatigue lifetime profile at 20 and 30 Hz will be reported in our future paper. The S-N curves at 1, 5 and 10 Hz indicate that the use of higher cyclic frequency reduces the *fatigue lifetime* of composite specimens at a given stress range in a progressive manner (Figures 4, 5 and 6). Fiber-matrix debonding was observed in very early stage of fatigue life (sometimes several cycles) throughout the stress range tested. Consequently the increase of frequency appeared to have negligible influence on the number of cycles for the onset of debonding. However, the *fatigue endurance limit,* below which no damage accumulation occurs, was clearly lowered by the use of higher cyclic frequency (Figures 1 and 3). Under the frequency of 1 Hz, the endurance limit at 10^7 cycles was around 5 MPa which is about 20% of the static tensile strength. The endurance limit was lowered to 3 MPa under 10 Hz.

In addition to shorter fatigue life at a given stress amplitude and lower fatigue endurance limit, the use of higher cyclic frequency resulted in the increase of S-N curve slope which reflects higher rate of strength degradation (Figures 4 and 7). Consistently shorter fatigue lifetime and lower endurance limit of aircraft tire carcass composite under higher frequency could be attributed to a greater rate of *heat* generation due to hysteretic loss of constituent materials. As shown by Figures 8 and 9, the increase of specimen surface temperature was relatively small in the case of 1Hz, but the temperature could reach 100°C easily under the frequency of 10 Hz. The specimen temperature became lower as the stress amplitude is reduced. At low values of stress amplitude, the specimen temperature reached a plateau since the generated heat was balanced by the surrounding environment. It should be noted that the temperature is higher inside the specimen than the surface because of poor heat conduction characteristics of elastomers. Some local hot spots are expected to form and degrade the materials.

It was also noted that, the fatigue life of nylon fiber-reinforced composite specimens is relatively sensitive to the wave form used for cyclic loading. In the preliminary phase of our study, cyclic testing utilized a wave form which was not perfectly sinusoidal and instead was closer to triangular shape. As a result, the specimens were subjected to cyclic loading with noticeably shorter duration near the maximum stress than perfectly sinusoidal loading mode. In this situation, the fatigue lifetime of composite specimen at a given stress range was greatly extended. Our previous report (*30*) showing longer fatigue life under the frequency of 5 Hz compared with the case of 1 Hz was found to be an erroneous result caused by this effect.

Damage Accumulation As observed in the case of model composites, the process of damage accumulation in aircraft tire carcass composite was accompanied by a continuous increase of *cyclic strain* as well as temperature. The change of cyclic strain (either maximum strain or strain range) underwent three stages after an *initial stepwise* increase (Figure 10). At first, the strain increased at a progressively lower rate until reaching a steady-state region. In the steady-state region, the increase of cyclic strain at a constant rate allows the estimation of *dynamic creep rate*. Throughout the first and second regions which consume a major portion of fatigue life, the damage accumulation occurred in the forms of fiber-matrix debonding and matrix cracking. Towards the end of the second region, partial delamination appeared at the specimen edge. In the final region, the cyclic strain increased at a progressively higher rate eventually leading to a catastrophic *failure* in gross delamination mode. The duration of final region was relatively long in the case of aircraft tire carcass composite as discussed before.

The first two regions in the curve of cyclic strain vs time for nylon fiber composite strikingly resemble a static creep curve of four-element spring-dashpot model for typical polymeric materials (*31*). The model postulates that, at periods much longer than the retardation time, the only response to stress is due to the viscous flow. However, in our composite specimens, the increase of cyclic strain is apparently caused by crack growth from the debonded fiber ends into the matrix as observed visually. This effect of irreversible damage accumulation may be superimposed to the effect of viscoelastic properties of elastomer matrix and nylon fiber in controlling the degree of cyclic strain increase, i.e., dynamic creep. Further study is necessary to assess respective contributions of these two factors to the dynamic creep process. Both the resistance to damage accumulation and time-dependent nature of material properties can be influenced by cyclic frequency. As observed, the use of higher frequency results in more heat generation. Heat generation is expected to accelerate the process of irreversible degradation of materials including fiber-matrix interface or increase the time-dependence of constituent material properties.

Fracture Mechanisms Our study examined how the variations of cyclic frequency and stress amplitude influence dynamic creep behavior of aircraft tire carcass composite. As shown in Figure 11, the extent of *dynamic creep at gross failure*, which is defined as the increase of cyclic strain beyond initial elastic deformation (Figure 10), was roughly independent of stress range under the frequency of 1 Hz. Most specimens were found to fail when the maximum cyclic strain reaches approximately 30% above the initial strain level. The initial strain level was extrapolated from the steady-state region of dynamic creep curve. Obviously a critical level of dynamic creep exists for gross failure of the composites under cyclic loading of low frequency. This level appears to be independent of the stress range when heat-induced degradation of materials is minimal and/or the time-dependence of constituent properties are not seriously altered. The situation can be regarded as a *mechanical fatigue* where the level of cumulative damage or crack density will solely control the fatigue lifetime of composite structures.

One interesting fact is that, when the frequency is raised to 10 Hz, the extent of dynamic creep at gross failure decreases with higher stress range (Figure 12). In other words, *smaller* amount of damage accumulation or *less* creep of constituent materials is required to induce gross failure of the composites under cyclic loading with higher stress range. The result clearly indicates the involvement of *material degradation* rather than the increase of time-dependence of constituent material properties. When the frequency is high enough, heat generation due to hysteretic loss is expected to degrade fiber-matrix interface as well as elastomer matrix. The degradation of matrix elastomer is often accompanied by the increase of stiffness. Loss of matrix flexibility and the decrease of fiber-matrix bonding strength will certainly reduce a level of dynamic creep for gross failure of composites. The described effect should be more pronounced under cyclic loading of higher stress amplitude which generates more heat. This explanation will be verified independently by the measurement of residual stiffness and strength and chemical analysis of composite specimens after fatigue loading for various periods of time.

V. CONCLUDING REMARKS

Our study plans to investigate the deformation and fracture mechanisms of fiber-reinforced elastomer composite specimens which simulate the material elements of *bias aircraft tire carcass* in various critical regions. In order to identify the stress, strain or temperature parameters that control the process of fatigue damage accumulation, these composite specimens will be subjected to more clearly defined loading conditions which in turn represent individual components of complex loading in the actual tires. Eventually, it is hoped a fatigue law can be formulated that will serve to predict the life of the tire carcass on a more analytical basis even under highly complex loading history. In implementing the proposed plans, the initial research effort has been concentrated in examining the fatigue behavior of angle-plied carcass composite specimens under *uniaxial tension* which represents circumferential loading in the *shoulder* area of aircraft tires. For a better determination of the failure modes, the previous phase of our study (*1,2,10*) utilized mainly the *model* composite systems reinforced by steel wire cables at a smaller reinforcement angle.

The present study has assessed the fatigue fracture mechanisms and their dependence on cyclic loading frequency in the case of of nylon fiber-reinforced elastomer matrix composite which represents the *actual carcass* of KC-135 aircraft tire. Compared with the model composites, the aircraft tire carcass composite system exhibited a smaller magnitude of interply shear deformation at a given axial strain. However, as a result of much higher axial strain required for gross failure, the carcass composite specimens were subjected to a considerably large interply shear strain during fatigue loading before failure. The carcass composite specimens exhibited semi-infinite fatigue life when stress amplitude was below a threshold level, i.e. *fatigue endurance limit.*. Under cyclic stresses exceeding the endurance limit, localized damage in the form of fiber-matrix *debonding* was readily observed around the cut ends of reinforcing fibers. Debonding and associated matrix cracking were widened and developed into the *delamination* eventually leading to gross failure of the composites. One unique feature of the aircraft tire carcass composite was a relatively long period of time sustained after the onset of partial delamination.

Fatigue lifetime of aircraft tire carcass composite system was influenced by cyclic frequency and, to a much smaller extent, by the length-to-width ratio of the specimen. The S-N curves showed that the fatigue life of composite specimen at a given stress amplitude becomes shorter under higher cyclic frequency. The use of higher cyclic frequency also lowered the fatigue endurance limit. Consistently shorter fatigue life and lower endurance limit of carcass composite under higher frequency could be attributed to a greater rate of *heat* generation due to hysteretic loss of constituent materials. As observed in the case of model composites, the process of damage accumulation in aircraft tire carcass composite was accompanied by a continuous increase of *cyclic strain* as well as temperature. The strain increased at a progressively lower rate eventually reaching a steady-state region. Throughout these regions which consume a major portion of fatigue life, the damage accumulation occurred in the forms of debonding and matrix cracking. Partial delamination appeared near the end of the steady-state region. In the final region, the cyclic strain increased at a progressively higher rate eventually leading to a catastrophic failure.

The observed increase of cyclic strain (so called *dynamic creep*) is believed to result from irreversible damage accumulation as well as the effect of viscoelastic properties of constituent materials. The present study examined how the variations of cyclic frequency and stress amplitude influence dynamic creep behavior of aircraft tire carcass composite. The extent of *dynamic creep at gross failure*, which is defined as the increase of cyclic strain beyond initial elastic deformation, was roughly independent of stress amplitude under the frequency of 1 Hz, but decreased with higher stress amplitude when the frequency was raised to 10 Hz. Obviously a critical level of dynamic creep exists for gross failure of the composites and this level appears to be independent of the stress amplitude at low frequency. When the frequency is high enough, heat generation due to hysteretic loss is expected to degrade fiber-matrix interface as well as elastomer matrix. The resultant loss of matrix flexibility and the decrease of fiber-matrix bonding strength will lower a level of dynamic creep for gross failure of composites. The described effect should be more pronounced under cyclic loading of higher stress amplitude which generates more heat.

VI. RECOMMENDATIONS

The following recommendations can be made for our continuing study on the deformation and fracture behavior of nylon fiber-reinforced aircraft tire carcass composites:

(1) Broaden the data base showing the dependence of fatigue lifetime of aircraft tire carcass composite on stress, strain and temperature history. Define respective roles of *stress amplitude*, *minimum cyclic stress*, and *frequency* in controlling heat generation and a critical level of dynamic creep at gross failure.

(2) Assess the contribution of viscoelastic properties of constituent materials to the dynamic creep process of carcass composite by performing *static creep* loading experiments at various temperatures. Define the exact role of hysteretic heating in the determination of the fatigue lifetime of composite by performing *isothermal* cyclic testing. (Cyclic testing in isothermal condition will resolve the issue of the possible interaction between hysteretic heating and the progressive increase of strain due to damage accumulation.)

(3) Assess the mechanisms of *thermal fatigue* which may determine the lifetime of carcass composite under cyclic loading of ultra-high frequency (above 10 Hz). Determine how the process of damage accumulation interacts with the materials degradation by measuring residual strength or stiffness and chemical composition of the composite at various points of fatigue life.

(4) Derive empirical criteria for the prediction of the fatigue lifetime of carcass composite under the various combinations of stress amplitude and mean stress.

(5) Confirm that the extent of cumulative damage is independent of *load sequence* and *frequency sequence*. Establish the damage models for the prediction of fatigue life under random spectrum loading.

(6) Establish the measurement of local strain change, heat generation or acoustic emission (AE) as a viable experimental technique for *real-time monitoring* of the damage accumulation process. Correlate AE energy release rate with the corresponding strain energy release rate.

(7) Develop test methodologies for the laboratory simulation of other types of failure processes besides shoulder delamination of aircraft tires. (Other common types of failure processes include *bead* and *lower sidewall area failure*.)

(8) Examine the effects of footprint load, inflation pressure and speed on the mileage to failure, deflection and temperature history of aircraft tires, and assess the failure modes of tires in each condition. Correlate these results with the fatigue resistance data of laboratory composite specimens representing the tire carcass.

Acknowledgements We wish to thank the Air Force Systems Command and the Air Force Office of Scientific Research for sponsorship of this research. We sincerely appreciate continuing support and encouragement for our research program from Dr. A. H. Mayer, Mr. A. V. Petersons and Mr. P. C. Ulrich at the Vehicle Subsystems Division of Flight Dynamics Directorate of Air Force Wright Laboratory. We should thank to our colleagues at the Landing Gear Systems Group of Vehicle Subsystems Division, Dr. M. Chawla, Ms. D. A. Homa, Ms. D. A. Kennedy, Messrs. N. Lindsley, J. P. Medzorian and P. M. Wagner, for the discussion of this research work. Finally our sincere thanks should go to Dr. A. G. Causa, Dr. Y. M. Huang and Mr. R. Burgan at The Goodyear Tire & Rubber Company for providing composite specimens.

References

(*1*) B. L. Lee, J. P. Medzorian, P. M. Fourspring, G. J. Migut, M. H. Champion, P. M. Wagner and P. C. Ulrich, "Study of Fracture Behavior of Cord-Rubber Composites for Laboratory Prediction of Aircraft Tire Durability", SAE International Aerospace Technology Conference, Paper #901907, Long Beach, CA (1990).

(*2*) B. L. Lee, J. P. Medzorian, P. K. Hippo, D. S. Liu and P. C. Ulrich, "Fatigue Lifetime Prediction of Angle-Plied Fiber-Reinforced Elastomer Composites as Pneumatic Tire Materials", ASTM Second Symposium on Advances in Fatigue Lifetime Predictive Techniques, Pittsburgh, PA (ASTM Standard Testing Publication #1211 in print) (1992).

(*3*) Personal communications with the researchers in tire industry.

(*4*) S. N. Bobo, "Fatigue Life of Aircraft Tires", *Tire Science and Technology*, Vol. 16, No. 4, p.208 (1988).

(*5*) S. K. Clark, "Loss of Adhesion of Cord-Rubber Composites in Aircraft Tires", *Tire Science and Technology*, Vol. 14, No. 1, p.33 (1986).

(*6*) J. Padovan, "On Standing Waves in Tires", *Tire Science and Technology*, Vol. 5, No. 2, p.83 (1977).

(*7*) J. H. Champion and P. M. Wagner, "A Critical Speed Study for Aircraft Bias Ply Tires", AFWAL-TR-88-3006 (1988).

(*8*) J. H. Champion, S. K. Clark and M. K. Hilb, "A Study of Vibrational Modes in Rolling Aircraft Tires", WRDC-TR-89-3092 (1989).

(*9*) J. P. Medzorian, "Prediction of Aircraft Tire Critical Speed", WL-TR-92-3003 (1992).

(*10*) B. L. Lee, J. P. Medzorian, B. Ku, Y. M. Huang and A. G. Causa, "Fatigue Fracture of Fiber-Reinforced Composites with Compliant Matrix", In preparation.

(*11*) N. J. Pagano ed., Interlaminar Response of Composite Materials, Composite Materials Series Vol. 5, Elsevier Science Publ. Co., New York, NY (1989).

(*12*) A. Y. C. Lou and J. D. Walter, "Interlaminar Shear Strain Measurements in Cord-Rubber Composites", *Experimental Mechanics*, Vol. 18, 457 (1978).

(*13*) J. D. Walter, "Cord-Reinforced Rubber" in S. K. Clark ed. Mechanics of Pneumatic Tires, U.S. Department of Transportation, Washington D.C. (1982).

(*14*) D. O. Stalnaker, R. H. Kennedy and J. L. Ford, "Interlaminar Shear Strain in a Two-Ply Balanced Cord-Rubber Composites", *Experimental Mechanics*, Vol. 20, p.87 (1980).

(*15*) J. L. Ford, H. P. Patel and J. L. Turner, "Interlaminar Shear Effects in Cord-Rubber Composites", *Fiber Science and Technology*, Vol. 17, p.255 (1982).

(*16*) H. Rothert, B. Nguyen and R. Gall, "Comparative Study on the Incorporation of Composite Material for Tyre Computation", Composite Structures (Proc. of 2nd Int'l Conf. on Composite Structures, p.549, Applied Sci. Publ. (1983).

(17) R. J. Cembrola and T. J. Dudek, "Cord/Rubber Material Properties", *Rubber Chemistry and Technology*, Vol. 58, p.830 (1985).

(18) R. F. Breidenbach and G. J. Lake, "Mechanics of Fracture in Two-Ply Laminates", *Rubber Chemistry and Technology*, Vol. 52, p.96 (1979).

(19) R. F. Breidenbach and G. J. Lake, "Application of Fracture Mechanics to Rubber Articles Including Tyres", *Philosophical Trans. Royal Soc. London*, Vol. A299, p.189 (1981).

(20) Y. S. Huang and O. H. Yeoh, "Crack Initiation and Propagation in Model Cord-Rubber Composites", *Rubber Chemistry and Technology*, Vol. 62, 709 (1989).

(21) J. Kawamoto, "Fatigue of Rubber Composites", Ph.D. Thesis (Advisor: J. F. Mandell), M. I. T., Cambridge, MA (1988).

(22) R. S. Rivlin and A. G. Thomas, "Rupture of Rubber - I. Characteristic Energy for Tearing", *J. Polymer Science*, Vol. 10, p.291 (1953).

(23) A. N. Gent, P. B. Lindley and A. G. Thomas, "Cut Growth and Fatigue of Rubbers - I. The Relationship between Cut Growth and Fatigue, *J. Applied Polymer Science*, Vol. 8, p.455 (1964).

(24) G. J. Lake and A. G. Thomas, "The Strength of Highly Elastic Materials", *Proc. Royal Soc. London*, Vol. A300, p.108 (1967).

(25) A. N. Gent, "Strength of Elastomers", in Science of Technology of Rubber edited by F. R. Eirich, Chapter 10, Academic Press, New York, NY (1978).

(26) A. N. Gent, "Detachment of an Elastic Matrix from a Rigid Spherical Inclusion", *J. Materials Science*, Vol. 15, 2884 (1980).

(27) R. G. Stacer, L. C. Yanyo and F. N. Kelley, "Observations on the Tearing of Elastomers", *Rubber Chemistry and Technology*, Vol. 58, 421 (1985).

(28) D. G. Young, "Fatigue Crack Propagation in Elastomer Compounds: Effects of Strain Rate, Temperature, Strain Level and Oxidation", *Rubber Chemistry and Technology*, Vol. 59, 809 (1986)

(29) R. F. Lee and J. A. Donovan, "J-Integral and Crack Opening Displacement as Crack Initiation Criteria in Natural Rubber in Pure Shear and Tensile Specimens", *Rubber Chemistry and Technology*, Vol. 60, 674 (1987).

(30) B. L. Lee and D.S. Liu, "Fatigue Fracture Behavior of Cord-Reinforced Rubber Composites", Final Report on 1991 Research Initiation Program, Air Force Office of Scientific Research Contract S-210-11M6-088 (1991).

(31) L. E. Nielson, Mechanical Properties of Polymers, Chap.3, Reinhold (1962).

Table 1

Specifications of Aircraft Tire Carcass Composite Specimen
(Reinforcement: 1206/2 Nylon cord)
(Matrix: Proprietary elastomer compound)

Reinforcement Angle	+38, -38, -38, +38 deg
Cord Modulus	2.07 GPa (300X10^3 psi)
Matrix Modulus	5.51 MPa (800 psi)
Cross-Sectional Area of Cord	0.342 mm^2 (5.3X10^{-4} inch^2)
Specimen Width	19.05 mm (0.75 inch)
Specimen Thickness	6.35 mm (0.25 inch)

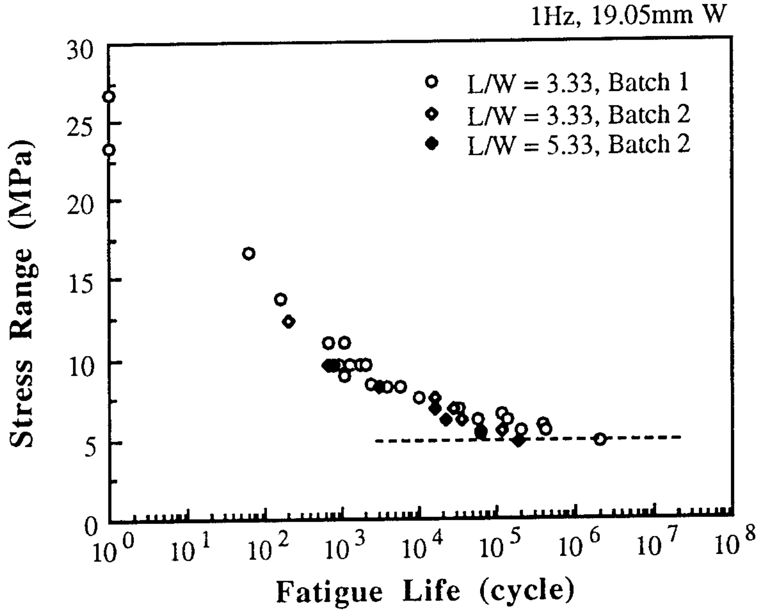


Figure 1

Stress Range vs Fatigue Life (S-N) Curve of Aircraft Tire Carcass Composites; 1 Hz Frequency; Semi-Log Scale.

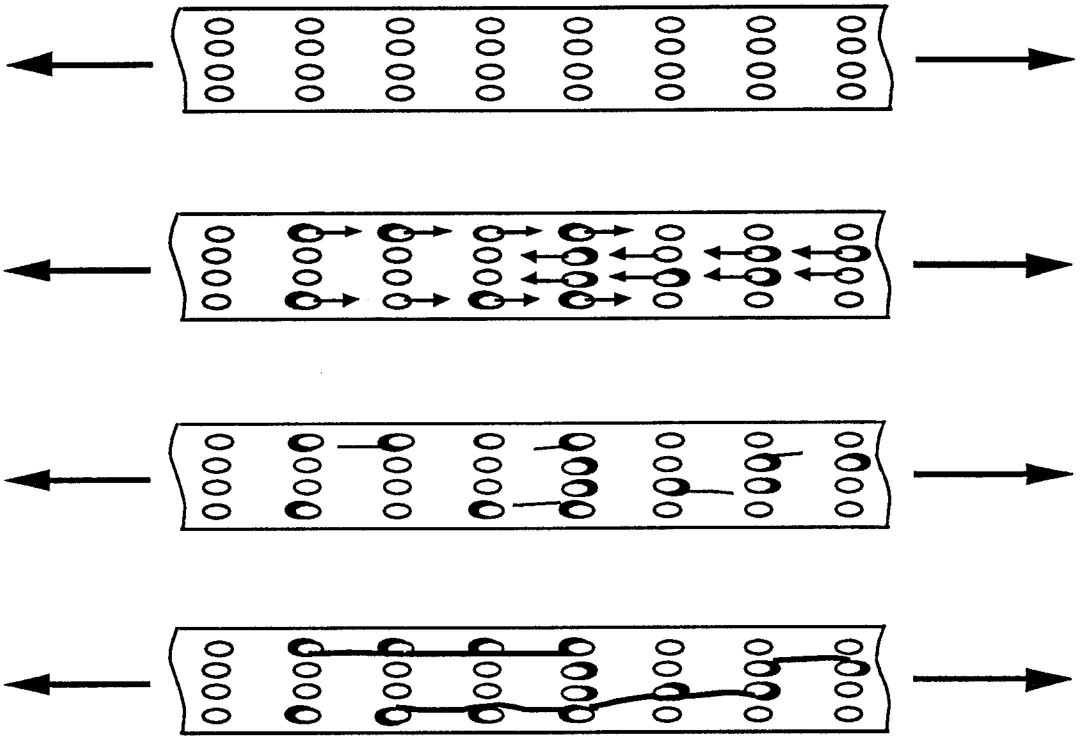


Figure 2

Failure Modes of Aircraft Tire Carcass Composites; +38/-38/-38/+38 Deg Reinforcement Angle.

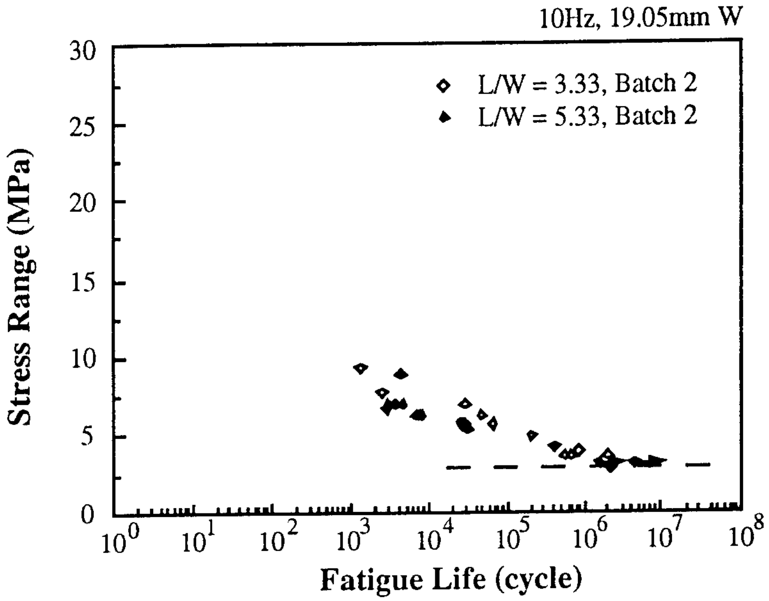


Figure 3

Stress Range vs Fatigue Life (S-N) Curve of Aircraft Tire Carcass Composites; 10 Hz Frequency; Semi-Log Scale.

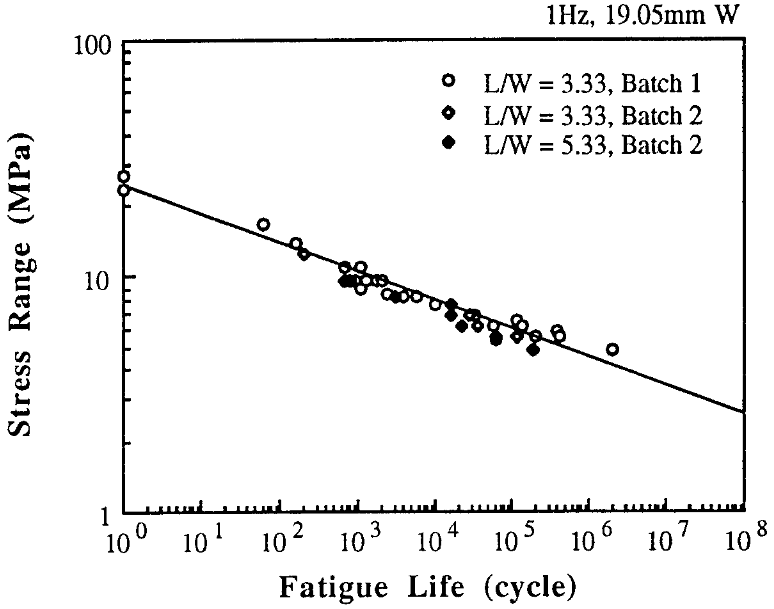


Figure 4

Stress Range vs Fatigue Life (S-N) Curve of Aircraft Tire Carcass Composites; 1 Hz Frequency; Log-Log Scale.

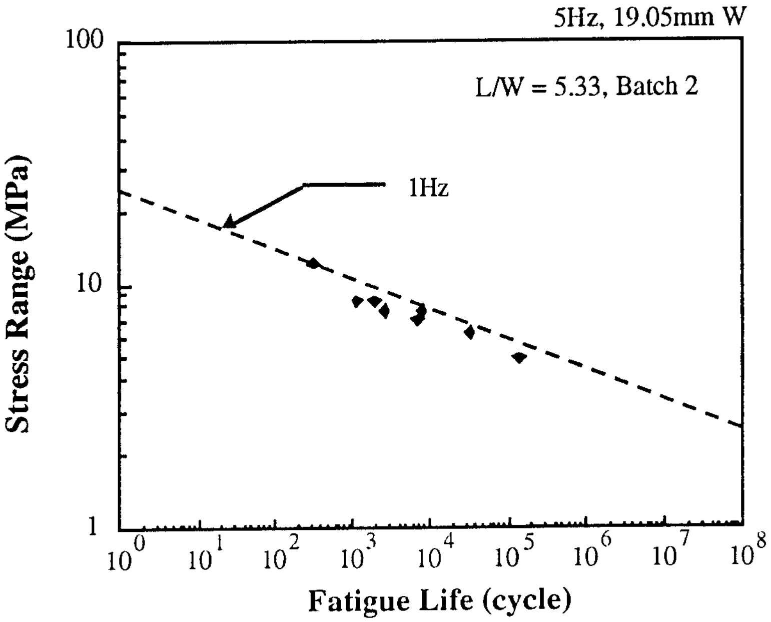


Figure 5

Stress Range vs Fatigue Life (S-N) Curve of Aircraft Tire Carcass Composites; 5 Hz Frequency; Log-Log Scale.

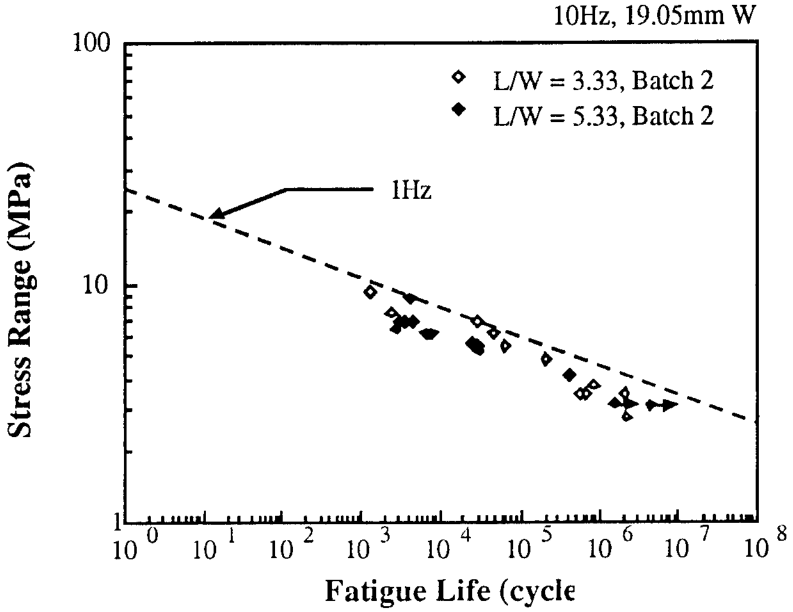


Figure 6

Stress Range vs Fatigue Life (S-N) Curve of Aircraft Tire Carcass Composites; 10 Hz Frequency; Log-Log Scale.

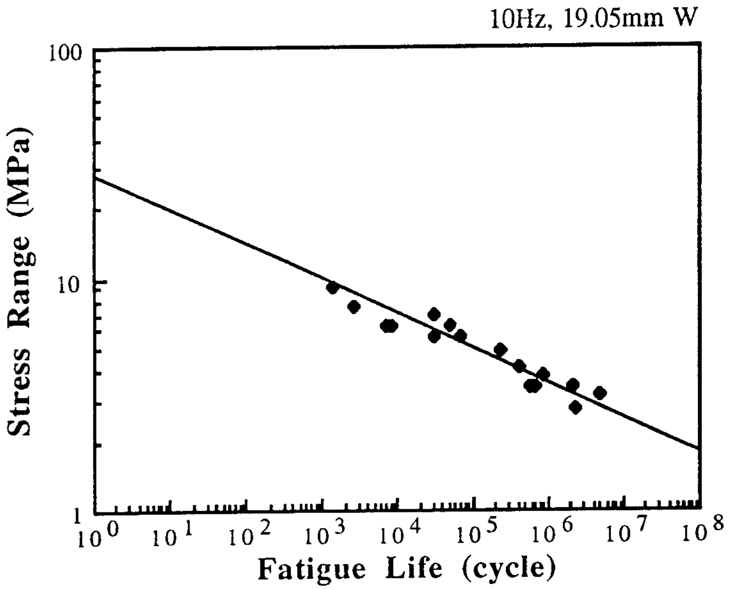


Figure 7

Stress Range vs Fatigue Life (S-N) Curve of Aircraft Tire Carcass Composites; 10 Hz Frequency; Log-Log Scale; L/W=3.3.

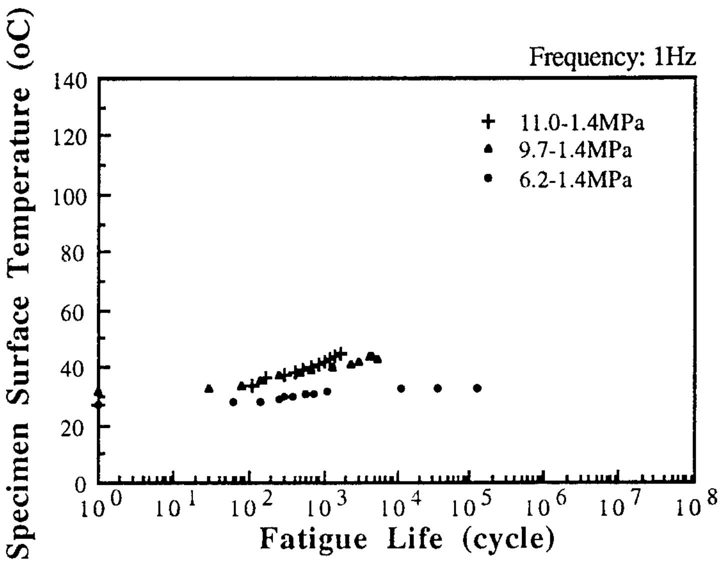


Figure 8

Specimen Temperature vs Fatigue Life of Aircraft Tire Carcass Composites; 1 Hz Frequency.

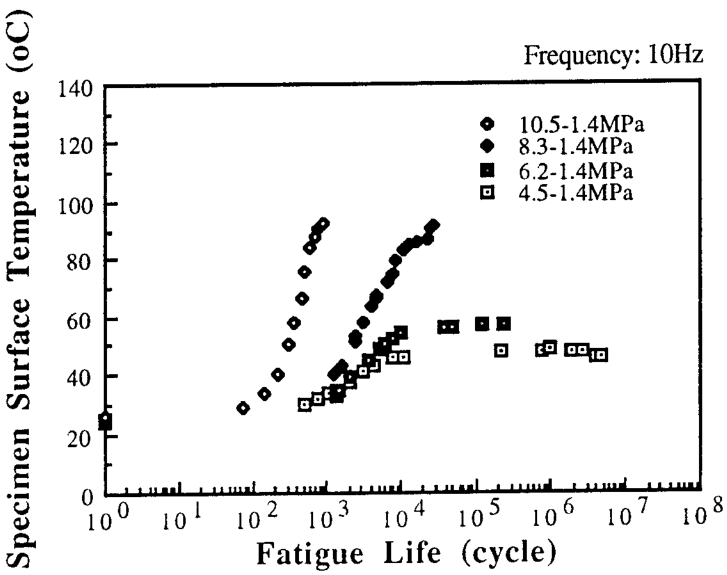


Figure 9

Specimen Temperature vs Fatigue Life of Aircraft Tire Carcass Composites; 10 Hz Frequency.

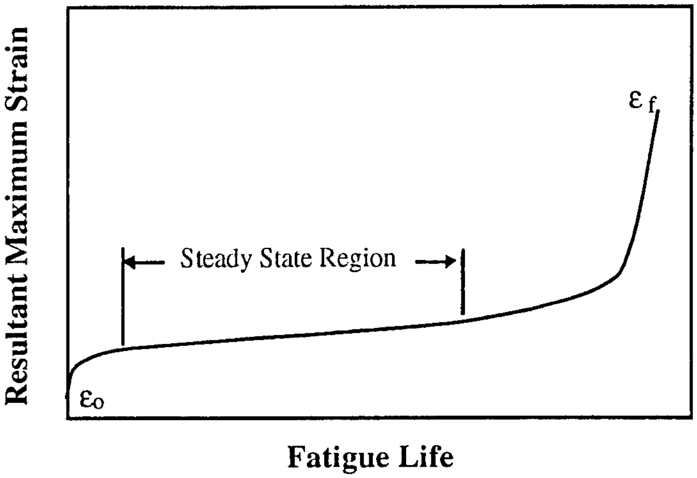


Figure 10

Increase of Resultant Maximum Strain vs Fatigue Life of Aircraft Tire Carcass Composites.

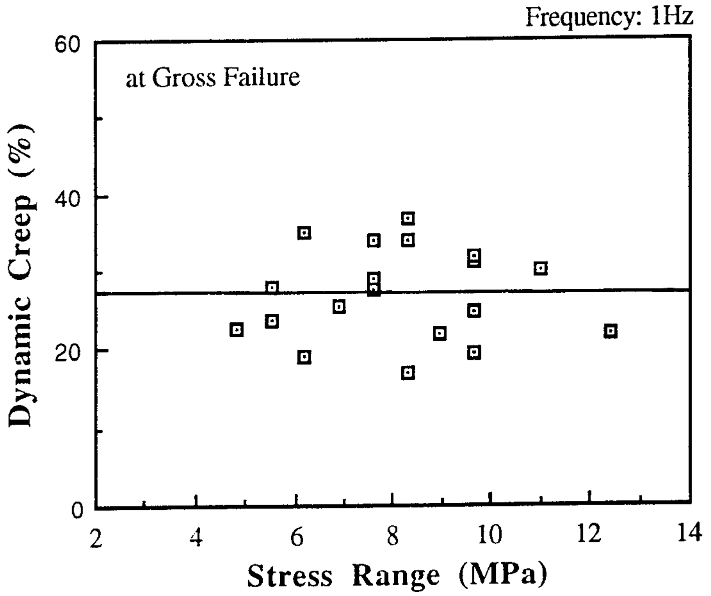


Figure 11

Dynamic Creep at Gross Failure vs Stress Range of Aircraft Tire Carcass Composites; 1 Hz Frequency.

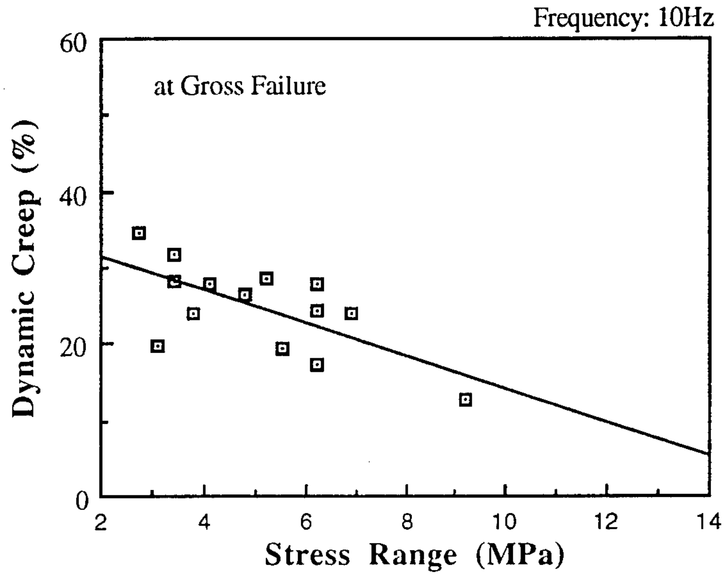


Figure 12

Dynamic Creep at Gross Failure vs Stress Range of Aircraft Tire Carcass Composites; 10 Hz Frequency.

TEXTURAL CHARACTERIZATION OF DEFORMED METALS

Jessica L. Mayes
Research Assistant
Department of Materials Science and Engineering

University of Kentucky
Lexington, KY 40506-0046

Final Report for:
Research Initiation Program
Wright Laboratory

Sponsored by:
Air Force Office of Scientific Research
Eglin Air Force Base, Florida

and

University of Kentucky

December 1992

TEXTURAL CHARACTERIZATION OF DEFORMED METALS

Jessica L. Mayes
Research Assistant
Department of Materials Science and Engineering
University of Kentucky

Abstract

This research project is in the broad general field of the theoretical understanding of plastic deformation in anisotropic metals. Two methods were implemented to determine grain orientation distributions for a specific metal. Pole figures were measured experimentally using an x-ray diffractometer and pole figures were calculated theoretically using a simulation program. Appropriate methodology is described with the intention of providing a basis for subsequent calculations (performed elsewhere) of anisotropic yield criteria and flow rules.

This project was carried out in cooperation with "Anisotropic Plastic Deformation of Metals", RIP number 12. The research was developed simultaneously under the direction of the same research supervisor and, therefore, the information contained within each report is applicable to both. For this reason, the background and general research plan are identical in the two reports which differ only in emphasis on theoretical modeling in this paper versus experimental data in the other report.

TEXTURAL CHARACTERIZATION OF DEFORMED METALS

Jessica L. Mayes

INTRODUCTION

It is frequently found that the mechanical properties of wrought metal products are not the same in all directions. The dependence of properties on orientation is called anisotropy [1]. Independence of properties on orientation is referred to as isotropy or the material is called isotropic. Very often anisotropy is troublesome, as in the formation of "ears" or non-uniform deformation in deep-drawn cups. However, it can also be beneficial, as when different strengths are required in different directions; for example, in a cylindrical pressure vessel. It is up to the ingenuity of engineers and materials scientists to control the deformation process of wrought metal products in order to derive the greatest benefit from the attendant structural alterations.

In striving to achieve this goal, the main emphasis of past development has been aimed toward isotropic materials. Deformation processing is frequently carried out at elevated temperatures (for example, hot rolling). Grain growth and recrystallization are enhanced at these temperatures, tending to keep the grain structure equiaxed and the material, therefore, relatively isotropic.

Two factors have promoted this past development. One has already been stated: very often anisotropy is troublesome. The other reason is the paucity of mathematical models available to the analyst attempting to understand and predict anisotropic plastic deformation.

In the early sixties, analyses such as those of Paul [2] and Greszczuk [3] opened the door to application of the work of Lekhnitskii [4] on anisotropic elasticity and interpretations such as the Primer [5]. Since that time, research interest in the necessarily anisotropic, composite materials has relied on these early elastic theories, even in the investigation of fracture. On the other hand, theories of plasticity have mainly assumed isotropic deformation.

The principal exceptions have been theories proposed by Hill [6-8] and by Hosford [9]. Of these, only the 1948 theory of Hill is fully three-dimensional. The other three are attempts to improve this early theory in the relatively simple case of plane stress. Plane stress is an approximation often applied to sheet metal deformation. The first theory of Hill, applied to plane stress deformation, leads to results that are inconsistent with experiments when simple tension is compared with balanced biaxial stretching. This discrepancy motivated the later theories, all of which are formulated to apply only to plane stress.

The second theory proposed by Hill is restricted to sheet materials that are anisotropic only with respect to the thickness direction; within the plane of the sheet, the material is assumed to be isotropic in its mechanical properties. This is an extremely restrictive idealization that severely limits applicability of this theory even though it incorporates a totally arbitrary "Hill Index" as an adjustable parameter that can be used to improve agreement between calculated and experimental results.

Both the theory proposed by Hosford and the most recent theory proposed by Hill accommodate planar anisotropy. Each also features a totally arbitrary, adjustable parameter. Consequently, it can be expected that reasonably good agreement can be obtained between theory and experiment using either of them. In fact, Hosford and co-workers [11-13] have shown some very favorable comparisons. At present, no one has yet published any calculations based on the recent Hill theory.

A substantially different approach to anisotropic plasticity theory is based on the seminal work of G.I. Taylor [13]. The so-called Taylor factor is a theoretical ratio of polycrystal to monocrystal yield strengths. This calculation, subsequently modified by Bishop and Hill [14] and Chin and Mammel [15], requires that the distribution of grain orientations in the polycrystal be specified. Taylor assumed a random distribution. He also assumed a polycrystalline deformation corresponding to homogeneous axial tension of an isotropic material. Material anisotropy can readily be introduced into such a calculation by prescribing a non-random distribution of grain orientations. Three-dimensionality can be dealt with through consideration of different deformation paths.

Current research into such anisotropic plasticity theories has been conducted by Kochs [16] and others at Los Alamos National Laboratory. Codes developed there are capable of predicting the fully three-dimensional yield surface, and corresponding plastic flow rule, for anisotropic, polycrystalline metals. The input required for such codes comprise uniaxial yield strengths and the grain orientation distribution function. With additional information concerning hardening, these codes can also predict subsequent flow surfaces and subsequent flow rules.

RESEARCH PLAN

Grain orientation distributions for a particular material of interest can be created in two ways: experimental measurement by an x-ray diffractometer and by theoretical calculations (codes). These codes can be further utilized to produce five-dimensional yield surfaces- an ultimate goal of mechanical analysis.

Some of the criteria for choosing a material for this type of research includes: workability, to attain several distinct levels of deformation; known anisotropy, so that the experimental grain orientation distributions are not always random; and other factors such as the ease of preparation of experimental specimens and availability of material After consulting with the focal lab, it was determined that a suitable material for establishing the desired methodology, which would simultaneously possess appropriate material properties for beneficial results in itself, would be OFE copper.

In order to facilitate determination of the experimental pole figures, modifications were necessary to the existing equipment. The University of Kentucky Department of Materials Science and Engineering has a Rigaku x-ray diffractometer outfitted with a Θ-2Θ goniometer. Because of recent "upgrades" to the x-ray capabilities of the equipment and software, the limited pole figure capabilities of the system were temporarily eliminated. In order to replace these options on the existing equipment with modern, convenient, PC-ready capabilities, it was necessary to send the unit to Japan for installation of new stepping motors and gears. In addition, the University acquired Rikagu's "Attachment Control" board and "Texture" software. One accomplishment of this research project was to leverage approximately $8000 of project funds into a $35,000 purchase of this new equipment, which is of central importance in follow-on research in this area.

On the computational side, Dr. Anthony Rollett has developed a very useful computer code [21] at Los Alamos National Laboratory for calculating theoretical pole figures. It is based upon postulated polycrystalline grain orientation distributions. A related code, developed there, can then take the postulated polycrystalline grain orientation distribution as the basis for computation of a texture-dependent polycrystalline yield surface. The University of Kentucky has obtained from Los Alamos copies of both popLA (preferred orientation package-Los Alamos) and LApp (Los Alamos polycrystal plasticity code). The basic research plan was to acquire the necessary x-ray equipment, determine pole figures for the selected material, and to apply these Los Alamos codes to the data generated.

METHODOLOGY

Having selected OFE copper as the project material, an initial stock of 8 mm (5/16 in.) diameter rod was obtained and fully characterized. Initially, tests were conducted to obtain yield strengths and other mechanical properties for specimens deformed at low strain rates (Standard Tensile Test). Because of the obvious applicability to technology, high strain rate deformation behavior data was also desired. To characterize this, both Taylor-Anvil Tests [23] and the Rod-on-Rod (ROR) Impact Tests [24] were performed using cylindrical projectiles.

In the Taylor Impact test, a cylindrical copper rod of 7.62 mm (0.3 inch) diameter is fired at high velocity into a large steel anvil, Figure 1a. In the Rod-on-Rod (ROR) Test, a similar impacting rod as that used in the Taylor test, 25 mm (1 inch) long, is impacted against a longer, 152 mm (6 inch), rod of the same material and cross-section, Figure 1b.

Special emphasis was placed on the ROR tests. By comparison to the anvil tests this test method eliminates rod-anvil compliance mismatch and frictional effects during impact. It also has the advantage of one continuous deformation medium throughout the impact sequence (especially beneficial for continuum code evaluations). For these reasons ROR testing became the principal technique in the study of high strain rate behavior.

After the ROR specimens were impacted and recovered they were sectioned axially, mounted, polished and etched for evaluation of their microstructural characteristics. A striking aspect of their microstructures was the occurrence of voids near the specimen axis near the impact edge. The severity of void formation depended upon the impact velocity and on the initial grain size. Investigation of this aspect of high strain rate material deformation behavior became a primary concern in the microstructural characteristic documentation of these specimens. This investigation resulted in a publication entitled "Void Formation in OFE Copper" that was presented at the 1992 Hypervelocity Impact Symposium and accepted for publication in the Journal of Impact Engineering [25].

The aforementioned microstructural characterization comprised a superficial analysis of the deformation. This is important in relation to damage models of material constitutive behavior for use in finite element, continuum codes. These voids dramatically indicate the occurrence of severe hydrostatic tension during what is commonly considered a compression test. This is a subject that certainly requires further study and must be taken into consideration in constitutive modeling.

However, the physical structure of the material on a much finer scale is what determines mechanical anisotropy and is the principal object of this project. Materials that show a preferred crystallographic orientation or texture exhibit a certain amount of mechanical anisotropy. Even cubic crystals such as copper, which have a high degree of symmetry become textured as a result of plastic deformation. The subsequent degree of anisotropy in a polycrystalline material is directly dependent on the properties of the individual crystals and the fraction oriented in a particular direction [26].

Texture can be quantitatively described by pole figure data. Once this data is obtained (and stored in ASCII format) it can be processed by appropriate software (popLA) to produce a corresponding grain orientation distribution, i.e., a three-dimensional Orientation Distribution Function (ODF). The ODF can subsequently be incorporated into a theoretical model of polycrystal plasticity (LApp) that will create a three-dimensional, anisotropic yield surface.

The experimental pole figures themselves are the backbone of the comparative analysis. They provide the baseline for development and fine-tuning of the data that comprises the eventual yield surface.

The test samples that were used in the Rigaku x-ray unit to produce pole figures were prepared as follows. An impact specimen was sliced perpendicular to its axis using a wafering saw with a thickness of 2.5 mm (0.10 inch). The resulting wafers were approximately 5 mm (0.2 inch) thick. They were ultrasonically cleaned to remove any materials introduced by the sectioning process. The wafers were then mechanically polished using 1 μm and 0.3 μm alumina powder to remove the deformation layer created by wafering. The specimen was ultrasonically cleaned after each polishing step to prevent cross-contamination between alumina particle sizes. After the polishing was complete the specimens were then chemically etched to remove the fine layer of deformation created by polishing. Careful attention must be paid to avoid overheating or plastic deformation during the polishing process. During chemical etching material must be removed uniformly and without pitting. The finished specimen may have a "matte" appearance, but surfaces must be flat and parallel [18].

Each prepared sample was secured in the mounting ring of the pole figure device using ordinary, clear pressure-sensitive tape. Note that the surface of the sample was not covered with tape. The tape was placed around the perimeter of the sample so that sufficient area is left uncovered for testing. Care must be taken so that the sample surface is in the same plane as the mounting ring of the pole figure device. Because of rotations and the lack of a reference "rolling direction" in rod-stock copper, it is important to initially identify a reference direction in the rod so that subsequent experimental figures are the same relative to each other.

GENERAL ANALYSIS OF POLE FIGURES

A pole figure is a stereographic projection, with a specified orientation relative to the specimen, that shows the variation of pole density with pole orientation for a selected set of crystal planes [27]. The basic method to display preferred orientation is to represent the orientation of a particular crystal direction [*uvw*] or the normal to a lattice plane (*hkl*) in a specimen coordinate system. The crystal direction is first projected onto a sphere of unit radius surrounding the crystal and the point on the sphere is defined by two angles, a co-latitude α, measured from the pole N, and an azimuth β, measured from an equatorial point E counterclockwise, Figure 2 [26]. This sphere is usually represented by its projection onto a plane. The pole figure shows where and to what extent (quantitatively) various orientations of crystals occur with respect to fixed directions within the specimen, for example, the rolling, transverse, and normal directions. Generally, a pole figure is derived from a spherical projection and lines are extended parallel to various crystallographic directions, or perpendicular to various crystallographic planes, in the crystal, from the center until they meet the surface of the sphere. These intersections of the various directions with the sphere are called poles and the reference points P and E in Figure 2 are usually selected as principal crystallographic axes. Every other pole can then be located by the two angles α and β as described above.

In order to transfer this three-dimensional sphere onto a two-dimensional sheet of paper, the sphere is projected onto a plane. Common ways of projecting spheres onto planes include cylindrical, conical, polyconic and circular projections [27]. There are two kinds of circular projections: stereographic and equal area. A stereographic net is constructed on a plane that slices through the center of the sphere perpendicular to the north-pole and south-pole axis.

Because of variations in intensities due to the way a crystal may be facing, its orientation can be found by measuring the intensities at different angles by x-ray diffraction and plotting their angular coordinates onto a net. Then, by laying a standard projection on top of the net, which has all the intensity points plotted onto it, one can tell which of the poles are appearing and relate that to the sample orientation [27]. To further simplify the pole map, equal intensity contours are drawn for various intensities from the lowest to the highest, so that the end result is something like a topographical map. The map is circular in outline corresponding to the stereographic plane within the sphere. The locations of peaks or mountains then correspond to the locations of certain crystallographic poles on the plane. The higher the peak, the greater number of such planes in the crystal. But, a crystal has only a specified number of each sort of plane. However, if the center sample is a polycrystal, an intensity peak indicates preferential alignment of certain planes with respect to poles N and E. Now, though, these reference points must relate to specimen orientation.

There are two methods generally used to make measurements for pole figures; one is transmission and the other is reflection. The central region of the pole figure is inaccessible to any transmission method and can be explored only by a reflection technique [27]. The most popular technique is the Schulz reflection method [27] because of the ease of sample preparation and accessibility to the central region of the pole figure. In this report, all pole figures and their analyses are based on the Schulz reflection method. The specimen must be of effectively infinite thickness (relative to the depth of penetration of the x-ray beam), or some rays will actually transmit through the sample, requiring extra data correction.

RESULTS AND DISCUSSION

The ODF is a theoretical model created from experimental pole figure data files. It is calculated in discreet steps (in this case, 10°) of $\Delta\varphi_1$, $\Delta\Phi$, $\Delta\varphi_2$, in each respective coordinate identified in Figure 3 [26]. These points are computer-generated in two-dimensional sections, then connected and smoothed by interpolation. Ideally, the texture will reach a relatively high value, as shown by the g_0 position, at certain points (preferred orientations), gradually attaining a maximum (ω area). To visualize the three-dimensional distribution function created, these lines can be imagined to be drawn on glass sheets which can be stacked on top of each other, producing forms like that shown in Figure 4 [26]. Often, the specific ODF code will be designed to approximate parameters to some extent to reduce the incredibly large number of coefficients for the spherical harmonics (the mathematical form of these models). This is not detrimental to the analysis and is usually necessary in recrystallization and plasticity calculations [28].

Such is the case in the popLA code. For this analysis, experimental pole figures are run and the data is stored as ASCII files. This data must be converted to popLA-compatible format by means of a conversion program provided by the manufacturer of the pole figure equipment (in this case, Rigaku). Next, the popLA code generates some initial ODF from which it calculates a theoretical pole figure. Comparison of calculated and experimental pole figure data enables the code to revise the ODF and repeat this calculation. When a satisfactory comparison is obtained between calculated values and experimental data, the theoretical ODF is presumed to describe the specimen material. This ODF is then processed by the LApp code to generate an isotropic yield surface for the material. Depending on the particular desired form of output, several options are available to process and analyze the data. A complete account of the software is available in 'Operational Texture Analysis' [21] and the literature accompanying the popLA and LApp programs provided by Los Alamos.

The intention of this project was to analyze data obtained from the Rigaku x-ray unit here at the University of Kentucky. However, in the complex and time-consuming process of upgrading the equipment, the data-file conversion program has not yet been furnished by Rigaku. Therefore, an example ODF and corresponding yield surface have been generated using a limited data file obtained earlier on x-ray equipment at Eglin AFB.

The file consists of (111) pole figure data for a wafer taken from an ROR impactor, close to the impact face. Figure 5 shows the theoretical (111) pole figure, produced by popLA from the experimental data. Figure 6, produced by popLA from a second experimental data file, shows the theoretical (111) pole figure for the rod before impact. This shows the fiber texture of the undeformed rod. Comparison of the two textures shown in Figures 5 and 6 is of particular interest in damage analysis.

The ODF corresponding to the pole figure shown in Figure 5 for the deformed ROR specimen was processed using the LApp code. This produced the yield surface shown in Figure 7.

CONCLUSIONS

By establishing an appropriate methodology, more accurate comparisons between experimental and theoretical data can be made, allowing refinement of the existing anisotropic plasticity codes. A better theoretical model will allow a more accurate prediction of a fully three-dimensional yield surface, and corresponding plastic flow rule, for polycrystalline metals. Utilizing additional information, such as hardening data, significantly improves the accuracy of the theoretical model and prediction of subsequent flow surfaces is feasible.

It is also beneficial to develop a database containing produced theoretical (and the accompanying experimental) pole figures, ODFs, and yield loci for a variety of materials, deformed to different strain levels, for a fuller understanding (and continual improvement) of the applications of these codes.

ACKNOWLEDGMENTS

The author would like to thank Mr. Joel House and Mr. Thad Wallace at Eglin AFB, Florida and Prof. Peter Gillis, Mr. Larry Rice and Mr. Kenny Boggs at the University of Kentucky for their help and contributions to the continued success of this project.

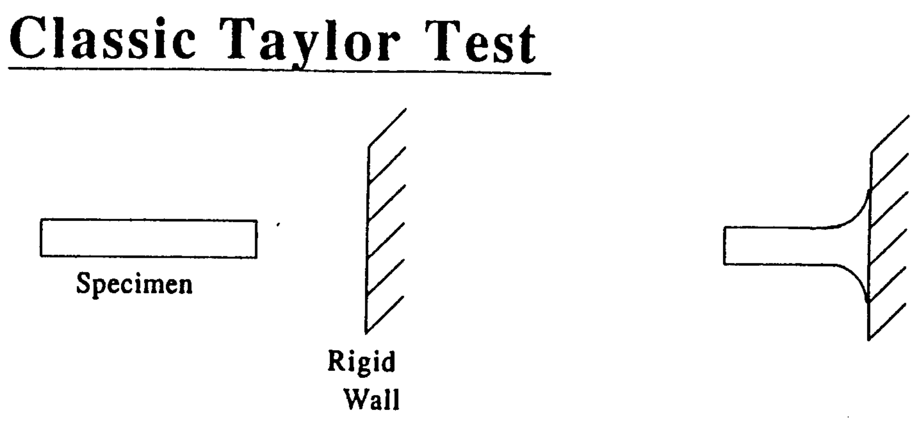


FIGURE 1a: Schematic Diagram of the Classic Taylor Test. ref [25]

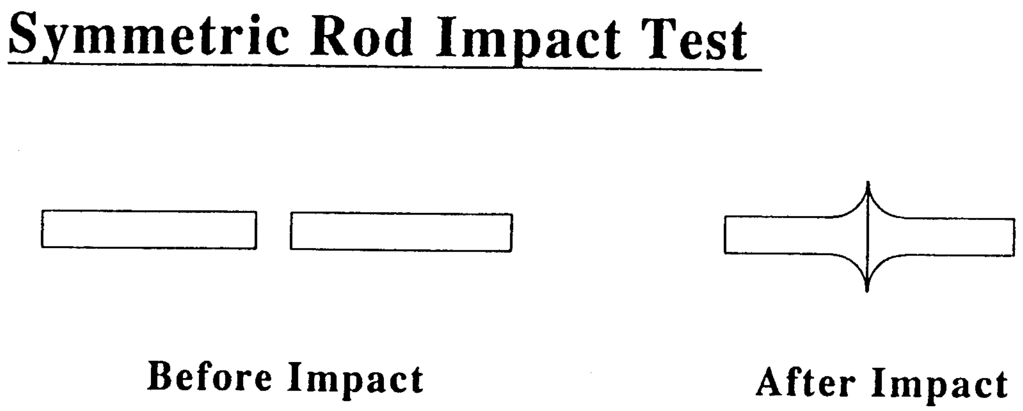


FIGURE 1b: Schematic Diagram of the Symmetric Rod (Rod-on-Rod) Impact Test. ref [25]

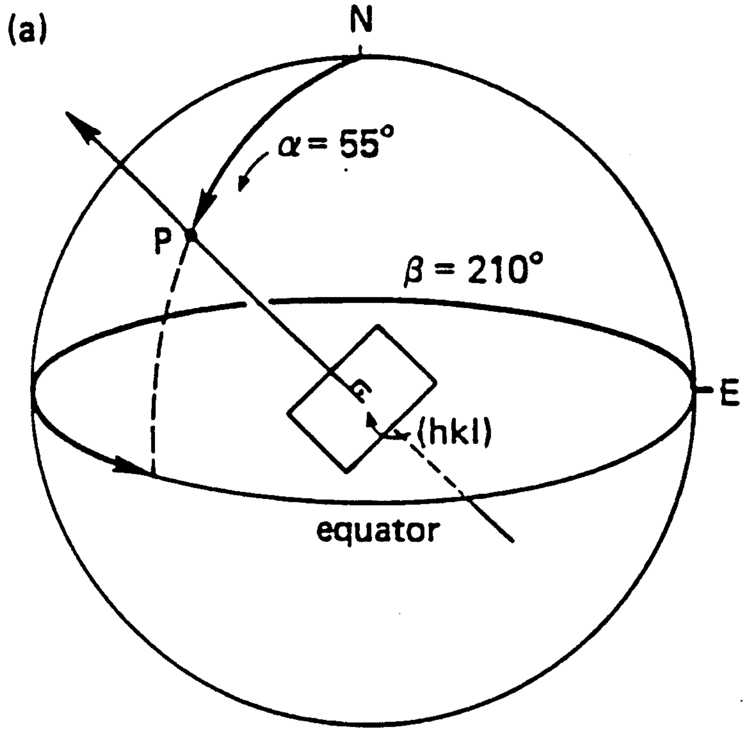


FIGURE 2: Spherical Coordinates for a Point P. ref [26]

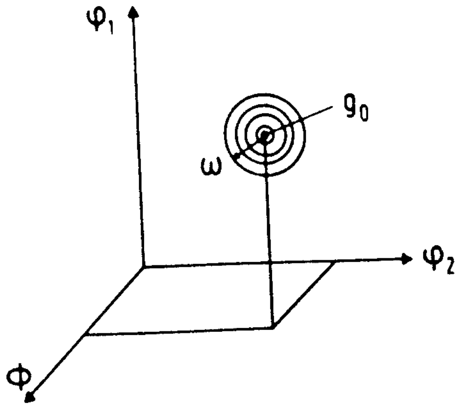


FIGURE 3: Ideal Orientation g_0 with a Certain Spread About It (ω). ref [26]

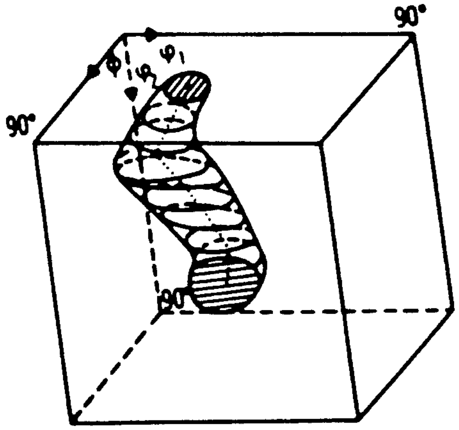


FIGURE 4: Equilevel Surface for Copper- Cold Rolled Texture. ref [26]

jhcu7-8 4-23-91 35 43.36 47

pole figures min = 429.; max = 5514.; last median = 999

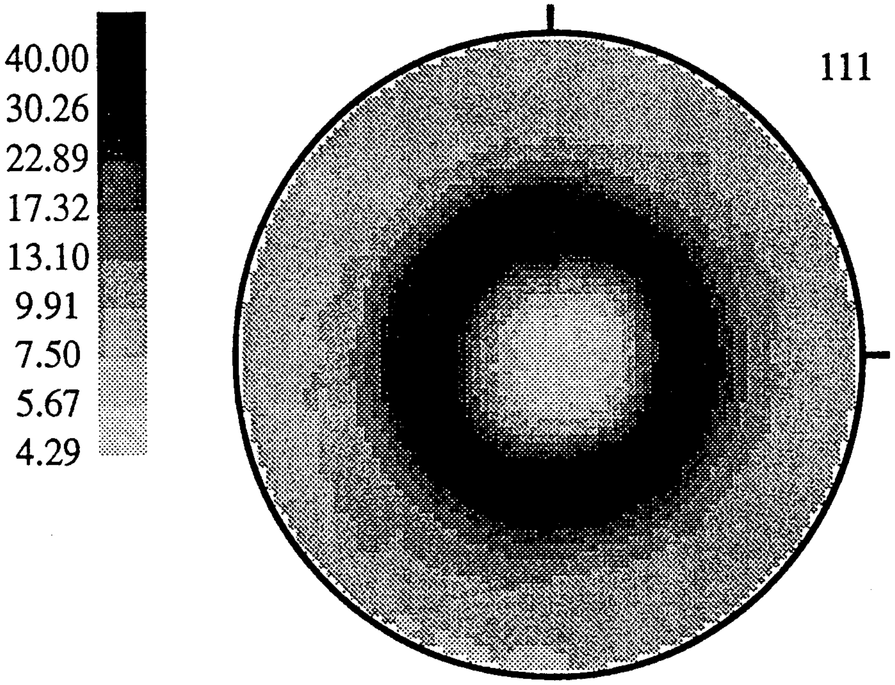


FIGURE 5: (111) Pole Figure Produced by Theoretical Calculatic s for a Deformed Copper Rod

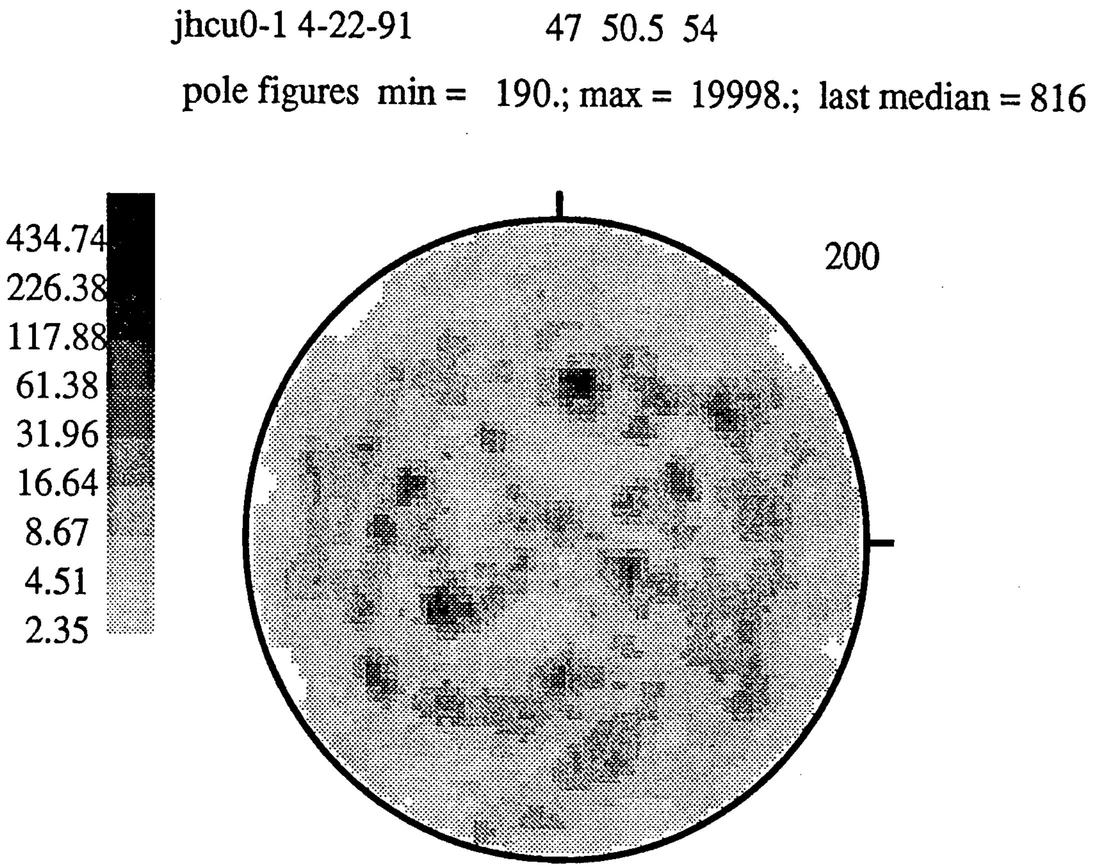


FIGURE 6: (111) Pole Figure Produced by Theoretical Calculations for an Undeformed Copper Rod

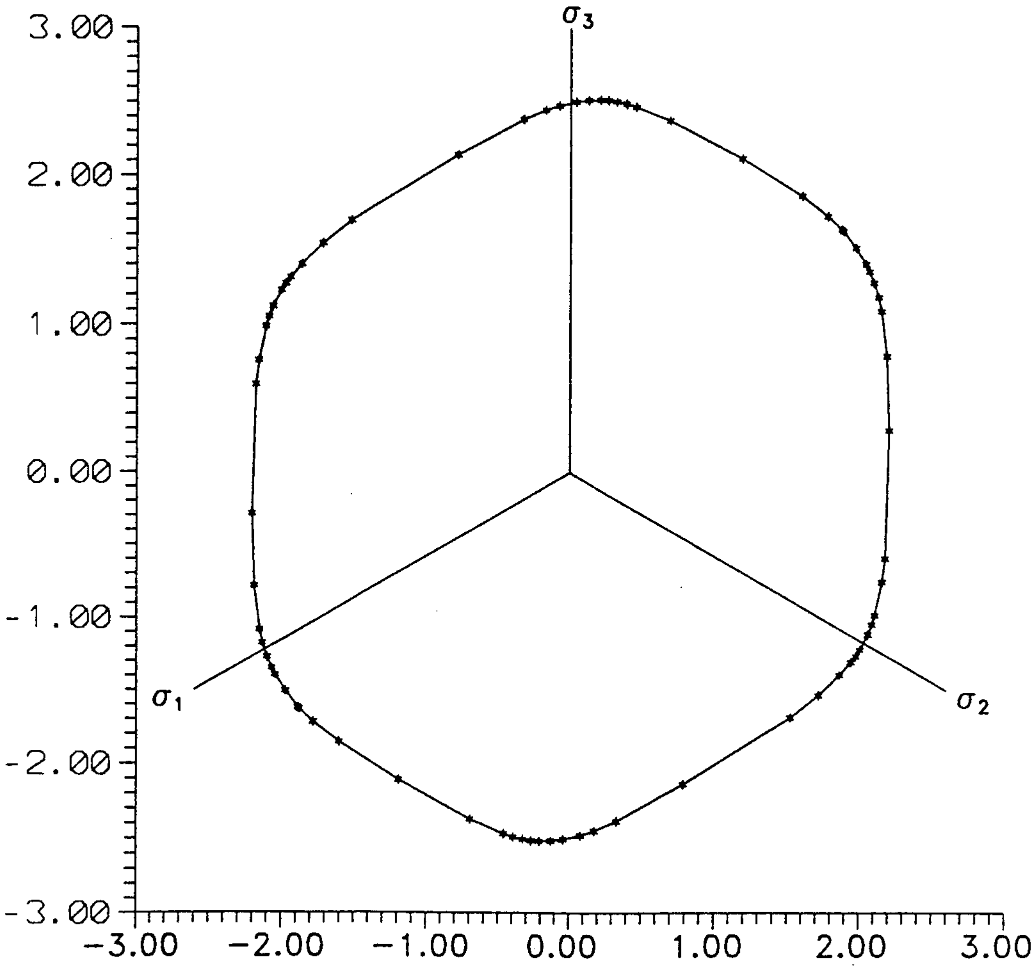


FIGURE 7: Theoretical Yield Surface for Deformed Coppe

REFERENCES

1. G. E. Dieter, Mechanical Metallurgy, McGraw-Hill, 3rd edition, 1986, p.322.

2. B. Paul, "Prediction of Elastic Constants of Multiphase Materials". AIME Trans. 218 36 (1960).

3. L. B. Greszczuk, "Elastic Constants and Analysis Methods for Filament Wound Shell Structures," Douglas Aircraft Missile & Space Systems Division Report SM-45849, 1964.

4. S. G. Lekhnitskii, Theory of Elasticity of an Anisotropic Elastic Body, Holden-Day, 1963.

5. J. E. Ashton, J. C. Halpin and P. H. Petit, Primer on Composite Materials, Technomic, 1969.

6. R. Hill, "A Theory of the Yielding and Flow of Anisotropic Plastic Metals", Proc. Roy. Soc. London A193 281 (1948).

7. R. Hill, "Theoretical Plasticity of Textured Aggregates", Proc. Camb. Phil. Soc. 85 179 (1979).

8. R. Hill, "Constitutive Modeling of Orthotropic Plasticity in Sheet Metals", J. Mech. Phys. Solids 38 405 (1990).

9. W. F. Hosford, "On Yield Loci of Anisotropic Cubic Metals", Proc. 7th North American Metalworking Conference, SME, Dearborn MI, 1980, p. 191.

10. W. F. Hosford, "Comments on Anisotropic Yield Criteria", Int. J. Mech. Sci. 27 423 (1985).

11. W. F. Hosford and R. M. Caddell, Metal Forming: Mechanics and Metallurgy, Prentice Hall, 1983.

12. A. Grof and W. F. Hosford, "The Effect of R-Value on Calculated Forming Limit Curves", Forming Limit Diagrams: Concepts, Methods and Applications, R. F. Wagoner, K. S. Chan and S. P. Keeler, Eds., TMS-AIME, (1989) p. 153.

13. G. I. Taylor, "Plastic Strain in Metals", J. Inst. Met. 62 307 (1938).

14. J. F. W. Bishop and R. Hill, "A Theory of the Plastic Distortion of a Polycrystalline Aggregate Under Combined Stresses" and "A Theoretical Derivation of the Plastic Properties of a Polycrystalline Face-Centered Metal", Phil. Mag. 42 414 and 1298 (1951).

15. G. Y. Chin and W. L. Mammel, "Generalization and Equivalence of the Minimum Work (Taylor) and Maximum Work (Bishop-Hill) Principles for Crystal Plasticity", Trans. Metall. Soc. AIME 245 1211 (1969).

16. A. D. Rollett, M. G. Stout and U. F. Kocks, "Polycrystal Plasticity as Applied to the Problem of In-Plane Anisotropy in Rolled Cubic Metals", Advances in Plasticity '89, A. F. Khan, Ed., Pergamon, 1989, p. 69.

17. ASTM Standard E8-89b

18. ASTM Standard E81-89

19. J. L. Mayes, "Material Characterization and Evaluation for Penetration Applications: Physical Properties", Report to AFOSR RDL Summer Research Program Office dated 16 August 1991.

20. S. L. Hatfield, "Material Characterization and Evaluation for Penetration Applications: Mechanical Properties", Report to AFOSR RDL Summer Research Program Office dated 16 August 1991.

21. J. S. Kallend, U. F. Kocks, A. D. Rollett and H. R. Wenk, "Operational Texture Analysis", Matls. Sci. Eng. A132 (1991).

22. A. D. Rollett, private communications.

23. G. I. Taylor, "The Use of Flat-Ended Projectiles for Determining Dynamic Yield Stress", Proc. Roy. Soc. (London), 194 289 (1948).

24. D. Erlich, D. A. Shockey and L. Seaman, "Symmetric Rod Impact Technique for Dynamic Yield Determination," AIP Conference Proceedings, No. 78, Second Topical Conference on Shock Waves in Condensed Matter, Menlo Park, CA, 1981, pp. 402-406.

25. J. L. Mayes, S. L. Hatfield, P. P. Gillis and J. W. House, "Void Formation in OFE Copper", accepted for publication on the Journal of Impact Engineering.

26. H. R. Wenk, Preferred Orientation in Deformed Metals and Rocks: An Introduction to Modern Texture Analysis, Harcourt, Brace and Jovanovich (1985).

27. B. D. Cullity, Elements of X-Ray Diffraction, Addison-Wesley, 2nd ed., (1978).

28. U. F. Kocks, J. S. Kallend and A. C. Biondo, "Accurate Representation of General Textures by a Set of Weighted Grains", to be published in the proceedings of the Ninth International Conference on Textures of Materials.

29. S. L. Hatfield, "Anisotropic Plastic Deformation of Metals", Report to AFOSR RDL Research Initiative Program (RIP no.12) dated December 1992.

TOLERANCE REASONING FOR A GENERATIVE PROCESS PLANNER

Joseph H. Nurre
Assistant Professor
Department of Electrical and Computer Engineering

Ohio University
College of Engineering and Technology
Athens, Ohio 45701

Final Report for:
Research Initiation Program
Wright Laboratory

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

and

Ohio University

December 1992

TOLERANCE REASONING FOR A GENERATIVE PROCESS PLANNER

Joseph H. Nurre
Assistant Professor
Department of Electrical and Computer Engineering
Ohio University

Abstract

The current ANSI Y 14.5 design standard is intended to ensure that geometrical dimension and tolerance requirements relating to part assembly and functionality, are specifically stated and thus carried out. This shifts the burden of understanding how to manufacture the geometry to the process planner. This report will discuss an algorithmic method of achieving designer specified tolerance in a generated manufacturing process plan. The method investigated and the software developed could be incorporated into a fully integrated design and manufacturing software product, such as the Rapid Design System.[1]

The primary issues addressed in this report regarding process plan tolerances are interpretation of design specifications and machine tolerance stackup. Interpretation requires an understanding of the syntax of the ANSI Y 14.5 standard, and its relationship to the manufacturing environment. Design tolerance is achieved when the tools, machines and fixtures are analyzed for tolerance build up in the manufactured part.

TOLERANCE REASONING FOR A GENERATIVE PROCESS PLANNER

Joseph H. Nurre

INTRODUCTION

The manufacturing process plan details the various production operations needed to create a part. The geometry of a part is specified by the designer using the ANSI Y 14.5 design standard. Successful manufacturers are able to create economical process plans, which fulfill the designer's intent. Research in the area of automated process planners has the potential to shorten product development times and improve product quality for U.S. manufacturers.

Approaches to automated process planners are usually separated into group technology methods and generative technology methods.[2] Group technology process planners handle geometric tolerances by grouping them into similar classes. Generative process planners handle manufacturing tolerances with imbedded rules and algorithms. This report will discuss an algorithmic method of achieving designer specified tolerance in a generated manufacturing process plan. To achieve this goal, two issues are addressed: interpretation of design specifications and machine tolerance stackup.

Interpretation of design deals with the conversion of ANSI Y 14.5 tolerance frame information to linear axis tolerances. Maximum and minimum tolerances are based on the function of a part. They are specified by the designer to satisfy constraints on interference and fit. The tolerances are, therefore, not usually related to orthogonal axes. The designer understands the need to allow for part variation in his design. Freedom in the geometric specification of a part ultimately leads to a reduction in the manufacturing costs.[3] Although, the designer may not be concerned with linear axes for his geometry, the machinist is. ANSI Y 14.5 tolerances must be converted to machine axes tolerances to allow for proper manufacturing.

Conversion of a tolerance from a functional datum to a manufacturing datum has been described by several researcher.[4][5] The ANSI Y 14.5 standard defines the following tolerance specifications[6]: Form, Orientation, Profile, Location and Runout. The Form tolerance is usually applied to a single feature or a portion of a feature. Form defines intrinsic properties of the features such as straightness, flatness, circularity, etc. Orientation tolerances control the relationship of surfaces to one another. One example would be to specify that two surfaces must be perpendicular with respect to each other. Profile tolerancing is a method of specifying deviations from a desired profile along the surface of a feature. Profile tolerances may refer to, for instance, the orthographic projection of a part's profile. Runout is a composite of both Form and Orientation tolerance. Runout controls permissible errors in a part surface during a complete revolution around a datum axis. The Runout tolerance may be specified as either total or circular. The two tolerances of the Location are Position and Concentricity. Both Position and Concentricity tolerances provide permissible variation in the specified location of a feature or a group of features. The relationship is typically specified with respect to a datum reference.

In regards to the manufacturing process planner, tolerance of Form, Orientation, Profile and Runout are functions of machine operation, tooling and sequencing. For instance, the roundness of a hole is controlled by the drill bit, and the speed and feed of the machine. To achieve a tighter roundness on the shape of a hole, one may need to follow the drilling operation with a boring operation. On the other hand, Location tolerances, which describe the relationships among features, require a more sophisticated analysis.

Location tolerance zones are specified using a control frame. The frame has information on geometric characteristics, diameter, tolerance and datum reference. Location zones are typically circular and positioned by nominal axes. The center of a feature or a group of features must fall somewhere within the specified zone.

The Interpretation software developed under this grant converts a circular tolerance zone, specified with a control frame, into a linear tolerance zone, for a single hole or a pattern of holes. The pattern of holes may be defined with respect to each other or an imaginary center. The software can be extended to address features other than holes. Calculations required to interpret the tolerance specification on a hole are typical of other features. The interpretation takes into account hole size, tolerance and datum reference.

As stated earlier, the manufacturing process plan details the various production operations needed to control Form, Orientation and Runout. A Tolerance Chart can be used to assess the Location tolerance for a tentative process plan. Machining operations, called for by the process plan, can produce geometric shapes to a quantified precision. The Tolerance Chart lists each operation in the plan, clearly indicating the locating and machined surfaces, as provided by the interpretation software. The Chart becomes a vehicle for calculating the accumulated error of multiple machining operations needed to manufacture the part. If the resultant process tolerances from the machining operations conform to the designer specified values, then the process plan is feasible. The accuracy of each machining operation, therefore, is controlled using the Tolerance Chart technique.

It was Wade[7] who first recorded and formalized the adhoc tolerance charting technique being used in industry. The technique is logical and deterministic, lending itself to computer application. General Motors was one of the first to develop a computer program that implemented the tolerance charting technique[8]. A graphical Tolerance Chart program was later presented by Ahluwalia[9] in 1986. Lehtlhet and Joshi[10] describe a method for using the Tolerance Chart to continuously monitor a manufacturing process. Xiaoqing and Davies[11] describe a tolerance program which is part of an interactive computer graphic system. Irani et al[12] presented a directed graph methodology described later in this report.

In the ideal case, datum surfaces used by the designer to position geometric features are based on the locating surfaces used by the machinist. This alleviates the problem of tolerance stackup and diminishes the need for a Tolerance Chart analysis. There are, however, several deviations from the ideal case. Datums that are assigned functionally, may not be conveniently located for the machinist. A functional datum may be a manufactured feature created later in the process plan, necessitating the need for a change in datums. Finally, features which must be symmetrical to each other can be handled within process measurements, special tooling or a tolerance analysis, before production.

The mathematical theory behind the Tolerance Chart is simple. The addition or subtraction of length dimensions is always characterized by the adding of tolerances for the individual lengths. Tolerance stackup problems are easily recognized by calculating the resultant mean and tolerances of multiple operations, and comparing them with the designer's specifications. Furthermore, the Tolerance Chart can be used to increase the maximum permissible tolerances of machining cuts, potentially reducing manufacturing costs.

Another function of the Tolerance Chart is to determine if the stock removed during a particular machining operation is appropriate. Tolerances assigned to machining cuts that are not constrained by the design specifications, may impact stock removal. Mean stock removal depends on the process capability. The worst case tolerance buildup should be taken into account before assigning the mean stock removal for a particular operation. At this time, mean stock removal is assigned by the process planner. An automated method for assigning mean stock removal is discussed later in this report.

To summarize, achieving a designer specified Location tolerance is accomplished in a generative process planner, by interpretation and tolerance charting. The system described in this report ensures that the maximum possible tolerance is allocated to each machining cut in the process and the stock removals for creating a part are practical. Any improper dimensions or tolerances that may exist in the product design are eliminated. The remainder of the this report is broken into a Software Development Section, Results Section and a Discussion Section. Each section will discuss the Interpretation software, followed by the Tolerance Chart Software developed.

SOFTWARE DEVELOPMENT

In order to describe the Interpretation software, it is necessary to review part of the ANSI Y 14.5 standard. A complete description of the standard can be found in several trade texts, including Punchochar.[6] The feature control frame provides position and tolerance information about a single feature or a group of features on a part. Figure 1, shows a control frame and its resulting tolerance for a hole. The symbol, ⊕, specifies a position tolerance. The shape of the position tolerance zone is circular, as specified by the ∅ symbol. The nominal center of the hole is measured from

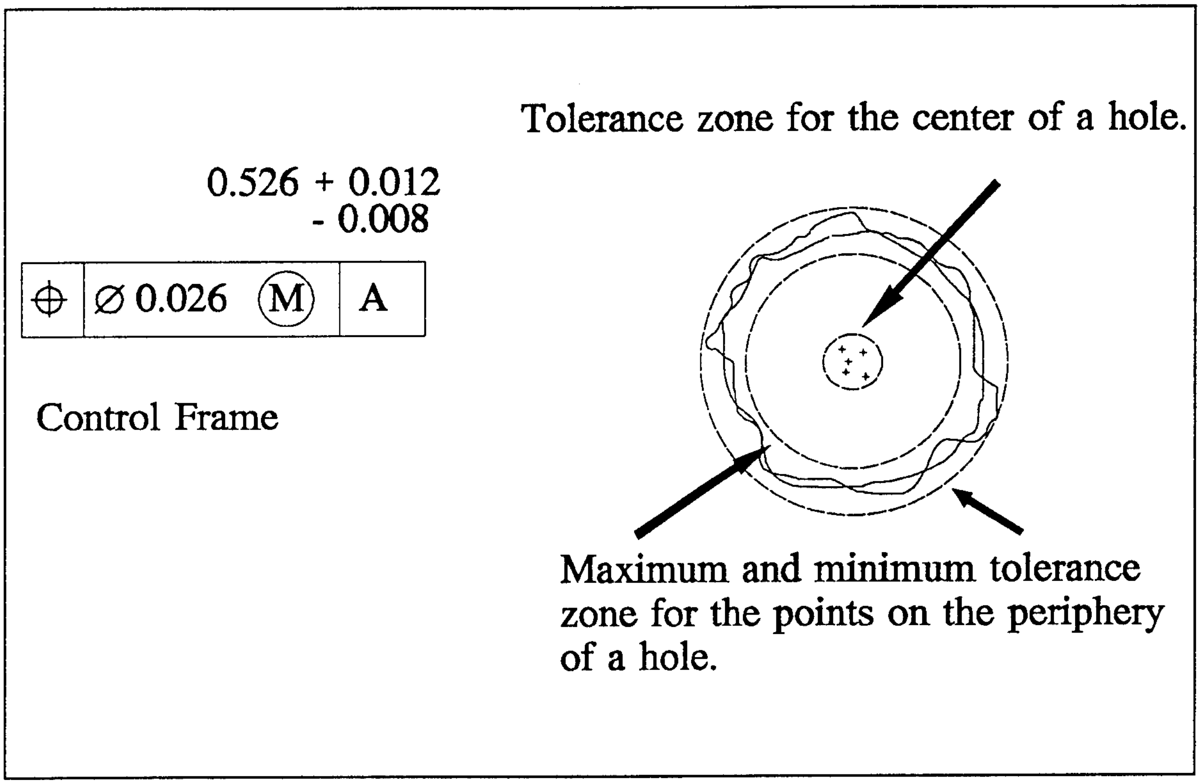

Tolerance zone for the center of a hole.
0.526 + 0.012
- 0.008
⊕ | ∅ 0.026 Ⓜ | A
Control Frame
Maximum and minimum tolerance zone for the points on the periphery of a hole.

Figure 1.

Datum 'A'. The hole and Datum 'A' would be shown on a schematic. The hole size is specified with a nominal value of 0.526, a positive variance of 0.012 and a negative variance of 0.008.

The symbol, Ⓜ, appearing in the control frame is a modifier and plays an important role in interpreting the tolerance zone. Three modifiers are available:

Ⓜ - Maximum Material Condition

Ⓛ - Least Material Condition

Ⓡ - Regardless of the Feature Size

Because a hole is created by removing material, the Maximum Material Condition, as specified in the figure, exists when the hole is at a diameter of 0.518. The control frame then specifies that a hole of diameter, 0.518, must be centered on a circle of radius 0.026.

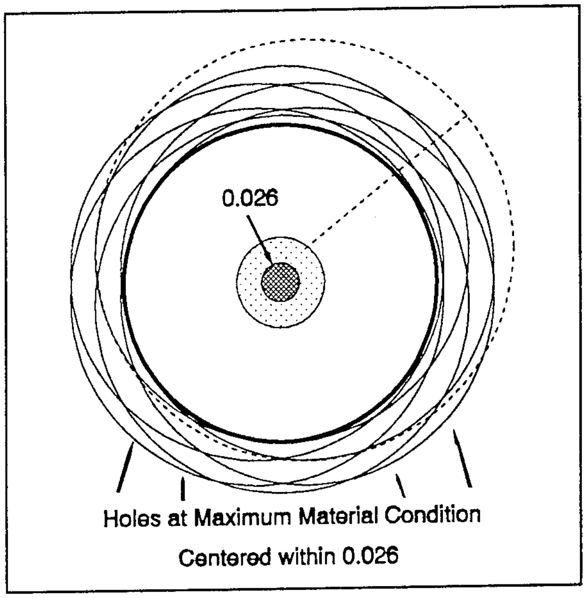


Figure 2.

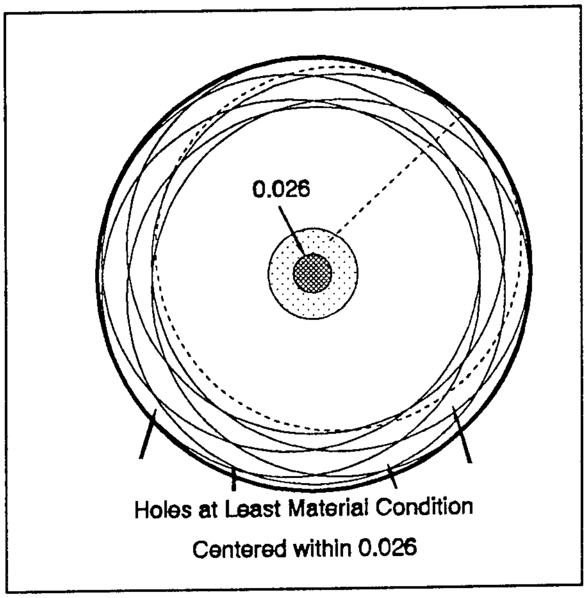


Figure 3.

The intent of the modifier given in Figure 1 is to ensure a minimum amount of material is removed as shown by the thick circle in Figure 2. If the size of the hole is larger than 0.518, (shown in Figure 2 with a dotted line) then its center can fall outside of 0.026 tolerance zone without violating the minimum amount of material to be removed. Therefore, interpretation c the control frame is dependent on the modifier and the size of the hole. In the case of Least Material Condition, the hole size is maximum, leaving the least amount of material on the part. Figure 3 demonstrates that the intent of this modifier is to ensure the hole falls within a bounded region shown, once again, by a thick circle. As the size of the hole decreases from 0.538, (shown in Figure 3 with a dotted line) its center position tolerance zone may increase. The third modifier, Regardless of the Feature Size, specifies the position tolerance zone should remain the same regardless of the hole's size.

The input to the Interpretation software is the tolerance control frame. A maximum and minimum size for the hole is specified, as well as, a tolerance zone and modifier. The size of the tolerance zone depends on the size of the hole, as explained above. The hole size that gives the tightest tolerance on its center position is always chosen. The circular tolerance zone is then converted to a rectilinear tolerance zone by simply enclosing a square region in the circular zone, as shown in Figure 4. The program outputs tolerance

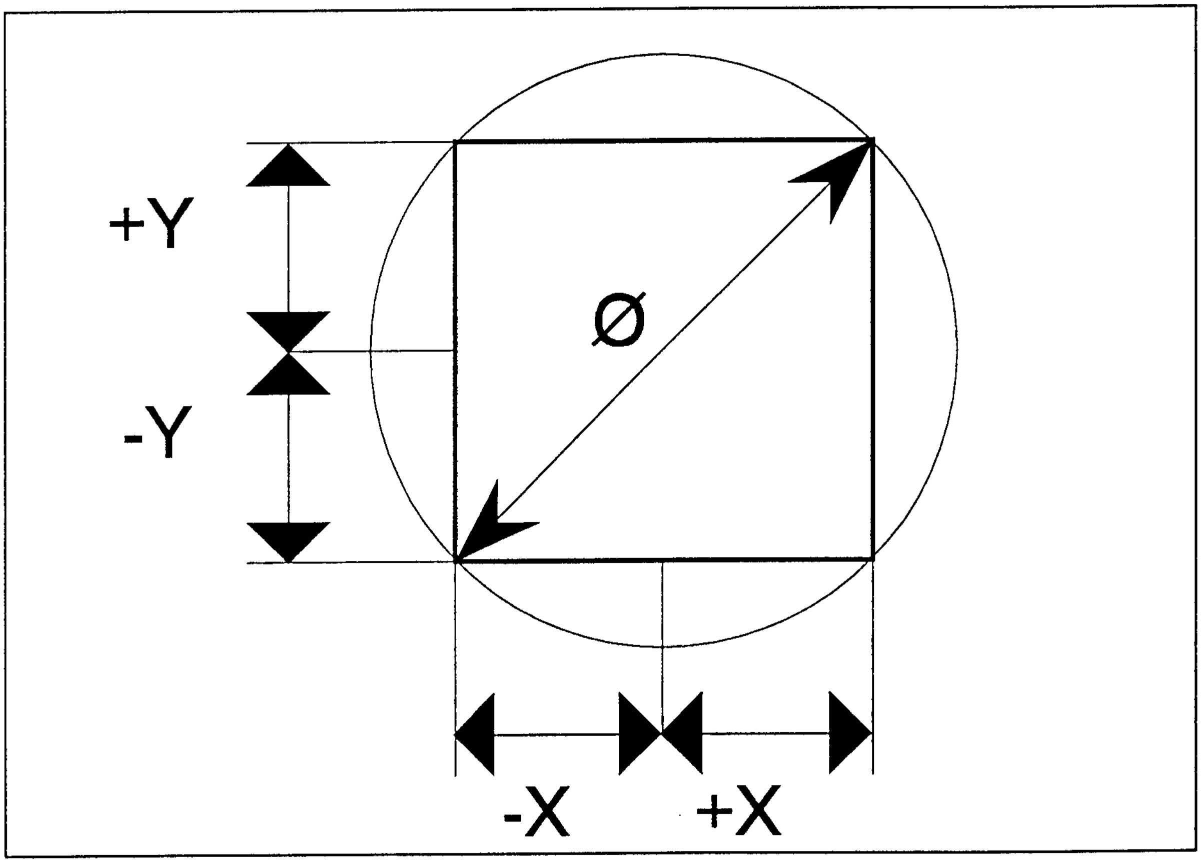


Figure 4.

values for the X and Y directions.

If a change in datum is necessary to simplify manufacturing, a feature may be dimensioned with respect to a datum's datum. In other words, to change datums, the feature's original datum must be positioned by a primary datum. The primary datum can then be used by the feature if the feature's tolerance zone is subtracted from the original datum's tolerance zone. A change in datums results in a tighter tolerance.

In the case of symmetrically placed features, an imaginary reference point is needed as a datum. This reference point would then have a nominal position and tolerance. Figure 5 shows an example of a widely used hole pattern, where an imaginary center point is used as a datum. The tolerance zone of each individual hole, with respect to the edge of the part, is calculated by a change in datums, as explained above. The interpretation program has been developed to handle feature patterns positioned with polar coordinates.

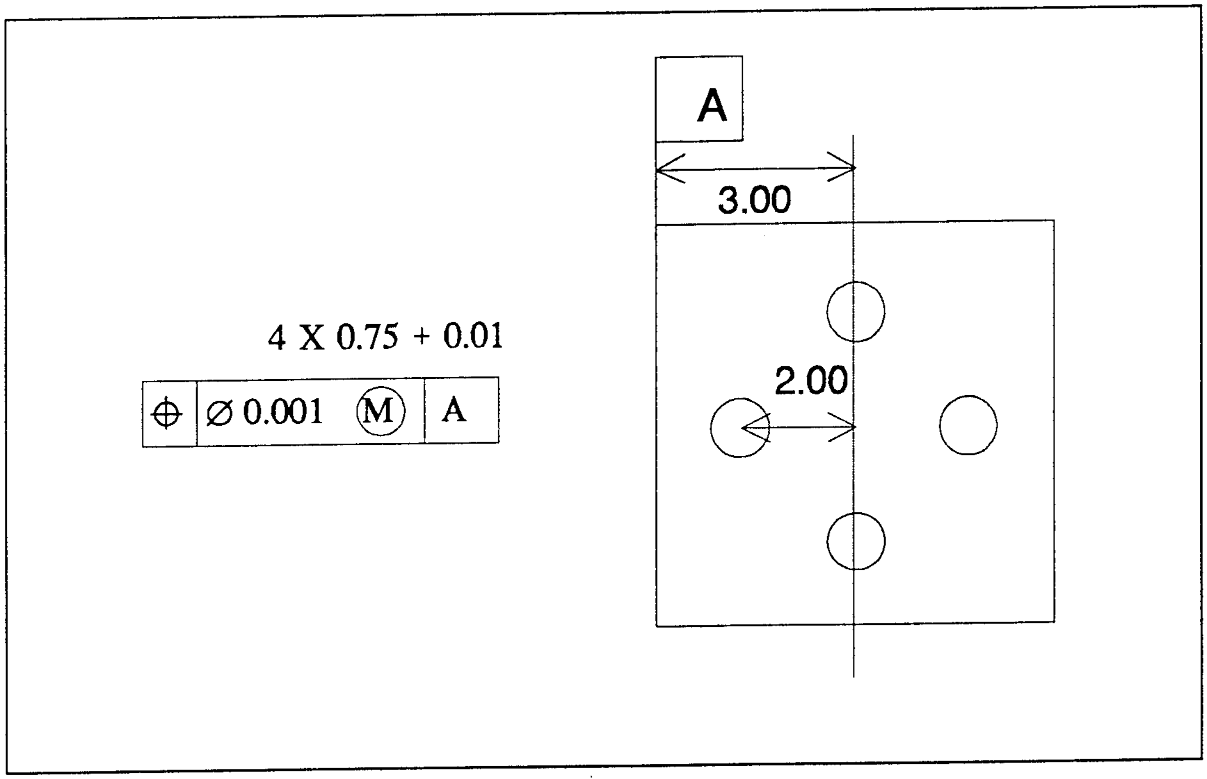


Figure 5.

With the ANSI control frames converted to linear axes tolerances, the manufacturing process plan can be represented as a directed graph. The arcs of the directed graph correspond to machining cuts. All datum and machining surfaces are ordered and numbered from left to right. Each machining cut is described in terms of two numbers representing a datum surface and a machine surface. The node or dot end of the arc corresponds to the locating datums. The arrow head corresponds to the resulting machined surface. An arc number indicates the sequence of the machining cut in the process plan in ascending order. A Search algorithm has been developed which identifies machining cuts that contribute towards the designer specified tolerance. This in turn determines the order in which the machining cuts appear in the Tolerance Chart. The algorithm is an improvement on the method presented by Irani et al[12].

Nodes on the directed graph represent surfaces. Let *p1* and *p2* represent two surfaces that have a position tolerance with respect to each other. The search algorithm transverses the arc with the largest arc number (sequence number) which has a node at either *p1* or *p2*. *P1* or *p2* is then updated with the second node value of the arc. When *p1* equals *p2*, the process is terminated and all the arcs traveled contribute to final tolerance between the specified surfaces.

Referring to Figure 6 as an example, consider the case of determining the tolerance between nodes 'c' and 'e'. Traversal of arc 9, the largest arc number, results in *p2* being updated from 'e' to 'a'. Traversal of arcs 7 and 6 will result in *p1* being updated from 'c' to 'b' and then from 'b' to 'a'. Since *p1* and *p2* are then equal, the process terminates. Applying the Search Algorithm, results in machining cuts 9, 7 and 6 contributing to the final tolerance between the surface represented by nodes 'c' and 'e'.

The Search algorithm is applied repeatedly to determine the machine cuts responsible for all designer specified tolerances. The machining cut sequence is arranged as a matrix of order *m x n* where *m* is the number of machining cuts contributing to the tolerance. Complete information pertaining to a particular machining cut is stored by row, where *n* corresponds to a column in the

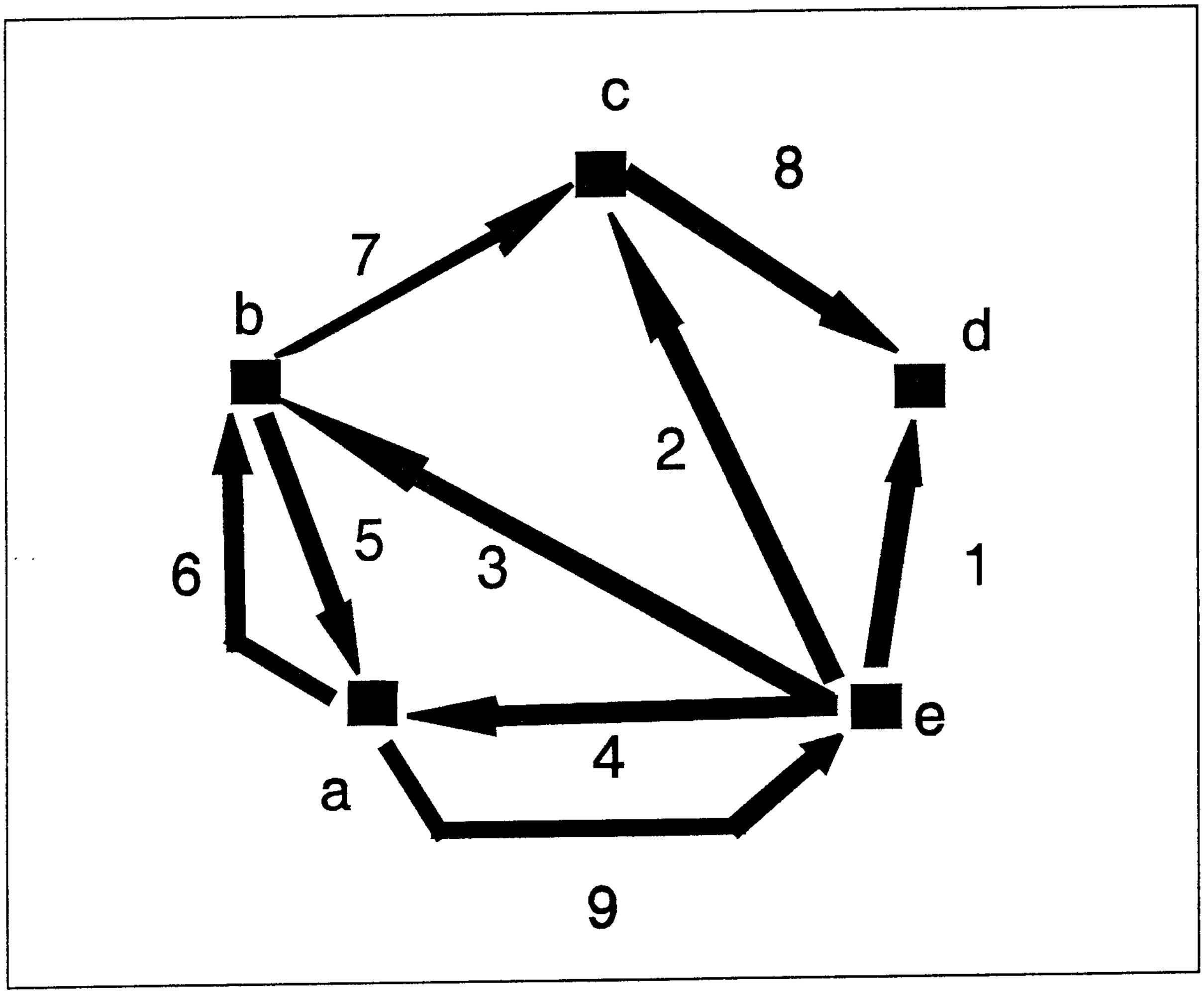
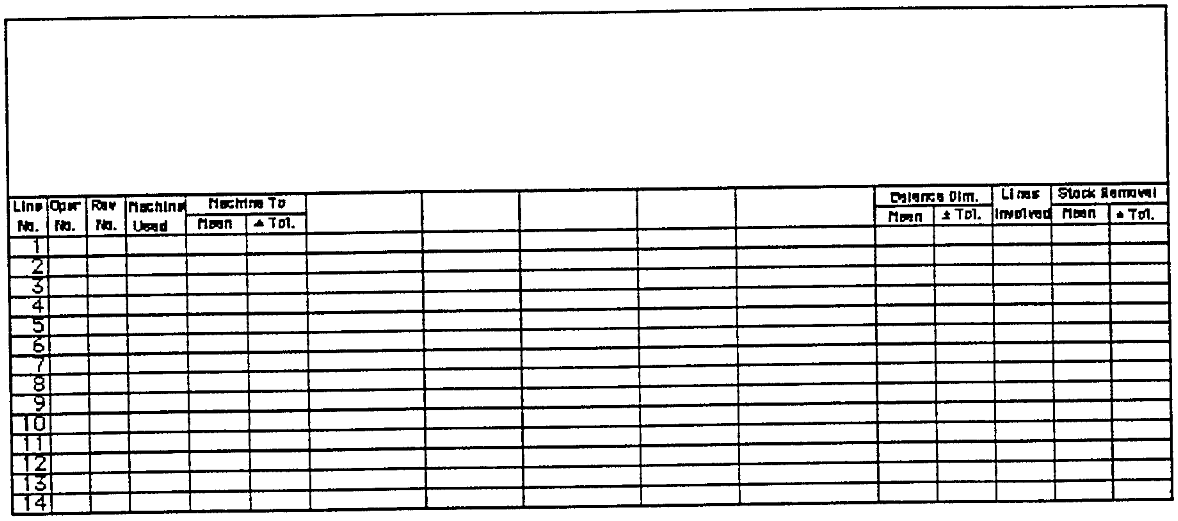


Figure 6.

Tolerance Chart. Figure 7 shows a Tolerance Chart with its different column titles. The first column contains the line numbers for the chart itself. The second column contains all the operation numbers for a particular part. The third column is used to track revisions in the process plan. Column Four contains the names of all the machines used for a corresponding operation. The fifth and sixth columns contain the mean dimensions and tolerances of the machining cuts. The seventh column which contains a graphic representation of the machining cuts and balance dimensions, is not used by the developed software. The eighth and ninth columns are used for the resulting mean and tolerance of a balance dimension. A balance dimensions is added after each



Line	Oper	Rev	Machine	Machine To							Balance Dim.		Lines	Stock Removal	
No.	No.	No.	Used	Mean	± Tol.						Mean	± Tol.	Involved	Mean	± Tol.
1															
2															
3															
4															
5															
6															
7															
8															
9															
10															
11															
12															
13															
14															

Figure 7.

m x n array is generated. The tenth column is the lines involved column, depicting the relationship between the different lines of the Tolerance Chart, as determined by the Search Algorithm. The eleventh and twelfth columns contain the stock removal mean and tolerances respectively. The combination of all machining cut sequences and balance dimensions, as determined by the Search Algorithm, results in an ordered Tolerance Chart. Completing the columns of such a chart is a well structured, deterministic operation, described by Wade.[7]

An additional feature being added to the software is its ability to distribute tolerances among machining cuts. The criteria for distribution of tolerances is based on typical tolerance ranges of machining operations presented by Trucks[13]. Currently, the software allocates the designer specified tolerance equally among the machining cuts. However, tolerances are automatically loosened on machining cuts that are unduly restrictive for achieving designer intent.

RESULTS

As stated in the Introduction, the software described in this report is intended to be part of a Computer Integrated Design and Manufacturing System.

The Rapid Design System (RDS) is one such fully integrated design and manufacturing software product which uses knowledge-based programming paradigms. The designer interfaces with the computer using a library of standard design features that help maintain the intent of the design, as well as, free the user from tedious drafting chores. The RDS will eventually use MetCAPP, a commercially available generative process planner.

Shown in Figure 8, is a flow chart of the future integration of the Interpretation Software, Tolerance Chart Software, MetCAPP and the RDS. First, the conceptual design is inputted by the user into the RDS using the Feature Based Design Environment. A descriptive file is generated for the part containing information about the number of features, their locations, sizes and any Positional Tolerances. The descriptive file will be read by the Interpretation Module, which will convert all ANSI Y 14.5 tolerance frame information to linear tolerances. A part description is also available to MetCAPP which generates a process plan for manufacturing each of the part's features. The process plan is transferred from MetCAPP to the Tolerance Chart software. By integrating the descriptive file and process plans a Tolerance Chart is generated.

The Interpretation and Tolerance Chart modules developed were intended to prove concepts. They were kept modular to enhance portability to a Computer Integrated Manufacturing system, such as the RDS. All software was written in Lisp[14][15] and operates under Wisdom Systems® Concept Modeller. The Interpretation software will be tested first with the part shown in Figure 9. The example given is a rectangular block with assorted holes requiring multiple control frames. The software converts the ANSI Y 14.5 circular tolerances information to the linear tolerances, as explained in the previous section. Output from the software is in two parts, as shown in Table 1. The first part is the linear tolerance zone for the linear axes. In the Table, only the X direction is shown. The second part of the data is the ordering and labeling of surface nodes, needed for the Tolerance Chart program.

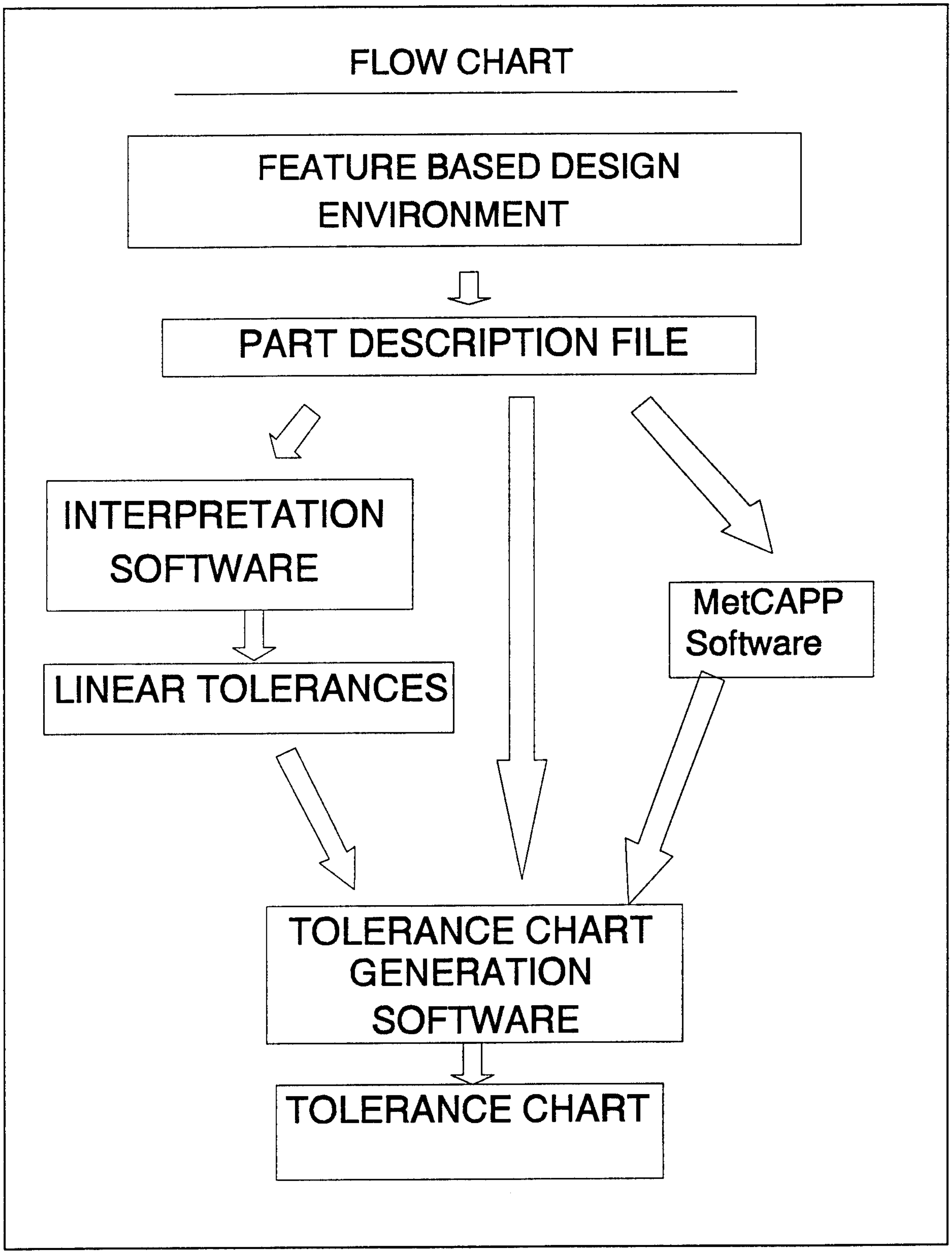

FLOW CHART
FEATURE BASED DESIGN ENVIRONMENT
PART DESCRIPTION FILE
INTERPRETATION SOFTWARE
LINEAR TOLERANCES
MetCAPP Software
TOLERANCE CHART GENERATION SOFTWARE
TOLERANCE CHART

Figure 8.

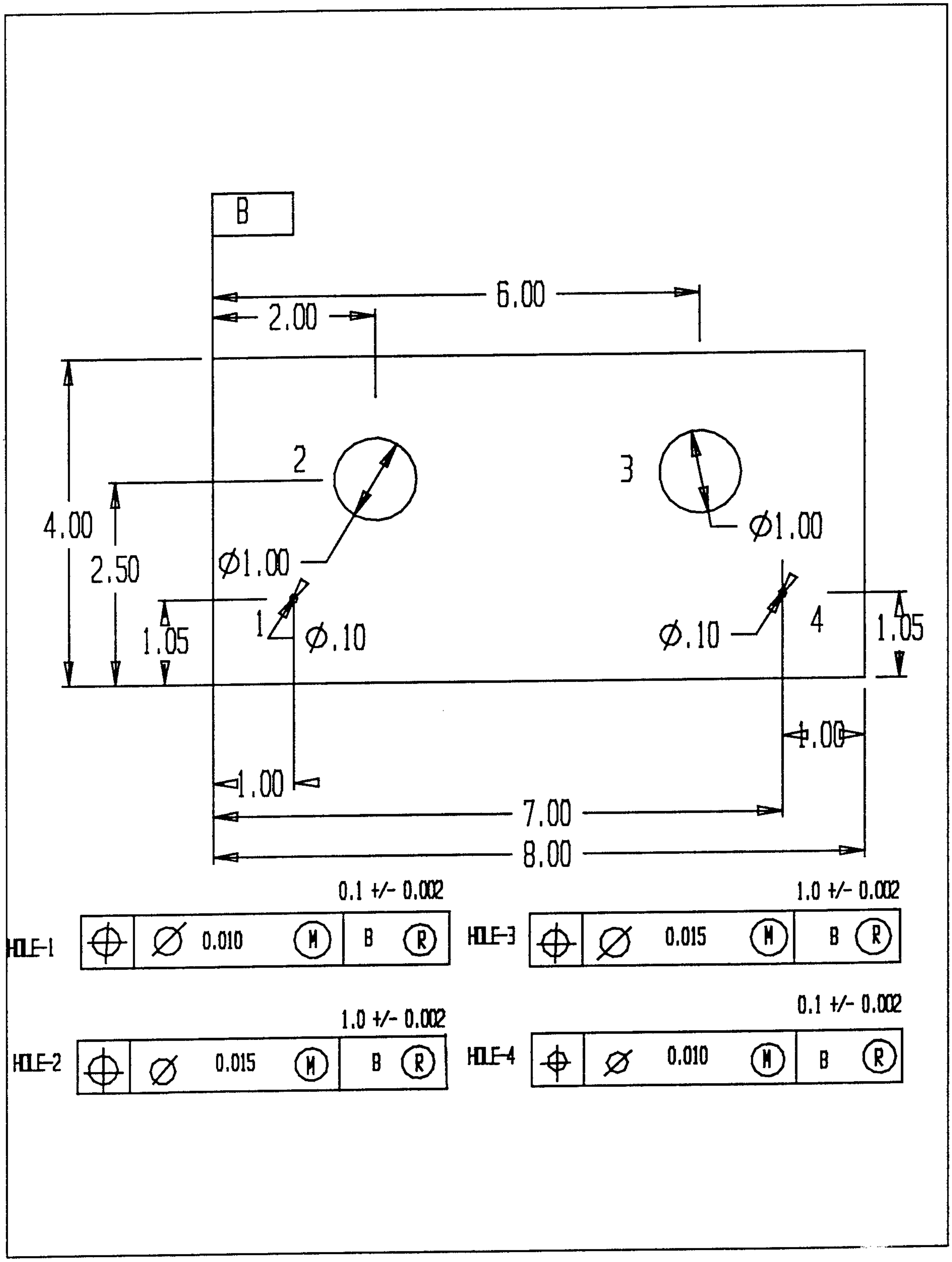

B
6.00
2.00
4.00
2.50
1.05
2
Ø1.00
3
Ø1.00
1
Ø.10
Ø.10
4
1.05
1.00
1.00
7.00
8.00
0.1 +/- 0.002
HOLE-1
⌖ Ø 0.010 Ⓜ B Ⓡ
1.0 +/- 0.002
HOLE-3
⌖ Ø 0.015 Ⓜ B Ⓡ
1.0 +/- 0.002
HOLE-2
⌖ Ø 0.015 Ⓜ B Ⓡ
0.1 +/- 0.002
HOLE-4
⌖ Ø 0.010 Ⓜ B Ⓡ

Figure 9.

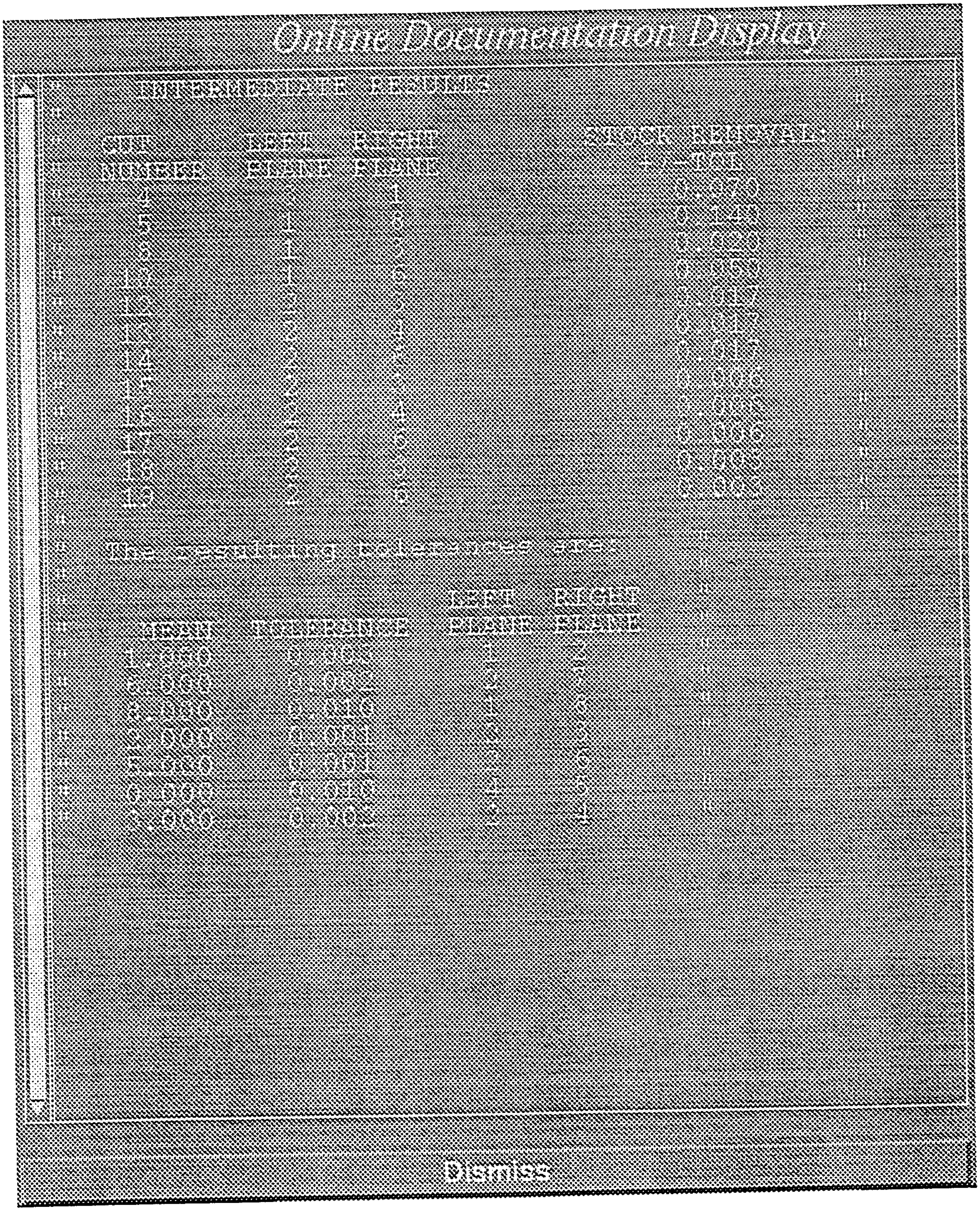
Table 1.

HOLE	DIMENSIONS		NODE PLANES	
	Mean	Tolerance	Datum	Machined
#1	1.0000	0.0035	1	2
#2	2.0000	0.0053	1	3
#3	3.0000	0.0053	1	4
#4	4.0000	0.0035	1	5

The information presented in the Table above is passed to the Tolerance Chart Program through a file. The example part presented is similar to a part described in Drozda and Wick[16] for which the authors have completed a Tolerance Chart. Drozda and Wick, however, fail to fully specify their example. It was necessary, therefore, to use their process plan, rather than generate one from MetCAPP, in order to test the Tolerance Chart program.

Information about machining cuts such as the locating surfaces is passed to the Tolerance Chart program. Tolerances associated with each machining cut can be generated as described in the previous section or provided by the user. The Tolerance Chart program completes the Chart in two steps. First the accumulated mean and tolerance for all cuts are calculated and compared to the designer specified dimensions. Stock removal tolerances are also calculated at this time. Figure 10 shows the intermediate results of the program.

The second step in the program is to calculate the mean machine position for intermediate cuts. The first step of the program has already verified that the process plan meets the designer dimensions for the part. Machine positioning requires that the user input mean stock removals values for the non-solid secondary machine cuts. Automated generation of Mean stock removals is an important area of research and will be discussed in the next section. The output from the second step of the Tolerance Chart generation software is shown in Figure 11. With mean stock removals specified, the Tolerance Chart generation software can loosen any machine cut tolerances so as to utilize the total dimensional freedom allowed by the designer.



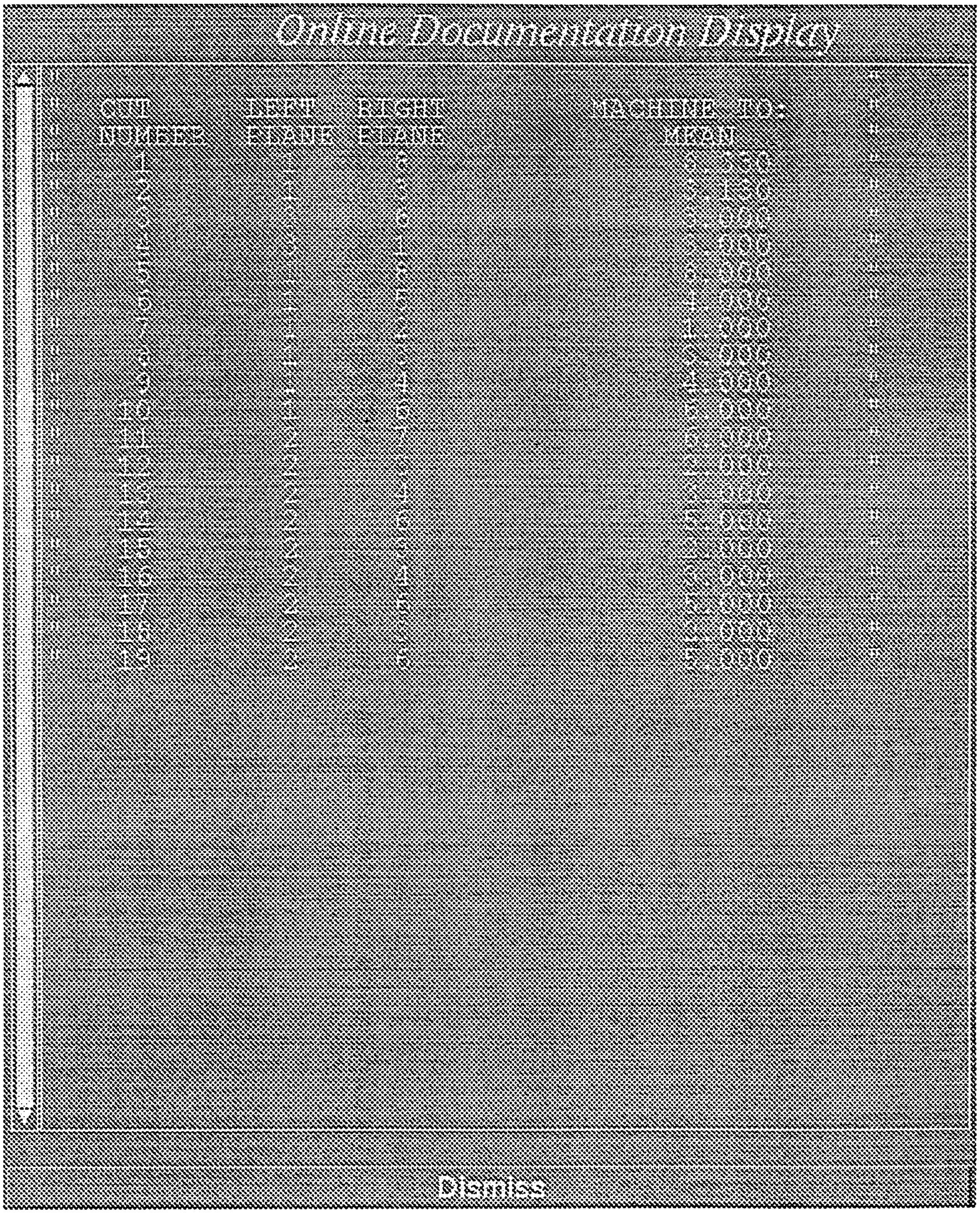
INTERMEDIATE RESULTS

CUT NUMBER	LEFT PLANE	RIGHT PLANE	STOCK REMOVAL: [illegible]
[illegible]	[illegible]	1	[illegible]
5	1	[illegible]	0.143
[illegible]	1	[illegible]	[illegible]
[illegible]	1	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
14	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

The resulting tolerances are:

MEAN	TOLERANCE	LEFT PLANE	RIGHT PLANE
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

Figure 10.



Online Documentation Display

CUT NUMBER	LEFT PLANE	RIGHT PLANE	MACHINE TO: MEAN
1	[illegible]	[illegible]	[illegible]
2	[illegible]	[illegible]	[illegible]
3	[illegible]	[illegible]	[illegible]
4	[illegible]	[illegible]	[illegible]
5	[illegible]	[illegible]	[illegible]
6	[illegible]	[illegible]	[illegible]
7	[illegible]	[illegible]	[illegible]
8	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]
10	[illegible]	[illegible]	[illegible]
11	[illegible]	[illegible]	[illegible]
12	[illegible]	[illegible]	[illegible]
13	[illegible]	[illegible]	[illegible]
14	[illegible]	[illegible]	[illegible]
15	[illegible]	[illegible]	[illegible]
16	[illegible]	[illegible]	[illegible]
17	[illegible]	[illegible]	[illegible]
18	[illegible]	[illegible]	[illegible]
19	[illegible]	[illegible]	[illegible]

Figure 11.

DISCUSSION

The Interpretation software was developed with relatively few obstacles. One difficulty was determining a method for handling the infinite variety of hole patterns possible. Holes could be located individually or symmetrically. In the case of a symmetrical layout, dimensioning among holes could be in cylindrical, rectangular or a non-orthogonal axes system. Symmetry could also be about an axis, a point or another feature. The software developed easily handled individually specified holes. Symmetry patterns were limited to rectangular and cylindrical coordinates. Interpretation of other dimensioning coordinates is an area of current research.

In the case of multiple tolerances on a single feature, consistency had to be assured. For tolerancing information to be inconsistent, a feature had to have at least two tolerance call outs which were not equal along a linear axis. Checking for consistency was simply a matter of checking for equivalence after interpretation.

The Tolerance Chart program worked as expected. It analyzed machining cut sequence and calculated the resultant feature positional tolerance, as well as, the stock removal tolerances. Furthermore, it calculated the mean machining position provided that the user inputs mean stock removal values. Automating the mean stock removal assignment process requires further investigation. Several factors are taken into account by the process planner when assigning mean stock removal. These factors include: process capability, variations in positioning a part relative to a tool, material hardness, surface finish and material removal needed to eliminate tool marks, tears, and otherwise clean up surface defects caused by previous operations. The program does provide a simple method for assigning tolerances to the machining cuts and is currently being revised to take into account additional machine data, presented by Trucks.[13] The Tolerance Chart generation software can loosen any machine cut tolerances so as to utilize the total dimensional freedom allowed by the designer.

Investigation into the tolerance charting technique for generative process planners is an important step in improving the usefulness of any Computer Integrated Manufacturing system. Future research is directed towards developing a system that considers the costs associated with tolerance. Alternative process plans and tolerance allocation will be considered. Linear programming may be one method to optimize for cost. The Tolerance Chart is an old technique which is being adapted to a new manufacturing environment.

REFERENCES

1. S.R. LeClair, "The Rapid Design System: Memory-Driven Feature-Based Design", *Proc. IEEE International Conf. on Systems Engineering*, Dayton, Ohio, August, 1991.

2. G. Boothroyd and W.A. Knight, *Fundamentals of Machining and Machine Tools*, 2nd Edition, Marcel Dekker, Inc., New York, New York, 1989.

3. L.W. Foster, *Geometric Dimensioning and Tolerancing - A Working Guide*, Addison-Wesley Publishing Co., Inc., Reading, Massachusetts, 1970.

4. L.E. Farmer and A.G. Harris, "Change of Datum of the Dimensions on Engineering Design Drawings", *International Journal of Machine Tool Design and Research*, Vol. 24, No. 4, 1984.

5. M.M. Sfantsikopoulos and S.C. Diplaris, "Coordinate Tolerancing in Design and Manufacturing", *Robotics & Computer-Integrated Manufacturing*, Vol. 8, No. 4, 1991.

6. D. Punchochar, *Interpretation Of Geometrical Dimensioning And Tolerancing*, Industrial Press Inc., New York, 1990.

7. O.R. Wade, *Tolerance Control in Design and Manufacturing*, Industrial Press, New York, 1967.

8. J.E. Nicks, *Basic Programming Solutions for Manufacturing*, Society of Manufacturing Engineers, Dearborn, Michigan, 1982.

9. R.S. Ahluwalia and A.V. Karolin, "CACT - A Computer Aided Tolerance Control System", *Journal of Manufacturing Systems*, Vol. 3, No. 2, 1986.

10. E.A. Lehtlhet and S. Joshi, " Framework For Dynamic Tolerance Control in Discrete Part Manufacturing", *Winter Annual Meeting of ASME*, Vol. 45, 1990.

11. T. Xiaoqing and B.J. Davies,"Computer Aided Dimensional Planning", *International Journal of Production Research*, Vol. 26, No. 2, 1988.

12. S. Irani, R. Mittal and E.A. Lehtlhet, "Tolerance Chart Optimization", *International Journal of Production Research*, Vol. 27, No. 9, 1989.

13. H.E. Trucks, *Designing for Economical Production*, 2nd Edition, Society of Manufacturing Engineers, Dearborn, Michigan, 1987.

14. *Common Lisp, The Reference*, Franz Inc., Addison-Wesley Publishing Co., Inc., 1988.

15. J.A. Moyne, *Lisp - A First Language Of Computing*, Van Nostrand Reinold, New York, N.Y., 1991.

16. T.J. Drozda and C. Wick, *Tolerance Control, A Selection from the Tool and Manufacturing Engineers Handbook*, Society of Manufacturing Engineers, Dearborn, Michigan, 1988.

17. J.H. Nurre, "Geometric Reasoning for Process Planning", *Proc. Third International Conference on Computer Integrated Manufacturing*, Troy, New York, 1992.

18. M. Schoonaver, J. Bowie and W. Arnold, *GNU Emax, Unix Text Editing And Programming*, Addision-Wesley Publishing Company Inc., 1991.

19. P. Hoffman, "Analysis of Tolerance and Process Inaccuracies in Discrete Part Manufacturing", *Computer Aided Design*, Vol. 14, 1982.

20. P. Gavankar, "Obstacles in the True Integration of CAD and CAM", *Proc. Third International Conference on Computer Integrated Manufacturing*, Troy, New York, 1992.

JACOBIAN UPDATE STRATEGIES FOR QUADRATIC AND NEAR-QUADRATIC CONVERGENCE OF NEWTON AND NEWTON-LIKE IMPLICIT SCHEMES

Dr. Paul D. Orkwis
Assistant Professor
Department of Aerospace Engineering and Engineering Mechanics

University of Cincinnati
Mail Location 70
Cincinnati, Ohio 45221-0070

Final Report for:
Research Initiation Program
Wright Laboratory

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

and

University of Cincinnati

March 1992

JACOBIAN UPDATE STRATEGIES FOR QUADRATIC AND NEAR-QUADRATIC CONVERGENCE OF NEWTON AND NEWTON-LIKE IMPLICIT SCHEMES

Dr. Paul D. Orkwis

Assistant Professor

Department of Aerospace Engineering and Engineering Mechanics

University of Cincinnati

Abstract

Several Jacobian matrix simplification ideas for Newton and Newton-like implicit Navier-Stokes equation solvers were evaluated. The Jacobian matrix simplifications involve updating only selected parts of the matrix with the most recently computed variables, freezing the entire Jacobian matrix after a specific number of iterations, and combinations of the two approaches. Numerical experiments were performed with these methods by computing supersonic flat plate and flat plate/wedge test cases. It was found that the approximate methods can give quadratic or better convergence rates if properly implemented.

Introduction

Before the introduction of the latest generation of high-speed supercomputers it was highly impractical to use memory intensive methods. However, today's machines like the Cray 2 and Y-MP are allowing researchers to reconsider memory intensive implicit schemes like Newton's method, which offer convergence rates that are significantly greater than typical Computational Fluid Dynamics (CFD) solvers. Newton's method has been applied to a variety of equation sets in recent years, including the potential equation [1], the Euler equations [2,3,4,5], and the Navier-Stokes equations [3,5,6,7]. Venkatakrishnan [7], Orkwis and McRae [8,9] and Orkwis [10] have shown that quadratic convergence rates are possible for state-of-the-art steady state Navier-Stokes discretizations like Roe's flux difference splitting. In addition, variations of the method [7,11] have demonstrated convergence rates greater than quadratic using a frozen Jacobian matrix. However, the price for improved convergence rates is high CPU time and memory requirements due to the complexity of the Jacobian matrix formation and solution processes. Improvements in these areas are required before the advantages offered by these schemes can be realized.

Many possibilities exist for improving the performance of Newton's method solvers. Several approaches have been tried to simplify the Jacobian matrix formation and solution processes, however, these simplifications did not retain the desirable convergence rates. The results obtained by Orkwis [10] and Liou and Van Leer [3] suggest that quasi-Newton's method solvers (methods that use approximate Jacobian matrices) do not exhibit quadratic convergence, and that an exact Jacobian matrix and matrix system solution are required.

Fortunately, the previous research did not exhaust all potential simplification ideas. It was the goal of this research to continue exploring this area to find approaches that would improve the efficiency of the Newton's method solver and retain high convergence rates.

One way to improve performance is to approximate the entries of the Jacobian matrix with a matrix that is "easier" to solve, as demonstrated by Orkwis and Liou and Van Leer. Another idea is to freeze the entire matrix for a select number of iterations or after a select iteration, ala Venkatakrishnan and Bailey and Beam. A third approach is to update the matrix entries only from points in the flow field with relatively large changes. These possibilities lead to the question of exactly how "correct" the Jacobian matrix must be in order to obtain quadratic or better convergence.

The following sections describe the approaches used in this research to answer the above question. They discuss the governing equations that were solved, describe the basic Newton's method procedure, present the Jacobian matrix simplifications ideas, discuss the graphical tools used to analyze the results, evaluate the results obtained with these methods, and finally give some concluding remarks.

Governing Equations

The equations which were discretely solved are the two-dimensional laminar compressible Navier-Stokes equations. They were transformed into generalized coordinates and discretized, as described previously by the second author [8,9,10], using Roe's flux difference splitting (FDS) [12] and the Spekreijse/Van Albada [4,13] continuous limiter. This set of equations was then used to form the Newton's method system.

Numerical Method

Newton's Method

Orkwis and McRae's Newton's method solver was derived by noting that the discrete governing equations may be written in the form:

$$\mathcal{F}\left(\bar{U}\right) = 0 \tag{1}$$

where $\bar{U}$ is the vector of conserved variables at each discretization point. Newton's method is then formed by creating the Jacobian matrix $\frac{\partial \mathcal{F}}{\partial \bar{U}}$ and writing:

$$\left(\frac{1}{\Delta t}[I] + \frac{\partial \mathcal{F}}{\partial \bar{U}}\right)^n \Delta^n \bar{U} = -\mathcal{F}^n(\bar{U}) \tag{2}$$

The Jacobian matrix is formed by using the MACSYMA symbolic manipulation expert system to differentiate $\mathcal{F}(\bar{U})$ and to output the FORTRAN code. The solution is obtained by solving for the iterate $\Delta^n \bar{U}$ and updating the solution variable vector via

$$\bar{U}^{n+1} = \bar{U}^n + \Delta^n \bar{U}$$

until equation (1) is satisfied.

The advantage to solving the exact system (equation 2 with $\Delta t = \infty$) is that quadratic convergence can be obtained if a "close enough" initial guess is supplied and an exact matrix inversion routine is employed. Unfortunately, a "close enough" initial guess is not often known a priori. For this reason, the $\frac{1}{\Delta t}$ modification is introduced which allows the scheme to iterate even if poor initial conditions are supplied.

However, linear convergence is obtained with this approach unless $\Delta t \rightarrow \infty$ as the computation nears convergence. An acceptable way of determining when this has occurred is to monitor a norm of the right hand side of equation 2. The quasi-timestep, Δt, is then varied inversely with this norm. This allows the correction to be removed as the final solution is approached.

Matrix Approximation Strategies

The Jacobian matrix formation process requires a significant amount of time because of the complexity of the exact Jacobian matrix entries. Considerable savings can result if this updating is avoided.

Global Freezing

The first method used to simplify the matrix formation process was "global freezing" of the entire matrix after a select number of iterations. This idea is similar to that used by Bailey and Beam [11]. The modified method can be written

$$\left(\frac{1}{\Delta t}[I] + \frac{\partial \mathcal{F}}{\partial \bar{U}}\right)^n \left(\bar{U}^{n,i} - \bar{U}^{n,i-1}\right) = -\mathcal{F}^{n,i-1}\left(\bar{U}\right) \qquad (3)$$

where i=1,...,k

$$\bar{U}^{n+1} = \bar{U}^{n,k} \qquad (4)$$

Several forms of this method exist. The approach used in this work was to employ the original modified Newton's solver, equation 2, for a select number of iterations and then apply equation 3. The approach fully updates the Jacobian matrix until a select iteration at which point it is frozen until convergence. Another approach is to use equation 3 throughout the computation, thereby updating the Jacobian matrix at select intervals. The latter method is the composition of one complete Newton step with $k-1$ simplified Newton steps. Both approaches offer a way of generating greater than quadratic convergence rates.

It should be noted that the convergence rates for the global freezing approximation report the residual only for the iterations at which the Jacobian matrix was updated. In this way the convergence plot better reflects the total amount of necessary computer resources since the work involved in the simplified Newton's steps is minimal as compared to that of the exact Newton's step. A better comparison could be provided by plots of residual versus CPU time. However, these plots are highly subjective as they depend upon the machine used and individual coding style. It was felt that the former approach to comparing the results could be more easily applied by others.

Partial Freezing

The second matrix approximation idea was "partial freezing" of the Jacobian matrix. This idea comes from consideration of the numerical situation that occurs in many problems. Consider a supersonic flat plate. The Jacobian matrix entries from points in the freestream region above the bow shock wave do not affect the convergence rate of the scheme, since little or no changes occur in the freestream region. The strategy of "partial freezing" uses this idea by searching the flow field for regions in which the Jacobian matrix can be altered without affecting the convergence rate. The two approaches used in this work to find these regions were updating of Jacobian matrix entries from the points with the greatest local residual values and updating the entries from preselected "important" regions.

In the case of "residual" partial freezing a tolerance is set based on the minimum and maximum local residuals. The value of the tolerance determines which Jacobian matrix entries are updated and which are frozen. The tolerance equation was

$$Tol = |residmin| + X|residmax - residmin| \tag{5}$$

where *residmin* is the minimum local density residual calculated during the last iteration, *residmax* is the corresponding maximum density residual, and X is a number between 0 and 1. Whenever the local density residual is greater than Tol the corresponding entries in the Jacobian matrix are updated. Those points not satisfying this criteria are frozen.

In the case of "regional" partial freezing only the Jacobian matrix entries from points that lie in a preselected region are updated. This approach is less flexible than the residual partial freezing idea but avoids the extra tolerance calculations.

Combined Approaches

The above methods were also applied as composite methods. The first composite method employs residual partial freezing for a select number of iterations and then switches to global matrix freezing. The second idea uses regional partial freezing for a select number of initial iterations and then also employs global freezing.

Matrix System Solution

The partial update ideas described above reduce the overall work involved in Jacobian matrix formation but do nothing to improve the efficiency of the resulting matrix system solution. An idea that can provide this improvement when small numbers of matrix updates are required is the Sherman-Morrison formula [14]. This formula states that if a matrix $\mathcal{A}$ is changed as per

$$\mathcal{A} \rightarrow \left(\mathcal{A} + \bar{U} \otimes \bar{V}\right) \tag{6}$$

then the inverse changes as per

$$\mathcal{A}^{-1} \rightarrow \left(\mathcal{A}^{-1} - \frac{\left(\mathcal{A}^{-1} \otimes \bar{U}\right)}{1 + \bar{V} \otimes \mathcal{A}^{-1} \otimes \bar{U}}\right) \tag{7}$$

To change an entire row of $\mathcal{A}$, $\bar{V}$ would be chosen as a unit vector and $\bar{U}$ the delta changes to that row. For a nonsparse matrix $\mathcal{O}(N^2)$ floating point operations are required to compute the inverse for each row/column update. This should be compared to $\mathcal{O}(N^3)$ operations needed to recompute it completely. However this approach requires the actual computation of the matrix inverse rather than using the simpler LU decomposition and back-solves.

The goal of this study was to first determine the update requirements for quadratic convergence. With this information the efficiency of the proposed schemes could then be determined by a careful comparison of operation counts between the partial update Sherman-Morrison matrix inverse approach and the usual matrix system solution approaches. It should be noted that the Sherman-Morrison formula was not implemented in this work since the goal of this effort was to determine the update requirements for the desired convergence rate. Therefore, the full savings inherent in the approximation strategies were not obtained since the usual matrix solution technique was still employed. Implementation of the Sherman-Morrison formula is planned for a later work.

Graphical Analysis

A useful tool in analyzing the results obtained with the methods and choosing the optimal partial update locations was graphic visualization of the iterative change in the local density residuals. This technique aided the search for the location of maximum residuals and pinpointed the necessary update locations. The tool also provided strong visual confirmation of the success or failure of a given update strategy, since one need only compare the local residual changes of the approximate approach to that of the exact solver to see how well the simplified strategy mimicked the exact Jacobian method. These results of the approximate strategy can be displayed next to the exact result using video animation or in the "storyboard" form shown in figures 7 to 14.

Results

This section describes the results obtained when the above Jacobian matrix approximations were implemented in the Newton's method code developed by Orkwis and McRae [8,9]. The grids and test cases are described first and then a discussion is given of the results obtained for the flat plate and flat plate/wedge test cases.

In this work, the Jacobian matrix approximation methods were tested by calculating the $M_\infty = 2$, $R_e = 1.65 \cdot 10^6$ flat plate and a $M_\infty = 2$, $R_e = 1.65 \cdot 10^7$ flat plate/15°

wedge computed by Orkwis [10]. Similar 40x40 flat plate and 79x40 flat plate/15° wedge meshes were used. Both grids have equal spacing in the x-direction and are described in the y-direction by the equation

$$y(j) = y_{min} + (y_{max} - y_{min}) \left(1 - s + \frac{2s}{1 + \left[\frac{s+1}{s-1}\right]^{\frac{ny-j}{ny-1}}} \right)$$

Where $s = 1.002$ for the flat plate and 1.001 for the flat plate/wedge, and $y_{max} - y_{min} = 1$ for both grids. Initial quasi-timesteps were $\Delta t_o = 4000$ for the flat plate and $\Delta t_o = 80$ for flat plate/wedge, however, neither of these values should be interpreted as optimal. Figures 1 and 2 respectively illustrate the density contours for the converged flat plate and flat plate/wedge solutions.

The tested matrix update schemes include;

1. full matrix updates for all iterations,
2. full matrix updates for a select number of iterations followed by global matrix freezing,
3. residual partial matrix updates,
4. regional partial matrix updates using a variety of preselected regions,
5. residual partial matrix updates for a select number of iterations followed by global matrix freezing,
6. regional partial matrix updates for a select number of iterations followed by global matrix freezing,

Flat Plate

Table 1 illustrates the iteration count and CPU time results obtained for the flat plate test case. It shows that the baseline full Jacobian matrix solution took 17 iterations and 414 Cray Y-MP CPU seconds to converge. Figure 3 illustrates the quadratic convergence

Update Strategy	Its	CPU Time (sec)	Savings %	# Update Pts	%
Full Update	17	414.0	0	6400	100
Full Update/Global Freezing	12	302.0	27.1	6400	100
Residual Partial Update	21	509.9	-22.2	2746	42.9
Regional Partial Update I	17	409.0	1.2	4248	66.4
Regional Partial Update II	18	425.9	-2.9	1440	22.5
Resid. P. Update/G. Freeze	12	324.9	21.5	2794	43.6
Region. P. Update I/G. Freeze	12	297.4	28.2	4248	66.4
Region. P. Update II/G. Freeze	12	293.0	29.2	1440	22.5

Table I: Flat Plate Results

of the density residual for this case and figure 7 depicts the iterative history of the local density residual contours. Figure 7 shows that the density residual decreases each iteration and illustrates the general residual reduction trends. This plot plays an important role in determining the details of the partial matrix update schemes.

The second case used a combination of full Jacobian matrix updates followed by global freezing. The global freezing iteration was chosen by referring to the baseline case which had a linear convergence rate for 12 iterations followed by quadratic convergence. Hence, the Jacobian matrix was fully updated for the first 12 iterations and frozen for the remaining iterations. Figure 4 shows the density residual for this case. The convergence rate matched that of the baseline solution during the early iterations but became greater than quadratic afterward. Recall that only residual values from iterations with full matrix updates are reported since they require considerably more work than the subiterations. In this case 9 additional subiterations were required after the linear region. The local density residual storyboard was not included for this case since it is nearly identical to the baseline case. The CPU time for this run was 302 seconds, which represents a 27.1% reduction as compared to the baseline. This provides further proof that the CPU time for subiterations is significantly less than that of full iterations.

An important variable for this case was the iteration defining the start of Jacobian matrix freezing. It was found that waiting until the solution was near the quadratic convergence region was optimal. Freezing the matrix before that point gave an increase in the number of iterations and CPU time. When the matrix was globally frozen immediately after the first iteration the solution diverged.

The third case was residual partial updating with $X = 0.01$. In this approach the Jacobian matrix was fully updated for the first iteration and the updating strategy was implemented afterwards. On average 2746 entries were updated out of a possible 6400, which represents a 57.1% reduction. Figures 3 and 8 illustrate that the density residual did not decrease as smoothly for this case as it had for the baseline case and that it was not quadratic. During the fifteenth and sixteenth iterations the density residual contours show an increase at some locations. This was significantly different from that obtained with the baseline case. It is felt that the sudden increase of the density residual occurred because some of the "correct" matrix entries were not updated during earlier iterations.

This run required 21 iterations, 4 more than the baseline case, and 509.9 CPU seconds, a 22.2% increase as compared to the baseline case.

The fourth case was regional partial updating of the Jacobian matrix. Two different regional partial updating methods were tested. The first was called regional partial update I, in which all points except those in the freestream region were used to update the Jacobian matrix. Figure 3 shows that the density residual matched that of the baseline case. It took 17 iterations and 409 CPU seconds, which represents a 1.2% saving. On average 4248 entries were updated out of a possible 6400, a 33.6% reduction.

The second regional updating scheme, regional partial update II, used only points from the boundary layer region to update the Jacobian matrix. Once again figure 3 shows the density residual. It is apparent that the convergence rate was much smoother than that obtained with residual partial updating. In addition, the local density residual contours, shown in figure 9, illustrate a residual reduction similar to that of the baseline case. The more dramatic result is that on average the number of Jacobian matrix entries being updated was 1440 out of 6400, a 77.5% reduction as compared to the baseline case and a 66.1% reduction as compared to regional partial update I. In addition, it took 18 iterations and 425.9 CPU seconds to reach the final solution, a 2.9% increase as compared to the baseline case. The above indicates that only the freestream region can be neglected in the partial updating strategy if quadratic convergence is to be maintained. However, the method can still be "competitive" in terms of CPU time when small numbers of points are updated.

Comparing the Jacobian matrix partial updating strategies based on residual and region II, it is evident that the location of the updated entries is more important than the quantity of entries being updated. Figure 10 shows the grid points being updated with the residual partial update scheme. It is apparent that many more points are updated in this case as compared to the boundary layer regional partial update case. In addition, the scheme did not converge as quickly as had the regional approach. This is significant because the number of matrix entries being updated is directly related to the computational work when the Sherman-Morrison formula is implemented.

Next, the composite Jacobian matrix approximation ideas were tested. The first combination case tested was residual partial updating while in the linear convergence region followed by global freezing of the Jacobian matrix until the method converged. Figure 4 shows that once again the convergence rate was greater than quadratic with global freezing. 12 iterations and 21 global freezing subiterations were required for convergence. In addition, 324.9 CPU seconds were needed, a 21.5% CPU reduction. On average, the number of entries being updated in the linear region was 2794 out of 6400, a 56.4% reduction. In these cases the local density residuals were the same as those obtained with the previous partial update schemes, hence they are not shown in storyboard form.

The second composite idea used the regional partial update I idea in the linear convergence region followed by global Jacobian matrix freezing until convergence. In this case the density residual matched that of the baseline global freezing case. 297.4 CPU seconds were required, a 21.5% reduction as compared to the baseline. The average number of matrix entries being updated was 4248 out of 6400, a 33.6% reduction.

The third composite case tested used the regional partial update II and global Jacobian matrix freezing. Figure 4 shows the density residual for this method, in which 12 iterations and 9 global freezing subiterations were required. 293 CPU seconds were needed for convergence, a 29.2% reduction as compared to the baseline. The number of average entries being updated in the linear region was 1440 out of 6400, a reduction of 77.5%.

Comparing the composite ideas based on residual and region II, the latter required less CPU time and less updated matrix entries. This result again shows that it is more important to know which entries to update than it is to update a large number of entries. The first composite idea had more entries being updated than the regional partial update II idea and yet took longer to converge. This difference would be magnified if the Sherman-Morrison formula were implemented.

Flat Plate/15° Wedge

The results obtained for the flat plate/15° wedge case, shown in table 2, were similar to those obtained for the flat plate test case. The baseline solution required 95 iterations and 6485 CPU seconds to converge. Figure 5 illustrates the density residual for the baseline and partial update cases, and figure 11 shows the storyboard of the local density residual contours. It should be noted that the initial condition for this test case made use of the results from the flat plate test case.

The full update/global freezing case required 90 iterations to converge and 6088.7 CPU seconds, a decrease of 6.5%. Once again the global freezing point was chosen by referring

Update Strategy	Its	CPU Time (sec)	Savings %	# Update Pts	%
Full Update	95	6485.0	-	12640	100
Full Update/Global Freezing	90	6088.7	6.5	12640	100
Residual Partial Update	99	6811.2	-4.8	2296	18.2
Regional Partial Update I	95	6528.9	-0.7	6324	50
Regional Partial Update II	99	6820.6	-5.2	3340	26.4
Resid. P. Update/G. Freeze	90	6143.2	5.3	2274	18
Region. P. Update I/G. Freeze	90	6133.5	5.4	6324	50
Region. P. Update II/G. Freeze	90	6144.4	5.3	3340	26.4

Table II: Flat Plate/Wedge Results

to the baseline case residuals and the local residual contours.

Next, the $X = 0.01$ residual partial update scheme was tested. This time 99 iterations were required for convergence and 6811.2 CPU seconds, which represents an increase of 4.8%. Figure 12 shows that the density residual contours did not decrease as smoothly as had the baseline case. Coincidentally the number of extra iterations required for this approximate Jacobian method was identical to the corresponding flat plate case. The number of points updated during the linear convergence region was 2296, a reduction of 81.8% as compared to the baseline case. Figure 14 shows the corresponding grid points being updated.

The third tested method was regional partial updating. Two different regional partial updating methods were tested. In the first case, regional partial update I, the region associated with all of the relatively important flow features was updated. The updated region was that falling below the line

$$y = [x - x_{corner}] \frac{y_{h2}}{x_{max} - x_{corner}} + y_{h1}$$

where $y_{h2} = .25$, $y_{h1} = .15$, and $x_{max} = .5$, plus the points in the boundary layer of the flat plate. This method required 95 iterations and needed 6528.9 CPU seconds to converge, which represents a 0.7% increase in required computation time. Figures 4 shows that the density residual identically matches that of the baseline case. However, the disadvantage of this case was that a large number of matrix entries had to be updated as compared to the residual partial update case. The average number of points being updated was 6324, which only represents an 50% reduction as compared to the baseline case. However this was a 178.1% increase as compared to the residual update case.

In the regional update II case, a much smaller number of matrix entries was updated. This strategy diverged for several poorly chosen update regions. The updated region for the presented results was that falling below the line with $y_{h2} = .15$, $y_{h1} = .01$, and $x_{max} = .5$. This case required 99 iterations but needed 6820.6 CPU seconds to converge, which is a 5.3% increase in computation time as compared to the baseline. The average number of points updated while in the linear convergence region was 3340, a 73.6% reduction as compared to the baseline case, but a 46.8% increase as compared to the residual pdate case. Figure 13 shows the corresponding local density residual contours.

It is important to note that for the flat plate test case the regional partial update method required less update points than did the residual partial update scheme, and that this trend was reversed in the flat plate/wedge case. This reinforces the previous statement that the amount of matrix entries being updated is less important than choosing the correct matrix entries to update. Also it is important to note that if all the "relatively important" regions of the flow field are updated then the scheme will converge quadratically.

Figure 6 depicts the density residuals for the composite method test cases. Residual partial updating/global freezing required 90 iterations to converge and 6143.2 CPU seconds, a 5.3% reduction from the baseline. The regional partial update I/global freezing method required 90 iterations to converge and 6133.5 CPU seconds, a 5.4% computational requirement savings. The regional partial update II/global freezing method also needed 90 iterations to converge and a nearly identical 6144.4 CPU seconds, resulting in a similar 5.3% computational requirement savings.

In summary, the approximate Jacobian methods exhibited quadratic convergence when appropriate matrix entries were updated and all of the methods exhibited better than quadratic convergence when applied with global freezing. The regional partial update II strategy achieved better performance than the residual partial update strategy, both in CPU saving and numbers of updated matrix entries, for the flat plate case. Conversely, the residual partial update idea outperformed the residual partial update II approach for the flat plate/wedge test case. This indicates that the location of updating is more important than the number of points updated and suggests that a better choice of update region exists than that tested for the flat plate/wedge case.

Conclusions

Several approximate Jacobian matrix ideas were tested. These approximate strategies were compared to exact Jacobian matrix methods. Density residual, CPU time, and the number of updated points in the Jacobian matrix were analyzed so that the feasibility of applying these methods with the Sherman-Morrison formula could be evaluated. The following was determine;

1. global Jacobian matrix freezing after a select number of iterations can exhibit quadratic or better convergence rates and produce reduction in CPU time,
2. partial update strategies will produce quadratic convergence rates when all of the "important" flow regions are updated,
3. greater than quadratic convergence rates can be obtained when partial update strategies are applied together with global freezing,
4. updating the proper points is much more important than updating a large number of points.

The current results have encouraged the authors to believe that the application of the Sherman-Morrison inverse procedure with a partial update/global freezing approach has the potential to provide quadratic or better convergence in an efficient solution procedure.

Acknowledgments

This research was supported by the U.S. Air Force Office of Scientific Research , Bolling AFB, Washington, D.C. Computational resources were provided by the Eglin Air Force Base Supercomputer Center. The authors wish to thank Drs. Dave Belk and Kirk Vanden for their opinions regarding Newton's method solvers.

Bibliography

1. E.E. Bender and P.K. Khosla, Application of Sparse Matrix Solvers and Newton's Method to Fluid Flow Problems, AIAA Paper 88-3700-CP, in *Proceedings, 1st AIAA/ASME/SIAM/APS National Fluid Dynamics Congress, Cincinnati, OH, 1988*, p. 402.

2. M. Giles, M. Drela, and W.T. Thompkins, Newton Solution of Direct and Inverse Transonic Euler Equations, AIAA Paper 85-1530, in *Proceedings, 7th AIAA Computational Fluids Dynamics Conference, 1985*, P. 394.

3. M.S. Liou and B. Van Leer, Choice of Implicit and Explicit Operators for the Upwind Differencing Method, AIAA Paper 88-0624, Reno, NV, 1988.

4. S.P. Spekreijse, "Multigrid Solution of the Steady Euler Equations," Ph.D. Dissertation, Centrum voor Wiskunde en Informatica, Amsterdam, 1987.

5. H. Hafez, S. Palaniswamy, and P. Mariani, Calculations of Transonic Flows with Shocks Using Newton's Method and Direct Solver, Part II, AIAA Paper [illegible] 0226, Reno, NV, 1988.

6. P.D. Orkwis and D.S. McRae, A Newton's Method Solver for the Navier-Stokes Equations, AIAA Paper 90-1524, Seattle, WA, 1990.

7. V. Venkatakrishnan, Newton Solution of Inviscid and Viscous Problems, *AIAA J.* **27**, 885 (1989).

8. P.D. Orkwis and D.S. McRae. A Newton's Method Solver for High-Speed Viscous Separated Flowfields, *AIAA J.* **30**, 78 (1992).

9. P.D. Orkwis and D.S. McRae, A Newton's Method Solver for the Axisymmetric Navier-Stokes Equations, *AIAA J.* **30**, 1507 (1992).

10. P.D. Orkwis, A Comparison of Newton's and Quasi-Newton's Method Solvers for the Navier-Stokes Equations, AIAA Paper 92-2644, in *Proceedings, 10th AIAA Applied Aerodynamics Conference, Palo Alto, CA, 1992*, p. 410.

11. H.E. Bailey and R.M. Beam, Newton's Method Applied to Finite-Difference Approximations for the Steady-State Compressible Navier-Stokes Equations, *J. Comput. Phys.* **93**, 108 (1991).

12. V.N. Vatsa, J.L. Thomas, and B.W. Wedan, Navier-Stokes Computations of Prolate Spheroids at Angle-of- Attack, *J. Aircraft* **26**, 986 (1989).

13. G.D. Van Albada, B. Van Leer, and W.W. Roberts,Jr., A Comparative Study of Computational Methods in Cosmic Gas Dynamics, *Astronomy and Astrophysics* **108**, 1982.

14. W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes, The Art of Scientific Computing*, (Cambridge University Press, New York, 1989), p. 66.

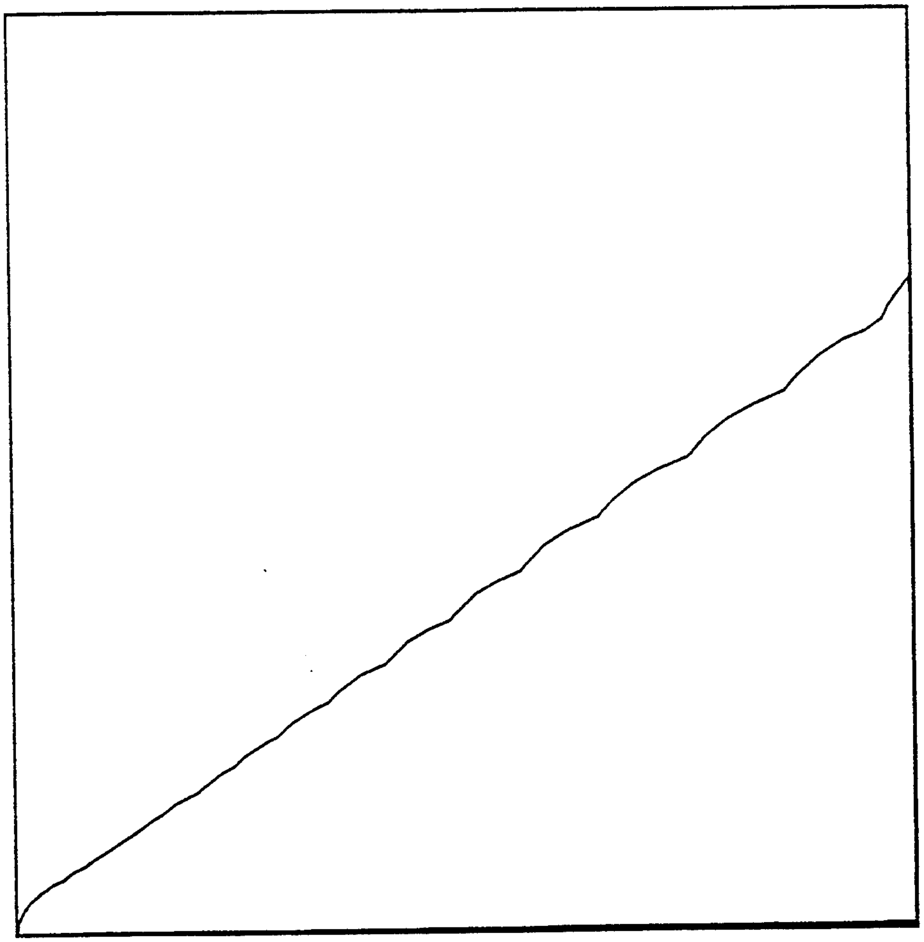


Figure 1: Flat Plate Density Contours.

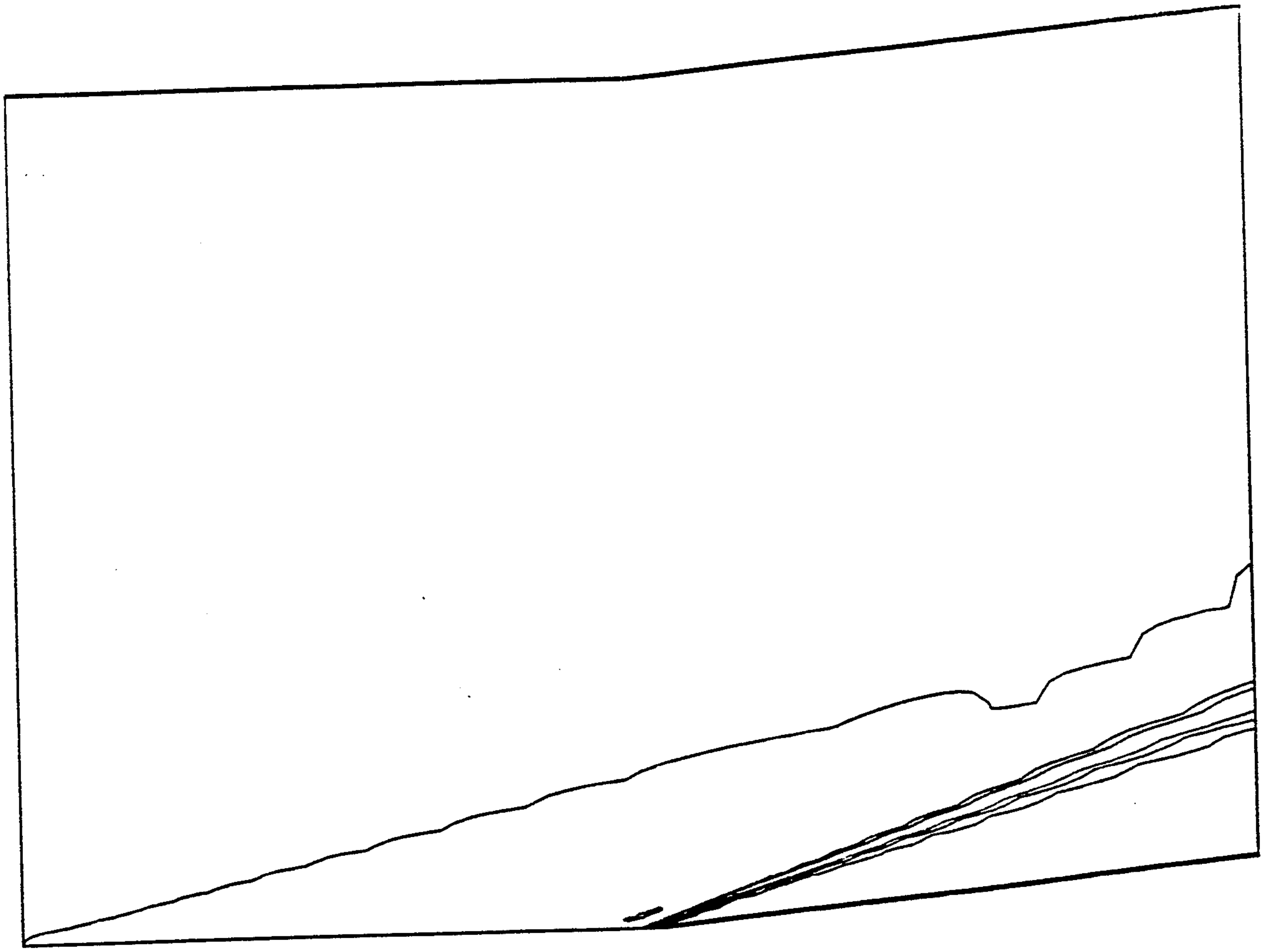


Figure 2: Flat Plate/15° Wedge Density Contours.

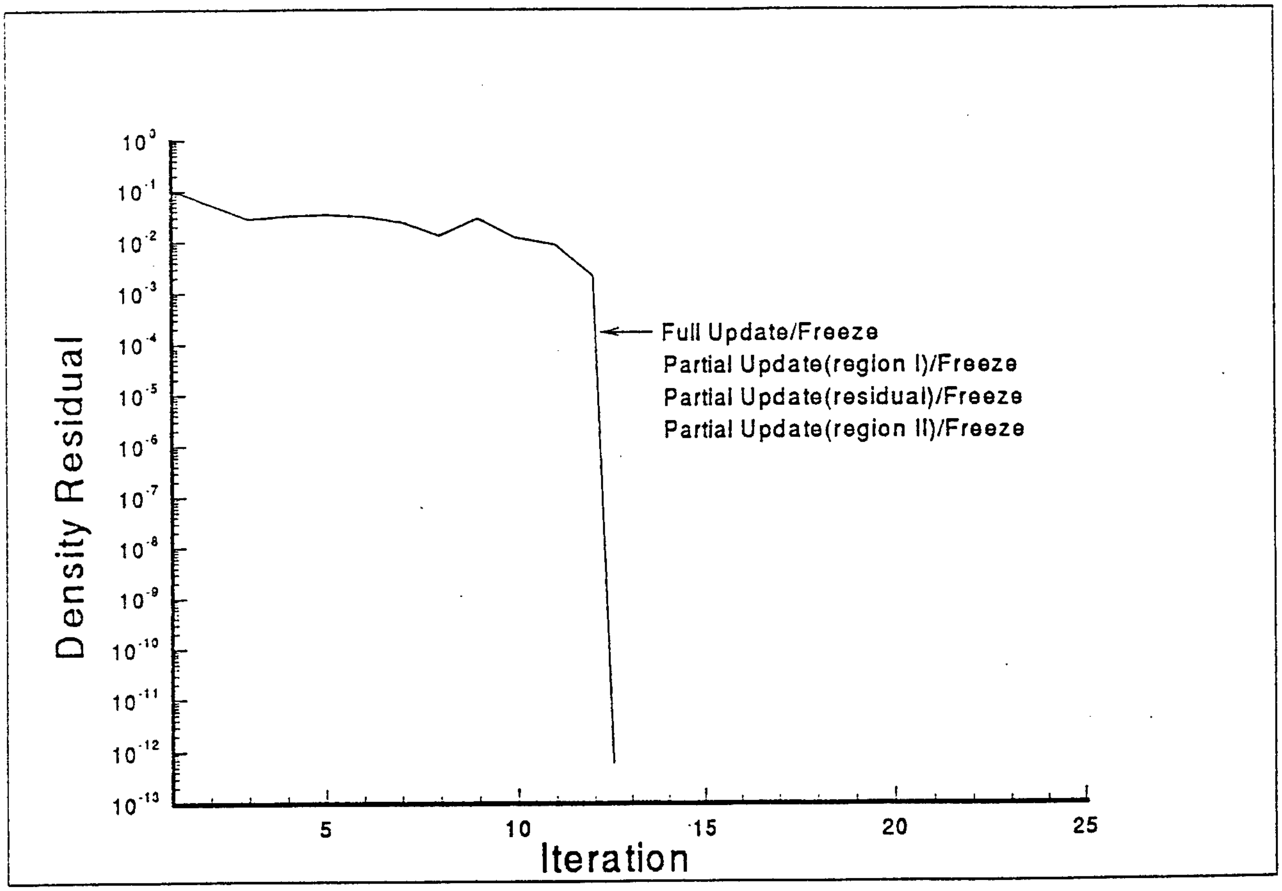


Figure 4: Flat Plate Density Residuals for the Full and Partial Update/Global Freezing Schemes.

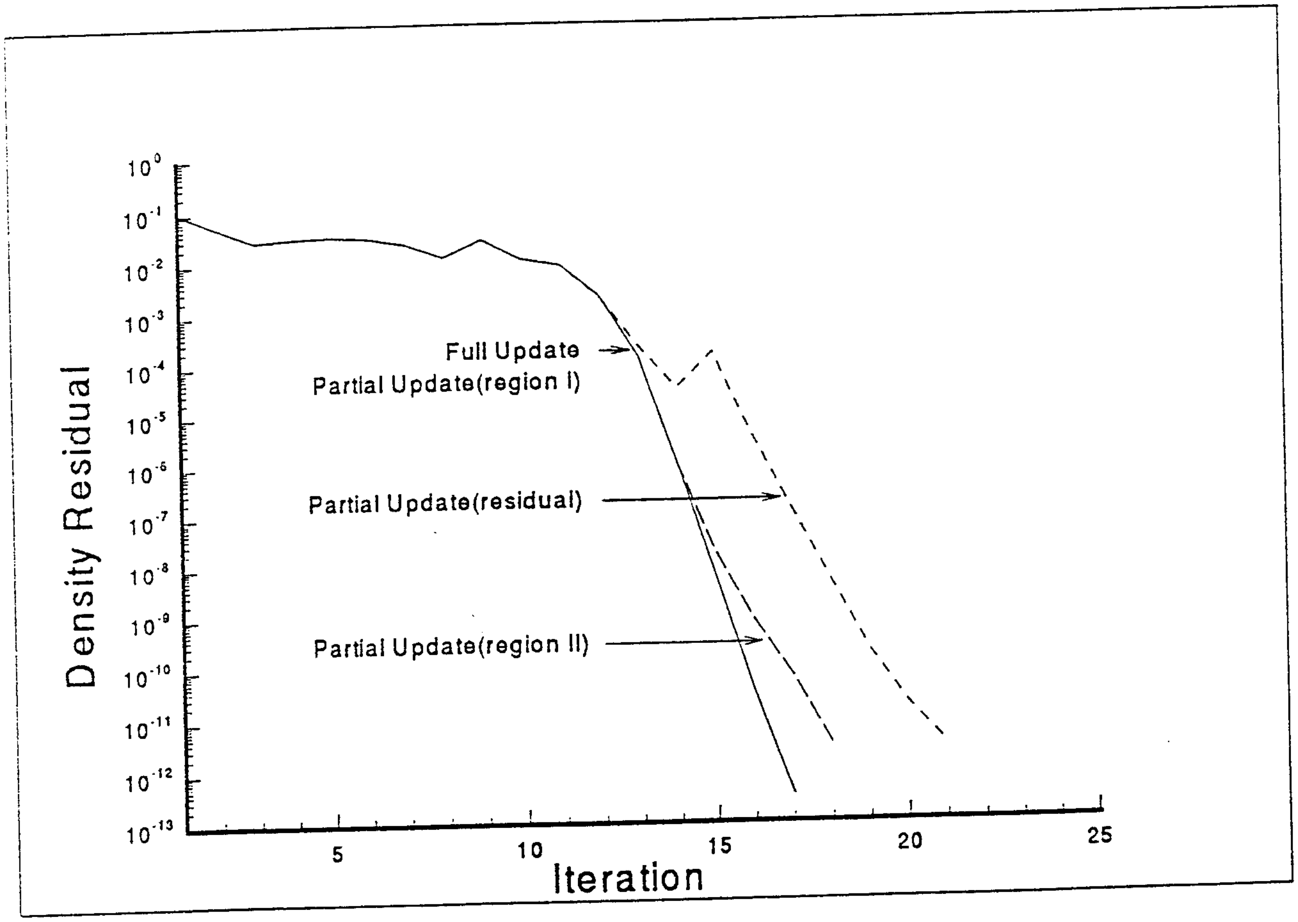


Figure 3: Flat Plate Density Residuals for the Full and Partial Update Schemes.

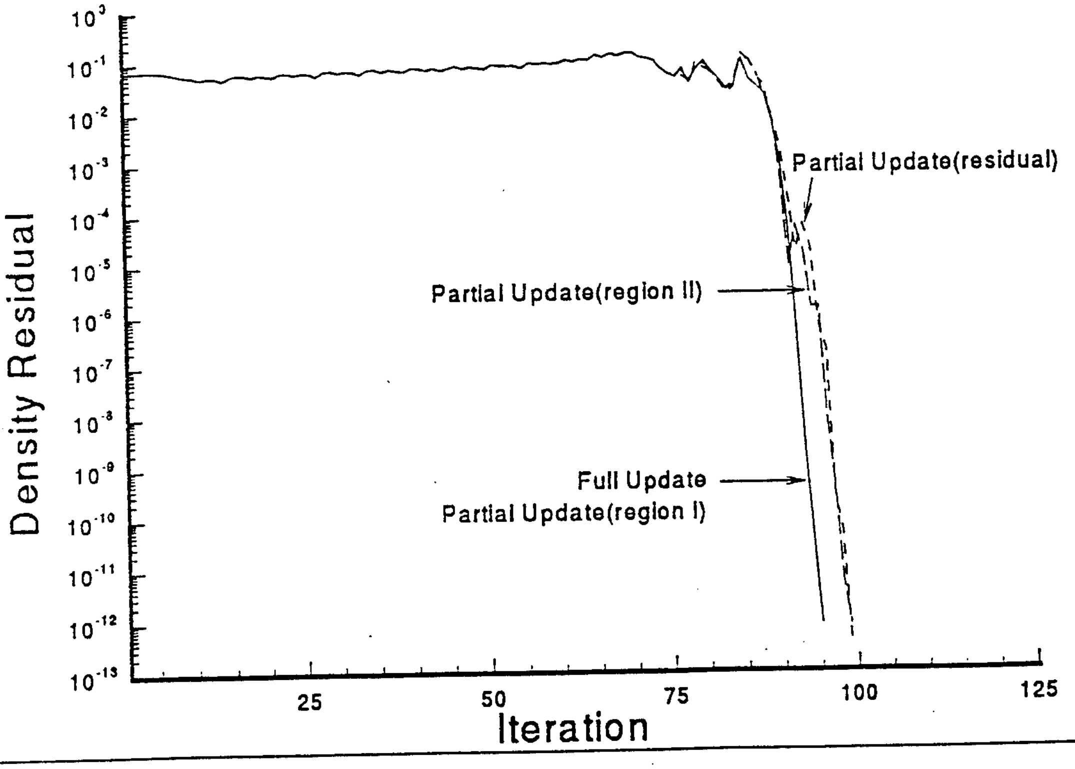


Figure 5: Flat Plate/15° Wedge Density Residuals for the Full and Partial Update Schemes.

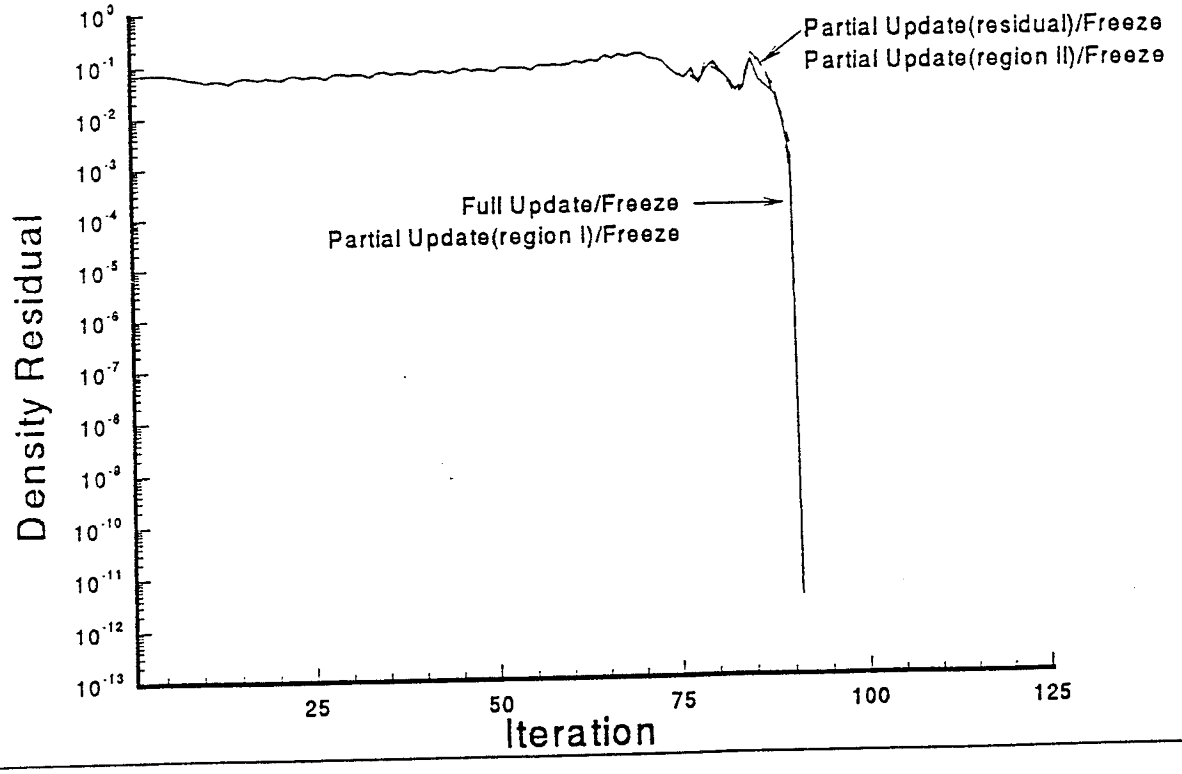


Figure 6: Flat Plate/15° Wedge Density Residuals for the Full and Partial Update/Global Freezing Schemes.

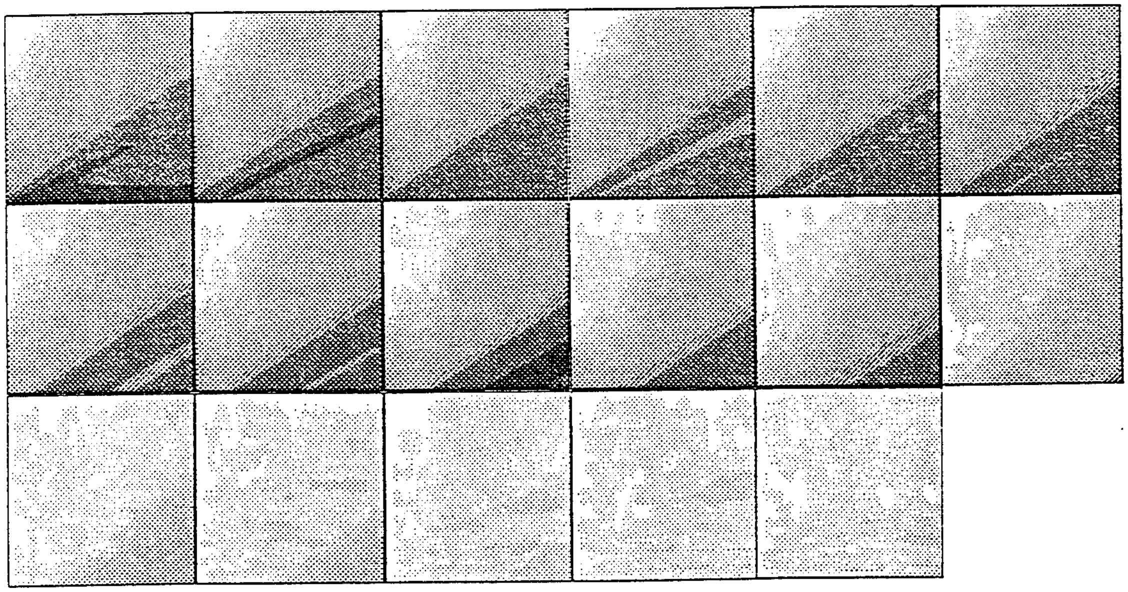


Figure 7: Flat Plate Local Density Residuals for the Full Update Scheme. (Darkest Region Indicates Greatest Residual).

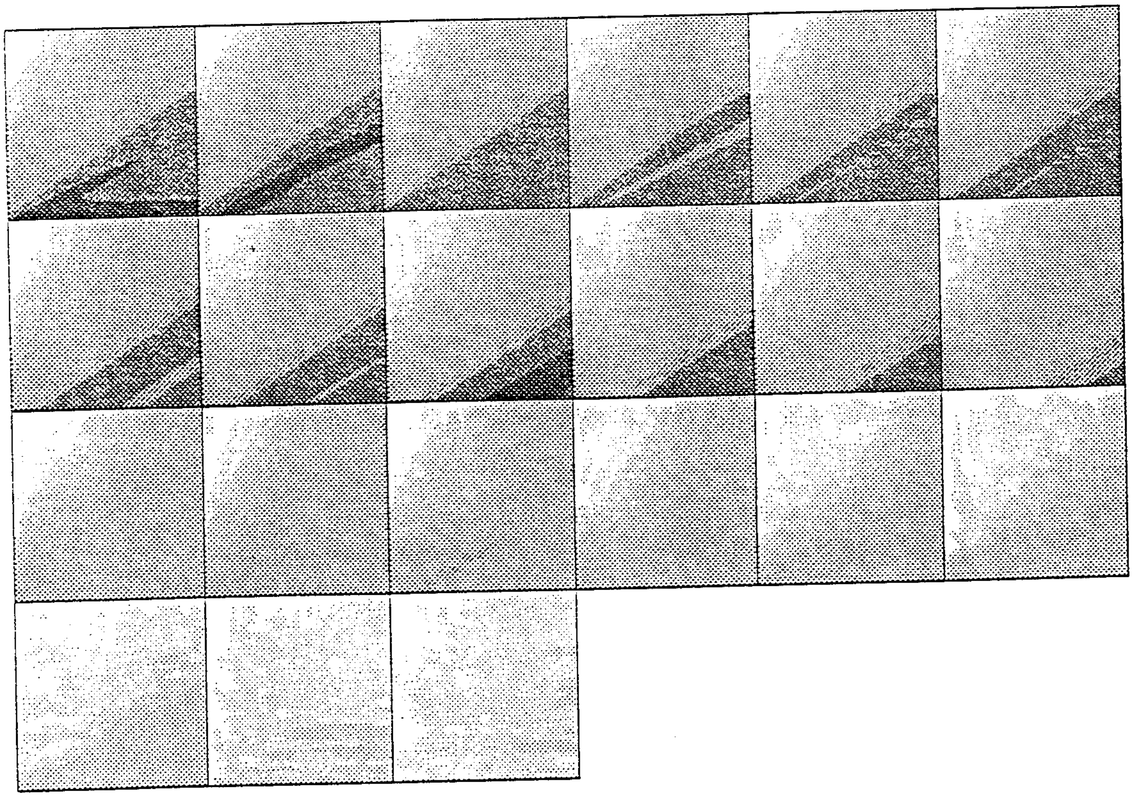


Figure 8: Flat Plate Local Density Residuals for the Residual Partial Update Scheme. (Darkest Region Indicates Greatest Residual).

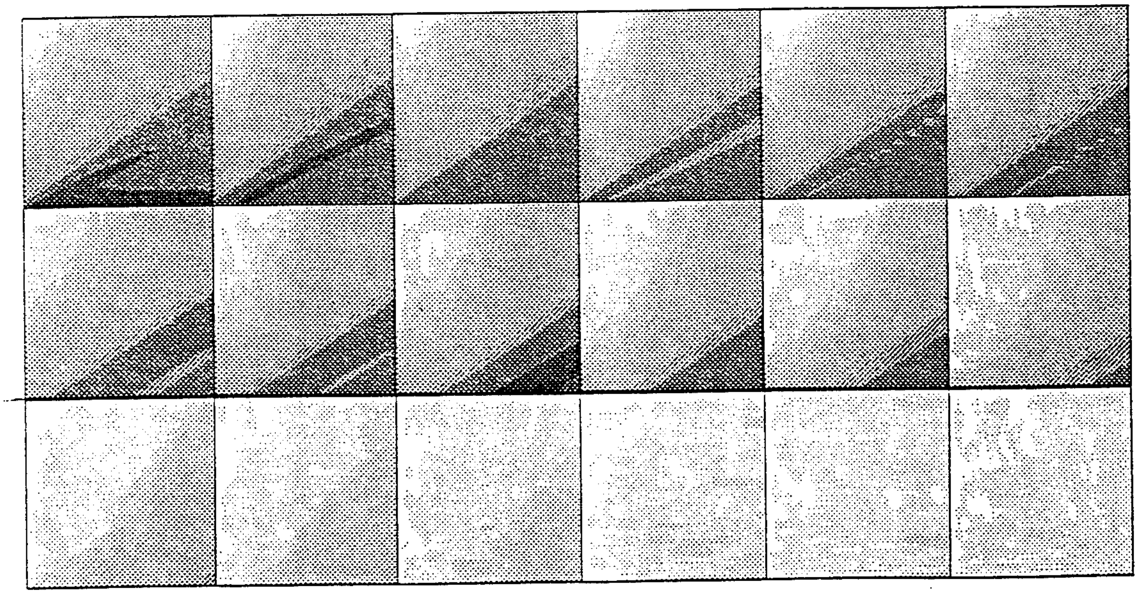


Figure 9: Flat Plate Local Density Residuals for the Regional Partial Update Scheme. (Darkest Region Indicates Greatest Residual).

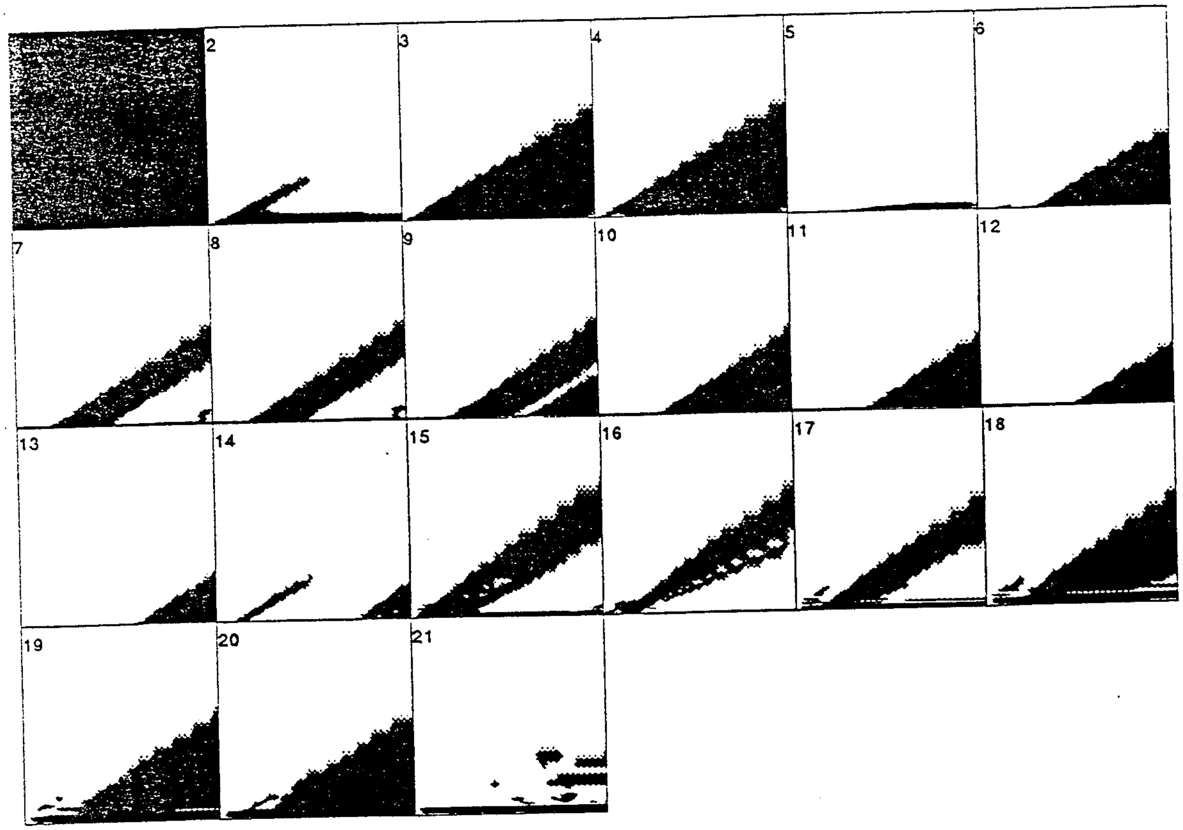


Figure 10: Flat Plate Update Locations for the Residual Partial Update Scheme.

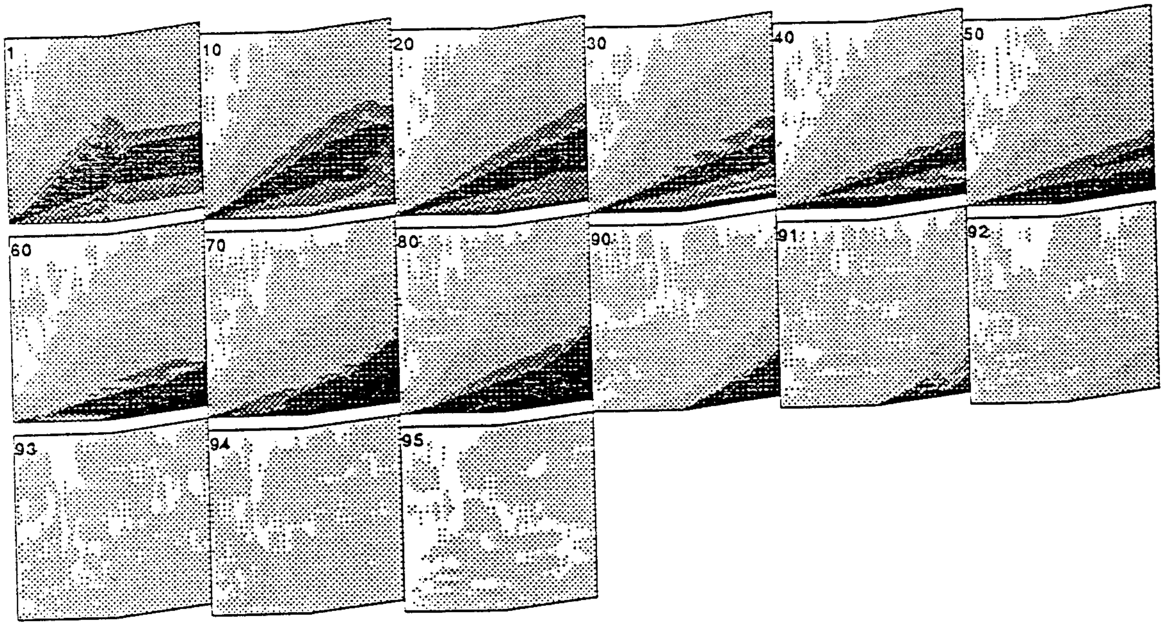


Figure 11: Flat Plate/15° Wedge Local Density Residuals for the Full Update Scheme. (Darkest Region Indicates Greatest Residual).

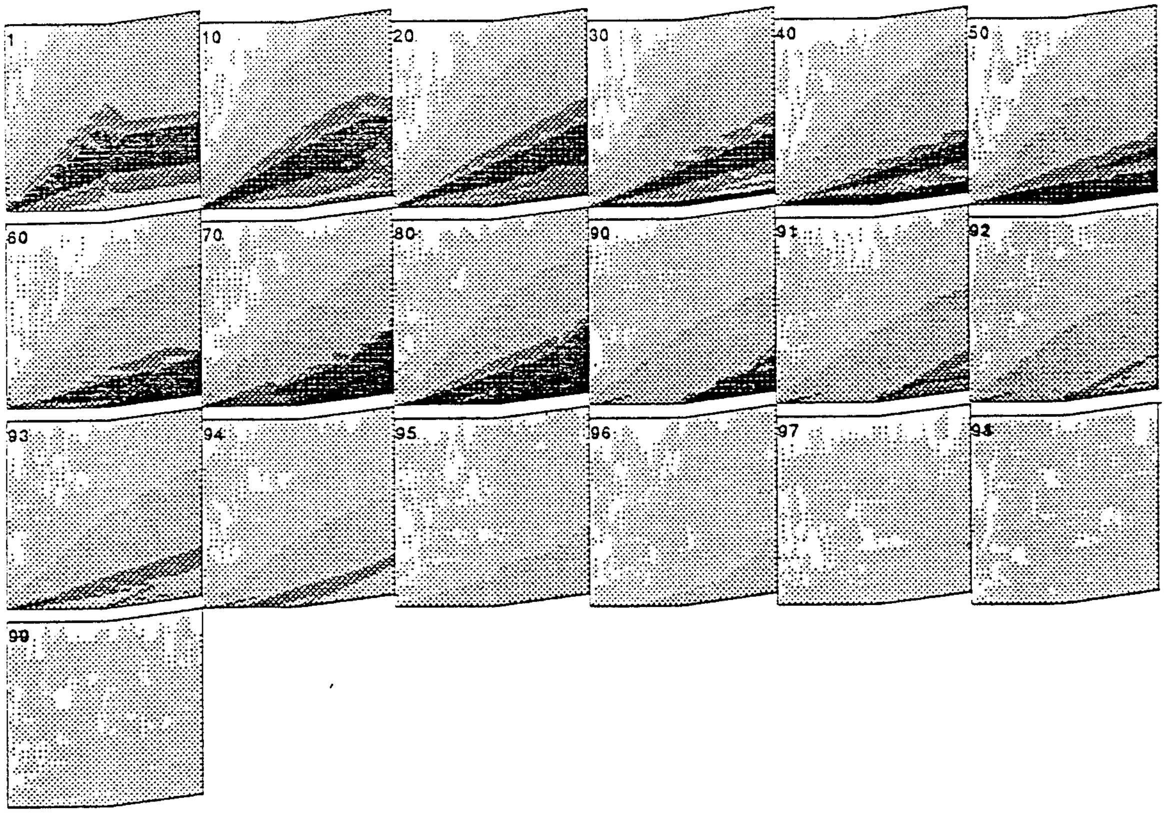


Figure 12: Flat Plate/15° Wedge Local Density Residuals for the Residual Partial Update Scheme. (Darkest Region Indicates Greatest Residual).

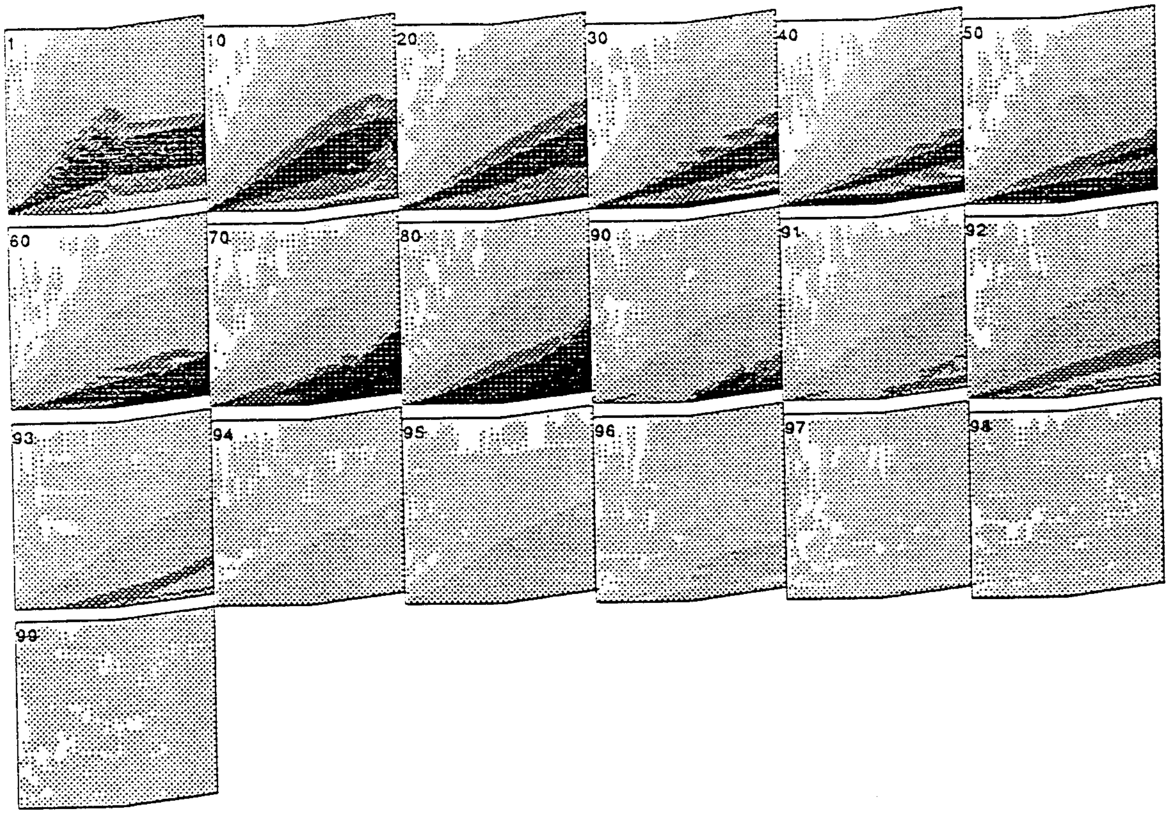


Figure 13: Flat Plate/15° Wedge Local Density Residuals for the Regional Partial Update Scheme. (Darkest Region Indicates Greatest Residual).

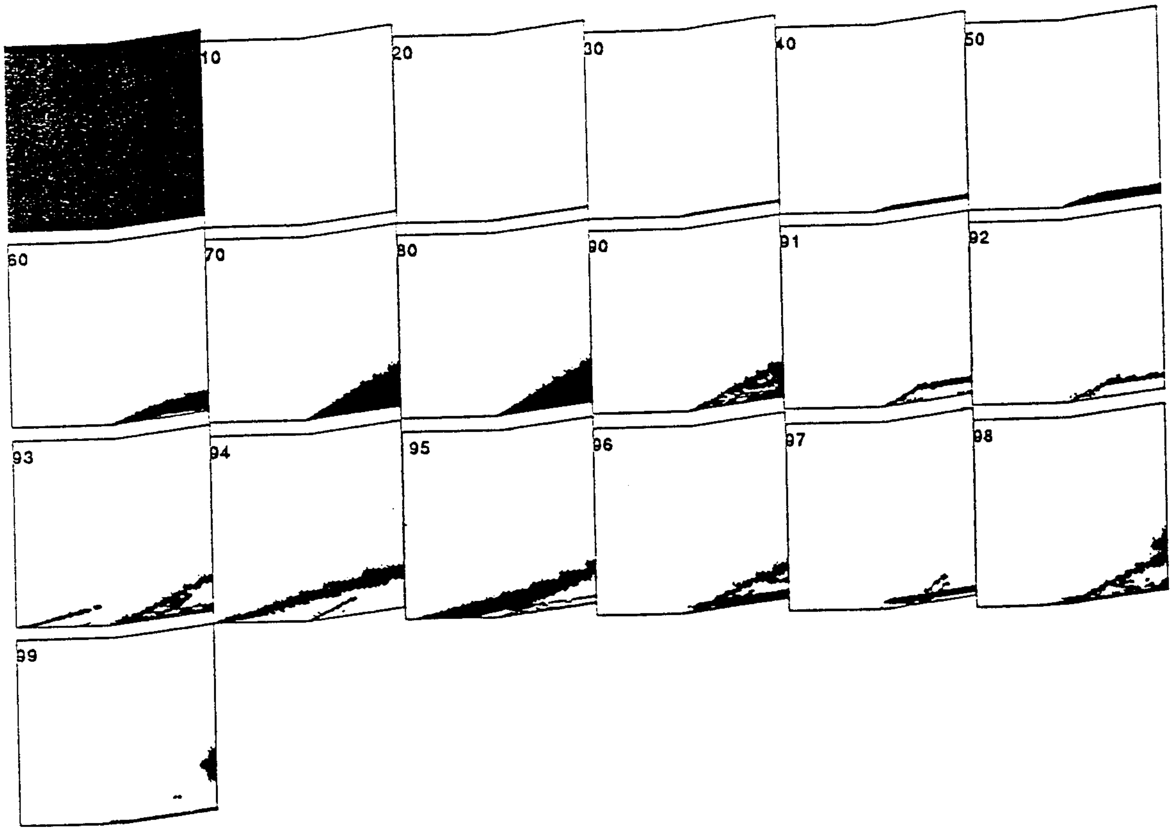


Figure 14: Flat Plate/15° Wedge Update Locations for the Residual Partial Update Scheme.

A MOLECULAR MODELING SYSTEM FOR PREDICTION OF SILOXANE LIQUID CRYSTALLINE POLYMER MICROSTRUCTURES

Steven K. Pollack
Assistant Professor
Department of Materials Science & Engineering

University of Cincinnati
493 Rhodes Hall
Cincinnati, OH 45221-0012

Final Report for:

Research Initiation Program
Wright Laboratories
Materials Directorate

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

and

The University of Cincinnati

March 1993

A MOLECULAR MODELING SYSTEM FOR PREDICTION OF SILOXANE LIQUID CRYSTALLINE POLYMER MICROSTRUCTURES

Steven K. Pollack
Assistant Professor
Department of Materials Science & Engineering
University of Cincinnati

Abstract

A molecular modeling facility was established in the PI's laboratory consisting of an IBM RS/6000 320H RISC based workstation and a number of commercial and public domain molecular modeling packages. Codes have been developed for the Polygraf™ package to allow for the efficient calculation of the structure and dynamics of rod-like molecules or assemblies of molecules in hexagonal close-packed arrangements. These codes have been shown to be approximate 5-10 time faster than full bulk density simulations of polymers. The system has also been utilized to model the NLO properties of a number of target molecules. Finally, torsional potentials for a number of flexible mesogens have been calculated using this system.

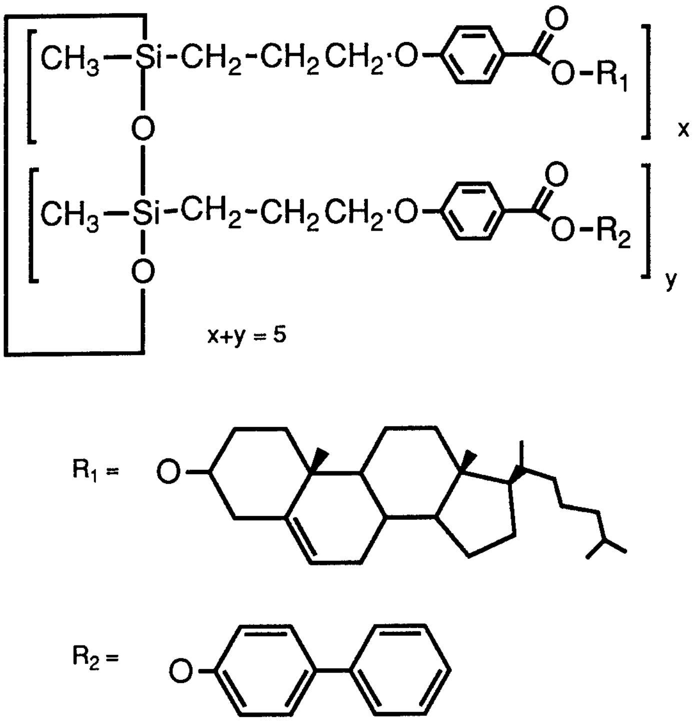
A MOLECULAR MODELING SYSTEM FOR PREDICTION OF SILOXANE LIQUID CRYSTALLINE POLYMER MICROSTRUCTURES

Steven K. Pollack

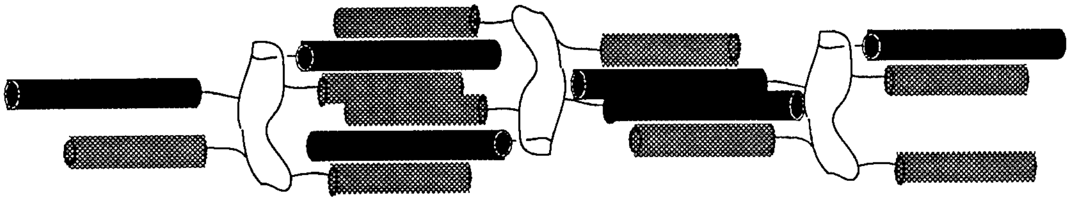
Introduction

Current research within the Laser Hardened Materials group at the Wright Laboratory(WL/MLPJ) has an effort directed towards the development of novel optical structures derived from a number of different polymer backbones including polypeptides, linear polysiloxanes and cyclic siloxanes. These materials are being developed for use as matrices for nonlinear optical chromophores. The ultimate use of these optically non-linear materials will be for protection of optical senors and personnel from laser weaponry as well as for optical signal processing (photonics).

As part of this research effort the author was involved, during his stay as a Summer Research Scientist, in the synthesis and characterization of novel chiral mesogens for use in cyclic siloxanes liquid crystals (CLC's). These materials can be fabricated into fibers and films in spite of their relatively low molecular weight.



This is thought to be due to a microstructure in which the attached mesogens (rigid groups capable of conferring liquid crystalline order) are packed in an interdigitated fashion between adjacent molecules .



Owing to the statistical substitution of mesogens, the system is amorphous in the solid state, but exhibits liquid crystalline behavior prior to melting.

We are interested in further examining the nature of the microstructure of these novel materials and how changes in molecular structure of the mesogens affect the packing and ordering of these materials. We are also interested in understanding how the incorporation of bound and free nonlinear optically active chromophores will effect packing and hence optical clarity. As there are a great many subtle changes that can be effected though changes in the chemistry of these materials, it is of great importance to develop good models for the nature of packing in the solid state. Once suitable models are in hand, these can be used to predict which molecular changes will be most effective in generating useful solid state structures. Synthesis and characterization of suitable model systems can further confirm the validity of these packing models.

Since these are relatively large molecules (from a computational point of view), the ability to examine a sufficient number of them to simulate the bulk state is computationally quite difficult. One way around this is to confine the molecules to packings that can simulate those most likely to be observed experimentally. Specifically, the packing of the currently known cyclic siloxanes seems to indicates a periodic structure with the interdigitation of the mesogens between adjacent molecules. This is consistent with the material's ability to produce long glassy fibers without the aid of molecular entanglement found in linear high polymers. If one could constrain the molecules to such a linear arrangement, then the details of the interdigitation could be more accurately modelled. When modelled as a single "chain" of CLC's, the models tend to deform in a manner that is not consistent with the experimental observations· This is chiefly due to the absence of an outer "shell" of chains confining the chain under examination. One way to accomplish is to use a full bulk description, with the chain of interest surrounded be a shell of fixed chains. This would still be computationally very costly. Alternatively, within the molecular mechanics formalism used for modelling one could incorporate a potential energy term which would constrain the chain in a cylindrical potential well or "pore", simulating the constraining effect of a shell of fixed chains. This term would be added to the existing potential energy terms due to bond stretching forces, torsional barriers, bending forces, and as well as those for atomistic long range interactions. This would remove the need for the calculation of a large number of chain-chain interaction terms which a full bulk model would require. Additionally, where experimental data is available, we can further impose other potential energy terms which will constrain the periodicity of the molecules to be that observed. The only variable would then be the interaction of the mesogens between adjacent units. This would again simplify the calculation procedure and lead to an intimate understanding of the nature of these interactions. These codes are readily added to existing commerical molecular modeling programs, alleviating the need to develop large complex programs.

Method

Fortran subroutines were developed to be incorporated into existing commerical molecular modeling software. To accomplish this the following hardware and software was purchased. We procured an IBM RS/6000™ Model 320H configured with 32 Mbyte of main memory and 800 MByte of disk storage. The system was also equipped with a 3D rendering graphics rendering engine. The system was tied into the University Ethernet™ system for ease of communication with the staff at WL/MLPJ. We chose the PolyGraf™ molecular dynamics/mechanics package for several reasons. The PI has had considerable experience with this package. Additionally, is it readily modified to include other potential energy terms into the basic molecular mechanics potential energy calculation. The potential energy terms utilized to create the cylindrical confinement was a soft repulsive wall of the functional form

$$E = A(r-r_0)^2$$

were r_0 is the radius of the "tube", r is the distance of the atom from the axis of the cylinder and A is the "strength" of the repulsive force. Additional subroutines were incorporated to allow the user to change the value of r and A prior to a simulation either in an interactive fashion or as part of a "macro". Although not utilized at this time, codes were also introduced to allow for a rectinlinear confinement. This could be utilized to model mechanical deformation or to observe the behaviour molecules trapped between two repulsive walls.

Testing of the suboutines has involved performing simulations of polyethylene chains confined in cylinders whose pore diameter is the same as that of a urea clatharate. Mattice and co-workers have performed similar simulations, but at full bulk density, that is, incorporating all the atoms of the clatharate crystal. Analysis of our trajectory files yield the same type of temperature behavior in both the full and the confined simulations. Based on the amount of time needed for the two forms of calculation, the confinement-based simulation ran some 5-10 time faster than the fully atomistic calculation. With this approach validated calculations are currenltly underway to study the packing behavior of the siloxane ring systems.

In addition to the molecular dynamics studies, we have utilized *ab initio* molecular orbital theory at the 3-21G and 6-31G* level to study the torsional potentials of diphenyl ethanes and benzyl phenyl ethers as models for flexible mesogens. The studues were conducted using the Spartan™ (Wavefunction, Inc.) package. Our studies indicate that the potential energy barriers are not significanly different from those of ethane and ethyl methyl ethers, indicating that the mesophase present in thises materaisl is due to intramolecular interactions rather than some special rigidity of these mesogens.

Finally, we have ported the public domain GAMESS and ZINDO packages to the system and are currently utilizing these programs for the calculation of

NLO and spectroscopic properties of potential target molecules used in the synthesis of frequency doubling polymer and liquid crystals.

Analytic Models for the Detection and Interception of Low Probability of Intercept (LPI) Communication Signals

Glenn E. Prescott
Associate Professor
Department of Electrical & Computer Engineering

The University of Kansas
1013 Learned Hall
Lawrence, KS 66045

Final Report for:
Research Initiation Program
Wright Laboratory (WL/AAAI)

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

December 1992

Analytic Models for the Detection and Interception of Low Probability of Intercept (LPI) Communication Signals

Glenn E. Prescott
Associate Professor
Department of Electrical & Computer Engineering
University of Kansas

Abstract

The communications engineer involved in the design of low probability of intercept (LPI) communication systems needs to be able to evaluate the detectability of a specific waveform design against a variety of intercept receivers. Because of the complex structure and wide bandwidth associated with LPI communication signals, intercept receivers usually have no other alternative than to resort to energy detection techniques when attempting to locate emitters. Unfortunately, conventional energy detectors (radiometers) are inherently nonlinear, and analysis of these systems can seldom be accomplished in closed form. Therefore in order to facilitate the evaluation of detectability for various LPI waveforms, we have developed a computer aided design tool, which is reported on here. This software tool is called Low Probability of Intercept Signal Detectability Analysis, or LPI/SDA. The program allows engineers to specify the LPI signal design and the candidate intercept receiver threat. A minimum signal detectability analysis is then computed for a given required level of intercept receiver performance. The program also performs a system quality factor analysis so the engineer can examine performance tradeoffs using a variety of LPI techniques, selecting the ones that are most cost effective in providing an effective covert communication system.

Analytic Models for the Detection and Interception of Low Probability of Intercept (LPI) Communication Signals

Glenn E. Prescott

1. Introduction

Recent emphasis in military communication systems has focused on the vulnerability of communication signals to interception. While in many instances, an anti-jamming capability is an essential feature for military communication systems, there are many situations in which communications covertness is more important. For example, the requirement for covert operation of military aircraft has led to the reduction of aircraft signatures in order to minimize aircraft detectability. One of the most critical aircraft signatures in this environment is its communication signals. Therefore, the emphasis on reduced aircraft detectability drives the requirement to limit the interceptability of its communication signals. The result is a low probability of intercept (LPI) communication system.

The characteristics of a communication system which is invulnerable to jamming are quite similar to those of an LPI communication system. The one notable difference is in the received signal-to-noise ratio. For an effective anti-jam (AJ) capability, large receiver signal-to-noise ratios and plenty of excess signal margin are desired. LPI communications, on the other hand, requires the minimum received signal-to-noise ratio necessary to provide the minimum level of acceptable performance.

1.1 LPI Signal Exploitation

Military RF communication systems must necessarily provide a high level of security against the exploitation of transmitted information by an unintended listener. This exploitation could be as simple as detecting the presence and location of a communications platform, or as complex as extracting the information contained in a transmitted signal. Nicholson [1] describes four sequential operations that exploitation systems attempt to perform:

1. Cover the signal – that is, a receiver is tuned to some or all of the frequency intervals being occupied by the signal when the signal is actually being transmitted.

2. Detect the signal – that is, make a decision about whether the power in the intercept bandwidth is a signal plus noise and interference or just noise and interference.

3. Intercept the signal – that is, extract features of the signal to determine if it is a signal of interest or not.

4. Exploit the signal – that is extract additional signal features as necessary and then demodulate the baseband signal to generate a stream of binary digits.

The probability that an interceptor can exploit an unknown communication signal is defined as Pr(E), which is given as:

$$Pr(E) = Pr(E|I)\, Pr(I|D)\, Pr(D|C)\, Pr(C)$$

where Pr(E|I) is the probability of exploitation given that the signal can be intercepted, Pr(I|D) is the probability of intercepting the signal given that it can be detected, Pr(D|C) is the probability of detecting the signal given that the signal is covered, and Pr(C) is the probability that the signal is covered. Everything that an unintended listener could conceivably want to do with a signal depends critically on having the ability to cover and detect the presence of the signal. Any subsequent actions are dependent upon signal detection.

Military communication system designers have traditionally employed spread spectrum waveforms to achieve covertness in a transmitted signal. These spread spectrum signals, in addition to permitting the use of code division multiple access (CDMA) for efficient bandwidth utilization, also incorporate significant anti-jam (AJ) and low probability of intercept (LPI) characteristics due to their low-level radiated power densities. The term "LPI" is used here as it is in much of the literature (e.g., [2]), although LPI signals are perhaps better described as low probability of detection (LPD) signals. LPI will be used in this report to describe signals which are difficult for an unintended receiver to detect.

The communications receiver in an LPI communication system possesses knowledge of the code which was used at the transmitter to spread the signal, and thus can de-spread the received signal by re-mixing it coherently with the code. This de-spreading operation allows the receiver to filter out a large portion of the noise power present within the spread bandwidth at the receiver front-end. An unintended receiver does not typically have knowledge of this spreading code and must make signal present decisions based solely on the received energy in some frequency band over some period of time. Furthermore, because the unintended receiver lacks the ability to de-spread the signal, it is unable to filter any of the noise power within the spread bandwidth. Receivers which make binary signal present decisions based on energy detection are called radiometric systems (radiometers), and represent the most common detection threat to LPI signals.

The inherent vulnerability of an LPI spread spectrum signal to detection by a particular radiometric system can be quantified in terms of the required carrier signal power to one-sided noise power spectral density ratio C/N_o required at the front end of the radiometer to achieve a specified probability of detection P_d and probability of false alarm P_{fa} performance level [3]. The LPI communication system designer uses this detectability information to select the spread spectrum modulation type and parameters to yield a signal which is minimally detectable by the most likely detection threat, in this case a particular type of radiometer system.

Analytical models have been developed which map the radiometer performance probabilities to the required front-end C/N_o. In this report we will develop several of the important analytical models for radiometric intercept receivers and use these to evaluate the detectability of LPI communication waveforms. We will also use these models to obtain a performance metric for the LPI communication system.

1.2 LPI System Evaluation

In order to evaluate the potential effectiveness of LPI communication systems, a common criteria is needed to aid in assessing the strengths and weaknesses of proposed techniques. Therefore, quality factors have been developed for LPI communication systems to provide a single, unified quantitative technique which allows the system engineer to evaluate LPI effectiveness in the presence of jammers and intercept receivers. We will concentrate on developing system quality factors for a Low Probability of Intercept (LPI) communication systems; and on describing a methodology for employing these quality factors for a variety of scenarios and systems.

The LPI system quality factors derived in this report originate from the system link equations which describe the signal and interference power gains and losses as a function of path losses, antenna gains, modulation type and interference rejection capability for any given scenario. Quality factors are developed for all major components of the LPI system which can provide some advantage to the cooperative transmitter and receiver over the jammer and intercept receiver.

2. LPI Techniques

An LPI communications capability for military communication systems is provided via an assortment of technologies and techniques. Many of these are briefly described below:

Power Control – Transmit power is increased until the receiver acknowledges reception. A feedback control link is required to adjust the transmit power to the minimum necessary for reliable communications.

Beam Pointing – Highly directional antennas are employed at the receiver and transmitter. Automatic tracking is required to maintain the received signal level, but spatial dispersion of the radiated signal energy is restricted.

Null Steering Antenna – If the receiver's antenna can place a null in the direction of a jammer, then less power will be needed from the friendly transmitter and thus it will be less detectable.

Low Sidelobe Antenna – Directional antenna radiates small amounts of power in directions other than the desired direction. However, to reduce spatial dispersion of signal energy to an absolute minimum, antenna sidelobes should be suppressed.

Frequency Control – Automatic selection of operating frequencies or frequency bands. For example, choose one transmit frequency for the near receiver, another transmit frequency for the distant receiver (e.g., 60 GHz which is highly attenuated, and 54 GHz which propagates further). This technique also includes hopping over several bands, such as HF, VHF, UHF, and L band.

Bandwidth Compression – Applies primarily to voice communications, but represents any technique which reduces the number of bits per second required for any given transmission. This means the receiver will require less signal strength since each bit can be processed longer.

Spread Spectrum Modulation – Using frequency hopping, phase shifting, time hopping or their combinations to spread the energy over a band of frequencies and reduce the power density. This makes the transmission less detectable.

Error Correction Coding – Error correcting codes reduce the signal energy required to maintain a specified level of receiver performance. Any technique which trades power for bandwidth can be used to enhance the LPI performance of a communication system.

Interference Suppression/Excision – The communication receiver employs a filter that can automatically (or adaptively) place a null at the frequency of a jammer. Therefore, less power will be needed from the communication transmitter to maintain a specified receiver performance level.

Signal Masking – The communications transmitter intentionally transmits at the same frequency as another radiator but with slightly less power. An intercept receiver will detect the stronger signal. However, the communications receiver will have processing gain and be able to detect the weaker transmission while rejecting the stronger one.

When signal exploitation (i.e., recovery of information from the transmitted signal) is required, the intercept receiver will be operating at a disadvantage to the communications receiver. The communications receiver can employ coherent processing on the spreading code, but the intercept receiver must rely on noncoherent processing, which means that the communications receiver needs less signal power to accomplish its primary task than does the interceptor. On the other hand, when signal detection or interception is required, the interceptor has the advantage in that it only needs to determine the presence of the signal (detection) or extract some characteristics or features from the signal (interception). Since information is not being recovered from the signal, then less received signal power is required for the intercept receiver to accomplish its job.

The effect of each of the techniques discussed above can be observed and evaluated by examining a communication system and intercept receiver in an operational environment, as described below.

2.1 The LPI Scenario

A typical LPI scenario is illustrated in Figure 1, representing any situation in which a cooperative transmitter and receiver are targeted by jammers – which disrupt the communications receiver, and intercept receivers – which attempt to detect and exploit the transmitted signal. Since all players are likely to be present in any realistic situation, both the communications receiver and the intercept receiver must be able to function in the presence of jamming.

The objective of any LPI communication system is to successfully conduct communications between a cooperative transmitter and receiver in such a manner as to minimize the probability of interception by an unauthorized receiver. The communication system is assumed to have a variety of techniques available for reducing the probability of interception. For example, the transmitter may employ steerable high gain antennas and emit a signal with low power density and large time bandwidth product. The communications receiver may possess null steering antennas, adaptive interference suppression filters, and employ coherent processing. On the other hand, the intercept receiver has similar available technologies - steerable antennas with low side lobes for eliminating inadvertent (or intentional) jamming, and adaptive filters for excising narrowband interference, for example.

The principle players in the LPI scenario each have critical performance parameters which can be easily evaluated and compared. For example, the communications transmitter is roughly characterized by its transmission power, type of modulation and antenna gain. The communications receiver characteristics are generally defined based on some minimum received bit energy per noise power density ratio, E_b/N_o at the receiver input required to provide some acceptable bit error performance, P_e. Receiver antenna gain, bandwidth and noise figure are also important parameters to be considered, all of which together determine the maximum communication range possible for a given transmitter/receiver pair.

The adversaries to the transmitter and receiver are the interceptor and jammer, respectively. The intercept receiver is typically a non-coherent energy detector whose performance is established by its probability of detection, P_d and probability of false alarm, P_{fa}. More sophisticated intercept receivers employing signal feature detectors may be employed, but they generally require a larger input signal to noise ratio for a given P_d and P_{fa}. Intercept receiver performance will also be influenced by the choice of modulation used by the communicator, the feature in the received waveform that is detected, the interceptor antenna characteristics, intercept receiver bandwidth and noise figure. For the communicator, LPI techniques must be effective in minimizing the maximum range between the communications transmitter and the intercept receiver – that is, requiring the interceptor to come unacceptably close to the transmitter in order to achieve signal detection.

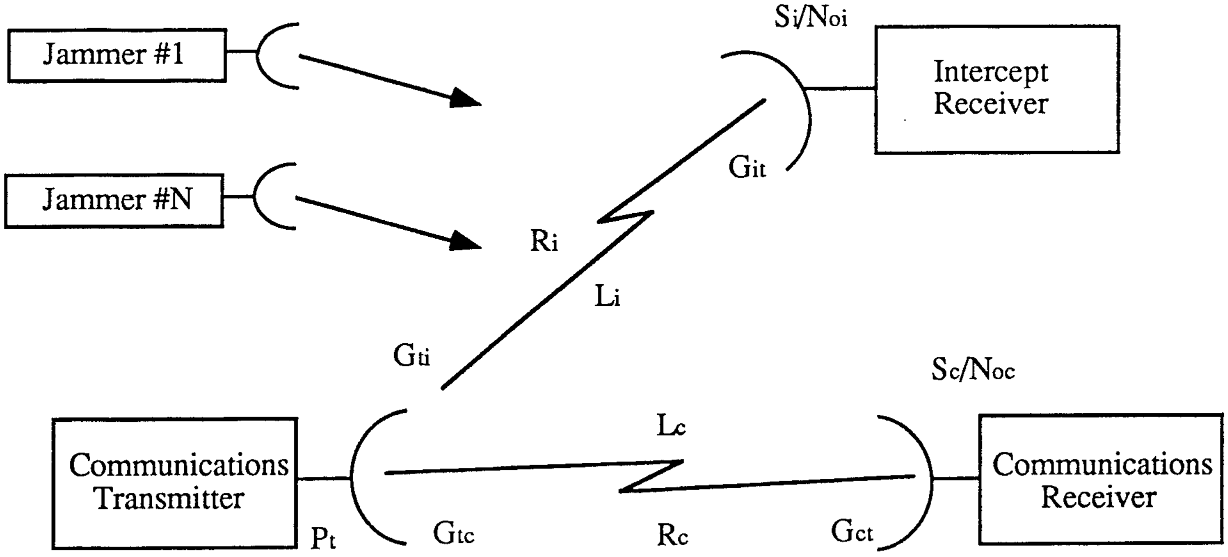


Figure 2-1: LPI Scenario

Another critical factor in the evaluation of LPI communications effectiveness is the jammer. The jammer targets the receiver, therefore the essential parameters in this case are the jammer operating frequencies, antenna gains and the jammer transmitted power spectral density - the amount of power that can be distributed over the operating frequency range of the communications receiver. The jammer can also impair the ability of the intercept receiver to detect the target signal. Therefore, the interceptor may require similar interference suppression techniques as used by the communication receiver.

We can analyze the relationships among these players and reveal potential trade-offs that may exist by performing a simple link analysis. The link analysis reveals strengths and vulnerabilities, and provides the system designer the insight to determine how to most effectively concentrate system resources.

2.1.1 The Communications Link

We begin by determining the ratio of signal power to noise power density, S_c/N_{sc} available at the communications receiver. Noise power density is of interest here instead of noise power so that the receiver bandwidth can remain unspecified. This will be important later when comparisons are made between the performance of intercept receivers versus communications receivers.

We can always assume that there is some performance requirement imposed on the communications receiver, expressed in terms of bit error probability. This requirement will dictate some minimum E_b/N_{sc} in order to conduct communications at some acceptable quality, P_e. Therefore, the required signal energy for a given receiver performance criterion can be expressed as a function of the received signal power as,

$$\frac{S_c}{N_{sc}} = M \frac{(E_b R_b)}{N_{sc}} \tag{2-1}$$

Where M is the communications link margin ($M > 1$), E_b is the bit energy and R_b is the transmitted bit rate. Normally the communication link will operate with some margin, M to assure quality and availability, often due to unanticipated atmospheric effects. The signal power available at the detector input of the communications receiver is obtained by performing a simple link analysis on the communication system based on the parameters illustrated in Figure 2-1:

$$S_c = \frac{P_t\, G_{tc}\, G_{ct}}{\left(4\pi R_c/\lambda\right)^2 L_c} \tag{2-2}$$

where in the numerator,

P_t	-	transmitter power
G_{tc}	-	gain of transmitter antenna in direction of the receiver
G_{ct}	-	gain of receiver antenna in direction of the transmitter

The denominator terms represent losses, where $(4\pi R_c/\lambda)^2$ is free space propagation loss and L_c accounts for atmospheric losses due to rain, water vapor and oxygen absorption. Expressing L_c in terms of propagation path length and path attenuation (expressed in decibels),

$$L_c = log^{-1}\left(\frac{1}{10}\, \xi_c R_c\right) \tag{2-3}$$

For simplicity, this loss is assumed to be linear as a function of range so that ξ_c is a generalized average loss factor expressed in units of dB/Km, for example. Such loss factors are not truly linear since the transmission medium is not homogeneous and will vary with season, temperature, time of day and altitude. However, the cumulative average effect is represented here by ξ_c.

To establish the pre-detection signal-to-noise ratio at the receiver input, the power spectral density of the interference, N_{sc} at the communications receiver (which is composed of the thermal noise power density, N_{oc} plus jammer power density, N_{Jc}), can be expressed as:

$$N_{sc} = kT_{ac} + kT_o(F_c - 1) + \sum_{n=1}^{N} \sum_{m=1}^{M} g_{cn}\, g_{cm} \frac{J_{nmc}}{B_c} \tag{2-4}$$

where,

$$N_{oc} = kT_{ac} + kT_o(F_c - 1) \tag{2-5}$$

$$N_{Jc} = \sum_{n=1}^{N} \sum_{m=1}^{M} g_{cn}\, g_{cm} \frac{J_{nmc}}{B_c} \tag{2-6}$$

and,

k	-	Boltzmann's constant
B_c	-	communications receiver bandwidth
T_{ac}	-	communications receiver antenna temperature
T_o	-	thermal noise (290° K)
F_c	-	communications receiver noise figure

The first term in N_{oc} represents noise at the antenna output, while the second term indicates the receiver sensitivity as expressed by the noise figure. The double summation component, N_{Jc} accounts for the power spectral density in each discrete frequency component transmitted by each jammer through the parameter J_{nmc}. The subscript n represents the jammer from the nth direction of arrival. The communication receiver is assumed to be jammed by N jammers originating from N unique directions (angles-of-arrival). The subscript m represents the mth frequency component from the nth jammer. Therefore, m is a frequency domain parameter which represents a spectral interfering tone much narrower in bandwidth than the desired signal.

The effects of the N jammers can be countered by null steering antennas (represented here by g_{cn}), and the m interfering tones can be eliminated by adaptive interference suppression techniques (represented here by g_{cm}). Both of these gain factors have the effect of reducing the interference level of the jammer by some amount. For example, with a 20 dB null in the direction of the nth jammer, $g_{cn} = .01$. In a dense jamming environment N_{sc} will be dominated by the double summation component, and $N_{sc} \approx N_{Jc}$; While in the absence of jammers the interference is primarily thermal, and $N_{sc} \approx N_{oc}$.

In summary, the communications link parameters which influence the ratio of received signal power to noise power density can be expressed as follows:

$$\frac{S_c}{N_{sc}} = M\,\frac{(E_b R_b)}{N_{sc}} = \frac{P_t\, G_{tc}\, G_{ct}}{L_c\, N_{sc}} \left(\frac{\lambda}{4\pi R_c}\right)^2 \tag{2-7}$$

Therefore, the maximum attainable range (while maintaining some specified minimum level of performance) for the communications link is:

$$R_c = \sqrt{\frac{P_t\, G_{tc}\, G_{ct}}{L_c\, N_{sc}} \left(\frac{\lambda}{4\pi}\right)^2 \frac{1}{M(S_c/N_{sc})}} \tag{2-8}$$

A similar development can be made for the link between the transmitter and the intercept receiver.

2.1.2 The Interceptor Link

The carrier power available to the intercept receiver is also a function of the link parameters:

$$S_i = \frac{P_t\ G_{ti}\ G_{it}}{\left(4\pi R_i/\lambda\right)^2 L_i} \tag{2-9}$$

where G_{it} and G_{ti} account for the gain of the transmitter antenna in the interceptor direction, and the gain of the interceptor receiver in the transmitter direction, respectively. R_i is the transmitter to interceptor path length and L_i accounts for the losses (other than free space loss) along that path. Also,

$$L_i = log^{-1}\left(\frac{1}{10}\,\xi_i R_i\right) \tag{2-10}$$

where ξ_i is the transmitter to interceptor path loss in dB/Km.

The interceptor must also operate within a hostile environment and be subject to the same types of interfering signals as the communications receiver. Therefore, the noise and interference at the intercept receiver can be expressed as,

$$N_{si} = kT_{ai} + kT_o(F_i - 1) + \sum_{n=1}^{N} \sum_{m=1}^{M} g_{in}\; g_{im} \frac{J_{nmi}}{B_i} \tag{2-11}$$

The antenna noise and receiver sensitivity (N_{oi}) are represented here, as are the interfering effects of jammers (N_{Ji}), where,

$$N_{oi} = kT_{ai} + kT_o(F_i - 1) \tag{2-12}$$

$$N_{Ji} = \sum_{n=1}^{N} \sum_{m=1}^{M} g_{in}\; g_{im} \frac{J_{nmi}}{B_i} \tag{2-13}$$

In this model, J_{nmi} represents the jammer power incident at the intercept receiver, and B_i is the interceptor bandwidth. Therefore J_{nmi}/B_i represents the jammer power density. As in the case described for the communicator. As with the communications link, the subscript n represents the jammer from the nth direction of arrival. The intercept receiver is assumed to be jammed by N discrete jammers each with a unique angles of arrival; and the subscript m represents the mth frequency component from the nth jammer.

To improve the effectiveness of the interceptor in a dense jamming environment, null steering antennas (represented by g_{in}) and adaptive interference suppression filters (represented by g_{im}) can be employed. Therefore, the intercept receiver can provide both spatial and spectral discrimination to gain an advantage over the communications system. Ideally, the combined effect of g_{in} and g_{im} is to whiten the interference environment.

For the interceptor link, some specified minimum acceptable P_d and P_{fa} will require a particular input signal-to-noise ratio for each type of intercept receiver. The interceptor link can be expressed as:

$$\frac{S_i}{N_{si}} = \frac{P_t\, G_{ti}\, G_{it}}{L_i\, N_{si}} \left(\frac{\lambda}{4\pi R_i}\right)^2 \tag{2-14}$$

This required S_i/N_{si} for some P_d and P_{fa} is reflected through the link equation resulting in some *maximum* intercept range (i.e., beyond this range, interception is unlikely), R_i:

$$R_i = \sqrt{\frac{P_t\, G_{ti}\, G_{it}}{L_i\, N_{si}} \left(\frac{\lambda}{4\pi}\right)^2 \frac{1}{(S_i/N_{si})}} \tag{2-15}$$

Taking the ratio of ranges, we can express the communications link versus the interceptor link as:

$$\left(\frac{R_c}{R_i}\right)^2 = \frac{G_{ct}\, G_{tc}}{G_{ti}\, G_{it}}\, \frac{L_i}{L_c}\, \frac{N_{si}}{N_{sc}}\, \frac{S_i/N_{si}}{S_c/N_{sc}}\, M \tag{2-16}$$

This expression establishes a starting point for developing LPI system performance metrics, or *quality factors*.

3. LPI System Quality Factors

In the evaluation of an LPI communication system, several parameters in (2-7) and (2-14) are not needed since we are interested only in the relative performance of the interceptor link with respect to the communication link. In particular, P_t and λ are not required since they are common to both links. We define LPI quality factors by first establishing performance requirements for both receivers - P_e for the communications receiver and P_d , P_{fa} for the intercept receiver. From these performance specifications some minimum S_c/N_{sc} and S_c/N_{si} are required. The required values of S_c/N_{sc} and S_c/N_{si} are then worked back through the LPI link equations until some ratio of ranges (LPI to interceptor range) can be identified. We begin this process by taking the ratio of required received signal powers to noise power densities required for a specified performance level, as described in (2-16) and repeated below:

$$\left(\frac{R_c}{R_i}\right)^2 = \frac{G_{ct}G_{tc}}{G_{ti}\,G_{it}} \cdot \frac{L_i}{L_c} \cdot \frac{N_{si}}{N_{sc}} \cdot \frac{S_i/N_{si}}{S_c/N_{sc}} M \qquad (3\text{-}1)$$

This expression for the ratio of ranges allows a comparison of the effectiveness of the LPI communication system in gaining an operating range advantage over the interceptor.

The essential parameters for evaluating the LPI system are now evident. With communication and intercept receiver performance requirements fixed, the system designer can select from a variety of LPI techniques (listed in the previous section) to determine the relative covertness of the communication system with respect to the interceptor, all of which may be operating in a hostile radio frequency environment. We can now define and discuss each quality factor element from the previous expression.

3.1 The LPI Quality Factor

The essential composite measure of quality for the overall LPI communication system is simply a function of the maximum allowable communication range and the maximum interceptor range required for detection. Expressed in decibels, the LPI quality factor is

$$Q_{LPI} = 20\ log\left(\frac{R_c}{R_i}\right) \qquad (3\text{-}2)$$

This expression indicates that any improvement in LPI effectiveness will either allow the communications system to operate over a longer range, or will require the intercept receiver to move closer to the transmitter in order to achieve some specified level of performance. Of course, we would like to maximize Q_{LPI} by requiring the interceptor to be unacceptably close to the target transmitter, or by allowing the receiver to operate effectively at some large range. We can control Q_{LPI} by designing LPI features into the communication system so that the individual terms in (3-1) are maximized. We can identify these individual terms as quality factors which relate to both the scenario (or environment) and the LPI communications system. Taken together, we can summarize the LPI quality factors as,

$$Q_{LPI} = 20\, log\left(\frac{R_c}{R_i}\right) = Q_{ANT} + Q_{ATM} + Q_{IS} + Q_{MOD} \tag{3-3}$$

where Q_{ANT} is the antenna quality factor, Q_{ATM} is the atmosphere (or propagation path) quality factor, Q_{IS} is the interference suppression quality factor, and Q_{MOD} is the modulation quality factor. These quality factors are described in detail below.

3.2 The Antenna Quality Factor

The antenna quality factor, expressed in decibels is

$$Q_{ANT} = 10\, log\left(\frac{G_{ct} G_{tc}}{G_{ti} G_{it}}\right) \tag{3-4}$$

and accounts for the advantage provided by steerable narrow beam antenna at the transmitter, receiver and interceptor. Therefore, G_{tc} , G_{ct} and G_{it} are designed to be large. On the other hand, the transmitter requires the transmitter sidelobe gain, G_{ti} to be as small as possible. Therefore, high gain, low sidelobe antennas at the communications transmitter and receiver keep the antenna quality factor large.

The LPI system has no control over G_{it}, which is the interceptor antenna gain. If G_{it} is large, the antenna quality factor will suffer. However, as the intercept receiver antenna gain increases, the beamwidth gets narrower and eventually the probability of intercept diminishes significantly due to missed coverage. The interceptor is now required to scan so slowly, in order to integrate enough energy in each of the directions of arrival, that a short transmitted pulse is likely to be missed. There is obviously a tradeoff in the amount of antenna gain the interceptor can effectively use with the time required to dwell on each look angle and the number of look angles to search.

3.3 The Atmospheric Quality Factor

The atmospheric quality factor is a function only of atmospheric effects and range. The atmospheric quality factor in decibels is,

$$Q_{ATM} = \xi_i R_i - \xi_c R_c \tag{3-5}$$

In most cases we are justified in making the assumption that the atmospheric path loses (excluding free space loss) for the communications link and the interceptor link are approximately equal, since the interceptor and the communications receiver are usually located close enough to one another that they operative under common atmospheric conditions.

3.4 The Interference Suppression Quality Factor

The interference suppression quality factor compares the ability of both the communications and intercept receivers to suppress, or minimize interference and noise. The interference suppression quality factor expressed in decibels is,

$$Q_{IS} = 10 \, log \left\{ \frac{kT_{ai} + kT_o(F_i - 1) + \sum_{n=1}^{N} \sum_{m=1}^{M} g_{in} \, g_{im} \frac{J_{nmi}}{B_i}}{kT_{ac} + kT_o(F_c - 1) + \sum_{n=1}^{N} \sum_{m=1}^{M} g_{cn} \, g_{cm} \frac{J_{nmc}}{B_c}} \right\} \tag{3-6}$$

When the receivers are successful in eliminating interference, Q_{IS} is dominated by the respective communications or intercept receiver noise figures and antenna noise temperatures. In most cases these filters are assumed to be adaptive interference suppression/excision filters which are intended to remove narrow band interference from the receivers before the demodulation or decision, and effectively "whiten" the interference environment. Any situation causing the intercept receiver to have more noise or interference, such as large noise figure or inability to null jammers, will cause

Q_{IS} to be large. On the other hand, if the interceptor can successfully null a jammer, the gain factor g_{in} becomes less than unity, reducing the received jammer power and reducing the quality factor. Likewise, the communications receiver's ability to null out and adaptively filter interference (accounted for by g_{cn}) provides an added advantage over the intercept receiver in the presence of jamming, thus increasing the quality factor.

3.5 The Modulation Quality Factor

The modulation quality factor, Q_{MOD} is one of the most significant of the figures of merit for the LPI communication system since it compares the quality of communications for some acceptable bit error rate P_e, with the quality of interception for some acceptable probability of detection, P_d and probability of false alarm, P_{fa}. Only the parameters of the signal and the method used to detect and demodulate the signal are important. Any factors causing the intercept receiver to require a larger signal to noise ratio to achieve a specified performance level increases the modulation quality factor. The modulation quality factor expressed in decibels is,

$$Q_{MOD} = 10\ log\left(\frac{S_i/N_{si}}{S_c/N_{sc}}\right) \tag{3-7}$$

Regardless of the scenario, the relative performance of the communications receiver compared to the intercept receiver is of critical importance. The communications receiver has the advantage of a priori knowledge of the signal waveform, which gives the communicator a significant signal processing advantage over the interceptor. On the other hand, since the interceptor only requires the detection of signal energy for a successful intercept, the interceptor requires far less signal information (and hence, less input signal-to-noise ratio) than does the communicator. The goal of the LPI system designer is to select a waveform that is effective for communication purposes but is relatively difficult for the intercept receiver to detect.

The denominator of the modulation quality factor, (Sc/N_{sc}) is calculated from the communication link parameters. The detection characteristics of intercept receiver are not a factor here. The generally applicable form of the calculation with all quantities evaluated in dB is [4]:

$$\frac{S_c}{N_{sc}} = \frac{E_b}{N_{sc}} + M + 10\ Log\,(R_b) - G_{code} \tag{3-8}$$

where,

E_b/N_{sc} is the theoretical ratio of uncoded bit energy to interference spectral density (in dB) required to attain the link's specified bit error probability.

G_{code} is the gain achieved by error-correction coding, evaluated in dB, at the link's specified bit error probability.

R_b is the burst bit rate, determined by the ratio of total bits (user information plus overhead) transmitted over some time interval to the duration of that interval.

M is the additional transmitter power margin (in dB) utilized in the communication link to compensate for short-term fading and other losses that the adaptive transmit power control cannot adjust rapidly enough to counteract.

At this point we can make some simplifying assumptions which will allow us to develop some insight into the use of the modulation quality factor. We will assume that the system is operating at the minimum input signal-to-noise ratio for some specified level of performance. This allow us to set $M = 1$. Furthermore, we assume that the adaptive interference suppression systems within the communications receiver and the intercept receiver are effectively whitening the background noise environment. In this case, both N_{jc} and N_{ji} are zero and S_c/N_{oc} and S_i/N_{oi} are the parameters of interest. Therefore, the input signal-to-noise ratio for the communication receiver can be expressed in normalized form as,

$$\frac{S_c}{N_{oc}} = \frac{E_b}{N_{oc}} R_b \tag{3-9}$$

Where E_b/N_{oc} is a function of the required P_e for a given modulation type and method of detection, which is defined as

$$\frac{E_b}{N_{oc}} = \zeta_c (P_e) \tag{3-10}$$

Therefore,

$$\frac{S_c}{N_{oc}} = \zeta_c\,(P_e)R_b \tag{3-11}$$

For the intercept receiver, the input signal to noise ratio is a function of P_d , P_{fa} as well as the integration time and bandwidth,

$$\frac{S_i}{N_{oi}} = \zeta_i\,(P_d, P_{fa}, T, W) \tag{3-12}$$

So that

$$Q_{MOD} = \frac{\zeta_i(P_d, P_{fa}, T, W)}{\zeta_c\,(P_e)\,R_b} \tag{3-13}$$

In order to use the modulation quality factor in an LPI analysis, we need to have the appropriate expressions for ζ_i and ζ_c. These will be discussed in the succeeding sections of this report.

4. Signal Detection Model

A simple detection model for LPI signals (such as suggested in [5]) is often useful in visualizing the complexities of the waveform and providing insight in how to best intercept it. The model employed here describes the frequency bandwidth versus time duration of the transmitted signal. It represents the transmitted waveform as a hierarchy of time-bandwidth units, proceeding from a course structure, which typically has a large degree of complexity (large time-bandwidth product), to progressively finer structures which eventually approach time-bandwidth products on the order of unity.

The JTIDS waveform is chosen as the candidate waveform for the present analysis. The JTIDS detection model employed here is based on the individual time slot (7.8125 msec.). Detection models can also be constructed at higher levels, based on the frame (12 sec.) or the epoch (12.8 min.). The JTIDS waveform detection model is illustrated in Figure 4-1. Further details on the JTIDS waveform structure are discussed in [6], and the time slot structure for both single and double pulse modes is described in Appendix A.

Note that as the structure of the waveform becomes more fine-grained, additional a priori information is required by the interceptor in order to implement the optimum intercept technique. The performance of the interceptor depends upon how much knowledge the interceptor has about the target signal before an intercept is attempted. Five levels of interceptor a priori knowledge are considered. At level 1 the interceptor has the least knowledge; and at level 5, complete knowledge of the waveform is assumed. In the case of the JTIDS waveform, we can roughly define these levels as follows [7]:

- Level 1 - The interceptor knows nothing about the signal.

- Level 2 - The interceptor has reasonable estimates of T1 and W1, as well as the transmission start and stop times.

- Level 3 - The interceptor knows T1 and W1, and the general time interva structure (as well as start and stop times), and has reasonable estimates on T

- Level 4 - The interceptor knows T1, W1, and T2 (as well as start and stop times); and has reasonable estimates on T3 and W3.

- Level 5 - The interceptor knows T1, W1, T2, T3 and W3.

We will assume the interceptor a priori knowledge to be at Level 3 for the analyses that follow.

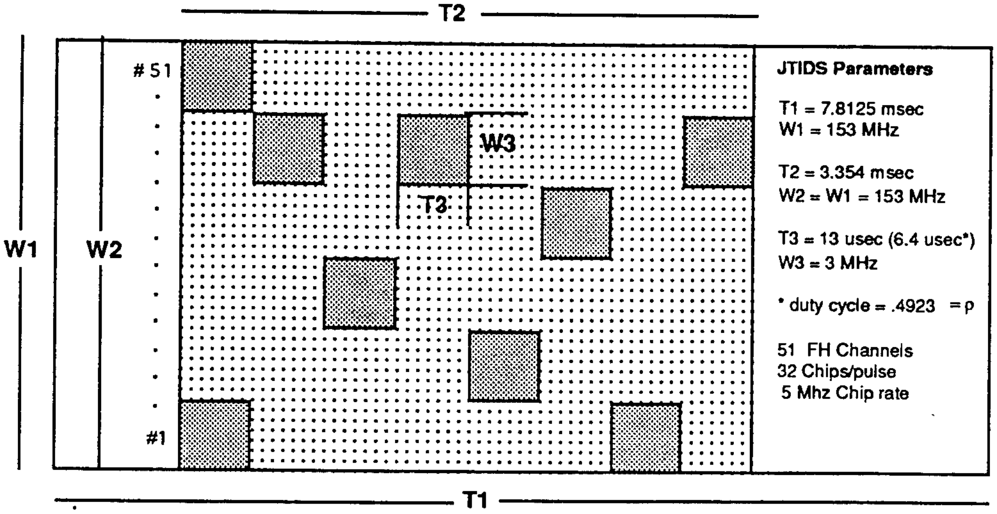


Figure 4-1: Detectability Model for the JTIDS Waveform

5. Nonlinear Intercept Receiver Models

The intercept receiver models considered in this analysis are assumed to perform some nonlinear operation (usually a squaring operation) on the received signal for the purpose of extracting signal energy, or some other waveform feature. We will assume that the interceptor has a priori knowledge at Level 3, but as usual, the spreading codes are unknown. We also assume that the received signal to noise ratio at the intercept receiver is small. This is usually a valid assumption since our LPI quality factor is based on the maximum communication range versus the maximum interception range. At the maximum interception range the signal to noise ratio will be very small. Under these conditions, (and assuming a suitably designed waveform) the optimum receiver has been shown to be a wideband total power radiometer [8]. If feature extraction is the goal of the interceptor, a higher signal to noise ratio is required resulting a smaller intercept range. This causes the LPI quality factor to increase significantly.

For signals having large time-bandwidth products, the output statistics of the radiometer can be assumed to be gaussian, and the detector performance can be completely characterized by the detectability factor δ, which has been defined at the square of the difference in the means of the output densities under noise and signal plus noise conditions [3]. The detectability factor δ, is a measure of the post detection, or output signal-to-noise power ratio of the detector.

$$\delta = \left[Q^{-1}(P_{fa}) - Q^{-1}(P_d)\right] = \frac{S_i}{N_{oi}\, W_i} \qquad (5\text{-}1)$$

Where W_i is the intercept receiver bandwidth. Five nonlinear radiometer models are assumed for use in the present LPI analysis. These models were suggested for this application in [1]

Wideband Radiometer Type	$\zeta_i\,(P_d, P_{fa}, T, W)$
Total Power Radiometer	$\delta\sqrt{\frac{W_1}{T_1}}$
AC Radiometer	$\sqrt[4]{\frac{2\delta^2 W_1^2}{T_1 T_3}}$

Pulse Rate Detector	$\frac{\delta}{\rho}\sqrt{\frac{2W_1}{T_2}}$
Hop Rate Detector	$\sqrt[4]{\frac{14.36\,\pi\,\delta^2 W_1^2}{T_2 T_3\,\rho}}$
Chip Rate Detector	$\delta\sqrt{\frac{W_1}{T_2}(316.2)}$

From these models it is apparent that the total power radiometer requires the least signal power for a successful intercept. Since the total power radiometer is the optimum receiver for detecting the presence of an unknown signal in a gaussian noise environment, we can use it as the standard against which other radiometers and feature detectors are measured.

6. Communication Receiver Models

For the communications receiver, the performance criterion is the probability of bit error. The probability of bit error can be related to the received E_b/N_{oc} for any particular type of waveform modulation and detection process. This relationship is expressed in the parameter $\zeta_c(P_e)$. Several popular modulation techniques are shown in the table below with the corresponding $\zeta_c(P_e)$:

Modulation Type	$\zeta_c(P_e)$
Noncoherent Binary FSK	$-2Ln\,(2P_e)$
Differentially Coherent Binary PSK	$-Ln\,(2P_e)$
Coherent Binary & Quadrature PSK	$\frac{1}{2}\left[Q^{-1}(P_e)\right]^2$

The JTIDS waveform employs 32-ary orthogonal noncoherent signaling with minimum shift keying as the modulation at the chip level. The expression for the probability of bit error for this modulation technique, which is shown below, is not easily expressed in the form specified for $\zeta_c(P_e)$:

$$P_e = \frac{1}{62} \sum_{n}^{32} \binom{32}{n} (-1)^n \, exp\left[-5 \frac{E_b}{N_o} \left(\frac{n-1}{n}\right)\right] \tag{6-1}$$

However, the expression can be inverted iteratively, or curves can be used, such as those shown in Appendix B.

7. Computer Aided Analysis

A low probability of intercept signal detectability analysis program has been developed [10] which calculates the detectability of certain spread spectrum signals by radiometric detection systems. This program assumes a model for the fundamental radiometer and then allows the computation of the minimum signal levels required for a particular level of intercept receiver performance. This program is entitled *Low Probability of Intercept Signal Detectability Analysis (LPI/SDA).* LPI/SDA is an efficient and highly accurate computer aided analysis tool to assist the LPI communications design engineer in determining the detectability of certain spread spectrum waveforms by radiometric detectors. Using this tool, the designer can construct a signal waveform which is least vulnerable to detection by the most likely detection threat. Further, the designer can analyze the covertness of the LPI signal in terms of scenario dependent and scenario independent factors via calculation of five different Quality Factor. LPI/SDA's results have been shown to be within .2 dB of similar analytical results presented in [1] and [5].

7.1 General Concepts in Computer Aided LPI Signal Analysis

Military communication system designers have traditionally employed spread spectrum waveforms to achieve a particular level of transmitted signal covertness. These spread spectrum signals, in addition to permitting the use of code division multiple access (CDMA) for efficient bandwidth utilization, also incorporate significant anti-jam (AJ) and low probability of intercept (LPI) characteristics due to their low-level radiated power densities.

The receiver in an LPI communication system possesses knowledge of the code which was used at the transmitter to spread the signal, and thus can despread the received signal by mixing it coherently with the code. An unintended receiver does not typically have knowledge of this spreading code and must make signal present decisions based solely on the received energy in some frequency band over some period of time. Receivers which make binary signal present decisions based on energy detection are called radiometric systems (radiometers), and represent the most common detection threat to LPI signals.

The inherent vulnerability of an LPI (spread spectrum) signal to detection by a particular radiometric system can be quantified in terms of the required carrier signal power to one-sided noise power spectral density ratio, C/N_{0I} required at the front end of the radiometer to achieve a specified probability of detection, P_d, and probability of false alarm P_{fa} performance level [5, 11]. The LPI communication system designer uses this detectability information to select the spread spectrum modulation type and parameters to yield a signal which is minimally detectable by the most likely detection threat, in this case a particular type of radiometer system.

Many researchers have developed analytical models which map the radiometer performance probabilities to the required front-end C/N_{0I}, e.g. [5]. LPI/SDA is a PC-based computer aided analysis system which automates the use of these models. The LPI communication system designer interacts with LPI/SDA's user-friendly hierarchical interface to describe a signal and radiometer system, and quickly determines the required C/N_{0I}. Signal and radiometer parameters can be easily modified to evaluate the effects of these changes on the required C/N_{0I}. LPI/SDA also has the ability to calculate five different LPI system Quality Factors. These Quality Factors describe the detectability of a signal separately in terms of scenario dependent and scenario independent factors [12].

7.2 LPI Signal Models

LPI/SDA models three standard and four hybrid types of spread spectrum signals. They are:

- Direct Sequence (DS)
- Frequency Hopped (FH)
- Time Hopped (TH)
- FH/DS
- TH/DS
- FH/TH
- FH/TH/DS

From an energy detection standpoint, these spread spectrum signals can be described in terms of relatively few parameters. These parameters are listed below and illustrated in the time-frequency plane shown in Figure 7-1. Note the similarity between this figure and Figure 4-1, which is the JTIDS message time slot structure. The notation for signal parameters shown in Figure 7-1 has been chosen to match that used in [5], since many of the radiometers modeled are taken from this reference.

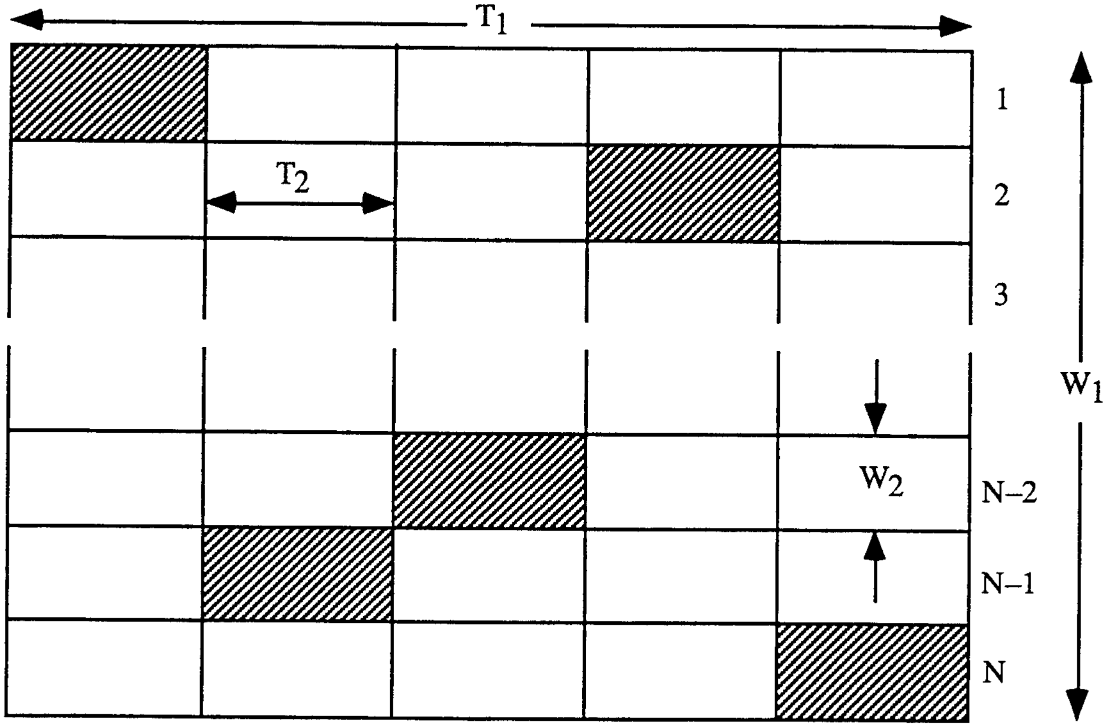


Figure 7-1: LPI/SDA Signal Parameter Notation

T1 — message duration (sec)
W1 — spread spectrum bandwidth (Hz)
T2 — pulse duration (sec)
b2 — number of pulses in T1
N — # of frequency hop bands in W1

To describe a DS signal, for instance, one would set N = 1, b2 = 1, and T2 = T1. T1 and W1 would be set to the desired message time and spread bandwidth, respectively.

7.3 Radiometer Models

The heart of all radiometric systems which LPI/SDA models is the wideband radiometer, also known as an energy detector or total power radiometer. This system, shown in Figure 7-2, filters a portion of the RF spectrum, squares this filtered signal to obtain signal power, and integrates from t - T to t to yield signal energy (typically this integration is implemented as integrate and dump rather than continuous integration). This signal energy is then compared to a threshold and, if the threshold is exceeded, a signal is declared present; otherwise no signal is declared present. Assuming ideal signals and filters, the wideband radiometer can equivalently be described as a system which observes a rectangular time-frequency *cell* with bandwidth equal to the bandpass filter bandwidth and time interval equal to the integration time *T*. It measures the total signal plus noise energy received in this cell and compares this energy to a threshold. A signal is declared present if the cell energy exceeds the threshold.

For signals which completely occupy a single time-bandwidth cell (e.g., DS signals) and are embedded in stationary white Gaussian noise, total power radiometers represent essentially the best performing detection systems which can be constructed [8]. If a signal is pulsed in time and/or frequency, the interceptor may be able to improve his detection performance significantly if he has knowledge of the pulse positions in both time and frequency and exploits this knowledge by using an appropriate radiometer system. These radiometer systems consist of one or more total power radiometers, each with a bandwidth and integration time matched to the time-bandwidth dimensions of a pulse. The binary signal present/not present decisions that each of these radiometers makes are processed in some manner to minimize the false alarm probability while maintaining a high detection probability. These systems are described further in [5].

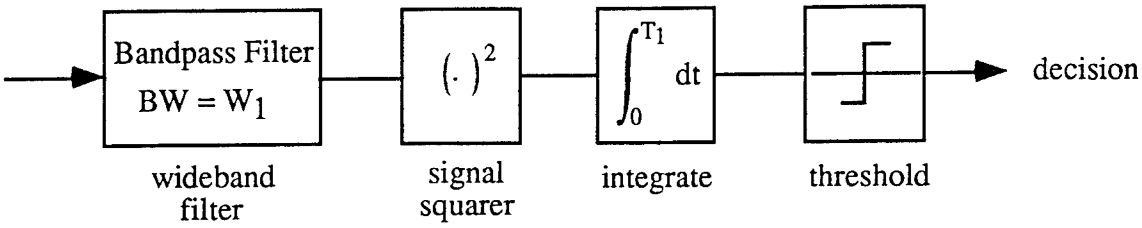


Figure 7-2: Wideband Radiometer System

7.4 The LPI/SDA Program

LPI/SDA contains a highly accurate analytical model for calculating the required C/N_{0I} at the front end of an intercept receiver given the time-bandwidth product *TW* of its observation cell and the desired P_d and P_{fa}. For pulse detection radiometer systems, the overall detection and false alarm probabilities can be mapped to single cell detection and false alarm probabilities, Q_d and Q_{fa}, and thus this detectability model is used for all types of radiometer systems.

The radiometer model is based on the following. When only noise is present at the input to a wideband radiometer, the output follows a chi-square probability density function (PDF) with $2TW$ degrees of freedom. With signal present, the output has a noncentral chi-square PDF with $2TW$ degrees of freedom and a noncentrality parameter of $2E_S/N_{01}$, where E_S is the signal energy received during a time interval of length T [11].

For large TW products, the output statistics approach Gaussian density functions for the noise and signal plus noise cases. Simple detectability models make this approximation and often further assume that the variance of the noise is equal to the variance of the signal plus noise (i.e., the signal is very weak). After signal detectability is calculated, a correction factor is typically applied to correct for the error introduced by making the Gaussian assumption (see for instance, [13 pg. 298]).

LPI/SDA calculates the required C/N_{01} using chi- square statistics except in the case of large TW products (> 500) where the Gaussian approximation is very good, and makes no assumption whatsoever about the signal and noise powers. A typical detectability calculation in LPI/SDA proceeds as follows. The user describes a signal and radiometer, and specifies a TW product, P_d and P_{fa} LPI/SDA maps P_d and P_{fa} to Q_{fa} and Q_d if necessary. The radiometer output PDF, given that the input is noise alone, is now fixed, and a detection threshold can be calculated by using a chi-square tail function routine to yield the correct Q_{fa}. With knowledge of this threshold, the noncentrality parameter of the output PDF given signal plus noise at the input can be varied until the correct Q_d is obtained. Q_d is calculated using a recursive generalized Q-function algorithm given in [14]. With the radiometer output PDFs determined, the noncentrality parameter of the noncentral chi-square PDF $(2E_S/N_{01})$ is multiplied by $1/2T$ to yield the required C/N_{01}. LPI/SDA modifies this number appropriately if the radiometer dimensions are not perfectly matched to the transmitted signal. This would be the case if, for example, a DS signal is specified along with a wideband radiometer whose bandwidth is something less than the spread bandwidth of the signal. For TW products greater than 500, the calculation proceeds similarly, except that the output PDFs are assumed to be Gaussian with the first and second moments equivalent to the corresponding chi-square moments, and thus Gaussian tail probabilities are used rather than chi-square.

7.5 Sample Detectability Calculations

The LPI/SDA user interface contains independent input pages for describing the spread spectrum signal and radiometer. Figure 7-3 illustrates the structure of the user interface and the paths which one may follow when going from one input page to another. On each of these pages, the user selects a signal or radiometer type via pop-up menus, and specifies the parameters that accompany the selected type. Another page allows the user to specify P_d and P_{fa} and calculate the requiredC/N_{01}. On yet another page, the user may input communications link information (path losses, antenna gains, receiver noise temperatures, etc.) which are required for Quality Factor calculations. Values for the five Quality Factors can be computed, and one of these Quality Factors, the Modulation Quality Factor, can be plotted against one of six system parameters, including P_d and P_{fa}. The user may move freely from page to page to change the signal type, radiometer type, and associated parameters and find immediately the effect of these changes on the required C/N_{01}. Compatibility and range checking are performed on all inputs to ensure that they are tenable.

Figure 7-4 illustrates a plot of the calculated C/N_{01} required as a function of Q_d for several values of Q_{fa} and a TW product of 10,000. These curves correspond within .2 dB to the curves of required E_S/N_{01} presented in [5 pp. 65- 70] (note that $E_S/N_{01} = C/N_{01}$ when $T = 1$).

Woodring and Edell [11] present several examples of calculating the detectability of spread spectrum signals by radiometers. In their first example, they describe a frequency hopped signal with a message time $T1 =$ 4 sec, a spread bandwidth $W1 = 2$ GHz, and a hop rate of 2000 hops/sec (which together with $T1$ corresponds to $b2 = 8000$ and $T2 = 500$ msec). The desired P_d and P_{fa} are 0.1 and 10^{-6} respectively. They report a required C/N_{01} of 48.9 dB-Hz, and given the same parameters, LPI/SDA returns 48.900 dB-Hz. In their second example, they calculate the detectability of a single hop or pulse. In this case, $T1 = T2 = 500$ msec is the pulse duration and $W1 = 2000$ Hz is the pulse bandwidth, yielding a TW product of unity. Note that $b2 = 1$. Woodring and Edell give a required C/N_{01} of 41.7 dB-Hz whereas LPI/SDA returns 41.740 dB-Hz.

8. Quality Factor Analysis

An LPI quality factor analysis was performed on the JTIDS waveform using the five candidate intercept receivers described in Section 5, with the communications receiver operating as described in Section 6, over a wide range of bit error rates. Results of the individual analyses on each intercept receiver was performed with detection probabilities of 0.7, 0.8, and 0.9; with the probability of false alarm in each case 10^{-4}. LPI/SDA was used to compute these performance curves, which are provided in Appendix C. A comparative summary of the modulation quality factor for all five intercept receivers is shown in Figure 8-1. Note that the wideband total power radiometer requires the smallest input signal-to-noise ratio for a successful intercept, as expected.

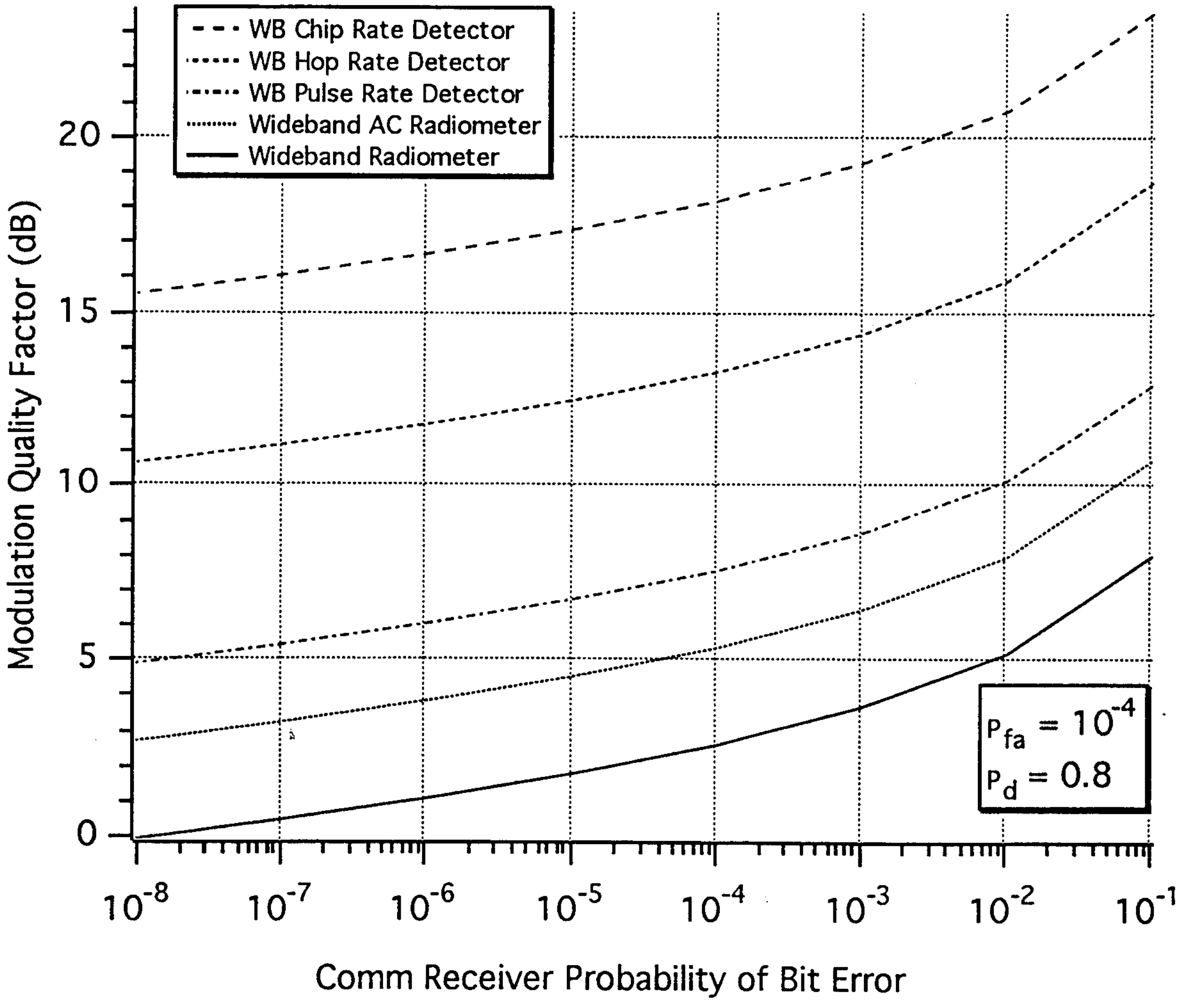


Figure 8-1: Modulation Quality Factors for Nonlinear Intercept Receivers

9. LPI Analysis of a Tactical Multiple-Access Radio Network

The Joint Tactical Information Distribution System (JTIDS) was originally developed to provide a tactical communications capability which would serve military command and control needs for the next decade. One of it's primary features is its high resistance to jamming which is achieved through the use of powerful error correcting codes, frequency hopping, direct sequence spreading and time hopping. The need for such diversity in spread spectrum techniques is due to the rigid network structure which makes the JTIDS signal a potentially attractive jamming target. The network structure is unavoidable because of the large number of users that must be serviced, and the throughput requirements of those users.

More than a decade after development of the JTIDS waveform its deficiencies are becoming apparent. Most notably, JTIDS is not well suited for use in military operations requiring covert communications. This deficiency is especially critical for aircraft operations that must be conducted within and behind the battle zone. Such aircraft require low RF signal emissions in their communications and navigation systems in order to avoid detection. Therefore, in this section we will examine possible variations on the design of the JTIDS waveform and multiple access structure which could result in a more jam resistant and covert tactical communication system.

9.1 JTIDS LPI Weaknesses

There are a number of deficiencies in the JTIDS waveform and signal structure which make it especially vulnerable to interception. A summary of each of these is provided below:

- Large transmitted output power - A C-130 airborne JTIDS terminal has a transmitted power output of 200 watts [15]. This large output power level is required to insure the minimum level of communications performance over the 300 nmi to 500 nmi range required by the JTIDS system. An intra-flight communication system should be able to function effectively with at least 10 dB less power since the communication range will be well under 100 nmi. Additional feature that would greatly improve the LPI performance of any communication system are adaptive power control (which is feasible for airborne communication systems), and steerable high gain antennas (which is probably not feasible within the frequency band of interest).

- Data rates are fixed - It is well known that low data rates provide more energy per bit to the detector (for fixed output power), which translates to less transmitted power required to maintain a specified level of performance. The JTIDS data rate is fixed and dependent upon the transmission mode selected. An effective LPI communication system requires the data rate to be variable in order to achieve a communications advantage (i.e., reduced Eb/No required) over the interceptor. High transmission rates can be used when interception is of no concern, and low transmission rates can be used to increase the LPI quality factor when the interceptor threat is present.

- Network structure is difficult to hide - The JTIDS network is designed to serve hundreds of simultaneous users having relatively large throughput requirements. This leads to a network structure which is necessarily complex, and also very predictable to an interceptor. An intra-flight communication system has more latitude to increase the level of uncertainty for the interceptor since the number of users to be serviced is typically less than 20. Time slots can be extended and transmission times can be reduced in order to achieve covertness without sacrificing throughput.

- Spreading of signal is limited - The JTIDS waveform consists of a hybrid DS/FH/TH signal operating within a bandwidth of 153 MHz. The spreading on the individual frequency hops is very limited, and the time-bandwidth product of each hop is only about 20. This makes the signal especially susceptible to channelized radiometers. The objective is to construct the signal so that the primary threat is the wideband radiometer. In order to achieve this objective the bandwidth of each hop can be extended as much as practical within the limited bandwidth available.

9.2 The Interceptor Threat

It is well known that the most effective intercept receiver against a 'noise-like' waveform in a stationary background of white gaussian noise is a wideband radiometer. On the other hand, if the communications waveform has any inherent structure that can be exploited, an intercept receiver can be designed to exploit this structure. Therefore, the engineer's objective in designing an LPI waveform is to reduce the signal structure to a level where the wideband intercept receiver is the primary threat. As a practical matter, there are limitations on the ability of a wideband radiometer to perform effectively in a hostile environment with multiple emitters. The actual sensitivity of a wideband radiometer in such an environment may be poor enough that a channelized radiometer or feature detector may actually be a more formidable threat.

Channelized radiometers are a formidable threat when narrow-band frequency hopping signals are employed. When the channelized radiometer is optimally matched to the hop bandwidth and duration, the interception threat is much greater than that posed by the wideband radiometer. Since the JTIDS waveform is a hybrid DS/FH/TH signal with relatively small time-bandwidth pulses, the radiometer threat must be considered. A simple rule-of-thumb evaluation of the relative detectability of some arbitrary waveform is provided by the *waveform detection factor* [16]:

$$Q_o = \frac{\zeta^2 R_h T_m}{N_f} \qquad (9\text{-}1)$$

where R_h is the hop rate, T_m is the minimum of either the message duration or interceptor observation time, N_f is the number of frequency hop cells and ζ is the pulse duty factor. For JTIDS:

$T_m = 3354\ \mu\text{sec}$
$R_h = 1/T_h = 76{,}923$ hops/sec
$N_f = 51$
$\zeta = T_o/Th = .4923$

resulting in Q_o = 1.06, which is a border-line case indicating that a channelized radiometer and wideband radiometer may be equally effective in detecting the signal. The preferable value of Q_o for any waveform is above 10 or 20.

Although the DS/FH/TH JTIDS signal structure has features which can be detected by channelized radiometers and feature detectors, the waveform can be modified so that a wideband radiometer is the most effective intercept receiver. Therefore, if the intercept receiver employs some other type of architecture, the advantage favors the communicator.

As an example of the use of the waveform detection factor, the JTIDS signal structure was modified by decreasing the number of hops, and then by varying the duty factor of the JTIDS pulses. The LPI/SDA program was then used to obtain the results illustrated in Figure 9-1 and Figure 9-2. The waveform becomes less vulnerable to the channelized radiometer as the number of frequency hopping cells decreases, and as the duty factor increases toward unity.

9.3 Waveform Design Strategy

The strategy for designing LPI waveforms can be derived from the LPI quality factor equations. Recall that for an LPI communication system, the LPI quality factor is a measure of the relative distances of the communicator and interceptor from the communications transmitter, as shown below:

$$Q_{LPI} = 20\,Log\left(\frac{R_c}{R_i}\right) = Q_{ANT} + Q_{ATM} + Q_{IS} + Q_{MOD} \qquad (9\text{-}2)$$

where Q_{ANT} is the antenna quality factor, Q_{ATM} is the atmospheric propagation quality factor, Q_{IS} is the interference suppression quality factor, and Q_{MOD} is the modulation quality factor. Each of these quality factors have been described elsewhere. In the design of LPI waveforms we are only interested in Q_{MOD}, and the other quality factors can simply be neglected or taken to be a constant, K. Therefore,

$$Q_{LPI} = 20\,Log\left(\frac{R_c}{R_i}\right) = K + Q_{MOD} \qquad (9\text{-}3)$$

The waveform can now be designed to maximize Q_{MOD}, which measures the performance of the intercept receiver against the communication waveform design. It depends on both the intercept receiver processing and the communication receiver processing. The modulation quality factor reveals the essential strategy for making any waveform more covert. We showed in Equation (3-13) that the modulation quality factor can be expressed as:

$$Q_{MOD} = \frac{\zeta_i\,(P_d, P_{fa}, T, W)}{\zeta_c\,(P_e)\,R_b} \qquad (9\text{-}4)$$

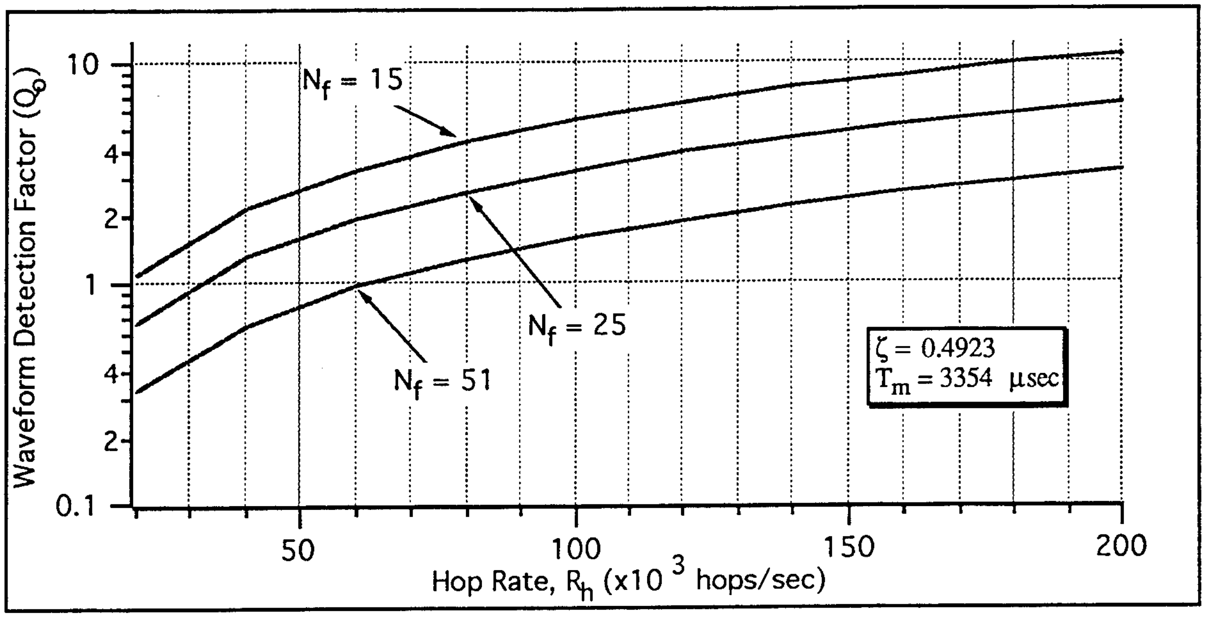


Figure 9-1: Waveform detection factor (varying number of frequency cells)

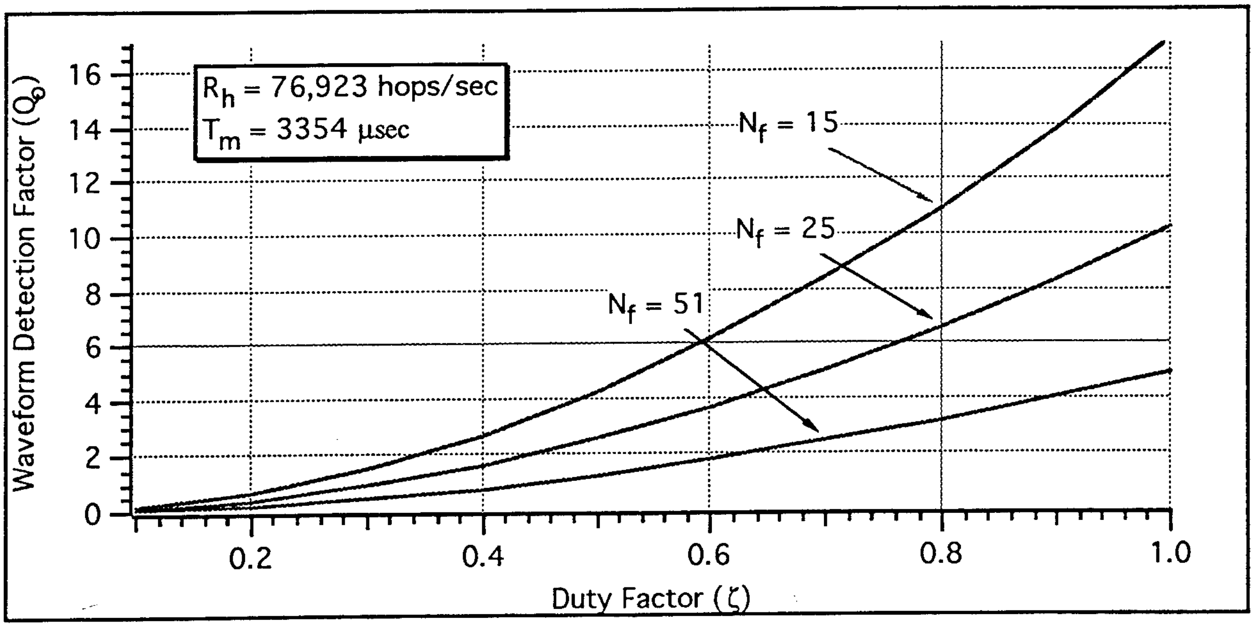


Figure 9-2: Waveform detection factor (varying duty factor of transmit pulse)

This expression gives rise to several observations regarding the design of LPI waveforms. The overall signal bandwidth should be as large as possible consistent with the operating frequency band and available frequency allocations. The data rate should be kept as low as possible consistent with mission objectives. Modern LPI systems will undoubtedly require data rate adaptation as well as use of efficient modulation and coding designs to reduce the E_b/N_0 required. This translates into reduced transmitted power and improved LPI performance.

9.4 Results of the Analysis

The first step in LPI waveform design is to structure the waveform for a baseline minimum detectability. Since we are examining variations on the JTIDS structure, we are limited in several important regards. Most importantly, the frequency allocation probably cannot be changed, and therefore the total bandwidth is not a design parameter. Since the JTIDS frequency band excludes the IFF bands, this limits the amount of pulse spreading allowed if the entire 153 MHz frequency band is to be used.

To achieve reduced detectability there are many parameters in the JTIDS waveform which can be changed. The generic TDMA signal detection model shown in Figure 9-3 defines the signal parameters.

Using the signal detection model and LPI/SDA, the baseline JTIDS detectability is computed and displayed in Figure 9-4. This graph shows the minimum detectable signal as a function of the wideband radiometer probability of detection, with the probability of false alarm as a parameter.

Finally, two variations are made on the JTIDS time slot structure by increasing the time slot and then decreasing the duty factor. The effect on the minimum detectable signal to noise ratio are shown in Figure 9-5 and Figure 9-6.

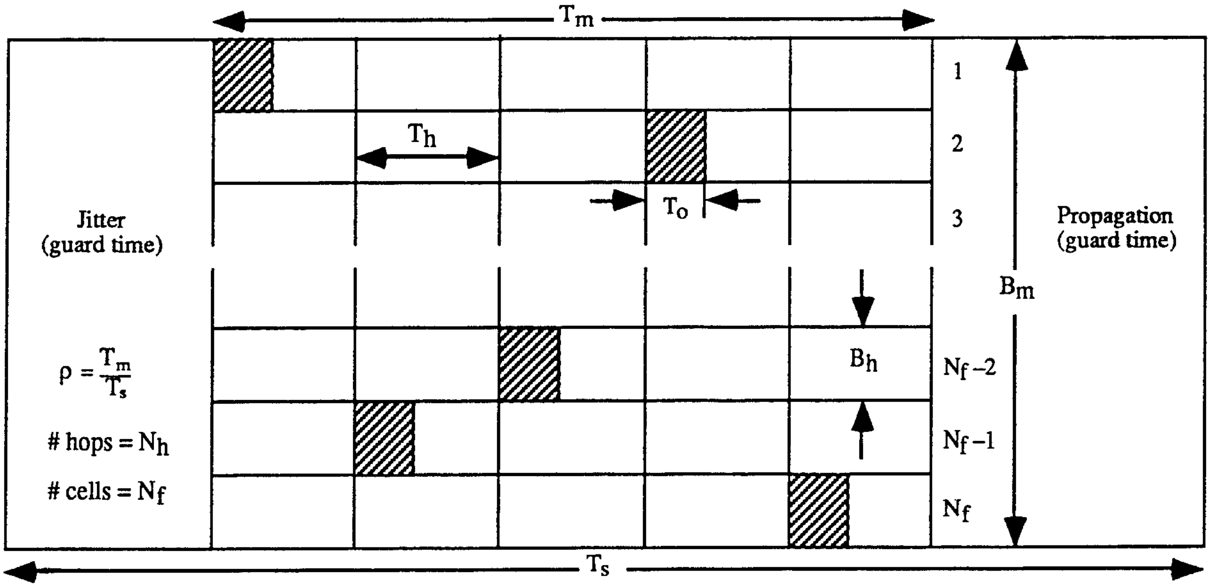


Figure 9-3: TDMA FH/DS/TH signal detection model

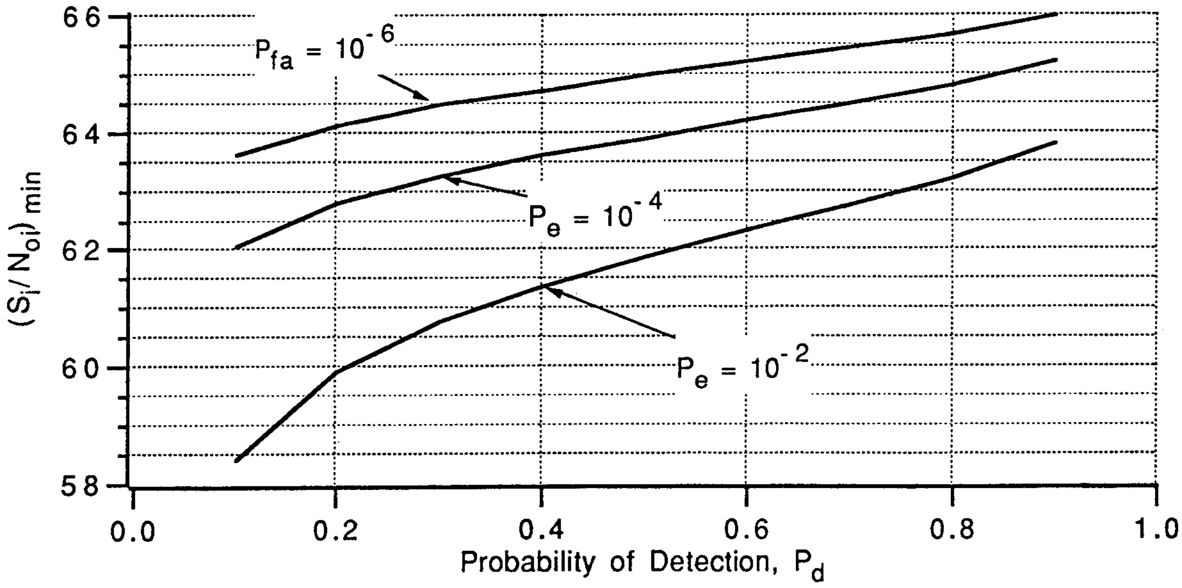


Figure 9-4: Baseline detectability of the JTIDS signal structure

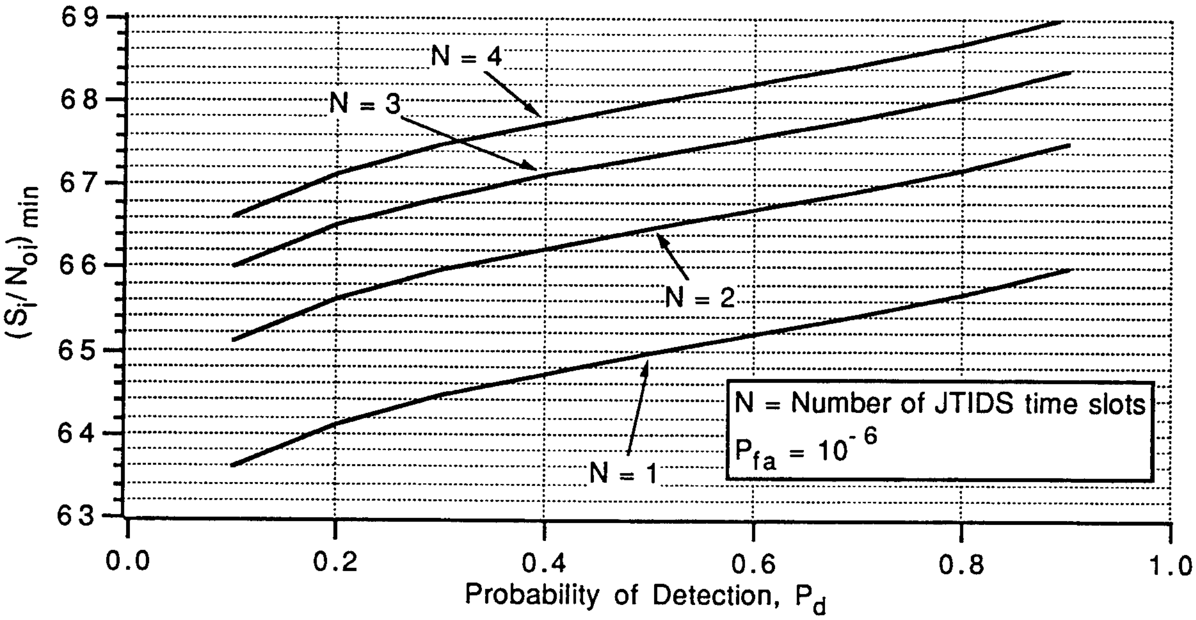


Figure 9-5: Minimum signal detectability with increased number of time slots

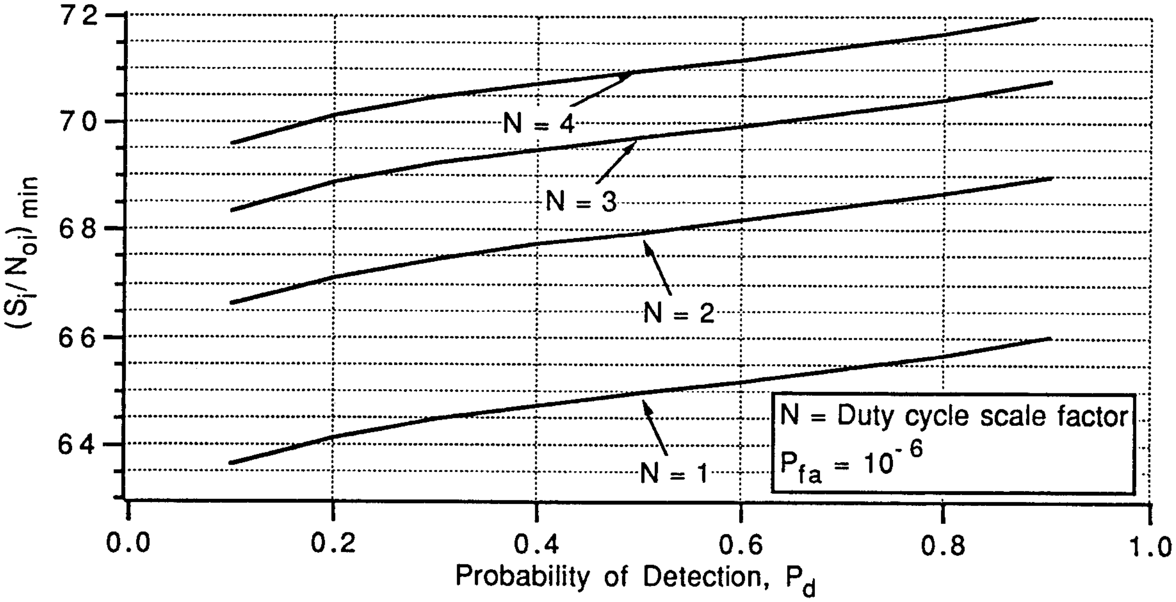


Figure 9-6: Minimum signal detectability with decreased duty factor

10. References

[1]. David L. Nicholson, *Spread Spectrum Signal Design: LPE and AJ Systems*, Computer Science Press, Rockville, Maryland, 1988.

[2]. Paul J. Crepeau, "LPI and AJ Modulation Quality Factors," NRL Memorandum Report 3436, Naval Research Laboratory, Washington DC, January 1977.

[3]. John D. Edell, "Wideband, Noncoherent, Frequency-Hopped Waveforms and their Hybrids in Low Probability of Intercept Communications," NRL Report 8025, Naval Research Laboratory, Washington DC, 8 November 1976.

[4]. Robert A Wright, "LPD Analysis of TDMA, Direct Sequence Spread Spectrum Signaling," Internal Memorandum, Paramax Systems Corp, Salt Lake City, UT, March 1992.

[5]. Robin A. Dillard, and George M. Dillard, *Detectability of Spread Spectrum Signals*, Artech House (ISBN 0-89006-299-4), Norwood MA, 1989.

[6]. TDMA JTIDS Overview Description, Mitre Corporation, MTR8413.

[7]. Edward W. Chandler, and George R. Cooper, "Development and Evaluation of an LPI Figure of Merit for Direct Sequence and Frequency Hop Systems," *Proceedings of the 1985 Military Communications Conference*, pp. 33.7.1 - 33.7.6, October 1985.

[8]. H. Urkowitz, "Energy Detection of Unknown Deterministic Signals," *Proceedings of the IEEE*, vol. 55, pp. 523 - 531, April 1967.

[9]. Robert A. Wright, "JTIDS Detectability," Unisys internal technical report, Unisys Defense Systems, Communications Systems Division, Salt Lake City, UT.

[10]. Scott P. Francis and Glenn E. Prescott, "Computer Aided Analysis of LPI Signal Detectability," *Proceedings of the 1991 Military Communications Conference*, Washington DC, October 1991.

[11]. D. Woodring and J. Edell, "Detectability Calculation Techniques," U.S. **Naval** Research Laboratories, Code 5480, Sept 1, 1977.

[12]. Lawrence L. Gutman, and Glenn E. Prescott, "System Quality Factors for LPI Communication," Proceeding of the 1989 IEEE International Conference on Systems Engineering. Dayton, Ohio, 1989.

[13] Simon, M., Omura, J., Scholtz, R., and Levitt, *Spread Spectrum Communications*, Vol. 3, Computer Science Press, Rockville, MA, 1985.

[14] George M. Dillard, "Recursive Computation of the Generalized Q-Function," *IEEE Transactions on Aerospace and Electronic Systems*, July 1973, pp. 614-615.

[15]. J. Patrick, "Effects of JTIDS on 1030 and 1090 MHz ATC Avionics," ECAC-CR-83-184, DoD ECAC, Annapolis, MD, December 1983.

[16]. F. Torre, R. DiFazio, and E. German, "LPI Waveform Study," RADC Final Technical Report, RADC-TR-90-24, Griffiss AFB, April 1990.

Appendix A

Joint Tactical Information Distribution System Time Slot Description

The component of the JTIDS signal that is of primary interest in the detection of the JTIDS waveform is the 7.8125 msec time slot, as shown below:

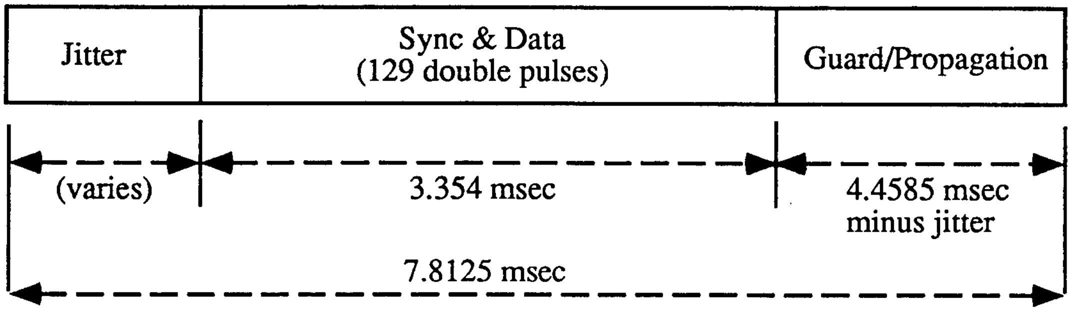


Figure A-1: JTIDS Time Slot

The time slot has 3.354 msec of sync and data transmission time, consisting of 129 double pulses. Each double pulse is comprised of two single pulses, with the data in the second pulse merely repeating the first pulse data. The 13 microsecond single pulse timing is shown below:

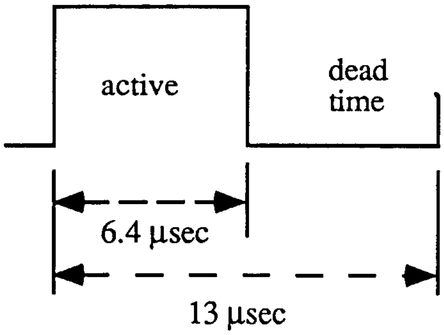


Figure A-2: JTIDS Single Pulse

During the active time of a single pulse, the waveform appears as a stream of 32 chips MSK-modulating a carrier. The chip rate is (1/6.4 µsec/32) = 5 MHz. As a JTIDS time slot progresses through its constituent 129 double pulses (258 single pulses), the carrier frequency is frequency hopped on a (single) pulse-by-pulse basis. The carrier is chosen by some pseudorandom algorithm from 51 possible frequencies. Since the carrier hops at 13 μsec intervals, the hop rate is 76.923 KHz. The center frequencies that comprise the hop cells are shown below:

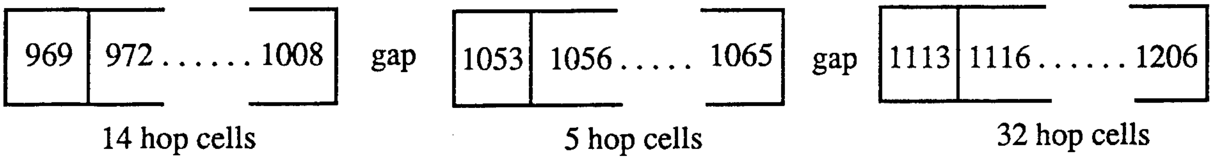


Figure A-3: JTIDS Hop Center Frequencies in MHz

The notable characteristics of this set are the 3 MHz hop cell spacing and the gaps surrounding 1030 MHz and 1090 MHz, which insure that JTIDS does not interfere with IFF transponder frequencies. The 3 MHz spacing is interesting in that it allows substantial overlap of adjacent cells. With MSK the null-to-null bandwidth is (1.5 Hz/bit/sec x 5 Mchips/sec) = 7.5 MHz. Note however, the JTIDS specification implies that the RF pulse bandwidth is 3 MHz. The power spectrum for MSK and QPSK is shown below for reference.

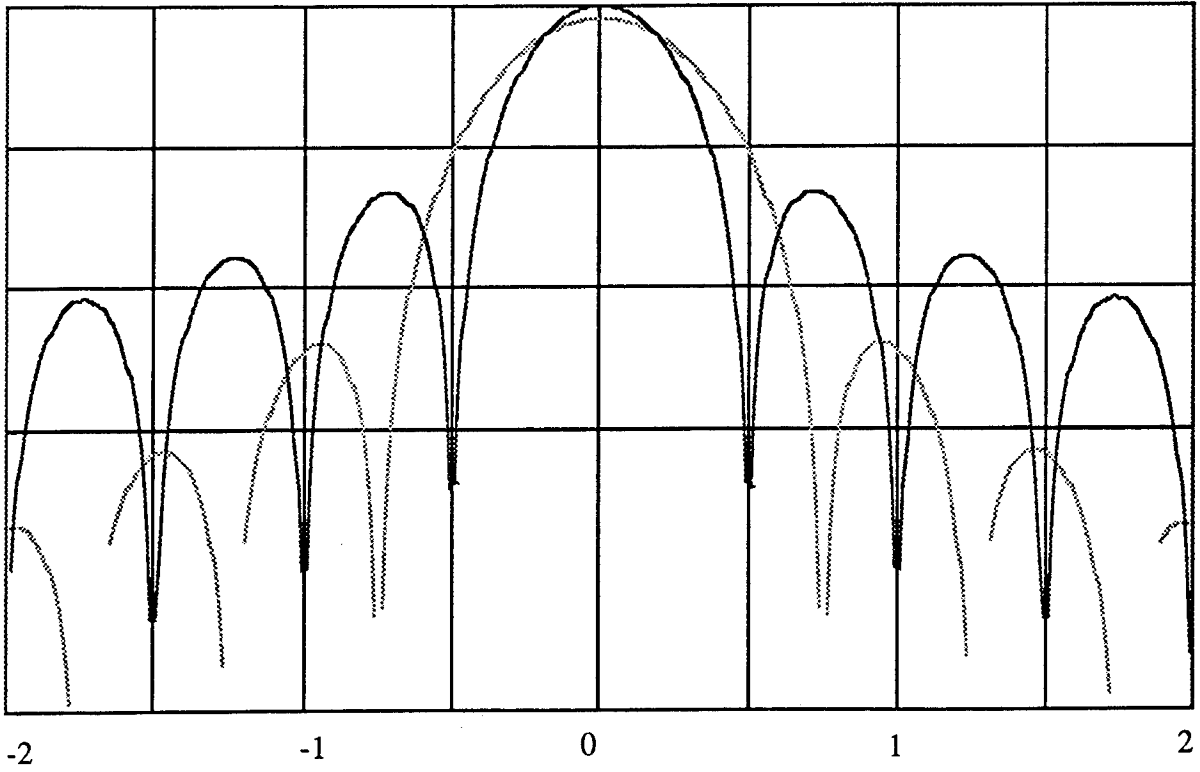
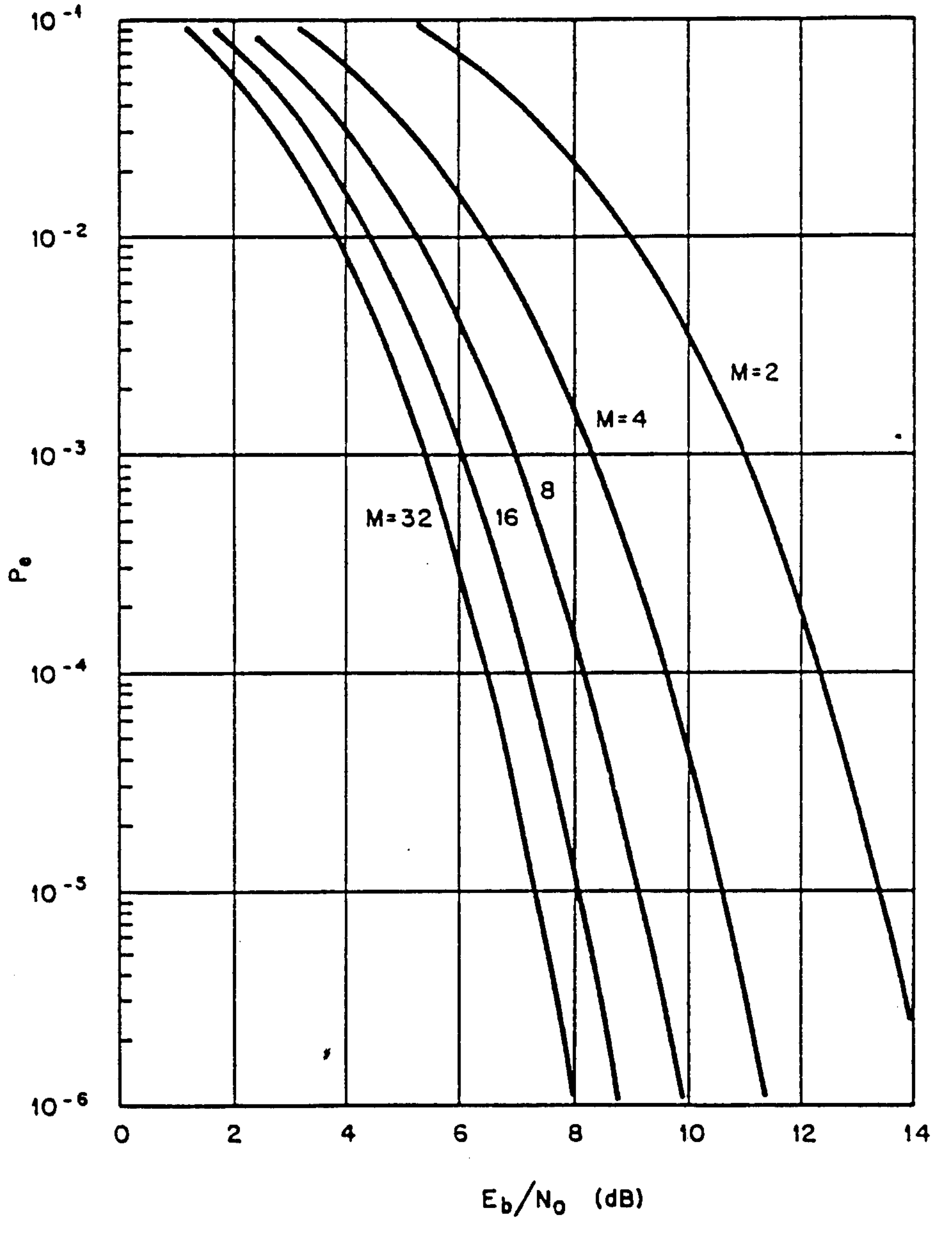


Figure A-4: Spectrum of Minimum Shift Keying (in green) and QPSK (in red) {Normalized frequency offset from carrier $(f-f_c)/R$ Hz/bit/sec}

Appendix B

Probability of Error Curves for Orthogonal Noncoherent Modulation



Appendix C
LPI Quality Factor Analysis with Wideband Radiometers

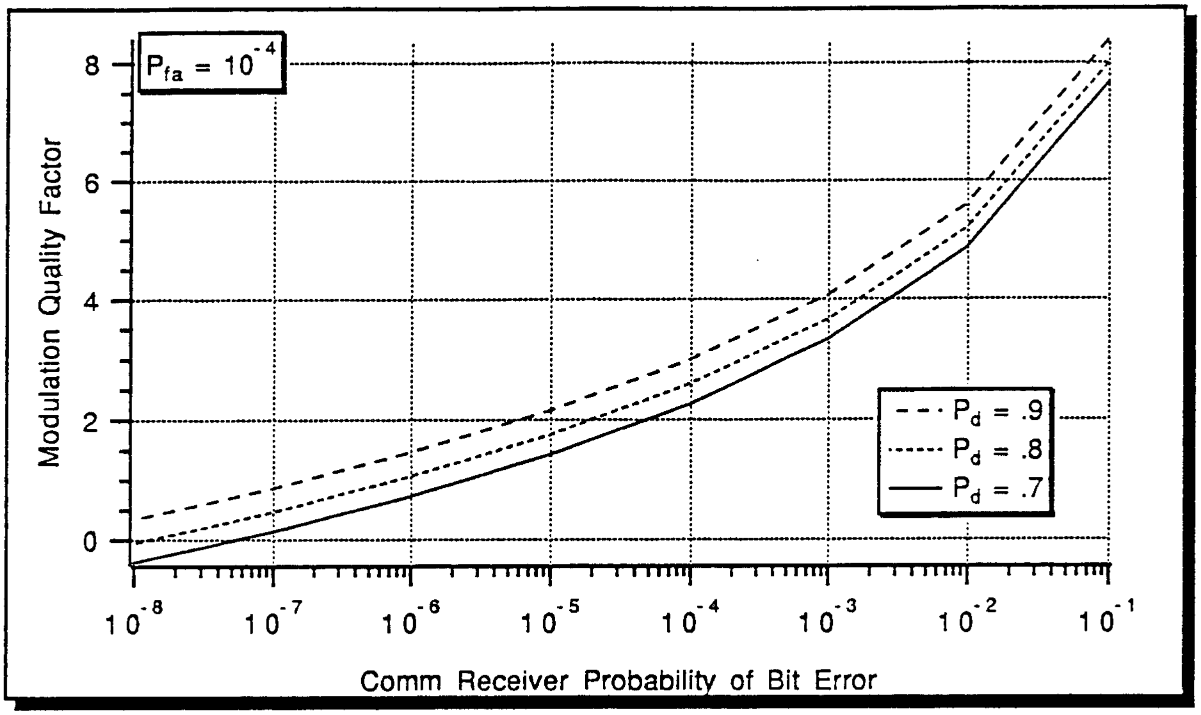


Figure C1 - Wideband Total Power Radiometer

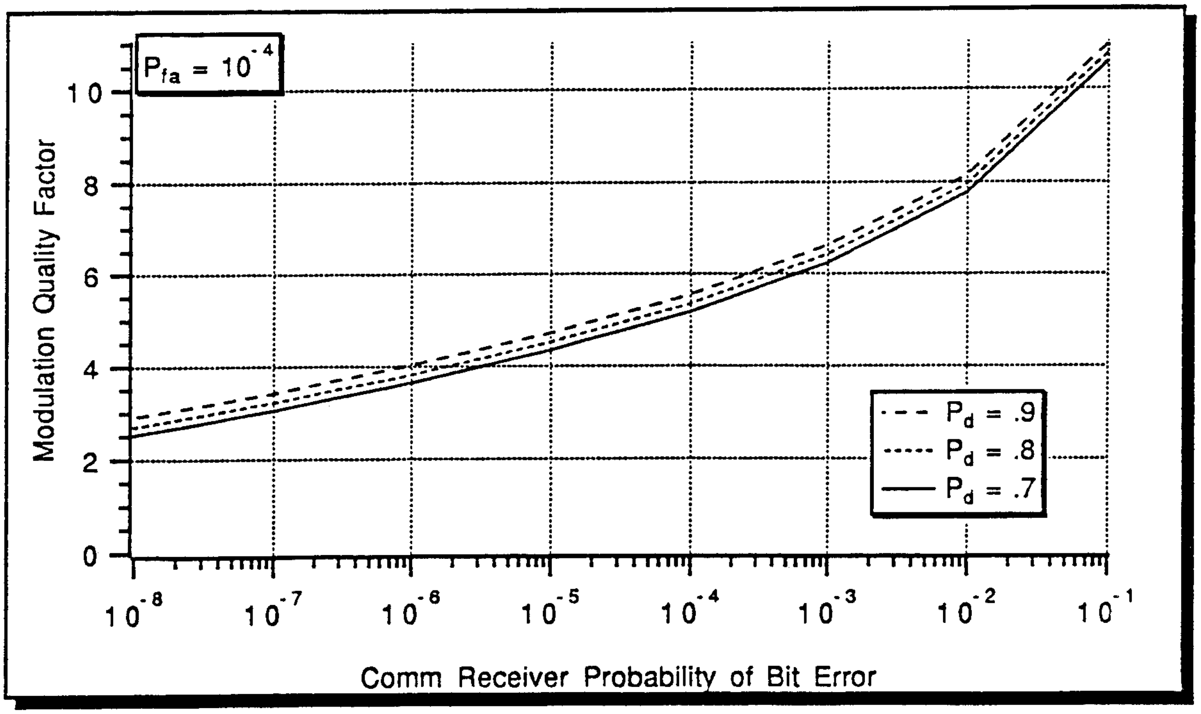


Figure C2 - Wideband AC Radiometer

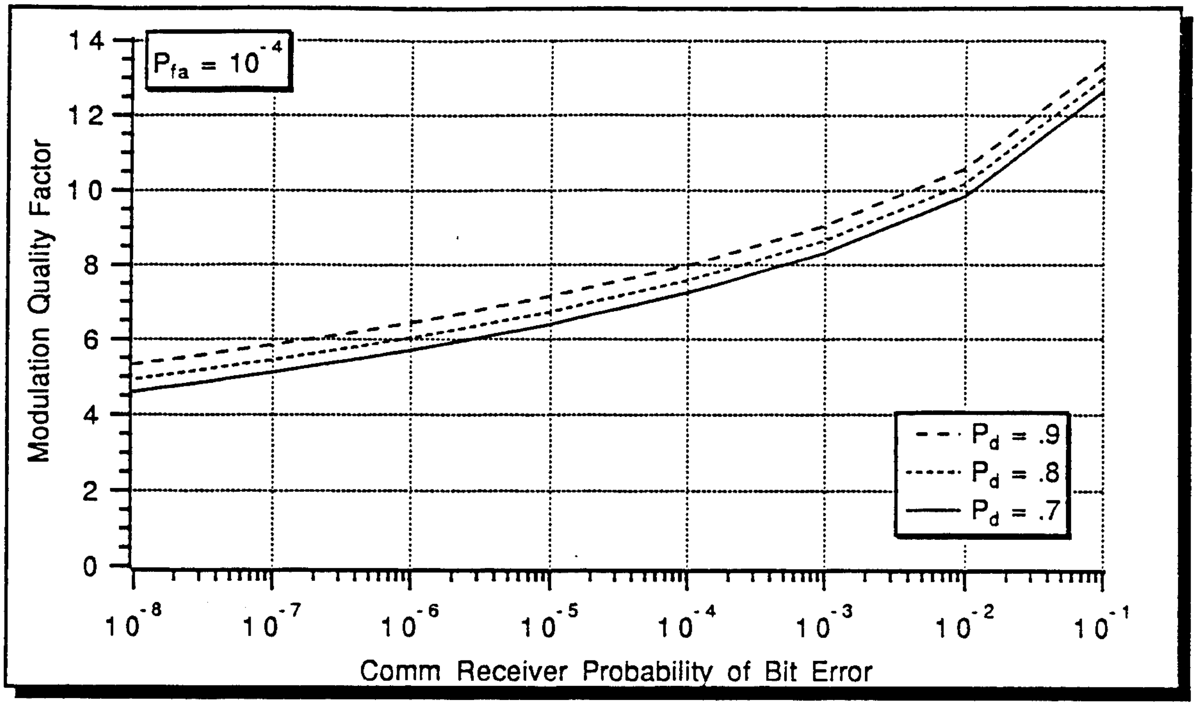


Figure C3 - Wideband Chip Rate Detector

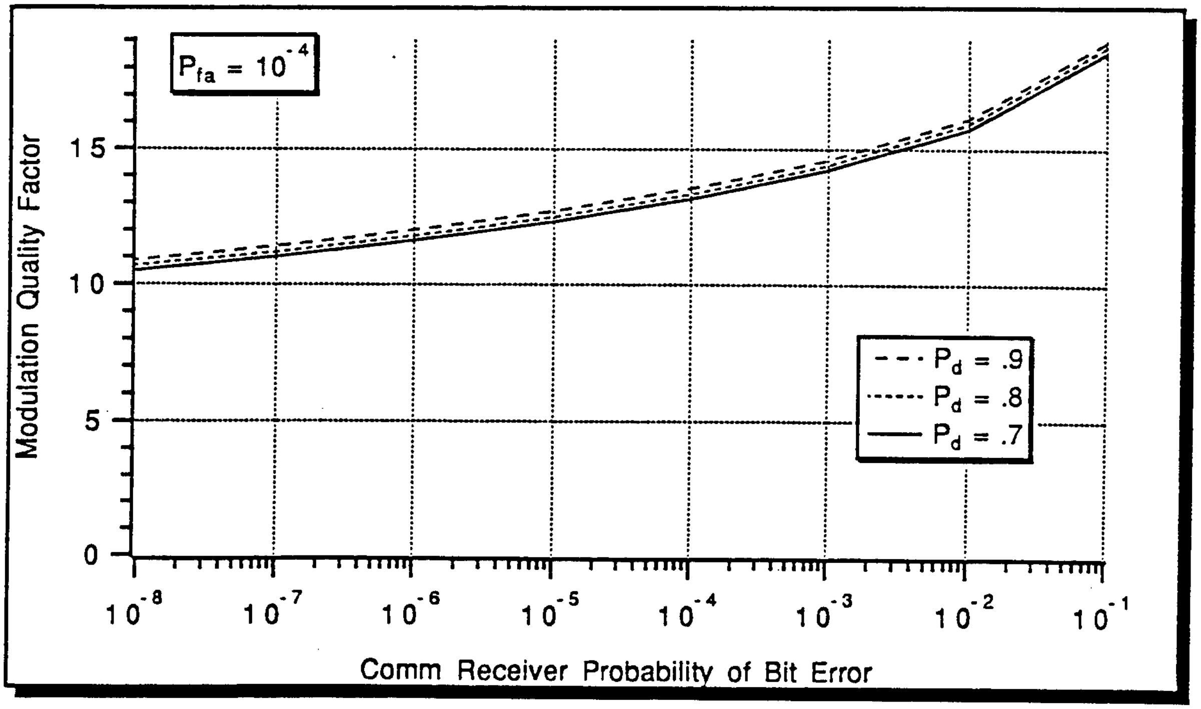


Figure C4 - Wideband Hop Rate Detector

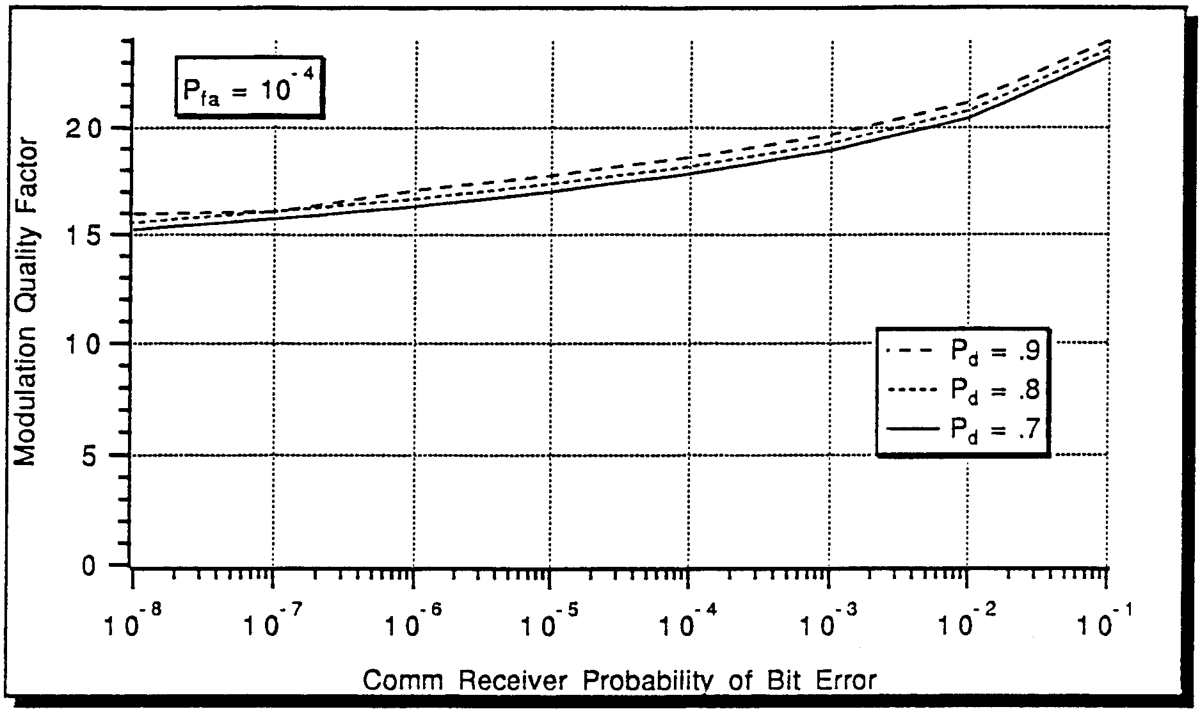


Figure C5 - Wideband Pulse Rate Detector

ROBUST LOW ORDER CONTROL DESIGN FOR UNCERTAIN SYSTEMS

Jenny L. Rawson
Assistant Professor
Electrical Engineering

North Dakota State University
Fargo, ND 58105

Final Report for:
Research Initiation Program
Wright Laboratory

Sponsored by:
Air Force Office of Scientific Research
Wright-Patterson Air Force Base

and

North Dakota State University

December 1992

ROBUST LOW ORDER CONTROL DESIGN FOR UNCERTAIN SYSTEMS

Jenny L. Rawson
Assistant Professor
Electrical Engineering
North Dakota State University

ABSTRACT

Two methods are presented for the design of low order observers which provide robust performance and stability for systems with real parameter uncertainties. These methods are based on Riccati inequalities which can be quickly constructed to incorporate multiple design goals. Dual systems are used to apply these design methods to a variety of plant configurations. An iterative process based on ideas from μ-synthesis takes advantage of adjustable parameters in the Riccati equations to maximize the robustness of the closed-loop system.

ACKNOWLEDGEMENTS

I would like to thank the Air Force Office of Scientific Research and Research and Development Laboratories for their support and administration of this research. I am grateful to Dr. Hsi-Han Yeh and Dr. Siva Banda for their assistance in selecting this research topic and for the technical guidance they provided. Former Capt. Andrew Sparks also helped me greatly in undertaking this project. Gratitude is due Dr. Chin Hsu for the technical discussions we had. And, special thanks go to Mr. Bau Tran without whose effort and cooperation the numerical examples could never have been completed.

ROBUST LOW ORDER CONTROL DESIGN FOR UNCERTAIN SYSTEMS

Jenny L. Rawson

I. INTRODUCTION

In Yeh et al [13] and Rawson [9], a method was presented for the construction of design equations for low-order observers to stabilize and provide disturbance attenuation for uncertain plants. It was shown that a single Riccati equation and a Riccati inequality resulted from the construction. While it was clear how the Riccati equation could be used directly in the design process, the design could be completed only in a certain special case where the Riccati inequality could be replaced by a Sylvester equation. In addition, there was no guidance for the selection of the values for the parameters included in the design equations.

The objectives of this report are to present new methods, based on the Riccati inequality, for completing the observer design, and to demonstrate how to select design parameters in order to maximize performance and stability robustness. An alternative to the Sylvester equation is developed for the special case, and a low-order Riccati equation for the general case. Duality is used to apply these results to systems subject to output disturbance and uncertainty in the plant output matrix; these were not included in [9, 13]. Construction of the Riccati equations introduces additional design parameters which can be exploited for the reduction of the conservatism of the stability and performance margins. This can be done systematically with a method reminiscent of the D-K iterations for μ-synthesis.

II. MATHEMATICAL BACKGROUND

This section covers the mathematical basics used in [9, 13] to derive the observer design Riccati equations. Because of the reliance on H_∞ theory, uncertainty was treated as complex and the knowledge of the structure of the uncertainty was not used to full advantage. However, the design equations serve as a basis for minimization of conservatism by use of a search method similar to μ-synthesis. Some of the material of [9, 13] is reviewed here for completeness and to point up some modifications that are needed in order to implement the search method in numerical examples.

2.1 Overview

An important result in [9, 13] is the machinery for formulating Riccati equalities or inequalities from design objectives. The basis for this are three theorems of Doyle [2, 3] used in μ-analysis. The first theorem defines the relationship between the H_∞ norm of a transfer function and disturbance rejection, the second uses the structured singular value (μ) for a non-conservative condition for the robust stability of an uncertain systems, and the third gives nonconservative conditions for a system to retain good performance when subjected to structured perturbations. Because of these theorems, design Riccati equations for multiple objectives can be constructed in a direct manner; a Riccati inequality can be written for the closed-loop system for each individual design goal, such as disturbance rejection or stabilization with respect to parameter perturbation, then the terms of the various inequalities can be consolidated into one overall Riccati inequality. Design equations follow upon the application of simple matrix inequalities. This method was applied to the design of a low-order (Luenberger) observer to give H_∞

disturbance rejection and stabilization for plants with additive perturbations of the state and input matrices. It was shown that a single Riccati equation is needed--for the selection of state feedback gains--and a Riccati ınequality must be satisfied for the disturbance rejection criterion to be met. Section IV is devoted to a review of the derivation of the former. Section V presents new methods for obtaining observer gains to satisfy the latter.

2.2 Basic Theorems

Consider the block diagram of Figure 1. The transfer functions G_{11} through G_{22} include plant and compensator dynamics along with any weighting functions needed for performance specifications. All uncertainty is placed in the feedback block Δ. Furthermore, define the block matrix:

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \tag{2.1}$$

The following three theorems from Doyle [2, 3] form the basis for our method of derivation of design equations. The first theorem is a definition of the H_∞ norm of a transfer function in terms of the L_2 norms of the input and output signals. It assumed that there is no uncertainty; i.e., $\Delta = 0$.

<u>Theorem A</u>: Let $\Delta = 0$, $\delta > 0$. Then, for any w_1 such that $\|w_1\|_2 < 1$, $\|z_1\|_2 < \delta$ iff $\|G_{11}\|_\infty \le \delta$.

■

The next theorem concerns the robust stability of the system for nonzero Δ. To take advantage of some of the structure in the uncertainty, structured singular value is used. This is defined as follows. Let

$$\underline{\Delta} = \{\Delta = \mathrm{diag}(\delta_1, \delta_2, \ldots, \delta_m, \Delta_1, \Delta_2, \ldots, \Delta_\ell) \mid \delta_i \in R, \Delta_j \in C^{k_j \times k_j}\} \qquad (2.2)$$

and,

$$\underline{B\Delta}_\delta = \{\Delta \in \underline{\Delta} | \bar{\sigma}(\Delta) < 1/\delta\} \qquad (2.3)$$

The structured singular value $\mu(H)$ of a transfer function H is defined as:

$$\frac{1}{\mu(H)} = \min_{\Delta \in \underline{\Delta}} \{\bar{\sigma}(\Delta) | \det(I - H\Delta) = 0\}, \qquad (2.4)$$

if there exists a $\Delta \in \underline{\Delta}$ such that $\det(I - H\Delta) = 0$. Otherwise, $\mu(H) = 0$. Also define another function of H:

$$\|H\|_\mu = \sup_\omega \mu[H(j\omega)] \qquad (2.5)$$

An important property of $\|H\|_\mu$ is that while it is not a norm, it is overbounded by the H_∞ norm. That is,

$$\|H\|_\mu \le \|H\|_\infty \qquad (2.6)$$

This will be used later in the development of design equations.

The second theorem gives a necessary and sufficient condition for stability of a system subjected to a combination of real and complex structured perturbations. Note that in many cases, the actual perturbations will be real with interdependencies. These will have to be covered by some complex and independent perturbations so that the condition for robust stability becomes merely sufficient. However, this robust stability theorem is usually less conservative than theorems based solely on norms.

Theorem B: The system of Figure 1 is stable for all $\Delta \in \underline{B\Delta}_\delta$ iff $\|G_{22}\|_\mu \le \delta$.

■

The third theorem gives conditions for stabilization and disturbance rejection in systems subject to structured perturbations. To include both requirements, a fictitious perturbation block Δ_1 between z_1 and w_1 is appended to the system as in Figure 2. This theorem is uses the lower linear fractional transformation of (G, Δ):

$$F_L(G,\Delta) = G_{11}G_{12}\Delta(I - G_{22}\Delta)^{-1}G_{12} \tag{2.7}$$

<u>Theorem C</u>: The system is stable and $\|F_L(G,\Delta)\|_\infty < 1$ for all $\Delta \in \underline{B\Delta}_\delta$ iff G is stable and $\|G\|_\mu \leq \delta$ where μ is taken with respect to the augmented set

$$\tilde{\underline{\Delta}} = \{\tilde{\Delta} = \mathrm{diag}(\Delta_1, \Delta) \mid \Delta \in \underline{\Delta}\}. \tag{2.8}$$

■

In order to obtain the desired Riccati equations, we will need a theorem first stated by Willems [11] and then extended by Veillette, Medanic and Perkins [10].

<u>Theorem D</u>: Let $\{A_{cl}, C_{cl}\}$ be a detectable pair, $\delta > 0$. Then,

$$\|C_{cl}(sI - A_{cl})^{-1}B_{cl}\|_\infty \leq \delta \tag{2.9}$$

and, A_{cl} is stable if there exists $P \geq 0$ such that

$$PA_{cl} + A_{cl}{}^TP + \delta^{-2}PB_{cl}B_{cl}{}^TP + C_{cl}{}^TC_{cl} \leq 0. \tag{2.10}$$

■

III. A RICCATI CONSTRAINT FOR ROBUST PERFORMANCE WITH PLANT UNCERTAINTY

The main result of [9, 13] is repeated below with some slight modifications. This theorem permits a Riccati inequality to be written for each individual design goal for a closed-loop system. Then, these inequalities can be combined in a straight forward manner into a single inequality. If this inequality has a positive semi-definite solution, and a detectability condition is met, then all of the design goals are met *simultaneously*.

Referring to Figure 3, let the transfer function L(s) be:

$$L(s) = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_\ell \end{bmatrix} (sI - A_{c1})^{-1} [B_1 \; B_2 \; \cdots \; B_3] \qquad (3.1)$$

Thus, the transfer function from w_1 to z_1 is the lower linear fractional transformation $F_1(L,\Delta)$. Blocks Δ_i, $i = 2, \ldots, \ell$ represent either system uncertainties or disturbance rejection criteria. The following theorem is the basis for the construction of Riccati equations for robust controller design.

<u>Theorem I</u>. If (A_{c1}, C_i) is detectable for some C_i, $1 \le i \le \ell$ and there exists a $P \ge 0$ and $d_i > 0$, $i = 1, \ldots, \ell$, such that

$$PA_{c1} + A_{c1}^T P + \delta^{-2} \sum_{i=1}^{\ell} d_i P B_i B_i^T P + \sum_{i=1}^{\ell} d_i^{-1} C_i^T C_i \le 0, \qquad (3.2)$$

then the system of Figure 3 is internally stable and $\|F_1(L,\Delta)\|_\infty < \delta$ for all Δ_i $i = 2, \ldots, \ell$, with $\bar{\sigma}(\Delta_i) \le 1/\delta$.

Proof of Theorem I: If inequality (3.2) is satisfied, then so are

$$PA_{c1} + A_{c1}^T P + \delta^{-2} P B_i B_i^T P + C_i^T C_i \le 0, \; i = 1, \ldots, \ell. \qquad (3.3)$$

Suppose that (A_{c1}, C_j) is the detectable pair. Then by Theorem D, A_{c1} is stable. Stability of A_{c1} trivially ensures the detectability of $(A_{c1}, [C_1^T \; C_2^T \; \ldots \; C_\ell^T]^T)$, so that by Theorem D, a positive semi-definite solution to (4.2) implies that $\|DLD^{-1}\|_\infty \le \delta$. The rest follows from Theorem C since $\|L\|_\mu \le \|DLD^{-1}\|_\infty$ for any $D = \text{diag}(d_1 I_{k1}, d_2 I_{k2}, \ldots, d_\ell I_{k\ell})$ [2,3].

■

Note the we also have $\|L_{ii}\|_\infty \leq 1$, $i = 1, \ldots, \ell$, where L_{ii} is the transfer function from $\mathbf{w}_i$ to $\mathbf{z}_i$ with the Δ_i, $i = 2, \ldots, \ell$, removed. These transfer functions have the state-space descriptions:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= A_{cl}\mathbf{x}(t) + B_i\mathbf{w}_i(t) \\ \mathbf{z}_i(t) &= C_i\mathbf{x}(t) \end{aligned} \tag{3.4}$$

Each inequality (3.3) can be used to design compensation to meet the i^{th} design goal. That is, by Thereom D, if $\{A_{cl}, C_i\}$ is detectable and there exists a $P \geq 0$ that satisfies (3.3), then A_{cl} is stable and $\|L_{ii}\|_\infty \leq 1$. If Δ_i represents uncertainty in the system, then by Theorem B, the uncertain system is also stable for all Δ_i with $\bar{\sigma}(\Delta_i) \leq 1$. Thus, a multiple goal design problem can be broken down into subproblems, each with its own Riccati inequality. Appropriate terms can be taken from each and combined into a singel Riccati inequality as suggested by Theorem I.

IV. PROBLEM I

Problem I is illustrated in Figure 4. The plant has bounded, real parameter uncertainties in the input and state matrices. Compensation is to be provided by a minimal order Luenberger observer. The design goals are: 1) to make the closed loop system asymptotically stable, and, 2) to bound by γ the H_∞ norm of the transfer function from $\mathbf{w}_1$ to $\mathbf{z}_1$ and $\mathbf{z}_2$; these goals are to be met for all real ΔA and ΔB falling within predetermined limits.

4.1 Plant Description

The plant is described by the following equations:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= (\mathbf{A} + \Delta\mathbf{A})\mathbf{x}(t) + (B + \Delta B)u(t) + G_1 w_1(t) \\ \mathbf{y}(t) &= \mathbf{Cx}(t) \\ \mathbf{z}_1(t) &= H_1\mathbf{x}(t) \\ \mathbf{z}_2(t) &= H_2 u(t) \end{aligned} \tag{4.1}$$

where, $\mathbf{x} \in R^n$ is the plant state vector, $u \in R^p$ the control input, $\mathbf{w}_1 \in R^{p1}$ a disturbance input, $y \in R^m$ the measured output, and $\mathbf{z}_1 \in R^{q1}$, $\mathbf{z}_2 \in R^{q2}$ are controlled outputs. In order to be sure of attaining closed-loop stability, it is assumed that:

A4.1 $\{\mathbf{A}, B\}$ is stabilizable.

A4.2 $\{\mathbf{A}, C\}$ is detectable.

A4.3 $\{\mathbf{A}, H_1\}$ is detectable.

A4.4 $\mathrm{Rank}(B) = p \le n$.

A4.5 $\mathrm{Rank}(C) = m \le n$.

A4.6 $\mathrm{Rank}(H_2) = q_2 \le p$.

The plant uncertainties will be decomposed as follows:

$$\Delta\mathbf{A} = \sum_{j=1}^{r} D_j \Delta_j E_j; \; \Delta B = \sum_{j=1}^{r} D_j \Delta_j F_j \tag{4.2}$$

with Δ_j e $R.^{k_j \times k_j}$ The matrices D_j, E_j and F_j will be fixed, with all of the **parameter** variation appearing in the Δ_j. Furthermore, the uncertainty will be bounded as

$$\bar{\sigma}(\Delta_j) \le \rho, \tag{4.3}$$

where D_j, E_j or F_j can be scaled so that for each j the perturbation of maximum norm achieves σ; this will reduce the conservatism of our approach. The conservatism will result mainly from using the maximum singular value, which covers a set of real-valued perturbations by a set of complex-valued perturbations (Khargonekar et al [5]).

It will also be convenient to stack the uncertainty matrices as:

$$\tilde{D} = [D_1\ D_2\ \ldots\ D_r];$$
$$\Delta = \mathrm{diag}(\Delta_1\ \Delta_2\ \ldots\ \Delta_r);$$
$$\tilde{E} = [E_1{}^T\ E_2{}^T\ \ldots\ E_r{}^T]^T;$$
$$\tilde{F} = [F_1{}^T\ F_2{}^T\ \ldots\ F_r{}^T]^T. \tag{4.4}$$

4.2 Minimal-Order Observer

Suppose that state feedback design has been performed to obtain a set of gains K_c such that if the states were available, the control law would be $\mathbf{u} = -K_c\mathbf{x}$. However, if the states are not available, an observer will have to be used. A minimal-order observer can be described by the following equations:

$$\begin{aligned} \dot{\mathbf{x}}_o(t) &= \mathbf{F}\mathbf{x}_o(t) + \mathbf{G}\mathbf{u}\ (t) + \mathbf{K}_f\mathbf{y}(t) \\ \mathbf{u}(t) &= -\mathbf{N}\mathbf{x}_o(t) - \mathbf{M}\mathbf{y}(t) \end{aligned} \tag{4.5}$$

where, $\mathbf{x}_o \epsilon R^{n-m}$. The observer states are an estimate of a linear combination of the plant states $\mathbf{Tx}$, where $\mathbf{T}$ is chosen so that $[T^T\ C^T]^T$ is nonsingular. An error signal can be defined as

$$\mathbf{e} = \mathbf{Tx} - \mathbf{x}_o \tag{4.6}$$

If $\Delta\mathbf{A} = 0$, $\Delta\mathbf{B} = 0$, the error $\mathbf{e}$ asymptotically approaches zero if $\mathbf{A} - \mathbf{BK}_c$ and $\mathbf{F}$ are stable, and if the following Luenberger constraint equations are satisfied (O'Reilly [8]):

$$\begin{aligned} \mathbf{TA} - \mathbf{FT} &= \mathbf{K}_f\mathbf{C} \\ \mathbf{G} &= \mathbf{TB} \\ \mathbf{NT} + \mathbf{MC} &= \mathbf{K}_c \end{aligned} \tag{4.7}$$

All observers satisfying these constraints may be parameterized in the following way.

If m by n matrix C has rank m, then there exists a nonsingular n by n matrix S such that

$$CS = [I_m \quad 0]$$

Let

$$\mathcal{A} = S^{-1}AS = \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{bmatrix}$$

$$\mathcal{B} = S^{-1}B = \begin{bmatrix} \mathcal{B}_1 \\ \mathcal{B}_2 \end{bmatrix}$$

where, $\mathcal{A}$ and $\mathcal{B}$ are partitioned conformably with CS. Then, any minimal-order observer satisfying the Luenberger constraints can be written as:

$$F = P^{-1}(\mathcal{A}_{22} - V_2\mathcal{A}_{12})P$$

$$T = P^{-1}[-V_2 \quad I_{n-m}]S^{-1}$$

$$G = TB = -P^{-1}V_2\mathcal{B}_1 + P^{-1}\mathcal{B}_2$$

$$K_f = P^{-1}(\mathcal{A}_{21} - V_2\mathcal{A}_{11} + \mathcal{A}_{22}V_2 - V_2\mathcal{A}_{12}V_2) \qquad (4.8)$$

$$N = K_cS\begin{bmatrix} 0 \\ I_{n-m} \end{bmatrix}P$$

$$M = K_cS\begin{bmatrix} I_m \\ V_2 \end{bmatrix}$$

where, P is an (n-m) by (n-m) nonsingular matrix and V_2 is an (n-m) by m matrix of parameters. Note that P effects only the internal structure of the observer, and V_2 can be chosen to assign F any self-conjugate set of n-m eigenvalues if $(\mathcal{A}_{22}, \mathcal{A}_{12})$ is observable.

4.3 Construction of Riccati Inequalities.

In [9, 13] it was shown in detail how to apply Theorem I to the task of obtaining Riccati inequalities to be used for the low-order observer design. This procedure is outlined here; more details can be obtained from the above references.

The system can be represented as in Figure 5. Additional noise inputs $\mathbf{w}_2$ and $\mathbf{w}_3$ emanate from, and additional controlled outputs $\mathbf{z}_3$ and $\mathbf{z}_4$ are fed into the perturbation blocks. Scaling blocks have been introduced to allow for the use of Theorem I. According to this theorem, Problem I can be broken down into three smaller problems: minimization of the effect of $\mathbf{w}_1$ on $[\mathbf{z}_1{}^T\ \mathbf{z}_2{}^T]^T$ (disturbance rejection), stabilization with respect to $\Delta\mathbf{A}$, and stabilization with respect to ΔB. Also, the procedure is greatly simplified by using states $[\mathbf{x}^T\ \mathbf{e}^T]^T$, where $\mathbf{e} = \mathbf{Tx} - \mathbf{x}_o$. A state-space description is obtained for each of the small problems. Then, the resulting data matrices are combined into an overall Riccati inequality for the solution of the robust design problem. The three small problems and their state-space descriptions are:

1. Disturbance Rejection

$$A_{cl} = \begin{bmatrix} \mathbf{A} - \mathbf{BK_c} & \mathbf{BN} \\ 0 & \mathbf{F} \end{bmatrix}$$

$$B_1 = \begin{bmatrix} \gamma^{-1}G_1 \\ \gamma^{-1}TG_1 \end{bmatrix}$$

$$C_1 = \begin{bmatrix} H_1 & 0 \\ -H_2K_c & H_2N \end{bmatrix}$$

By assumptions A4.3, A4.4 and A4.6, and the stability of $\mathbf{F}$, $\{Acl,\ C_1\}$ is a detectable pair.

2. Stabilization with Respect to $\Delta\mathbf{A}$.

$$A_{cl} = \begin{bmatrix} \mathbf{A} - \mathbf{BK_c} & \mathbf{BN} \\ 0 & \mathbf{F} \end{bmatrix}$$

$$B_2 = \begin{bmatrix} \tilde{\mathbf{D}} \\ \mathbf{T}\tilde{\mathbf{D}} \end{bmatrix}$$

$$C_2 = [\ \rho\tilde{\mathbf{E}} \quad 0\ ]$$

3. Stabilization with Respect to ΔB.

$$A_{c1} = \begin{bmatrix} A - BK_c & BN \\ 0 & F \end{bmatrix}$$

$$B_3 = \begin{bmatrix} \tilde{D} \\ T\tilde{D} \end{bmatrix}$$

$$C_3 = [\ \rho^{-1}\tilde{F}K_c \ \ \rho^{-1}\tilde{F}N\ ]$$

Theorem I is now applied to design an observer so that the closed-loop system is stable and the transfer function between w_1 and $[z_1^T\ z_2^T]^T$ is bounded by γ for all allowable perturbations of the plant state matrix A and input matrix B. The state, input and output matrices above are inserted into (3.2) as:

$$R(P) = PA_{c1} + A_{c1}^TP + \delta^{-2}P(d_1^2B_1B_1^T + d_2^2B_2B_2^T + d_3^2B_3B_3^T)P + d_1^{-2}C_1^TC_1 + d_2^{-2}C_2^TC_2 + d_3^{-2}C_3^TC_3 \leq 0, \qquad (4.9)$$

It is known already that (A_{c1}, C_1) is a detectable pair, so that if there exists a $P \geq 0$ and a set of observer and state feedback gains which satisfy (4.9), then the above goals are attained simultaneously. A design equation and an inequality to be tested can be derived from (4.9) by first assuming that $P = \mathrm{diag}(P_1, P_2)$ and making use of the matrix inequality:

$$XY + YX \leq aXQX + a^{-1}YQY \qquad (4.10)$$

for all $a > 0$, $Q > 0$. After setting:

$$K_c = [d_1^{-1}H_2^TH_2 + (d_3\rho^2)^{-1}\tilde{F}^T\tilde{F}]^{-1}B^TP_1 \qquad (4.11)$$

one obtains:

$$R(P) \leq \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}$$

where,

$$\begin{aligned} U_{11} = {} & P_1A + A^TP_1 + d_1{}^{-2}H_1{}^TH_1 + d_2{}^{-2}\rho^2\tilde{E}^T\tilde{E} \\ & + \delta^{-2}P_1(d_1{}^2(1 + a)\gamma^{-2}G_1G_1{}^T + d_2{}^2(1 + b)\delta^{-2}\tilde{D}\tilde{D}^T + d_3{}^2(1 + c)\delta^{-2}\tilde{D}\tilde{D}^T \\ & - B[d_1{}^{-2}H_2{}^TH5V2 + (d_3{}^2\rho^2)^{-1}F^TF]^{-1}B^T)P_1 \end{aligned} \qquad (4.12)$$

$$\begin{aligned} U_{22} = {} & P_2F + F^TP_2 + d_1{}^{-2}N^TH_2{}^TH_2N + (d_3{}^2\rho^2)^{-1}N^T\tilde{F}^T\tilde{F}N \\ & + P_2T[d_1{}^2(1 + a^{-1})\gamma^{-2}G_1G_1{}^T + d_2{}^2(1 + b^{-1})\tilde{D}\tilde{D}^T \\ & + d_3{}^2(1 + c^{-1})\tilde{D}\tilde{D}^T]T^TP_2 \end{aligned} \qquad (4.13)$$

for any a, b, c > 0. Also,

$$U_{12} = U_{21}{}^T = 0.$$

Observer design consists of finding P_1, $P_2 \geq 0$ such that U_{11}, $U_{22} \leq 0$. A design equation can be obtained from (4.11) and (4.12) since the 0 matrix is negative semi-definite:

$$\begin{aligned} & P_1A + A^TP_1 + d_1{}^{-2}H_1{}^TH_1 + d_2{}^{-2}\rho^2\tilde{E}^T\tilde{E} \\ & + \delta^{-2}P_1(d_1{}^2(1 + a)\gamma^{-2}G_1G_1{}^T + d_2{}^2(1 + b)\delta^{-2}\tilde{D}\tilde{D}^T + d_3{}^2(1 + c)\delta^{-2}\tilde{D}\tilde{D}^T \\ & - B[d_1{}^{-2}H_2{}^T3BH_2 + (d_3{}^2\rho^2)^{-1}\tilde{F}^TF]^{-1}B^T)P_1 = 0 \end{aligned} \qquad (4.14)$$

This equation is solved for $P_1 \geq 0$, which is then inserted into (4.11) to obtain gain K_c. Inequality (4.13) is discussed in the next section.

4.4 Observer Design Equations

In [13], the design procedure was halted after the gain K_c was chosen to satisfy equation (4.14), leaving the remaining observer gains undetermined. However, they must be selected so that in equation (4.13), $U_{22} \leq 0$. It was suggested that the procedure of Monahemi et al [6] could be used for their selection. Unfortunately, this procedure is applicable only when rank $([G_1\ \tilde{D}]) < m$, restricting the resolution of the structure of the uncertainties ΔA and ΔB.

An alternative approach is to apply the O'Reilly Parameterization, and derive an observer Riccati equation. Observer design consists of obtaining a positive definite solution to this equation, solving for the parameter V_2 in terms of this solution, and inserting V_2 into the parameterization equations to yield the observer gains.

First, consider equation (4.13) which can be written as:

$$U_{22} = P_2F + F^TP_2 + d^{-2}N^T\mathcal{F}^T\mathcal{F}N + P_2T\mathcal{D}\mathcal{D}^TT^TP_2 \le 0 \qquad (4.15)$$

where,

$$\mathcal{F} = \begin{bmatrix} d_1^{-1}H_2 \\ \rho d_3^{-1}\tilde{F} \end{bmatrix}$$

$$\mathcal{D} = [\ d_1\gamma^{-1}\sqrt{1 + 1/a}\ G_1 \quad d_2\sqrt{1 + 1/b}\ \tilde{D} \quad d_3\sqrt{1 + 1/c}\ \tilde{D}\ ]$$

The goal is to find $P_2 \ge 0$ and observer gains F, N and T so that $U_{22} \le 0$. The observer gains must also satisfy the Luenberger constraints. A method for meeting these goals is proposed in the following.

<u>Claim 1</u>: Let

$$S^{-1}\mathcal{D} = \begin{bmatrix} \mathcal{D}_1 \\ \mathcal{D}_2 \end{bmatrix}$$

$$\mathcal{N} = \mathcal{F}K_cS\begin{bmatrix} 0 \\ I_m \end{bmatrix}$$

If there exists a $Y \ge 0$ and a V_2 such that

$$Y(\mathcal{A}_{22} - V_2\mathcal{A}_{12}) + (\mathcal{A}_{22} - V_2\mathcal{A}_{12})^TY + \delta^{-2}Y(-V_2\mathcal{D}_1 + \mathcal{D}_2)(-V_2\mathcal{D}_1 + \mathcal{D}_2)^TY + \mathcal{N}^T\mathcal{N} \le 0 \qquad (4.16)$$

and, $\mathcal{A}_{22} - V_2\mathcal{A}_{12}$ is stable, then for any nonsingular P there exists a $P_2 \ge 0$ and a set of observer gains such that $U_{22} \le 0$ and all of the observer constraints are satisfied.

Proof of Claim 1: This follows from the definition of U_{22}. The parameterized gains are inserted in the proper places in (4.15):

$$U_{22} = P_2P^{-1}(A_{22} - V_2A_{12})P + P^T(A_{22} - V_2A_{12})^TP^{-T}P_2$$
$$+ d^{-2}P_2P^{-1}[-V_2 \quad I]\begin{bmatrix} D_1 \\ D_2 \end{bmatrix}[D_1{}^T \quad D_2{}^T]\begin{bmatrix} -V_2{}^T \\ I \end{bmatrix}P^{-T}P_2$$
$$+ P^T[0 \quad I]S^TK_c{}^TF^TFK_cS\begin{bmatrix} 0 \\ I \end{bmatrix}P \leq 0$$

The inequality holds if and only if:

$$P^{-T}U_{22}P^{-1} = Y(A_{22} - V_2A_{12}) + (A_{22} - V_2A_{12})Y$$
$$+ d^{-2}Y[-V_2 \quad I]\begin{bmatrix} D_1 \\ D_2 \end{bmatrix}[D_1{}^T \quad D_2{}^T]\begin{bmatrix} -V_2{}^T \\ I \end{bmatrix}Y$$
$$+ [0 \quad I]S^TK_c{}^TF^TFK_cS\begin{bmatrix} 0 \\ I \end{bmatrix} \leq 0$$

where $Y = P^{-T}P_2P^{-1} \geq 0$ if and only if $P_2 \geq 0$. The final version of this inequality results from noting that:

$$[-V_2 \quad I]\begin{bmatrix} D_1 \\ D_2 \end{bmatrix} = -V_2D_1 + D_2$$

and,

$$F^TFK_cS\begin{bmatrix} 0 \\ I \end{bmatrix} = N$$

■

Claim 1 does not give the necessary design equations. These are derived in the following two claims, each corresponding to one of two possibilities on rank(D). Claim 2 uses a Riccati equation to obtain the parameter V_2.

Claim 2: Suppose that rank$(D_1) = m$. If there exists a $X > 0$ such that:

$$[A_{22} - D_2D_1^T(D_1D_1^T)^{-1}A_{12}]X + X[A_{22} - D_2D_1^T(D_1D_1^T)^{-1}A_{12}]^T$$
$$+ D_2D_2^T + d^{-2}XN^TNX - XA_{12}^T(D_1D_1^T)^{-1}A_{12}X = 0 \qquad (4.17)$$

and V_2 such that

$$V_2 = XA_{12}^T(D_1D_1^T)^{-1} \qquad (4.18)$$

and $A_{22} - V_2A_{12}$ is stable, then for any nonsingular P, there exists $P_2 \geq 0$ and a set of observer gains satisfying all of the observer constraints and resulting in $U_{22} \leq 0$.

■

Proof of Claim 2: First, $F = P^{-1}(A_{22} - V_2A_{12})P$ is stable. To show that inequality (4.16) of Claim 1 is satisfied, recall that if matrix $Q = 0$, then $Q \leq 0$, and if $Q > 0$, then $Q \geq 0$. Next, manipulate equation (4.17):

$$0 = A_{22}X + XA_{22} - D_2D_1^T(D_1D_1^T)^{-1}A_{12}X - X[D_2D_1^T(D_1D_1^T)^{-1}A_{12}]^T$$
$$+ D_2D_2^T + 0808d^{-2}XN^TNX - XA_{12}^T(D_1D_1^T)^{-1}A_{12}X \qquad ()$$

Multiply both sides of () from left and right by δX^{-1} and let $Y = \delta^2X^{-1}$ to get:

$$0 = YA_{22} + A_{22}Y - YD_2D_1^T(D_1D_1^T)^{-1}A_{12} - [D_2D_1^T(D_1D_1^T)^{-1}A_{12}]^TY$$
$$+ \delta^{-2}YD_2D_2^TY - \delta^2A_{12}^T(D_1D_1^T)^{-1}A_{12} + N^TN$$
$$= YA_{22} + A_{22}Y - YD_2D_1^T(D_1D_1^T)^{-1}A_{12} - [D_2D_1^T(D_1D_1^T)^{-1}A_{12}]^TY$$
$$+ \delta^2A_{12}^T(D_1D_1^T)^{-1}D_1D_1^T(D_1D_1^T)^{-1}A_{12} - \delta^2A_{12}^T(D_1D_1^T)^{-1}A_{12}$$
$$+ \delta^{-2}YD_2D_2^TY - \delta^2A_{12}^T(D_1D_1^T)^{-1}A_{12} + N^TN$$

With $YV_2 = \delta^2A_{12}^T(D_1D_1^T)^{-1}$, this becomes:

$$0 = YA_{22} + A_{22}Y - YD_2D_1^TV_2^TY - YV_2D_1D_2^TY$$
$$+ \delta^{-2}YV_2D_1D_1^TV_2^TY - YV_2A_{12}$$
$$+ \delta^{-2}YD_2D_2^TY - \delta^2A_{12}^TV_2^TY + N^TN$$

Or,

$$0 = Y(A_{22} - V_2A_{12}) + (A_{22} - V_2A_{12})^TY$$
$$+ \delta^{-2}Y(-V_2D_1 + D_2)(-V_2D_1 + D_2)^TY + N^TN$$

■

Claim 3 is applied when it is possible with an observer to achieve exact recovery of the state feedback closed-loop transfer function. The procedures of Monahemi et al. [6] or Niemann et al [7] could also be used. Claim 3 does employ the same procedure but the proof is different.

Claim 3: Suppose that D_1 has full column rank and that $A_{22} - V_2A_{12}$ is stable, where $V_2D_1 = D_2$. Then, for any nonsingular P, there exists a $P_2 \geq 0$ and a set of observer gains complying with the constraints (), such that $U_{22} \leq 0$. ■

Proof of Claim 3: With V_2 as above, the inequality of Claim 1 becomes:

$$Y(A_{22} - V_2A_{12}) + (A_{22} - V_2A_{12})^TY + N^TN \leq 0$$

which is a Lyapunov inequality. Since $A_{22} - V_2A_{12}$ is stable, this will have a solution $Y \geq 0$. ■

V. PROBLEM II

It is possible to apply the procedures of Section IV to a system with perturbations in the measurement matrix and output disturbance by taking the dual of the system. What is sacrificed is the ability to include perturbations in the input matrix and control of the input signal u. The system for which this can be done is illustrated in Figure 6 and is described by:

$$\begin{aligned} \dot{x} &= (A + \Delta A)x + Bu + G_1w_1 \\ z_1 &= H_1x \\ y &= (C + \Delta C)x + G_2w_2 \end{aligned} \tag{5.1}$$

where $\Delta C = \tilde{G}\Delta\tilde{E}$ with Δ and $\tilde{E}$ given in (4.4) and $\tilde{G}$ having a conformable structure. Also, some assumptions must be made on the plant:

A5.1 $\{\mathbf{A},\ B\}$ is stabilizable;

A5.2 $\{\mathbf{A},\ C\}$ is detectable;

A5.3 $\{\mathbf{A},\ G_1\}$ is stabilizable;

A5.4 $\mathrm{Rank}(B) = p \leq n.$

A5.5 $\mathrm{Rank}(C) = m\ 8\Phi\ n.$

A5.6 $\mathrm{Rank}(G_2) = r_2 \leq m$

Compensation is provided by the dual observer [8]:

$$\begin{aligned} \mathbf{x}_c &= F^T\mathbf{x}_c - N^T\mathbf{w}_c \\ \mathbf{w}_c &= \mathbf{y} + CT^T\mathbf{x}_c \\ \mathbf{u} &= M^T\mathbf{w}_c + K_f{}^T\mathbf{w}_c \end{aligned} \tag{5.2}$$

The goal is to design this dual observer so that the closed-loop system is stable and that there is and upper bound of γ on the H_∞ norm of the transfer function between $[\mathbf{w}_1{}^T\ \ \mathbf{w}_2{}^T]^T$ and z_1 for all Δ_i with maximum singular values bounded above by ρ. This can be done by applying the design equations of the previous section to the dual system:

$$\begin{aligned} \mathbf{x}_d &= (\mathbf{A} + \Delta\mathbf{A})^T\mathbf{x}_d + (C + \Delta C)^T u_d + H_1{}^T x_1 \\ \zeta_1 &= G_1{}^T\mathbf{x}_d \\ \zeta_2 &= G_2{}^T u_d \\ y_d &= B^T\mathbf{x}_d \end{aligned} \tag{5.3}$$

It is clear that this is of the same form as system (4.1) of section IV. Therefore, the compensator can be designed by making the appropriate substitutions in the design equations, solving for the gains for the observer (4.5), then obtaining the dual of this observer to yield (5.2). This process is illustrated in a numerical example in section VI.

VI. SEARCH PROCEDURES FOR ROBUST DESIGN

In the previous sections, Riccati equations were devised to facilitate the design of an observer for robust performance and stability of a system subject to uncertainty. These equations contain three sets of adjustable parameters: $S_1 = \{d = [d_1\ d_2\ d_3]^T | d_1, d_2, d_3 > 0\}$, $S_2 = \{[a\ b\ c]^T |\ a, b, c > 0\}$ and $s_3 = \{\delta > 0\}$. The first set can be used to minimize an upper bound on $\|L\|_\mu$ as discussed in the proof of Theorem I; a design goal is to find the $d \epsilon S_1$, and the corresponding compensator, such that $\|DLD^{-1}\|_\infty$ is minimized. Set S_2 is used to trade off between the state feedback and Claim 2 design equations (4.12), (4.17); parameters a, b and c can be searched until there exists a solution for both equations. They may also be used to avoid high gains in the observer. It was found that adjustments in the remaining parameter δ were sometimes necessary in order for solutions to both design equations to exist. In cases where Claim 3 can be applied, the parameters in set S_2 can be dropped. The procedure for Claim 3 will be discussed first.

6.1 Observer Design with Claim 3

When Claim 3 can be applied for observer design, it is possible to achieve the same level of robust performance with an observer as with state feedback. Observer design thus consists of solving for state feedback gain K_c such that $\|DLD^{-1}\|_\infty$ is minimized. If, when there is exact transfer recovery, the resulting observer state matrix F is stable, the design is essentially complete.

A numerical search algorithm is outlined below. The underlying philosophy is similar to the D-K iteration of μ-synthesis [1, 2, 3], where a full-order compensator is designed to minimize an upper bound on μ for a closed-loop system. But, in this case, the iterative procedure is concerned with state feedback design only. It has also been found that for particular numerical examples, there are relative values of the parameters that nearly ensure that the state feedback Riccati equation has a positive definite solution; these relationships are expressible as inequalities and are included in the search algorithms to obtain starting values for the parameters. A drawback of using Claim 3 is that there may not exist any V_2 such that $V_2\mathcal{D}_1 = \mathcal{D}_2$ and $\mathcal{A}_{22} - V_2\mathcal{A}_{12}$ is stable, regardless of the parameter values. In this case, Claim 2 should be used instead.

<u>Algorithm for Claim 3</u>

1. Obtain a starting δ.
2. Obtain a starting d_1, d_2, and d_3.
3. Solve (4.12) for P_1 and (4.11) for K_c.
4. If there is no solution to () or $A - BK_c$ is unstable, adjust d_1, d_2, and d_3, and go to step 3.
5. Solve $V_2\mathcal{D}_1 = \mathcal{D}_2$ for V_2.
6. If $\mathcal{A}_{22} - V_2\mathcal{A}_{12}$ is not stable, adjust d_1, d_2, and d_3, and go to step 3.
7. Form L(s). Search for d such that $\|DLD_{-1}\|_\infty$ is minimized.
   If there is a significant change in d, go to step 3.
8. If the disturbance rejection and stability robustness are inadequate, select a lower δ and go to step 2.
9. Obtain the remaining observer gains using (4.9).

Note that the bound on the allowable uncertainties Δ_i has been scaled down by the $\|DLD^{-1}\|_\infty$. Similarly, the H_∞ norm of the transfer function between w_1 and $[z_1^T \;\; z_2^T]^T$ has been amplified by the same amount.

6.2 Observer Design with Claim 2

When Claim 3 cannot be used, the observer will cause some loss of robustness from that which is possible with state feedback. The amount of the loss will be dependent on the increase in δ needed for both Riccati design equations to have solutions. This relationship is made unclear by the inclusion of parameters a, b, and c in the design Riccati equations. These appeared in the derivation of the Riccati inequalities, and represent the addition of conservatism in the robustness conditions. Searches over these three parameters will have to be made until both Riccati equations (4.12) and (4.17) have solutions, if possible.

Algorithm for Claim 2

1. Obtain a starting δ.
2. Obtain starting d_1, d_2, and d_3.
3. Obtain starting a, b, and c.
4. Solve (4.12) for P_1 and (4.11) for K_c.
5. If there is no solution to (4.12) or $A - BK_c$ is unstable, adjust d_1, d_2, d_3, a, b, and c, and go to step 4.
6. Solve (4.17) for X and (4.18) for V_2.
7. If there is no solution to (4.17) or if $A_{22} - V_2A_{12}$ is unstable, adjust a, b, and c, and go to step 4.
8. Form L(s). Search for d such that $\|DLD^{-1}\|_\infty$ is minimized. If there is a significant change in d, go to step 3.
9. If the disturbance rejection and stability robustness are inadequate, select a lower δ and go to step 2.

There is a great deal of freedom suggested in steps 5 and 7. The manner in which these parameters are adjusted is dependent on the structure of the actual system and on any additional design constraints that might exist. System structure determines how the parameters affect the solution of the Riccati equations. Any extra freedom in parameter selection may be used to minimize observer gains.

6.3 Numerical examples

Two examples are discussed in this section. The first is observer design with Claim 3. The second design uses Claim 2 for a dual observer.

Example 1.

Suppose the plant of Figure 5 is described by:

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & -1 \\ -2-\Delta a & -1-\Delta a \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\mathbf{u}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\mathbf{w}(t)$$

$$\mathbf{z}_1(t) = [1 \quad 1]\ \mathbf{x}(t) \qquad (6.1)$$

$$\mathbf{z}_2(t) = \mathbf{u}(t)$$

$$\mathbf{y}(t) = [1 \quad 1]\ \mathbf{x}(t)$$

A low-order compensator is to be designed so that the closed-loop system is stable and the transfer function between $\mathbf{w}_1$ and $\mathbf{z} = [\mathbf{z}_1{}^T\ \mathbf{z}_2{}^T]^T$ is bounded above by $\gamma = 1$ for all $\Delta a \in [-0.5,\ 0.5]$.

For this problem, a first order observer may be designed. The perturbations in the state matrix may be separated out as:

$$\mathbf{A} + \Delta\mathbf{A} = \begin{bmatrix} 0 & -1 \\ -2 & -1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} D[-0.5 \quad -0.5]$$

with $|\Delta| < 1$. The state feedback Riccati equation (4.12) is then:

$$P_1\mathbf{A} + \mathbf{A}^T P_1 + Q_1 + P_1(R_1 - R_2)P_1 = 0$$

where

$$Q_1 = d_1{}^{-2}H_1{}^TH_1 + d_3{}^{-2}\rho^2\tilde{E}^T\tilde{E}$$

$$= d_1{}^{-2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + d_3{}^{-2}\rho^2\begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}$$

$$R_1 = \delta^{-2}[d_1{}^2\gamma^{-2}G_1G_1{}^T \quad d_2{}^2\tilde{D}\tilde{D}^T]$$

$$= \delta^{-2}(d_1{}^2\gamma^{-2} + d_2{}^2)\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$R_2 = d_1{}^2B(H_2{}^TH_2)^{-1}B^T$$

$$= d_1{}^2\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

It has been found that if the main diagonal elements of $R_1 - R_2$ are negative, it is more likely that (6.1) has a solution. Thus, a starting point can be used for the search algorithm:

$$\delta^{-2}(d_1{}^2\gamma^{-2} + d_2{}^2) - d_1{}^2 < 0$$

For this inequality to be satisfied, there is an additional condition that:

$$\delta\gamma > 1$$

It has also been determined that is is not necessary to search over both d_1 and d_2, so d_1 is arbitrarily fixed at $d_1 = 1$.

For the observer design, a nonsingular S must be selected so that $CS = [I \quad 0]$. It is also important to take into account that matrices D_1 and A are dependent on S^{-1}. For this example:

$$S = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$$

This gives:

$$A = S^{-1}AS = \begin{bmatrix} 2 & 2 \\ -2 & -3 \end{bmatrix}$$

$$D = S^{-1}[d_1c^{-1}G_1 \quad d_2\tilde{D}]$$

$$= \begin{bmatrix} -d_1\gamma^{-1} & d_2 \\ -d_1\gamma^{-1} & -d_2 \end{bmatrix}$$

In spite of D_1 not having full column rank, $D_1 = -D_2$, so that Claim 3 can be applied with $V_2 = -1$ and $F = A_{22} - V_2 A_{12} = -1$.

For fixed L(s), $\|DLD^{-1}\|_\infty$ was minimized with respect to d_2 (algorithm step 7) via a modified Powell algorithm [4]. The algorithm was stopped when d_2 changed less than 0.1% between passes through the algorithm. Results were as follows:

$\delta = .900$; $\gamma = 1.12$; $\rho = 1$;

$\|DLD^{-1}\|_\infty = 0.901$;

$d_1 = 1$; $d_2 = 0.0611$;

$K_c = [281 \quad 149]$;

$$\mathbf{A} - \mathbf{B}K_c = \begin{bmatrix} 0 & 1 \\ -279 & -150 \end{bmatrix};$$

$$D = \begin{bmatrix} 0.891 & 0.0611 \\ -0.891 & -0.0611 \end{bmatrix};$$

$\mathbf{N} = 133$; $\mathbf{M} = 149$; $\mathbf{F} = -1$; $\mathbf{T} = [1 \quad 0]$; $K_f = 1.00$; $\mathbf{G} = 0$.

The closed-loop eigenvalues are {-1, -1.89, -148}.

Example 2.

This design problem demonstrates Claim 2 and the dual system approach of section V. It is adapted from an example in Yeh et al [12]; the system is illustrated in Figure 6 with the following data:

$\mathbf{A} = \text{diag}[1, -1, -2, -1]$;

$\Delta\mathbf{A} = \tilde{\mathbf{D}}\Delta\tilde{\mathbf{E}}$;

where,

$$\tilde{\mathbf{D}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}; \ \Delta = \begin{bmatrix} \Delta_1 & 0 & 0 \\ 0 & \Delta_2 & 0 \\ 0 & 0 & \Delta_3 \end{bmatrix}; \ \tilde{\mathbf{E}} = \begin{bmatrix} 0.4 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0.2 & 0 \end{bmatrix};$$

with $|\Delta_i| < 1$, $i = 1, 2, 3$;

$$C = H_1 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

$$B = G_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix};$$

$G_2 = I.$

In [12], the surrogate system approach was used to design a full-order compensator so that the closed-loop system was stable and the transfer function T_{zw} between $w = [w_1{}^T \ w_2{}^T]^T$ and z_1 was bounded above by $\gamma = 1.7$ for all $|\Delta_i| < 1$. In this example, it is to be seen if the same level of performance can be achieved with a second order observer.

The first step here is to find the dual of the above system. This can be done by making the substitutions: $A := A^T$, $C := B^T$, $B := C^T$, $\tilde{E} := \tilde{D}^T$, $\tilde{D} := \tilde{E}^T$, $G_1 := H_1{}^T$, $H_1 := G_1{}^T$, $H_2 := G_2{}^T$.

The design process yielded the following:

$$S = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix}$$

$\delta = 3;\ \gamma = 1.7;\ \rho = 1;$

$\|DLD^{-1}\|_\infty = 1.24;$

$d_1 = 1;\ d_2 = 1.04;\ a = 12.51;\ b = 0.400;$

$$K_c = \begin{bmatrix} 5.632 & 0.373 & 0.065 & -0.041 \\ 0.168 & 0.745 & 0.408 & 0.270 \end{bmatrix}$$

$$A - BK_c = \begin{bmatrix} -4.632 & -0.373 & -0.065 & 0.041 \\ -0.168 & -1.745 & -0.408 & -0.270 \\ -0.168 & -0.745 & -2.408 & -0.270 \\ -5.632 & -0.373 & -0.065 & -0.960 \end{bmatrix}$$

$$D = \begin{bmatrix} 0.611 & 0.611 & 0.778 & 0.195 & 0.000 \\ 0.611 & 0.611 & 0.000 & 0.000 & 0.389 \\ 0.000 & 0.611 & 0.000 & 0.195 & 0.000 \\ 0.000 & 0.611 & 0.000 & 0.000 & 0.389 \end{bmatrix}$$

$$V_2 = \begin{bmatrix} -0.655 & -0.344 \\ -0.280 & -0.039 \end{bmatrix}$$

$$T = \begin{bmatrix} 0.655 & 1.655 & -0.344 & -0.344 \\ 0.280 & 0.280 & 0.961 & -0.039 \end{bmatrix}$$

$$F = \begin{bmatrix} -2.309 & 0.344 \\ -0.561 & -1.961 \end{bmatrix}$$

$$G = \begin{bmatrix} 0.311 & 1.311 \\ 0.241 & 1.241 \end{bmatrix}$$

$$N = \begin{bmatrix} -5.259 & 0.106 \\ 0.576 & 0.138 \end{bmatrix}$$

$$K_f = \begin{bmatrix} 0.108 & -1.405 \\ 1.527 & -0.939 \end{bmatrix}$$

$$M = \begin{bmatrix} 9.045 & -1.844 \\ -0.248 & 0.473 \end{bmatrix}$$

These gains are used in the dual observer of (5.2). Eigenvalues of A_{cl} are $\{-1.141, -1.313, -2.607, -4.683, -2.135 \pm j0.403\}$.

The bound on $\|T_{zw}\|_\infty$ is only guaranteed to be $\gamma\|DLD^{-1}\|_\infty = 2.11$ for any $|\Delta_i| < \rho/\|DLD^{-1}\|_\infty = 0.806$. Of course, there is conservatism due to covering a set of real perturbations with a set of complex perturbations and approximating $\|L\|_\mu$ by an upper bound.

VII. DISCUSSION

It has been shown how low-order observers can be designed for robust performance of uncertain systems. The procedure which was partially stated in [9, 13] has been completed by the use of another Riccati equation or a linear equation to be solved. A set of design parameters has been exploited in algorithms for maximizing a robustness measure in a process reminiscent of D-K iterations for full-order controller design. These algorithms have been demonstrated in numerical examples.

Conservatism of design results from the upper bounds used in the derivation of the Riccati inequalities and from H_∞ theory which assumes that all system perturbations are complex. This conservatism could be reduced by including more parameters in the design equations so that the structure of the uncertainties could be refined. It may also be possible to tighten up the bounds in the derivation by including an appropriate matrix Q in inequality (4.10).

Unfortunately, many systems cannot be accurately described using either of the models in Problems I and II. They may need to include both output disturbance and control of the input signal, or both ΔB and ΔC. It appears that the methods in this report could be adapted to some of these systems, but the order of the compensator would have to be increased. If the Problem I formulation were used, the minimal compensator order would be $n - m + \text{rank}[\tilde{G} \quad G_2]$. For Problem II this would be $n - p + \text{rank}[H_2^T \quad \tilde{F}^T]$. Neither method would be likely to produce a controller with low enough order. If the Riccati equation construction method of [9, 13] is to be used for a more general system, some alternative to the Luenberger observer is needed. This is an area for future research.

VIII. BIBLIOGRAPHY

[1] R. L. Dailey, *Lecture Notes for the Workshop on H_∞ and μ Methods for Robust Control: 1990 American Control Conference*, San Diego, CA, May 1990.

[2] J. C. Doyle, "Structured uncertainty in control system design," *Proceedings of the 24th Conference on Decision and Control*, Ft. Lauderdale, FL, pp. 260-265, Dec. 1985.

[3] J. C. Doyle, A. Packard "Uncertain multivariable systems from a state space perspective," *Proceedings of the 26th Conference on Decision and Control*, Los Angeles, CA, pp. 2147-2152, Dec. 1987.

[4] D. M. Himmelblau, *Applied Nonlinear Programming*, New York: McGraw-Hill, 1972.

[5] P. P. Khargonekar, I. R. Petersen and K. Zhou, "Robust stabilization of uncertain linear systems: quadratic stabilizability and H_∞ control theory," *IEEE Transactions on Automatic Control*, vol. AC-35, no. 5, pp. 356-361, Mar. 1990.

[6] M. M. Monahemi, J. B. Barlow, and D. P. O'Leary, "The design of reduced-order Luenberger observers with precise LTR," *Proceedings of the 1991 Guidance, Navigation and Control Conference*, Paper number AIAA-91-2731, August 1991, New Orleans.

[7] H. H. Niemann, P. Sogaard-Andersen, and J. Stoustrup, "Loop transfer recovery for general observer architectures," *International Journal of Control*, vol. 53, no. 5, pp. 1177-1203, 1991.

[8] J. O'Reilly, *Observers for Linear Systems*, London: Academic Press, 1983.

[9] J. L. Rawson, *A Report on Robust Control Design for Structured Uncertainties*, submitted to Flight Control Division, Flight Dynamics Laboratory, Wright-Patterson AFB, OH, August 1991.

[10] R. J. Veillette, J. V. Medanic and W. R. Perkins, "Robust stabilization and disturbance rejection for systems with structured uncertainty," *Proceedings of the 28th Conference on Decision and Control*, Tampa, FL, pp. 936-941, Dec. 1989.

[11] J. C. Willems, "Leasts squares stationary optimal control and the algebraic Riccati equation," *IEEE Transactions on Automatic Control*, vol. AC-16, no. 6, pp. 621-634, Dec. 1971.

[12] H. H. Yeh, S. S. Banda, A. C. Bartlett, and S. A. Heise, "Robust design of multivariable feedback systems with real parameter uncertainty and unmodelled dynamics," *Proceedings of the 1989 American Control Conference*, Pittsburgh, PA, pp. 662-670, June 1989.

[13] H. H. Yeh, J. L. Rawson, and S. S. Banda, "Robust control design with real parameter uncertainties," *Proceedings of the 1992 American Control Conference*, Chicago, IL, pp. 3249-3256, June 1992.

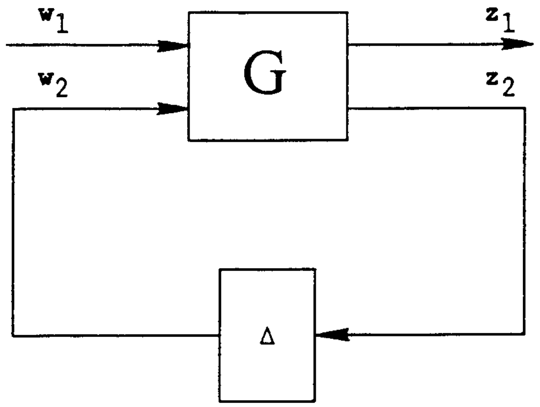


Figure 1. Perturbed System

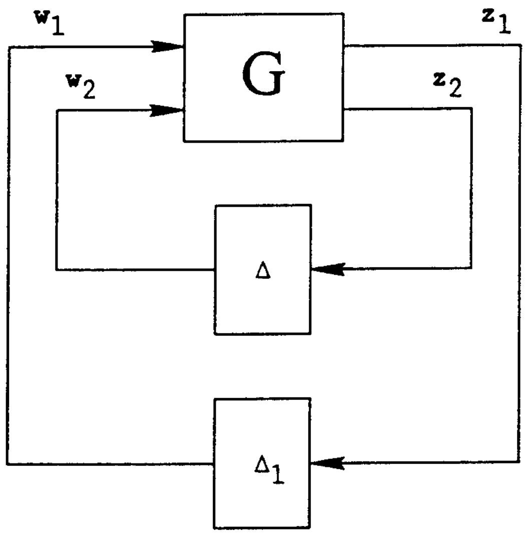


Figure 2. Perturbed System with Fictitious Uncertainty

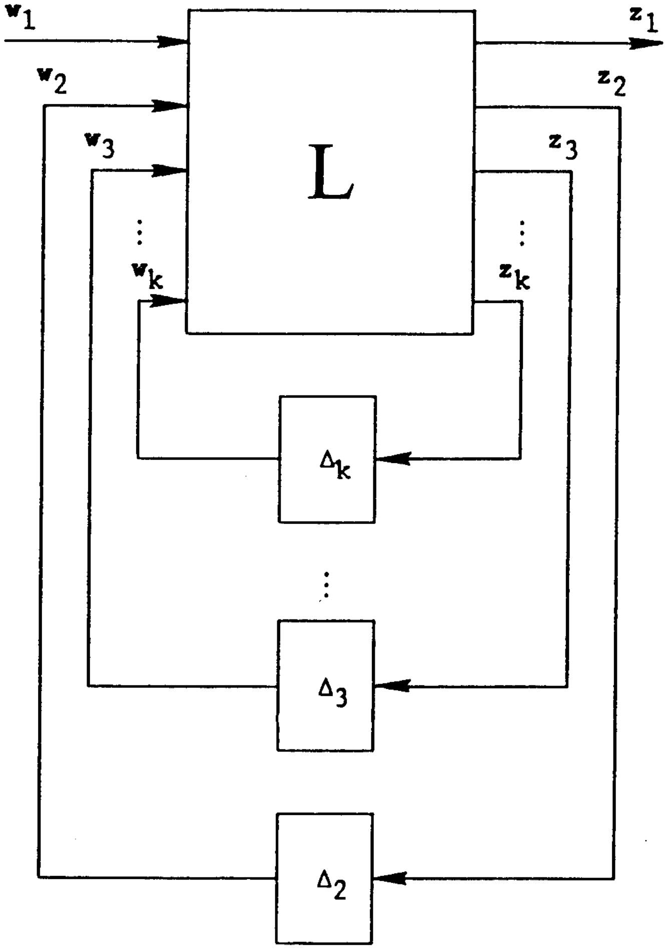


Figure 3. Perturbed System for Theorem I

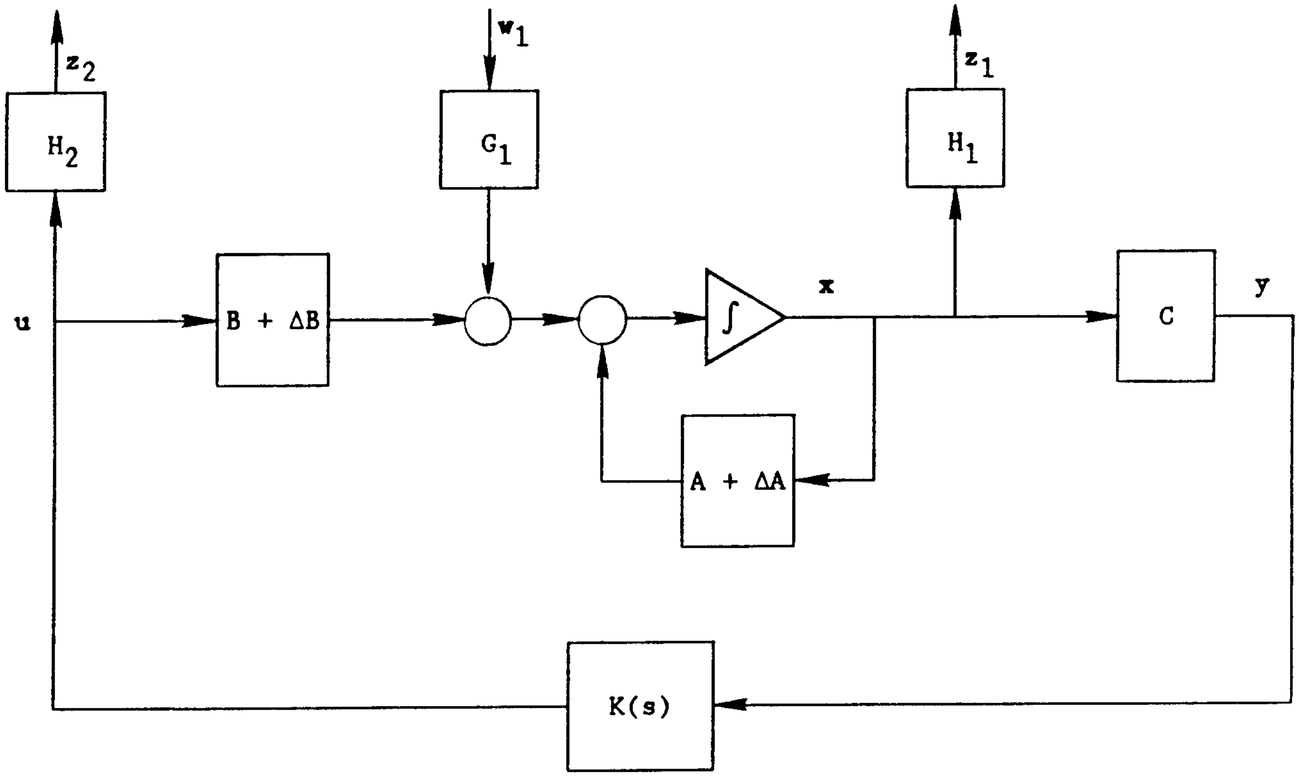


Figure 4. Closed-Loop Control System

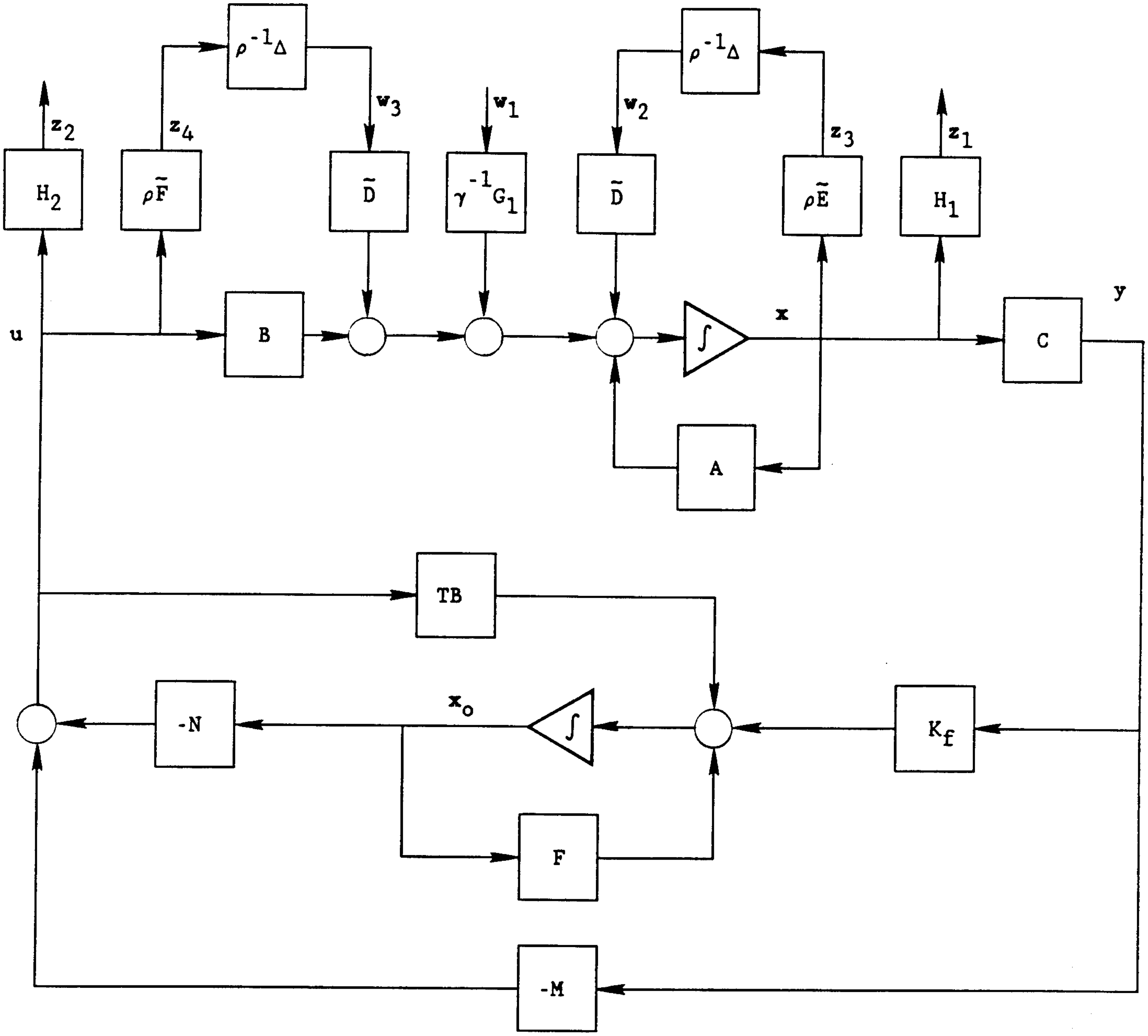


Figure 5. Perturbed System with Observer

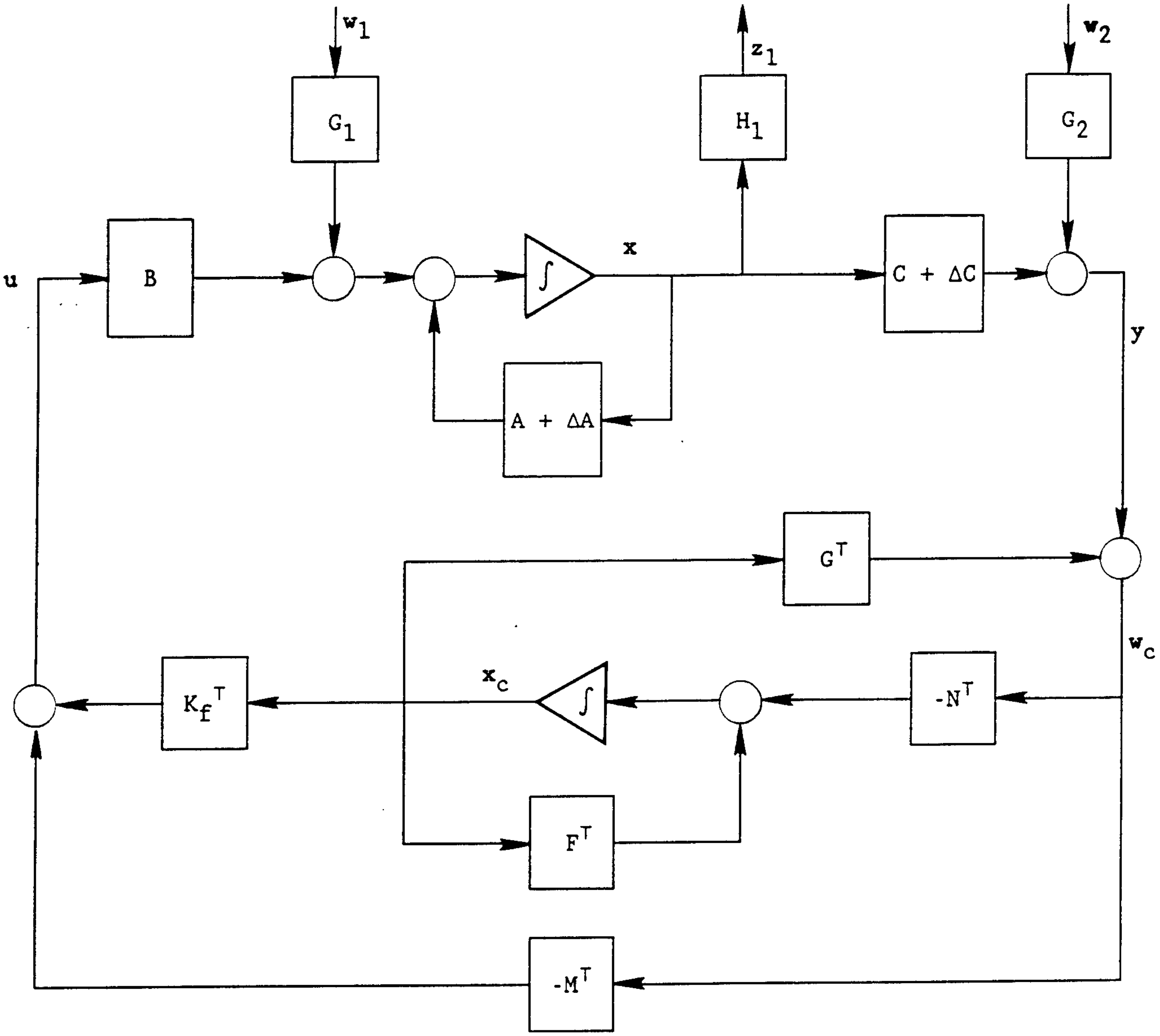


Figure 6. Dual System

CONFORMATIONAL STRUCTURE AND DYNAMICS IN PERFLUOROPOLYALKYLETHER LUBRICANTS

Martin Schwartz
Professor
Department of Chemistry

University of North Texas
215 W. Sycamore
Denton, TX 76203

Final Report for:
Research Initiation Program
Wright Laboratory

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

and

University of North Texas

December 1992

CONFORMATIONAL STRUCTURE AND DYNAMICS IN PERFLUOROPOLYALKYLETHER LUBRICANTS

Martin Schwartz
Professor
Department of Chemistry
University of North Texas

Abstract

Molecular Orbital calculations have been performed on perfluoroethylmethyl ether [PFEME] and perfluorodimethoxymethane [PFDMOM] in order to determine the geometries and energies of their equilibrium and transition state structures.

In PFEME it was found that the CCOC skeleton in the "trans" conformer is twisted by 17° from 180°, and there is a similar rotation of the perfluoromethoxy fluorines about the terminal CC bond, giving the molecule an overall helical structure. It was also determined that the energy of the gauche conformers (relative to the trans form) is unusually high,resulting in virtually 100% trans conformers at normal temperatures.

The structure of "trans" PFDMOM was also found to be helical with a twist of both the COCO skeletal and FCOC terminal dihedral angles from 180°. It was further observed that rotation about both of the internal CO bonds is much more facile (low barriers and roughly equal equilibrium energies) than the equivalent internal rotation in PFEME.

These results indicate that the rigidity and, hence, viscosity of linear perfluoropolyalkylethers should rise with an increasing proportion of -OCCO- linkages in the chain, a conclusion that is in agreement with experimental observations. These quantum mechanical data will be utilized in future molecular dynamics simulations of the molecular motions and bulk properties of PFPAE liquids.

CONFORMATIONAL STRUCTURE AND DYNAMICS IN PERFLUOROPOLYALKYLETHER LUBRICANTS

Martin Schwartz

I. INTRODUCTION

Perfluoropolyalkylether (PFPAE) fluids possess the viscoelastic, thermal and lubricity properties necessary to serve as effective, stable liquid phase lubricants.[1,2] No currently available commercial PFPAE lubricants, however, are capable of operation at the temperature extremes and oxidative conditions required for lubrication of high performance gas turbine jet engines.

The viscoelastic properties of polymer fluids such as the PFPAE's are, of course, intimately connected to the chain flexibility in these systems which is, in turn, dependent upon the potential energy barriers to internal rotation about single bonds in the polymer. During this past year, Ms. Christine Stanton (of my research group) and I have been working in collaboration with Dr. Harvey Paige (of the Materials Directorate at Wright Laboratory) in the application of quantum mechanical calculations to model the potential surface for internal rotation about the various C-C and C-O bonds in PFPAE homolog molecules. The goal of this research is to obtain a better understanding of chain mobility in the perfluoroether polymers. In this report, we describe the results of two investigations completed during the year which will be submitted for publication.[3,4] In addition, Chris Stanton and I, in collaboration with Dr. Paul Marshall of the UNT Chemistry Department, have completed a comparative *ab initio* and semiempirical quantum mechanical investigation of the rotational properties ethylmethyl ether. The results, to be submitted for publication,[5] will not be included in this report.

In the first investigation this year, we studied the torsional potential to internal rotation about the central C-O bond in perfluoroethylmethyl ether

(PFEME), $CF_3CF_2OCF_3$. The results have been compared to those from earlier calculations on perfluorobutane[6,7] (PFB) and ethylmethyl ether[8] (EME). We have also completed a second study of internal rotation about about the two central C-O bonds in perfluorodimethoxymethane (PFDMOM), $CF_3OCF_2OCF_3$. These results have been compared both to those reported for dimethoxymethane[9] (DMOM) and to our work on PFEME.

II. CALCULATIONS

Ab initio molecular orbital calculations were performed on a Cray X-MP/216 computer using the *Gaussian 90* MO program.[10]

A. PFEME

Equilibrium and saddle point geometries were gradient optimized[11] at the SCF level using the 6-31G(d)[12,13] basis set. Hartree-Fock energies were also obtained with the 6-311G(d)[14] basis, using the 6-31G(d) geometries. In addition, single point second order Møller-Plesset[15] (MP2) energy calculations were performed with both basis sets.

Vibrational frequencies were calculated for all conformers using the 6-31G(d) basis set. Transition state geometries were confirmed by the presence of a single imaginary frequency.

B. PFDMOM

Since the skeleton conformation of PFDMOM depends upon the dihedral angles of both of the central C-O bonds, the calculation was performed in two parts. First, a grid of energies was established with all combinations of ϕ_1 and ϕ_2 ranging from -180° to +180° in increments of 30°. Because of the large number of independent conformations (thirteen), the geometry optimizations were performed with the 6-31G*(C,O) basis set, which is the 6-31G basis[12,13] augmented by polarization functions on the carbon and oxygen atoms. This was followed by single point energy calculations with the 6-31G* basis using the 6-31G*(C,O) geometries; for brevity, the energies from these calculations are not contained in this report.

Second, once the approximate positions of the potential energy minima and barriers were located, a number of the precise stationary states were obtained using the 6-31G* basis set; these results are contained in Table 4. Additional single point MP2 energies were obtained for all grid points and stationary states. However, again for brevity, these are not shown.

RESULTS AND DISCUSSION

A. PFEME

Geometries

Selected geometric parameters for all equilibrium and transition state conformations of PFEME are displayed in Table 1. In the table and following discussion, the atoms are numbered as shown in Figure 1A.

The most striking result for the optimized geometries is that the skeletal dihedral angle, $\phi(C_1C_2OC_3)$, of the equilibrium trans conformer is twisted from 180°, by approximately 17°. The perfluoromethoxy group in this rotamer is similarly twisted [$\phi(F_{3A}C_3OC_2)=162°$] from a purely trans configuration, giving an overall helical structure to the molecule; this rotamer is labeled as "Twist-Trans" , TT, in Table 1. Similarly, the stable gauche conformer is labelled G, and the three transition states are (T)* [$\phi=180°$], (GT)* [$180° < \phi < 60°$] and (GG')* [$\phi=0°$].

Dixon[6,7] and Van Catledge[6] have earlier reported that the structure of the trans conformation of PFB is helical, characterized by a twist of the carbon skeleton, $\phi(CCCC)=164.6°$,[7] and rotation of the fluorines about both terminal CC bonds, $\phi(FCCC)=170.2°$.[7] They have suggested that the calculated helicity in PFB, and in other perfluoroalkanes, may result from a tendency to lower the repulsive interactions between CF dipoles on the 1 and 3 carbons, which are precisely parallel in the perfectly trans structure.

Most significantly, while fluorines on the perfluoromethoxy carbon of PFEME are twisted from 180° in the TT rotamer (*vide supra*), there is virtually no rotation of the fluorines on the C_1 carbon about the C_1C_2 axis

Table 1. Conformation Dependence of Selected Geometric Parameters in PFEME[a,b]

Parameter	TT	(T)*	(GT)*	G	(GG′)*
$R(C_1C_2)$	1.529	1.530	1.537	1.535	1.543
$R(C_2O)$	1.358	1.358	1.365	1.364	1.363
$R(C_3O)$	1.358	1.359	1.355	1.355	1.351
$R(C_1F_{1A})$	1.311	1.311	1.311	1.310	1.310
$R(C_1F_{1B})$	1.310	1.310	1.310	1.310	1.312
$R(C_1F_{1C})$	1.310	1.310	1.311	1.310	1.309
$R(C_2F_{2A})$	1.319	1.320	1.314	1.314	1.320
$R(C_2F_{2B})$	1.320	1.319	1.318	1.319	1.318
$R(C_3F_{3A})$	1.301	1.301	1.301	1.308	1.301
$R(C_3F_{3B})$	1.309	1.308	1.307	1.313	1.309
$R(C_3F_{3C})$	1.307	1.308	1.310	1.301	1.308
$<(C_1C_2O)$	107.9	107.5	114.4	115.1	119.8
$<(C_2OC_3)$	121.4	122.5	125.7	125.2	130.3
$<(C_2C_1F_{1A})$	109.1	109.0	108.9	109.0	107.9
$<(OC_3F_{3A})$	107.1	106.6	106.5	106.9	106.6
$\phi(C_1C_2OC_3)$	162.8	180.0	76.8	61.7	0.0
$\phi(F_{1A}C_1C_2O)$	179.1	180.1	165.8	171.1	169.2
$\phi(F_{3A}C_3OC_2)$	162.3	180.0	172.8	160.4	166.1

a) Bond lengths are in angstroms and angles in degrees.
b) See Figure 1A for atom numbering.

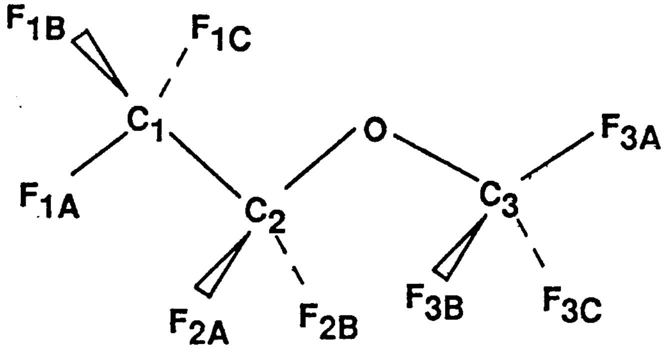


Fig. 1A. Structure and atom numbering in trans PFEME

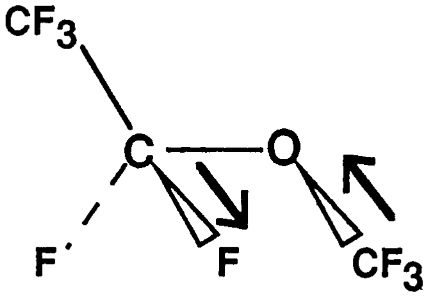


Fig. 1B. Structure of PFEME at the (GT)* transition state

[$\phi(F_{1A}C_1C_2O)$=179.1°]. This result lends supportive evidence for Dixon's explanation of the helicity in perfluoroalkanes since, as seen clearly in Figure 1A, 1,3 CF bond dipole repulsions are absent in trans PFEME.

Pacansky, *et al.*[16] have reported the geometry of the TT rotamer of PFEME determined with the 3-21G[17] basis set. Their CC bond length is shorter (by 0.03 Å), and their calculated CO and CF bonds are longer (by 0.01-0.02 and 0.01-0.03 Å, respectively) than those found here. Bond angles agree to within 1-2° on average. The calculated twist of the skeleton using the smaller basis was similar to that obtained here, with $\phi(C_1C_2OC_3)$=164.6°, but the rotation of the perfluoromethoxy group was substantially higher [$\phi(F_{3A}C_3OC_2$=147.6°].[16] Smart and Dixon[18] have also calculated the geometry of the TT equilibrium conformer using a DZ+D_C basis set, which is a double zeta basis with polarization functions on carbon. Their results agree even more closely with those obtained here, particularly for the dihedral angles [$\phi(C_1C_2OC_3)$=160.9° and $\phi(F_{3A}C_3OC_2$=161.2°]. The observed differences, which are relatively small, are expected when comparing geometries calculated with and without polarization functions on the various atoms.

Not surprisingly, since the (T)* energy barrier is quite low (*vide infra*), the geometric parameters of the TT equilibrium conformer and the (T)* saddle point agree closely (Table 1), with the exception of the dihedral angles, which are all 180° in the transition state.

From the table, it is observed that both the COC and CCO angles vary in the order, $(T)^* \approx TT < (GT)^* \approx G < (GG')^*$, reflecting the increased 1,4 atomic repulsions with decreasing CCOC dihedral angle. Similarly, $R(C_1C_2)$ and $R(C_2O)$ exhibit modest increases with diminishing dihedral angle; this trend is not, however, observed for $R(C_3O)$, which is slightly longer in the TT and (T*) rotamers.

Significantly, it may be seen in the last column of the table that, unlike in PFB,[6,7] the terminal fluorines remain twisted from 180° in the (GG')* conformation of PFEME. This may reflect greater F-F repulsions in the syn conformation of the ether, resulting from the shorter CO bond lengths.

It is of interest to compare the geometry of PFEME with that reported earlier for ethylmethyl ether (EME),[8] using the same [6-31G(d)] basis set. Both of the CO bond lengths in the fluorinated ether are substantially shorter than in EME, in which $R(C_2O)=1.417$ Å and $R(C_3O)=1.390$ Å in the trans conformer; in contrast, the CC bond lengths are approximately the same [$R(C_1C_2)=1.516$ Å in EME]. Too, the COC bond angle is markedly greater in PFEME [$<(C_2OC_3)=114.2°$ in trans EME], whereas the CCO bond angles are approximately equal in the two species [$<(C_1C_2O)=108.6°$ in EME].

The shortening of the CO bond lengths and increase in the COC angle in the fluoroether may be explained on the basis of relative bond polarities. The calculated difference between the carbon and oxygen Mulliken charges [q(C)-q(O)] is 1.7-2.1 in PFEME,[19] whereas it is only 0.6-0.8 in EME.[20] Therefore, one expects a more ionic and, hence, shorter CO bond in the fluorinated compound. Consequently, the COC bond angle should increase to moderate the otherwise enhanced electrostatic and/or van der Waal's interactions between fluorines on the 2 and 4 carbons. The same trends in bond lengths and angles and in Mulliken charges are found in calculations on dimethyl ether and its fluorinated analogue.[21]

Energies

Total HF and MP2 energies (in au's) of the five equilibrium and saddle point conformations of PFEME, calculated with the two basis sets (using the 6-31G(d) geometries) are presented in Table 2A. Energies (in kcal/mol) relative to the TT rotamer are given in 2B. For comparison, relative conformational energies for EME[8] and PFB[7,22] are also contained in the latter half of the table.

One sees from the table that the (T)* transition state is only slightly higher in energy than the TT minimum, with $\Delta E \approx 0.3$-0.4 kcal/mol at the SCF

Table 2. Calculated Conformational Energies in PFEME[a]

Method	TT	(T)*	(GT)*	G	(GG′)*
A. Total Energies (hartrees)					
HF/6-31G(d)	-984.007 346	-984.006 812	-984.002 264	-984.002 692	-983.995 404
HF/6-311G(d)	-984.268 451	-984.267 804	-984.262 704	-984.263 112	-984.255 622
MP2/6-31G(d)	-985.904 547	-985.903 717	-985.900 419	-985.901 509	-985.894 362
MP2/6-311G(d)	-986.470 460	-986.469 558	-986.465 422	-986.466 529	-986.458 868
B. Relative Energies (kcal/mol)					
PFEME/HF/6-31G(d)	0.00	+0.34	+3.19	+2.92	+7.49
PFEME/HF/6-311G(d)	0.00	+0.41	+3.61	+3.35	+8.05
PFEME/MP2/6-31G(d)	0.00	+0.52	+2.59	+1.91	+6.39
PFEME/MP2/6-311G(d)	0.00	+0.57	+3.16	+2.47	+7.27
EME/HF/6-31G(d)[b]	--	0.00	+2.56	+1.67	+6.84
PFB/HF/DZ+P[c]	0.00	+0.15	+2.35	+1.47	+8.30
EME/MP2/6-31G(d)[b]	--	0.00	+2.67	+1.40	+7.00
PFB/MP2/DZ+P[c]	0.00	+0.38	+2.41	+1.48	+8.02

a) All energies were calculated using the HF/6-31G(d) geometries.
b) From Ref. 5.
c) From Ref. 4.

level, and ΔE≈0.5-0.6 cal/mol with correlation energy corrections.

One observes, also, that relative energies of the G, (GT)* and (GG′)* conformations are all lower at the MP2 than the HF level, within a given basis set. This trend may be explained by analysis of Table 2A, from which it is found that, with the 6-31G(d) basis for example, the correlation energy correction (in kcal/mol) varies in the order, (T)* [-1190.2 < TT [-1190.4] < (GT)* [-1191.0] < G [-1191.4] < (GG′)* [-1191.5]. It is reasonable that inclusion of electron correlation will preferentially stabilize the more structurally congested conformations [lower ϕ(CCOC)] in which the electronic repulsions are greatest.

The principal effect of increasing the size of the basis is to increase relative energies of the G, (GT)* amd (GG′) conformations. Again, this trend arises from analysis of Table 2A, where it is found that the larger basis sets preferentially stabilize the TT and (T)* states in comparison to the above conformations. Thus, the effects of increasing basis size and introducing electron correlation tend to offset one another.

A principal focus of this investigation is to perform a comparison of the torsional potential for rotation about the central bond in PFEME to those obtained earlier in the fluoroalkane and in the nonfluorinated ether. The three energy curves, calculated at the HF/6-31G(d) level for PFEME and EME and at the equivalent HF/DZ+P[22] level for PFB, are shown superposed (displaced from one another for clarity) in Figure 2. As noted, the SCF and MP2 energies of the stationary state conformers of the latter two molecules are also tabulated at the bottom of Table 2B.

One observes from both figure and table that the rotational potential of PFEME is markedly different from that of either of the other molecules in the vicinity of the G and (GT)* conformations. Most striking is that the dihedral angle of the (GT)* barrier is shifted by over 40° below the nominal angle of 120° (see also Table 1). In order to verify this highly unusual behavior, we

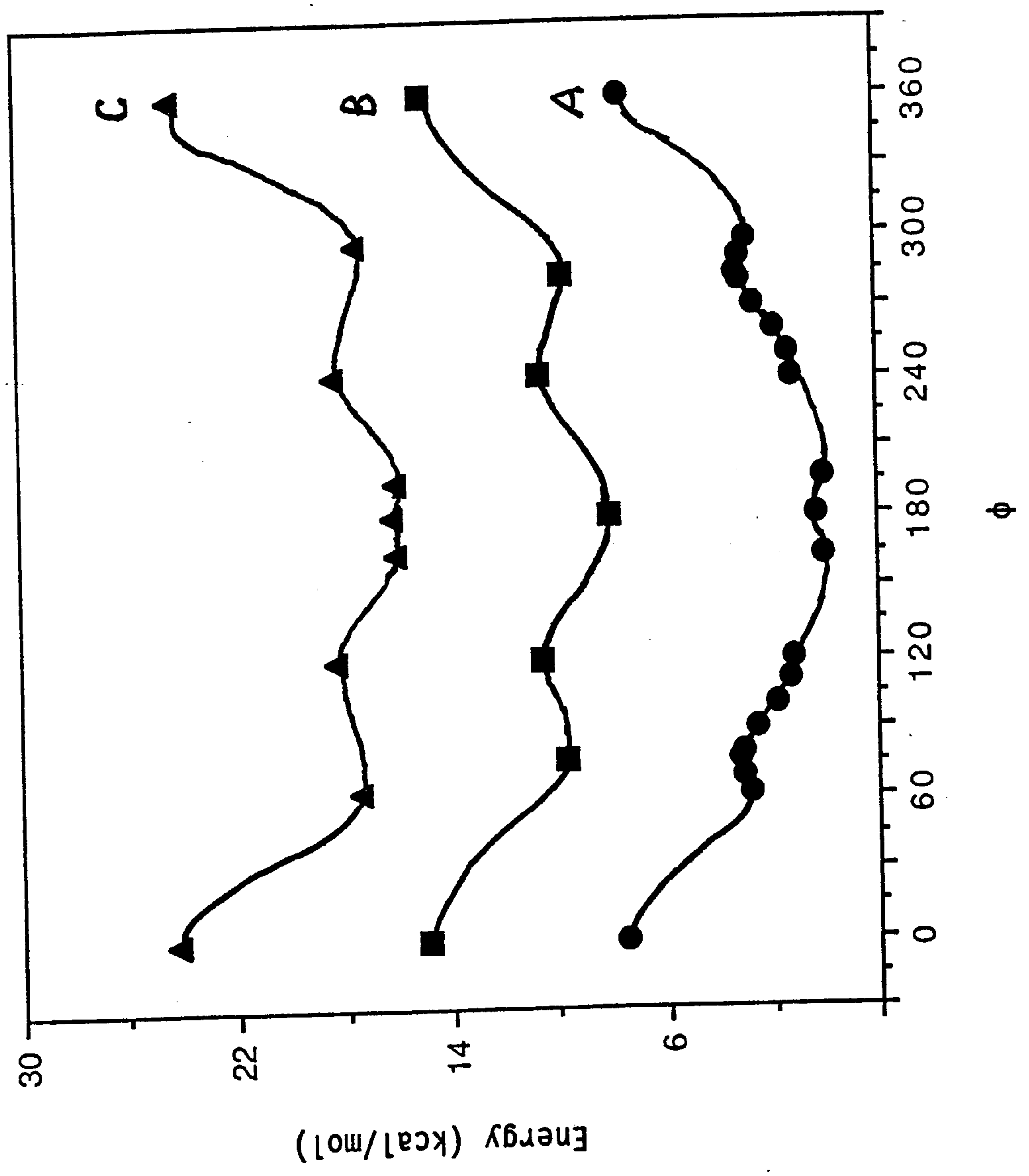


Fig. 2. The torsional potential in (A) PFEME - Circles; (B) EME - Squares (from ref. 5); (C) PFB - Triangles (from ref. 4).

The potential curves for EME and PFB are displaced upward by 8 and 16 kcal/mol, respectively, for clarity of presentation.

have performed additional HF/6-31G(d) optimizations at various fixed values of $\phi(C_1C_2OC_3)$ ranging from 70° to 120°. The resultant energies, plotted also in Figure 2, confirm the position of the (GT)* transition state.

A possible explanation for the large shift in the torsional angle of this transition state may be found by examination of the structure of PFEME at $\phi(C_1C_2OC_3)=120°$ (Figure 1B). One observes from the figure that, at this angle, the C_3O and one of the C_2F bond dipoles are precisely antiparallel, which would lead to a stabilization of this configuration, not expected in either PFB or EME, and thus result in a shift of the position of the energy maximum.

A very important feature of the potential energy diagram of PFEME is that the relative energy of its G equilibrium conformers is greater than that in either of the two other molecules. The explanation for this difference may reside in the closer approach distance between terminal fluorine atoms, resulting from the lower CO and greater CF (compared to CH) bond lengths, which would destabilize the gauche conformation in PFEME.

The consequence of the unusually high energy of the two G conformations of PFEME is that, from the Boltzmann distribution, one would have more than 98% of the molecules in the T conformation. Thus, if a PFPAE has numerous $-O-CF_2CF_2-O-CF_2-O-$ linkages, its conformation in the liquid would tend to be elongated (higher percentage of trans bonds) and stiff (since only one of the three conformations about the CO bond is thermally accessible).

As seen clearly in Figure 2, the overall effect of the displaced position of the (GT)* transition state and the high energy of the G conformer is to result in a very shallow, narrow energy minimum for the gauche conformation of PFEME. From Table 2B, the energy barrier, E[(GT)*]-E[G], varies from only 0.26 to 0.69 kcal/mol, dependent upon the level and method of calculation.

Vibrational frequencies for all five stationary states of PFEME were calculated using the 6-31G(d) basis set and are displayed, for reference, in

Table 3; these values have been multiplied by the normal 0.90 scale factor to account for the effects of vibrational anharmonicity and electron correlation.

B. PFDMOM

Selected geometric parameters (and energies) of various of the stationary states in PFDMOM are displayed in Table 4. As observed in the above comparison of PFEME and EME, the CO bond lengths are shorter (by approximately 0.03 Å) and COC bond angles are greater (by roughly 8°) than those reported by Wiberg and Murcko[9] in their investigation of dimethoxymethane (DMOM). Again, these differences can be attributed directly to the greater bond ionicity and steric repulsions in the perfluorinated ether.

Both bond lengths and angles are relatively insensitive to the molecular conformation with the exception of $<(O_1C_2O_2)$, which rises if either or both of the torsional angles are 60° or 0°. This effect is due to repulsions between the two terminal perfluoromethyl groups.

It is most interesting that, as found for PFEME, the Trans-Trans form (nominally with ϕ_1=ϕ_2=180°) of PFDMOM is actually helical in character, with a rotation of the two terminal CF_3 groups [ϕ(FCOC)$\neq$180°] as well as a twist of both of the COCO dihedral angles from 180°, the value found in DMOM.[9] Significantly, there are two separate helical conformations for the Trans-Trans conformer [180,180(A) and 180,180(B)], which are close to one another in energy. Analysis of the two structures reveals that in the second configuration, there is an actual reversal in the direction of the helix between the two ends of the molecule. This helix reversal is also observed in the Trans-Gauche [180,60] conformation which, too, has two distinct conformations of similar energy (Table 4).

In Figure 3 is shown the potential energy curve for rotation of the second C-O dihedral angle in PDMOM while the first torsion is held constant at (A) ϕ_1=60° and (B) ϕ_1=180°, respectively. One observes quite clearly that when the first C-O bond is trans (180°), then the second bond is free to assume the

Table 3. Vibrational Frequencies in PFEME[a,b]

Vib. No.	TT	(T)*	(GT)*	G	(GG′)*
1	35	25i	28i	41	74i
2	54	57	41	53	48
3	83	109	113	102	89
4	122	117	174	158	148
5	202	211	189	197	199
6	215	218	221	213	218
7	301	301	308	312	318
8	327	323	335	332	338
9	343	344	344	342	346
10	364	364	363	365	366
11	419	420	440	434	421
12	507	515	457	458	489
13	515	515	514	516	509
14	542	541	546	547	539
15	587	588	585	585	590
16	612	614	606	608	615
17	644	649	650	647	617
18	669	667	706	709	682
19	745	745	724	718	717
20	837	834	793	789	776
21	905	905	929	932	941
22	1119	1121	1113	1114	1120
23	1224	1225	1232	1231	1238
24	1244	1240	1259	1261	1251
25	1280	1281	1271	1269	1262
26	1288	1288	1288	1284	1268
27	1291	1290	1290	1291	1292
28	1302	1301	1301	1305	1306
29	1343	1342	1349	1345	1332
30	1463	1463	1434	1437	1430

a) Frequencies are in units of cm.$^{-1}$

b) Calculated frequencies have been scaled by the factor 0.9.

Table 4. Stationary State Geometries and Energies in PFDMOM.[a]

	Nominal Dihedral Angles						
Param.	180,180(A)	180,180(B)	180,120	180,60(A)	180,60(B)	180,0	60,60
$R(C_1O_1)$	1.358	1.357	1.357	1.359	1.357	1.360	1.356
$R(C_1O_2)$	1.354	1.355	1.359	1.361	1.361	1.358	1.362
$R(C_2O_2)$	1.354	1.355	1.357	1.354	1.355	1.356	1.362
$R(C_3O_2)$	1.358	1.357	1.355	1.358	1.357	1.356	1.356
$<(O_1C_2O_2)$	105.0	104.8	106.4	109.7	108.6	110.8	114.2
$<(C_1O_1C_2)$	121.1	121.2	121.5	121.4	121.4	121.4	122.6
$<(C_2O_2C_3)$	121.1	121.2	122.7	122.2	121.8	124.9	122.6
$\phi(F_{1A}C_1O_1C_2)$	163.3	196.5	196.5	197.0	163.0	163.4	188.3
$\phi(F_{3A}C_3O_2C_2)$	163.2	163.5	179.2	157.7	195.8	178.5	188.3
$\phi(C_1O_1C_2O_2)$	160.2	195.7	191.7	196.4	161.6	161.1	68.6
$\phi(O_1C_2O_2C_3)$	160.2	164.4	123.5	47.7	84.8	2.4	68.6
E(au)+1058	-0.885960	-0.885408	-0.883838	-0.885956	-0.885426	-0.883248	-0.884607
ΔE(kcal/mol)[b]	0.00	0.35	1.33	0.00	0.34	1.70	0.85

a) Bond lengths are in anstroms and angles in degrees.
b) Energy relative to the 180,180(A) stationary state.

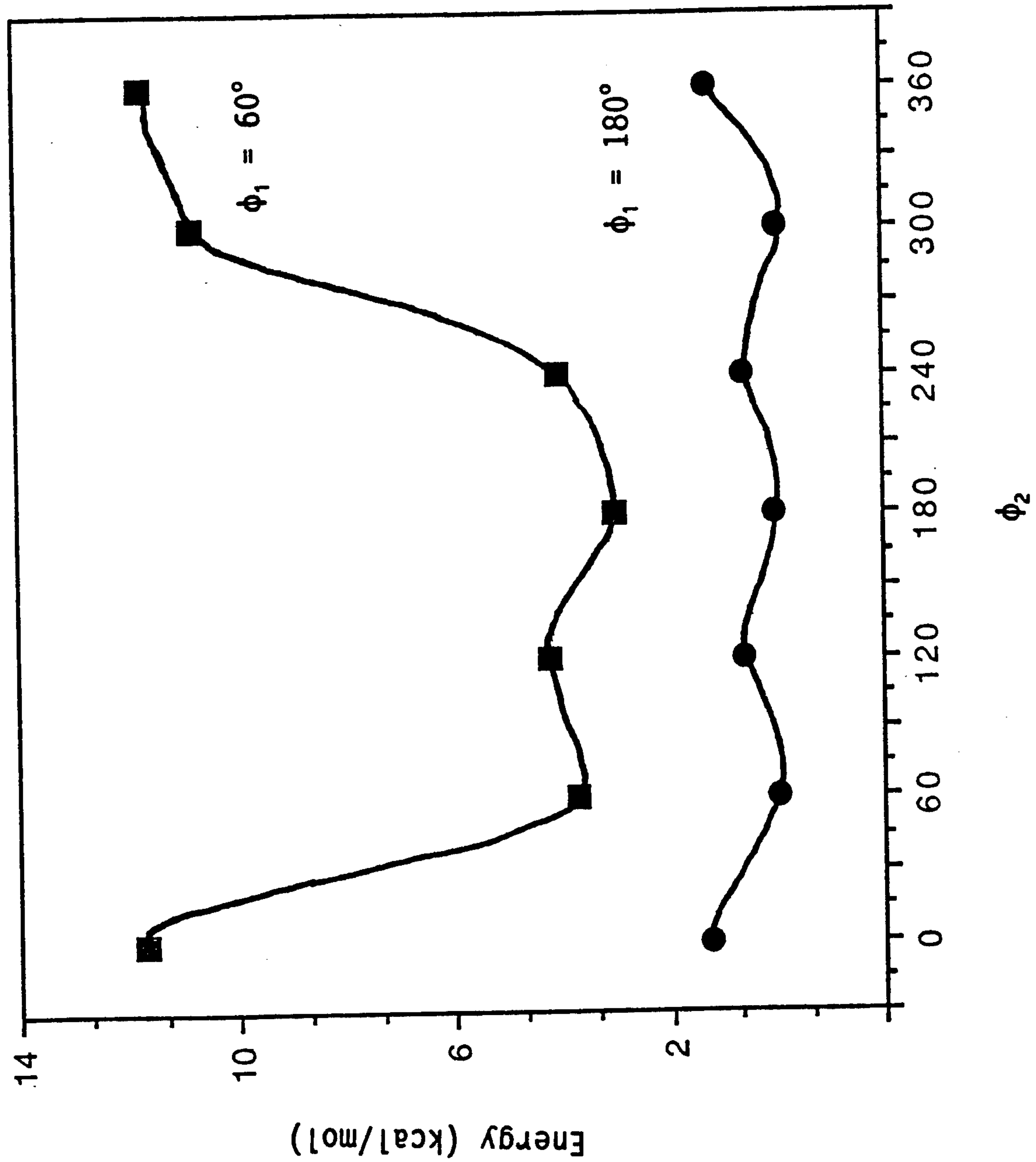


Fig. 3. The torsional potential for the rotation of ϕ_2 with (A) ϕ_1 = 60° and (B) ϕ_1 = 180°.

The potential curve for (A) has been displaced upward by 3 kcal/mol for clarity of presentation.

trans or either of the gauche conformations with equal facility [$\Delta E_{GT} \approx 0$]. Even when ϕ_1=60°, then ϕ_2 can be either 60° or 180°, although ϕ_2=-60° (=300°) is thermally forbidden [$\Delta E \approx 8$ kcal/mol]. The cause of the very high energy when ϕ_1=60° and ϕ_2=-60° is that in this conformation, the two terminal perfluoromethyl groups lie on the same side of the central O-C-O plane, which causes severe steric repulsions in the molecule.

A comparison of the potential curves for PFEME and PFDMOM reveals that, whereas only one of the three equilibrium conformations is populated in the former molecule, the CO bond conformation in the latter species can be in either two or all three of the equilibrium states. This result has substantial ramifications on the conformational flexibility of perfluoropolyalkylether polymers (*vide infra*).

IV. SUMMARY AND CONCLUSIONS

Molecular geometries and energies of all equilibrium and transition state conformations of perfluoroethylmethyl ether were determined by *ab initio* molecular orbital calculations using the 6-31G(d) basis set. The CCOC skeleton in the "trans" conformer is twisted by 17° from 180°. There is a similar rotation of the perfluoromethoxy fluorines about the terminal CC bond (by 18°), but no twisting of fluorines on the other terminus of the molecule. These results indicate that, as suggested earlier for perfluorobutane, the helical structure of the trans conformer in PFEME is induced by dipolar repulsions between CF bonds on alternant carbons.

The energy of the gauche conformers of PFEME (relative to the twisted trans structure) is unusually high [2.9 kcal/mol], from which it may be determined that one would have virtually 100% trans molecules at thermally accessible temperatures.

The structures and energies of perfluorodimethoxymethane [PFDMOM] were determined as a function of the dihedral angles about both of the central C-O bonds. As found for PFEME, the CO bonds were shorter and the COC angles greater than in the non-fluorinated species. It was also determined that, here too, the molecules adopt a helical configuration in the nominally Trans-Trans state.

It was also determined that in PFDMOM, if one of the two skeletal dihedral angles is 180° (trans), then the second angle can adopt any of the three equilibrium values with equal facility; if the first angle is 60°, then the second CO torsional angle can still adopt two of the three values. Thus, this molecule is substantially more flexible in PFEME, in which all of the CO bonds are trans.

The most significant conclusion is that if a perfluoropolyalkylether polymer has a high percentage of $-O-CF_2CF_2-O-$ linkages, one expects that it will be very rigid. If, in contrast, the majority of the linkages are of the type $-O-CF_2-O-$, then the polymer will exhibit a much higher degree of flexibility and, hence, lower viscosity. These conclusions can be used to explain the general empirical observation that, for a given molecular weight, the viscosity of a PFPAE fluid increases with rising C:O ratio, since this ratio is a measure of the relative number of the former type of linkage.

VI. FUTURE INVESTIGATIONS

We intend to extend the above studies to the molecule, perfluoro-1,2-dimethoxyethane, $CF_3OCF_2CF_2OCF_3$. This will permit us both to verify the results obtained this summer on PFEME and to determine the potential surface for rotation about C-C bonds in linear PFPAE's. We also wish to model the internal rotation in branched perfluoroethers (such as Krytox). Therefore, we shall study the molecules, $CF(CF_3)_2OCF_3$ and $CF_3OCF_2CF(CF_3)OCF_3$.

We plan to use the results of these quantum mechanical investigations to obtain accurate potential energies for bond stretching, bending and internal rotation, which will then be utilized in molecular dynamics simulations of perfluoroether polymers in order to obtain *a priori* predictions of the dependence of bulk properties such as viscosity, density, thermal expansion and isothermal compressibility as a function of molecular structure in PFPAE lubricants.

REFERENCES

1. Snyder, C. E., Jr.; Dolle, R. E., Jr. *ASLE Trans.* **1975**, *19*, 171.

2. Snyder, C. E., Jr.; Gschwender, L. J.; Tamborski, C. *Lubr. Eng.* **1981**, *37*, 344.

3. Stanton, C. L.; Paige, H. L.; Schwartz, M., "*Ab Initio* Study of Molecular Geometry and the Torsional Potential in Perfluoroethylmethyl Ether," *J. Phys. Chem.* (Submitted).

4. Stanton, C. L.; Paige, H. L.; Schwartz, M. A manuscript on the rotational potential surface of perfluorodimethoxy methane is in preparation for submission to *J. Am. Chem. Soc.*

5. Stanton, C. L.; Marshall, P.; Schwartz, M., "Comparison of *Ab Initio* and Semiempirical Methods in Determination of the Molecular Geometry and Rotational Barriers in Ethylmethyl Ether." To be submitted to *THEOCHEM.*

6. Dixon, D. A.; Van Catledge, F. A. *Int. J. Supercomput. Applic.* **1988**, *2*, 52.

7. Dixon, D. A. *J. Phys. Chem.* **1992**, *96*, 3698.

8. Tsuzuki, S.; Tanabe, K. *J. Chem. Soc., Faraday Trans.* **1991**, *87*, 3207.

9. Wiberg, K. B.; Murcko, M. A. *J. Am. Chem. Soc.* **1989**, *111*, 4821.

10. *Gaussian 90*, Revision F; Frisch, M. J.; Head-Gordon, M.; Trucks, G. W.; Foresman, J. B.; Schlegel, H. B.; Raghavachari, K.; Robb, M.; Binkley, J. S.; Gonzalez, C.; Defrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Martin, R. L.; Kahn, L. R.; Stewart, J. J. P.; Tolpiol, S.; Pople, J. A.; Gaussian, Inc.: Pittsburgh, PA, 1990.

11. Pulay, P. In *Applications of Electronic Structure Theory*; Schaefer, H. F., III, Ed.; Plenum Press: New York, 1977; p 153.

12. (a) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1982**, *56*, 2257. (b) Hariharan, P. C.; Pople, J. A.; *Theor. Chim. Acta* **1973**, *28*, 213.

13. Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265.

14. (a) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650. (b) McLean, A. D.; Chandler, G. S. *Ibid.* **1980**, *72*, 5639.

15. Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

16. Pacansky, J.; Miller, M.; Hatton, W.; Liu, B.; Scheiner, A. *J. Am. Chem. Soc.* **1991**, *113*, 329.

17. Pietro, W. J.; Francl, M. M.; Hehre, W. J.; Defrees, D. J.; Pople, J. A.; Binkley, J. S. *J. Am. Chem. Soc.* **1982**, *104*, 5039, and references contained therein.

18. Smart, B. E.; Dixon, D. A. "Heterolytic C-F Bond Energies and Stabilities of Poly(perfluoroethers," article preprint. **CHANGE** - Harvey, can we get a better reference?

19. Calculated Mulliken charges on the skeletal atoms of PFEME (the TT conformation) are: $q(C_1)=$ +1.08; $q(C_2)=$ +1.00; $q(O)=$ -0.70; $q(C_3)=$ +1.36.

20. Calculated Mulliken charges on the skeletal atoms of EME are: $q(C_1)=$ -0.49; $q(C_2)$=0.02; $q(O)=$ -0.60; $q(C_3)=$ -0.16. Williamson, C. L.; Marshall, P; Schwartz, M., unpublished results.

21. Williamson, C. L.; Paige, H. L.; Schwartz, M., unpublished results.

22. Conformational energies for PFB, taken from ref. 7 and contained in Table 2B, were obtained by single point calculations (using DZ+D_c geometries) with a DZ+P basis set, which is a double zeta basis with polarization functions on all atoms.

USE OF THRESHOLDED SPECKLE IN TESTING CCDS

Dr. Glenn D. Boreman
Principal Investigator
Associate Professor, EE

Martin Sensiper
Co-Principal Investigator

Alfred D. Ducharme
Graduate Research Assistant

University of Central Florida
Department of Electrical Engineering
Center for Research in Electro-Optics and Lasers
Orlando, FL 32816

Final Report for:
AFOSR Research Initiation Program (RIP)
# F49620-90-C-0076

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

October 1992

USE OF THRESHOLDED SPECKLE IN TESTING CCDS

Dr. Glenn D. Boreman, Martin Sensiper, and Alfred D. Ducharme
Center for Research in Electro-Optics and Lasers
University of Central Florida

Abstract

Spatial-frequency filtering of laser speckle patterns has proven to be a useful tool in the measurement of MTF for focal plane arrays. Intensity thresholding of the laser speckle patterns offers nearly an order of magnitude savings in digital storage space. The effect of this thresholding on the spatial-frequency power spectral density of the speckle pattern is investigated. An optimum threshold level is found that minimizes distortion of the power spectrum for the classes of speckle data used for MTF testing.

USE OF THRESHOLDED SPECKLE IN TESTING CCDS

Dr. Glenn D. Boreman, Martin Sensiper, and Alfred D. Ducharme

1. Introduction

This report describes the findings of research completed as a continuation of an ongoing collaboration between the Air Force Armament Lab, Eglin Air Force Base, and the University of Central Florida, Center for Research in Electro-Optics and Lasers (CREOL). The goal of this collaboration is to investigate the performance charaterization of focal-plane arrays (FPAs) using laser speckle. The use of laser speckle as a random test target allows a series of measurements to be performed which are intended to completely charaterize state-of-the-art solid-state cameras currently being used by the Air Force.

The measurement we are currently investigating is modulation transfer function (MTF). MTF describes the spatial frequency response of a CCD FPA for all frequencies. Each point in an MTF measurement quantifies the response of the CCD FPA for a sine-wave of a single spatial-frequency. MTF is used to measure the resolution capabilities of FPAs. Our previous work on FPA MTF[1-3] has shown that laser speckle provides a suitable test target for this measurement.

The current algorithm used to calculate the MTF for CCD FPAs requires a considerable amount of data for accurate results. For an MTF with 100 points, the algorithm uses 100, 512-by-512 pixel, 8-bit images. This equates to approximately 27 mega-bytes of digital storage space. Over time the storage of raw data could become expensive making it desirable to reduce the amount of space used for each measurement.

The size of the data arrays used in the calculation are fixed by the size of the CCD array being tested, meaning only two ways to reduce the data exist. First, the number of points used to plot the MTF could be minimized. This would reduce the accuracy of the calculation because the spacing of data points dictates the size fluctuation in MTF that can be detected. Second, the speckle intensity could be thresholded eliminating all but 2 of the 256 gray levels used to represent the digitized speckle field (Fig. 1). This binarization would transform the 8-bit data into 1-bit data, reducing the total amount of data by nearly an order of magnitude.

This report shows that the fidelity of the spatial-frequency power spectral density of the laser speckle intensity is maintained after a thresholding operation is performed. The effects of the intensity thresholding will be investigated and an optimal threshold value will be determined for the narrowband laser speckle used in the MTF calculation.

The second section of this report discusses the derivation of the spatial autocorrelation function for the continuous and thresholded intensity of the laser speckle. The derivation will be exemplified using two different cases. The first case will be a simple square aperture and the second will be a double-slit aperture used in the calculation of MTF.

The relationship between the autocorrelation and power spectral density of the thresholded laser speckle will be discussed in the third section. The same two examples will be used, and an optimal threshold level will be determined for each case based on a minimum-mean-squared-error analysis in their power spectral densities. A computer simulation of these two examples was performed to check the validity of this analysis and its results are given in section 4 in terms of power spectrum.

2. Autocorrelation of thresholded laser speckle

In this section we will derive the spatial autocorrelation function for intensity thresholded laser speckle. To simplify the reading, this function will be referred to as the *thresholded autocorrelation*. The derivation will begin with the spatial autocorrelation function for intensity distribution of nonthresholded laser speckle, which will be referred to as the *continuous autocorrelation*.

The continuous autocorrelation is measured using two observations of the speckle intensity. Ideally, these observations are made using two point detectors that sample the speckle field simultaneously at different points in a common plane (x,y). The autocorrelation is formed by finding the expected value of the product of the two observations for each unique separation distance. This is expressed as,[4]

$$R_I(x_1,y_1;x_2,y_2) = \langle I(x_1,y_1)(x_2,y_2) \rangle . \qquad (2.1)$$

In reality the detectors are of finite size, which changes the overall shape of the autocorrelation by convolving Eq. (2.1) by the spatial autocorrelation of a single detector. This paper omits the added effect of finite detector size so that the singular effect of thresholding could be investigated.

The only factor contributing to the form of the spatial autocorrelation is the amplitude of the field emanating from the generating aperture, $P(\xi,\eta)$. This relationship is described by Goodman[5] as

$$R_I(\Delta x,\Delta y)=\langle I\rangle^2\left[1+\left|\frac{\iint_{-\infty}^{\infty}|P(\xi,\eta)|^2\exp\left[i\frac{2\pi}{\lambda z}(\xi\Delta x+\eta\Delta y)\right]d\xi\, d\eta}{\iint_{-\infty}^{\infty}|P(\xi,\eta)|^2\, d\xi\, d\eta}\right|^2\right]. \tag{2.2}$$

From this we can see that the continuous autocorrelation is basically the magnitude squared of the Fourier transform of $P(\xi,\eta)$. As a result, using Eq. (2.2) we can calculate an analytical expression for the continuous autocorrelation for any generating aperture which exists physically.

Our goal is to determine the thresholded autocorrelation, $R_I^{(b)}$. We must determine a relationship between the continuous and thresholded autocorrelations. The speckle field is thresholded using the following operator,

$$I^{(b)}(x,y)=\begin{cases}1, & \text{if}\, I(x,y)\geq b\,\langle I(x,y)\rangle\\ 0, & \text{if}\, I(x,y)< b\,\langle I(x,y)\rangle\end{cases}. \tag{2.3}$$

The thresholded autocorrelation can be determined from the continuous autocorrelation using a relationship calculated by Barakat,[4]

$$R_I^{(b)}(\Delta x,\Delta y)=\frac{\Gamma^2(\alpha,b\,\alpha)}{\Gamma^2(\alpha)}+\frac{(b\,\alpha)^{2\alpha}\exp(-2b\,\alpha)}{\Gamma^2(\alpha)}\sum_{n=1}^{\infty}\frac{(R_I(\Delta x,\Delta y))^n}{\binom{n+\alpha-1}{n}}(L_{n-1}^{(\alpha)}(b\,\alpha))^2, \tag{2.4}$$

where Γ is the complimentary incomplete gamma function and $L_{n-1}^{(\alpha)}$ is the associated Laguerre polynomial.[7] The parameter α is equivalent to the mean of the intensity squared divided by the variance. In this paper we are assuming point detectors meaning that the value of α is equal to 1. Using α=1, Eq. (2.4) can be simplified to

$$R_I^{(b)}(\Delta x,\Delta y) = \exp(-b)\left[1 + b^2\sum_{n=1}^{\infty}\frac{(R_I(\Delta x,\Delta y))^n}{n^2}(L_{n-1}^{(1)}(b))^2\right]. \qquad (2.5)$$

This equation was computed utilizing the recursion formula for Laguerre polynomials[7] over the range of possible correlation values (0.0 to 1.0). The result is given in Fig. 2 in the form of a mapping. A value of continuous autocorrelation, R_I, can be mapped to the corresponding value of thresholded autocorrelation, $R_I^{(b)}$, using Fig. 2. As an example, for a threshold value of b = 3 a value of continuous autocorrelation, R_I = 0.4, would map to a value of thresholded autocorrelation, $R_I^{(3)}$ = 0.25. Repeating this procedure for all values on a continuous autocorrelation function would yield the thresholded autocorrelation for a particular threshold value.

Two cases will be used to exemplify this mapping technique. One, a simple square aperture whose continuous autocorrelation can be found in Goodman.[5] Two, a double-slit aperture that was used in Ref. 3 to produce narrowband filtered laser speckle (Fig. 1).

The amplitude of the field emanating from a square aperture is expressed as

$$|P(\xi,\eta)|^2 = \mathrm{rect}\frac{\xi}{L}\,\mathrm{rect}\frac{\eta}{L}, \qquad (2.6)$$

where L is the measure of each side of the square aperture. Substituting Eq. (2.6) into eq. (2.2) we find the continuous autocorrelation to be

$$R_I(\Delta x,\Delta y) = \langle I \rangle^2 \left[1 + \mathrm{sinc}^2 \frac{L\,\Delta x}{\lambda z}\, \mathrm{sinc}^2 \frac{L\,\Delta y}{\lambda z} \right]. \tag{2.7}$$

This is illustrated in Fig. 3 (solid line) for a single dimension with the bias level, $\langle I \rangle^2$, subtracted out. The autocorrelation was normalized so the correlation values would range from 0.0 to 1.0. Each value on the solid line in Fig. 3 is transformed by finding the corresponding value in Fig. 2. This was done to form the three new curves (dotted lines) which represent the thresholded autocorrelations for different threshold values.

For the second case the amplitude of the field emanating from the double-slit aperture is expressed as

$$|P(\xi,\eta)|^2 = \mathrm{rect}\frac{\xi}{l_1}\, \mathrm{rect}\frac{\eta}{l_2} * \left[\frac{2}{L} \delta\delta\left(\frac{2}{L} \xi \right) \right], \tag{2.8}$$

where l_1 and l_2 are the width and height, respectively, of each rectangle and L is their separation distance in ξ. The continuous autocorrelation is given by Eq. (2.2);

$$R_I(\Delta x,\Delta y) = \langle I \rangle^2 \left[1 + \mathrm{sinc}^2\left(\frac{l_1 \Delta x}{\lambda z} \right) \mathrm{sinc}^2\left(\frac{l_2 \Delta y}{\lambda z} \right) \cos^2(\pi L\, \Delta x) \right]. \tag{2.9}$$

Once again, the thresholded autocorrelation can be determined using Eq. (2.9) and the correlation mappings in Fig. 2. The resulting thresholded autocorrelations for three different threshold values are shown in Fig. 4. Only the

Δx axis is plotted because the cosine term in Eq. 2.9 is in 'Δx'.

The predominant change seen in the transformation of continuous-to-thresholded autocorrelation is the addition of a correlation bias, which results from a reduction in the uniqueness of the individual speckles. Consider the speckle intensity on a single dimension. The continuous speckles differ from each other in maximum intensity, length, and the rate that the intensity rises to and falls from the maximum intensity. Because each speckle is unique, the correlation tends to zero for large separation distances. The thresholded speckle intensity appears like a 'boxcar' signal with each speckle represented by a single box. The speckles all have the same intensity value and only differ in length. Since, each box 'looks' like every other box except for scaling, there will always be some bias correlation. This effect can be seen in Figs. 3 and 4. The amount of bias is equal to exp(-b) which can be determined by evaluating Eq. (2.5) for $R_I(\Delta x,\Delta y) = 0$

3. Power spectrum of thresholded laser speckle

The power spectral density is the frequency domain counterpart of the autocorrelation function. This relationship is described by the Wiener-Kinchine theorem[6] and is expressed as

$$S_I(\nu_x,\nu_y) = \mathcal{F}\{R_I(\Delta x,\Delta y)\}\,. \tag{3.1}$$

The power spectral density describes the amount of power that exists at each spatial frequency in the laser-speckle intensity distribution. Thresholding the speckle intensity may distort this distribution considerably if the correct threshold level is not chosen. An optimal threshold level occurs when the mean-squared error between the continuous and thresholded power spectrum is minimized. Where the continuous and thresholded power spectrums follow the terminology convention defined in the second section.

The continuous power spectrum for the square scattering area can be calculated by Fourier transforming Eq. (2.7) which yields

$$S_I\left(\nu_x,\nu_y\right) = \langle I\rangle^2\left[\delta(\nu_x,\nu_y) + \left(\frac{\lambda z}{L}\right)^2 \Lambda\left(\frac{\lambda z}{L}\nu_x\right)\Lambda\left(\frac{\lambda z}{L}\nu_y\right)\right]. \tag{3.2}$$

This expression is the reference from which the mean-squared error will be determined. The reference curve is shown as a solid line in Fig. 5. To find the thresholded power spectrum the thresholded autocorrelation is Fourier transformed numerically. This was done for three different threshold values

(Fig. 5, dotted lines). The impulse function at the origin in Fig. 5 comes from the correlation bias described at the end of the section 2.

An analysis of Fig. 5 shows that the thresholded curves approach the shape of the continuous reference curve as the threshold level is increased from b=1.0 to b=1.45. As the threshold level is increased further, the thresholded curve moves away from the reference. To determine it exactly, the mean-squared error in the power spectral density from nx = 0 to nx = 200L/lz was calculated for threshold values between b=1.0 and b=3.0 (Fig. 6) . A value of b=1.45 was determined as the absolute minima and subsequent optimal threshold for the square aperture.

The same technique was applied to the double-slit aperture. From Eq. (2.9) the reference power spectrum is

$$\begin{aligned} S_I(\nu_x,\nu_y) = \langle I \rangle^2 \Bigg\{ & \delta(\nu_x,\nu_y) + \frac{1}{2}\frac{(\lambda z)^2}{l_1 l_2}\Lambda\left[\frac{\lambda z}{l_1}\nu_x\right]\Lambda\left[\frac{\lambda z}{l_2}\nu_y\right] \\ & + \frac{1}{4}\frac{(\lambda z)^2}{l_1 l_2}\Lambda\left[\frac{\lambda z}{l_1}\left(\nu_x - \frac{L}{\lambda z}\right)\right]\Lambda\left[\frac{\lambda z}{l_2}\nu_y\right] \\ & + \frac{1}{4}\frac{(\lambda z)^2}{l_1 l_2}\Lambda\left[\frac{\lambda z}{l_1}\left(\nu_x + \frac{L}{\lambda z}\right)\right]\Lambda\left[\frac{\lambda z}{l_2}\nu_y\right]\Bigg\}. \end{aligned} \tag{3.3}$$

The thresholded power spectrums for three threshold values are given in Fig. 7. Although the lines are closely spaced, further inspection shows that the thresholded power spectrum exhibits the same behavior seen in the first case. The mean-squared error calculation in the power spectral density was calculated from ν_X = L/2λz to ν_X = 3L/2λZ. This limited range was used so that the optimal threshold value would be associated with the least amount of distortion in the outer triangle. The result was an optimal threshold value of b=1.535 (Fig. 8).

4. Thresholded laser speckle simulation

To verify the shape of the power spectral density for the optimal threshold values obtained in section 3, a Monte-Carlo simulation was performed. The simulation began by numerically propagating light, with unit magnitude and uniformly distributed $(-\pi,\pi)$ phase, from each generating aperture. This created a large continuous one-dimensional data record containing simulated laser speckle. The record was then separated into several segments of equal length. The power spectrum was calculated for each segment and averaged together to form an estimate of the continuous power spectrum for the entire record.

To calculate estimates for the thresholded power spectrums the original data record was first thresholded at the optimal threshold value. The resulting binarized record was segmented and the estimates were calculated as before.

The results of the simulation are given in Figs. 9 and 10. The raw data was plotted as single points in both figures. The solid lines are the results determined in Section 3 for the optimal threshold values. The data show an increase of deviation with a decrease in frequency. This effect was expected, because the record is of finite length and there are less low frequency speckles with which to form an average. A decrease in deviation is often seen in simulations of this type as the number of segments averaged is increased.

5. Conclusions

The results presented here show that thresholding performed at the optimal level is a viable method for reduction of data volume. We have shown that, for the MTF application, the important information is contained in the placement and size of the speckles and not in the intensity fluctuation between them. This spatial frequency signature is preserved under the threshold operation.

The thresholding technique was implemented on a PC class computer. The results very closely matched those calculated on a Sun 4/330 using laser speckle with a continuous intensity distribution. Thresholding the laser speckle intensity at the optimal threshold value reduced the amount of raw data by a factor of 8. This is a significant factor in the volume of data which is required for the test.

References:

1. G. Boreman and E. Dereniak, "Method for measuring MTF of CCDs using laser speckle," Opt. Eng. **24**(1), 148-150 (1986).

2. M. Sensiper, "Implementation of a system for evaluation the MTF of CCDs using laser speckle," 1991 Final Summer Report, AFOSR Summer Research Program.

3. M. Sensiper, G. Boreman, A. Ducharme, D. Snyder, "MTF Testing of Detector Arrays Using Narrowband Laser Speckle," accepted for publication in Opt. Eng.

4. R. Barakat, "Clipped Correlation functions of Aperture Integrated Laser Speckle," Appl. Opt. **25**, 3885 (1986).

5. J. Goodman, *Laser Speckle and Related Phenomena*, J.C. Dainty, ed., pp. 35-40, Springer-Verlag, Berlin (1975).

6. G.R. Cooper, C.D. McGillem, *Probabilistic Methods of Signal and System Analysis* (2nd ed.,Holt, Rinehart and Winston, Inc., New York, 1971), p.253.

7. L.C. Andrews, *Special Functions for Engineers and Applied Mathematicians* (Macmillan Pub. Co., New York, 1985), p. 179.

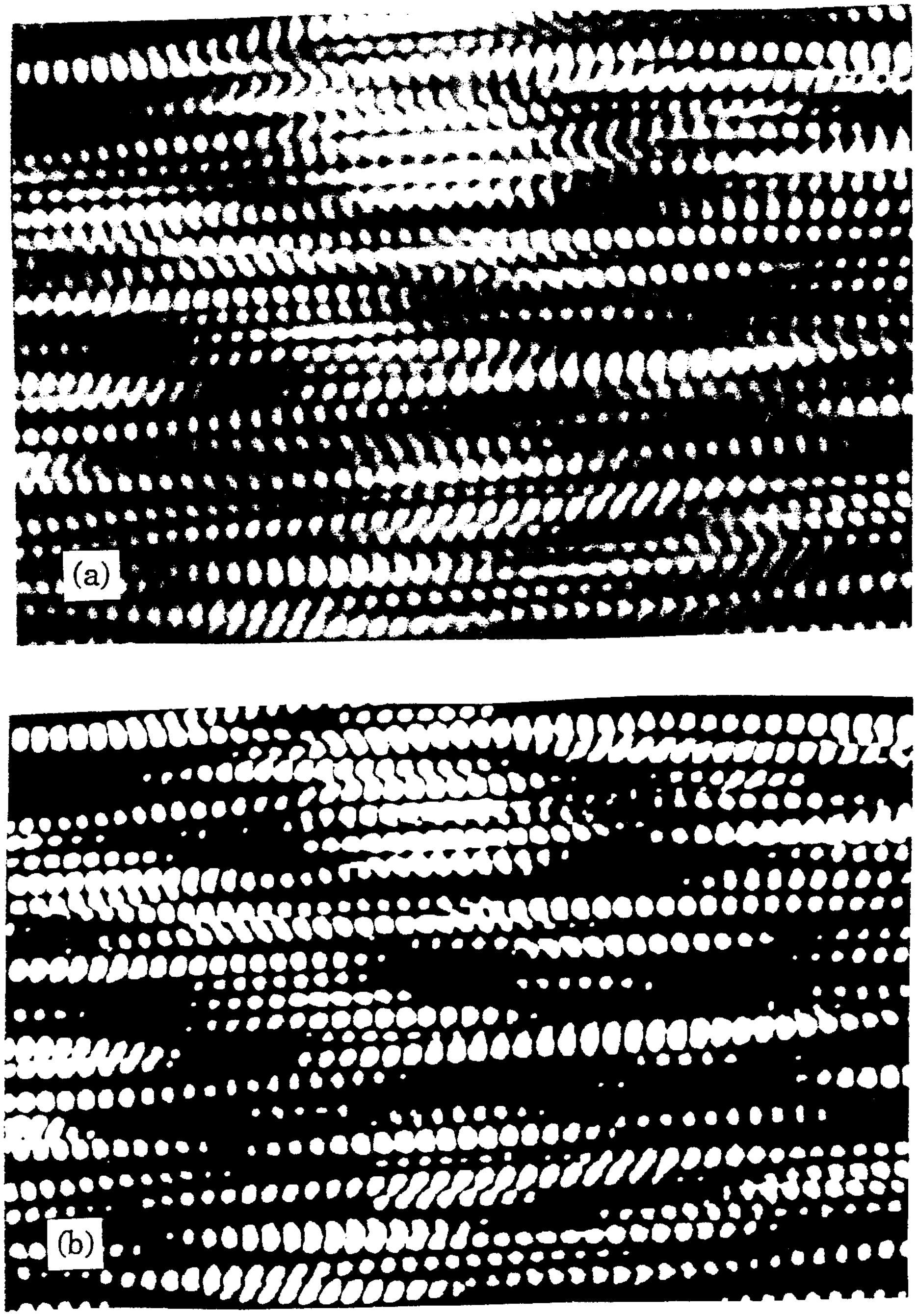


Fig 1 Examples of narrowband laser speckle. (a) Continuous laser speckle with 256 gray levels. (b) Laser speckle reduced to 2 gray levels by thresholding at the mean intensity.

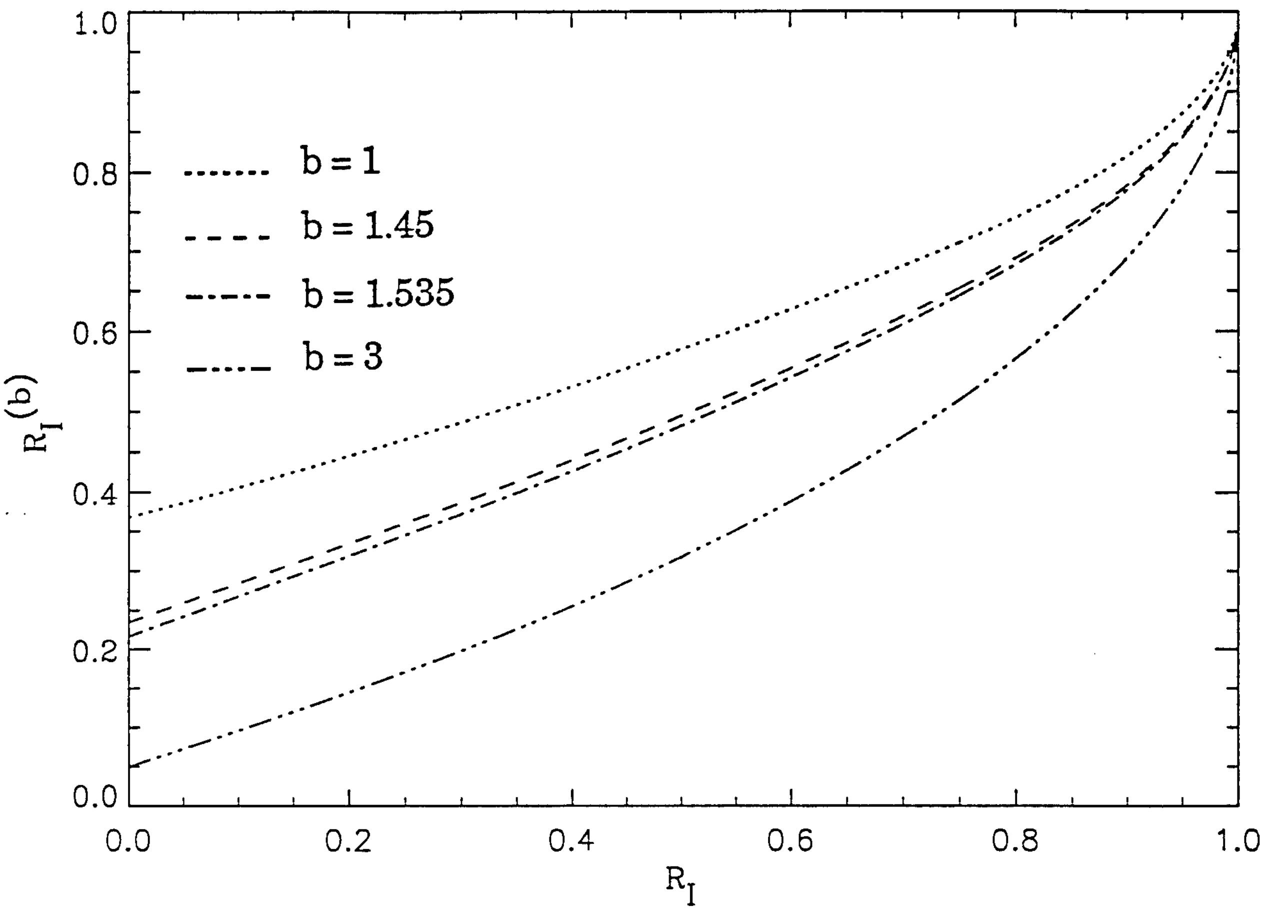


Fig. 2 Mapping for continuous autocorrelation (R_I) to thresholded autocorrelation ($R_I^{(b)}$) for threshold levels: b = 1, 1.45, 1.535, and 3.

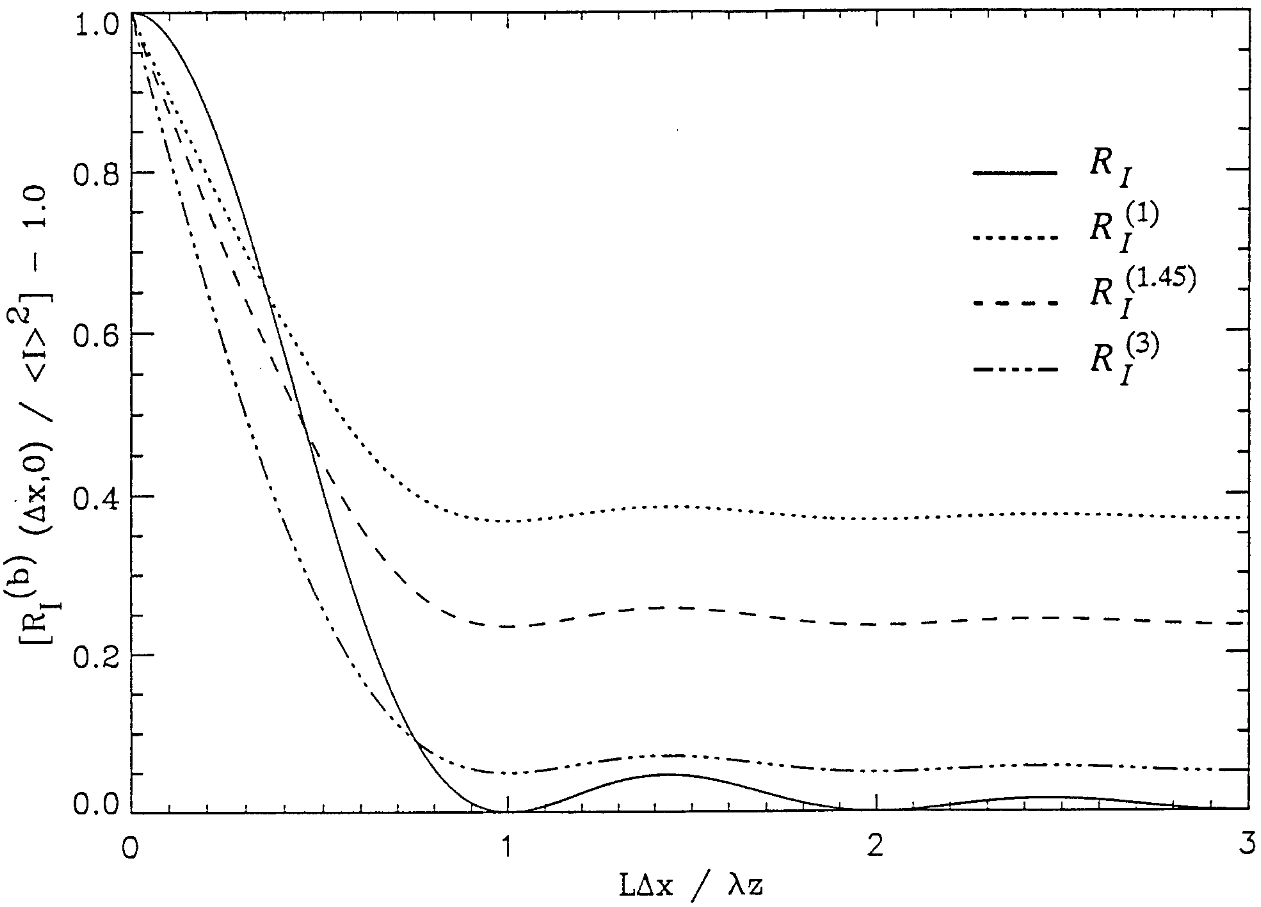


Fig. 3 Autocorrelation function of the speckle intensity resulting from a square aperture.

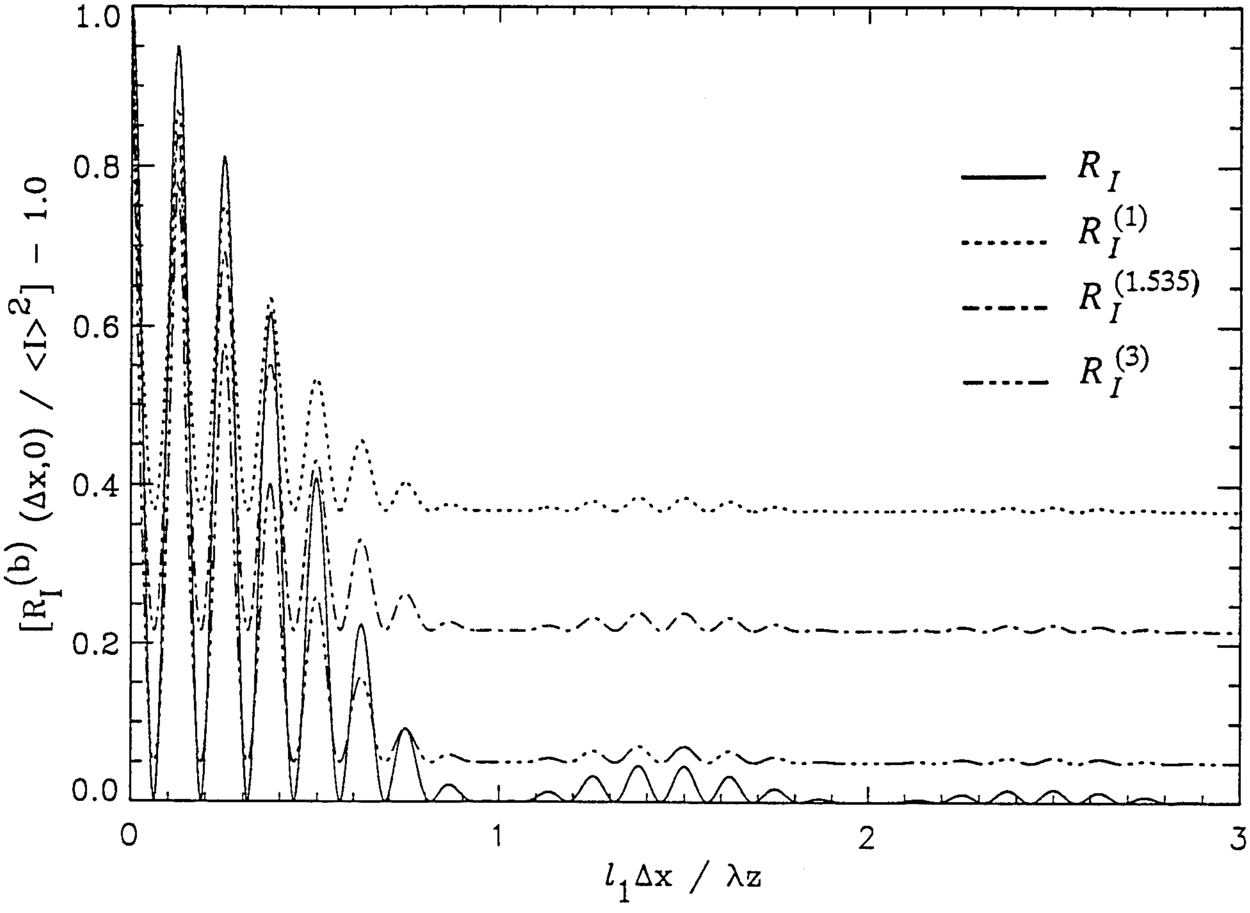


Fig. 4 Autocorrelation function of the laser speckle intensity resulting from a double-slit aperture.

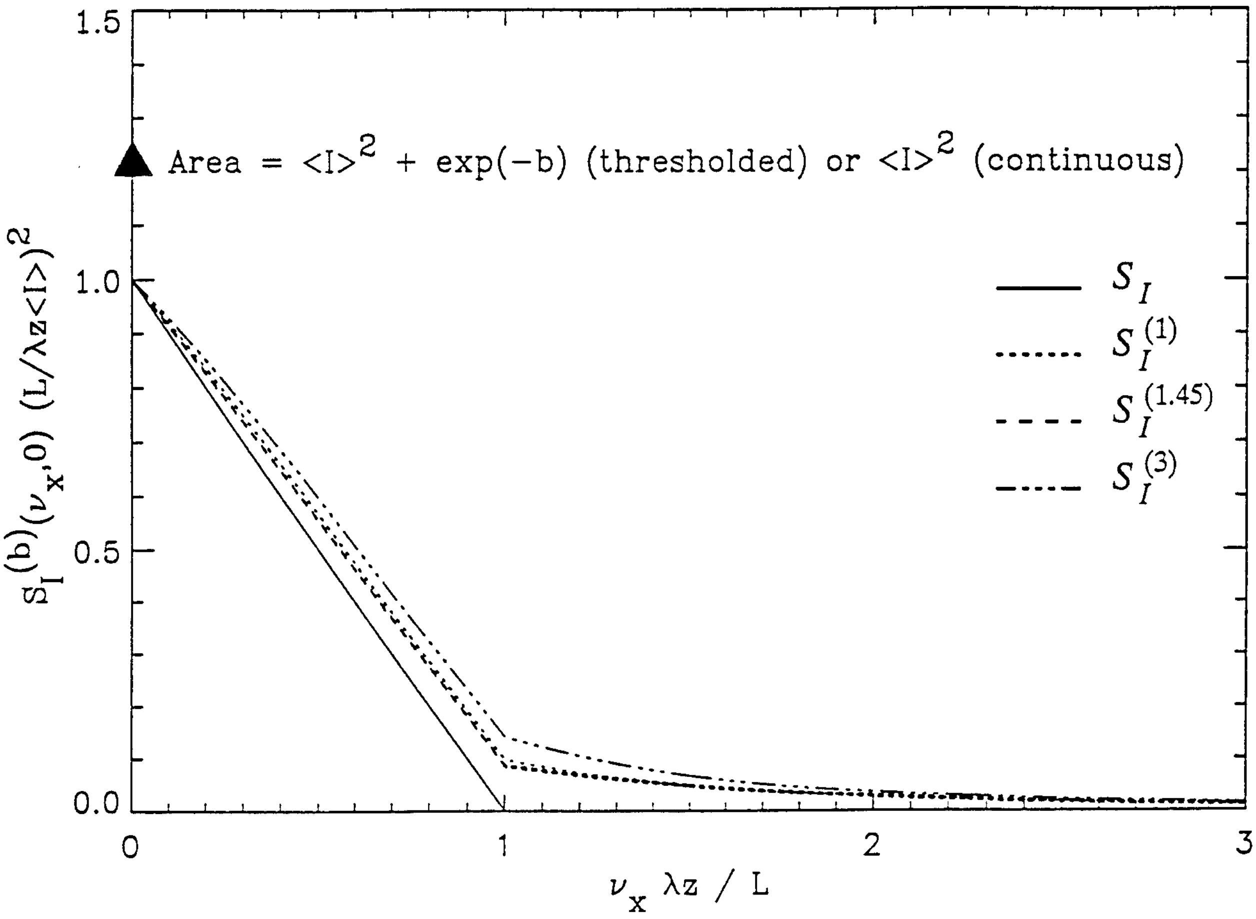


Fig. 5 Power spectral density of the laser speckle intensity resulting from a square aperture. Area refers to the strength of the impulse function.

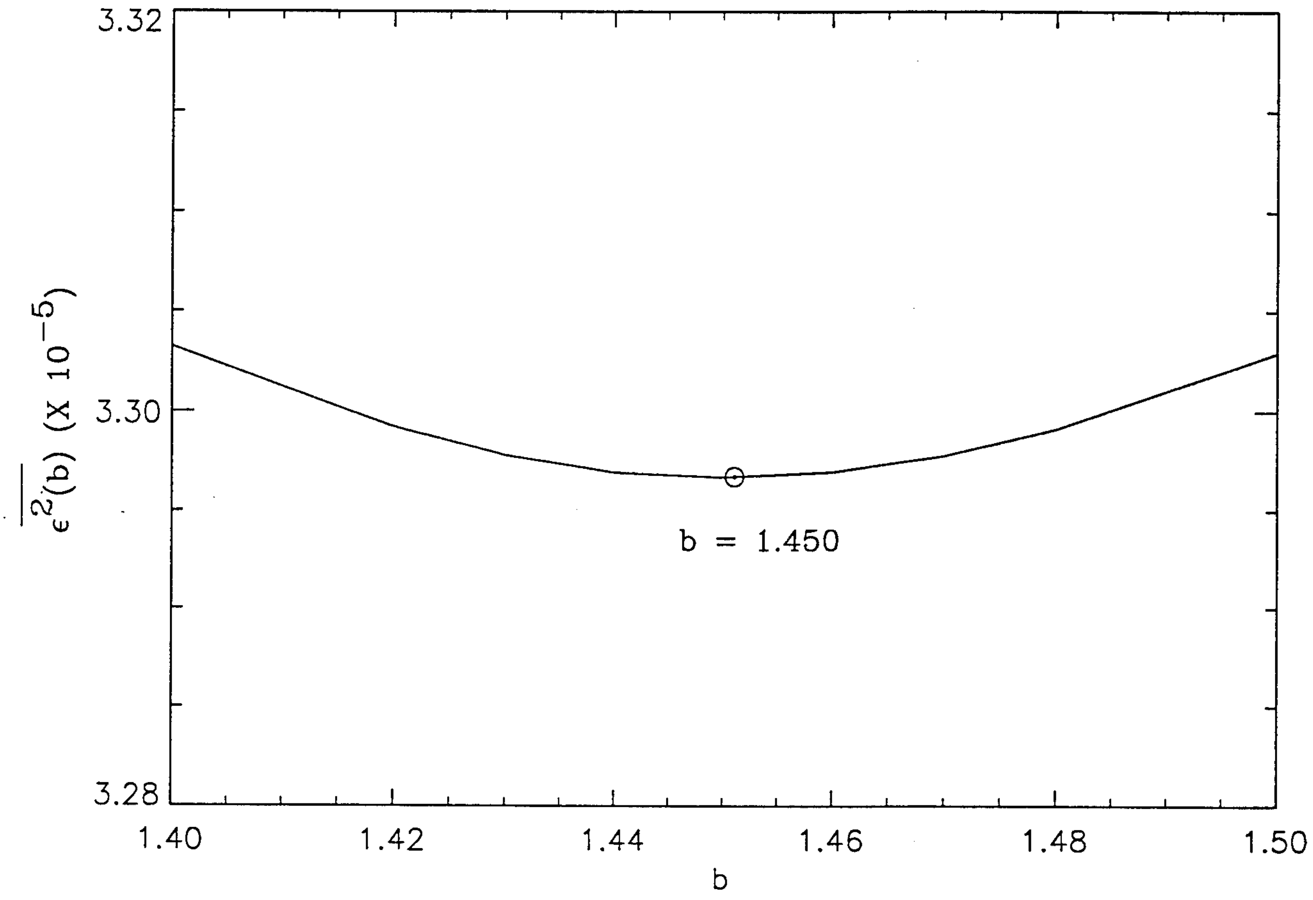


Fig. 6 Mean-squared error found in power spectral density of square aperture as a function of threshold value, b.

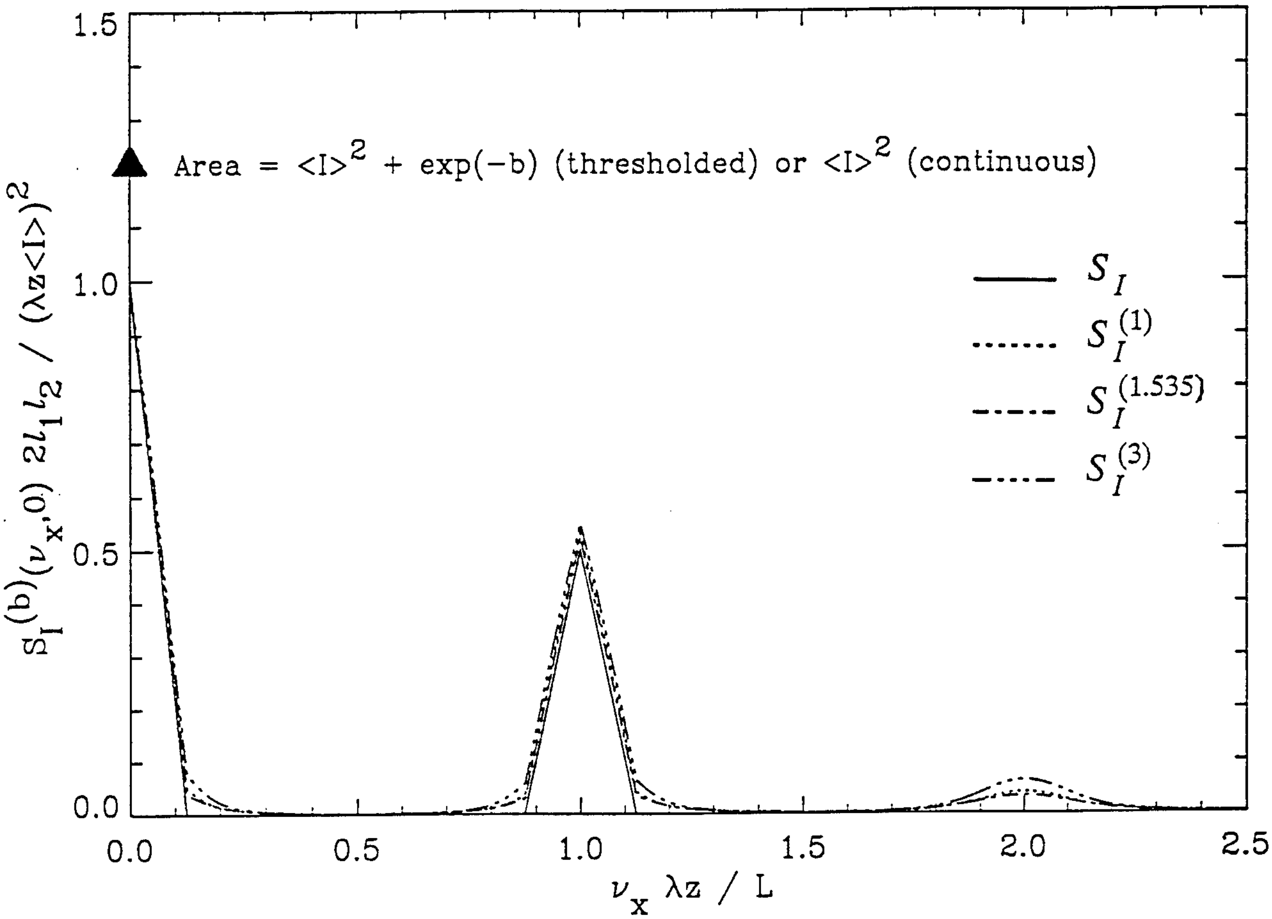


Fig. 7 Power spectral density of the laser speckle intensity resulting from a double-slit aperture. Area refers to the strength of the impulse function.

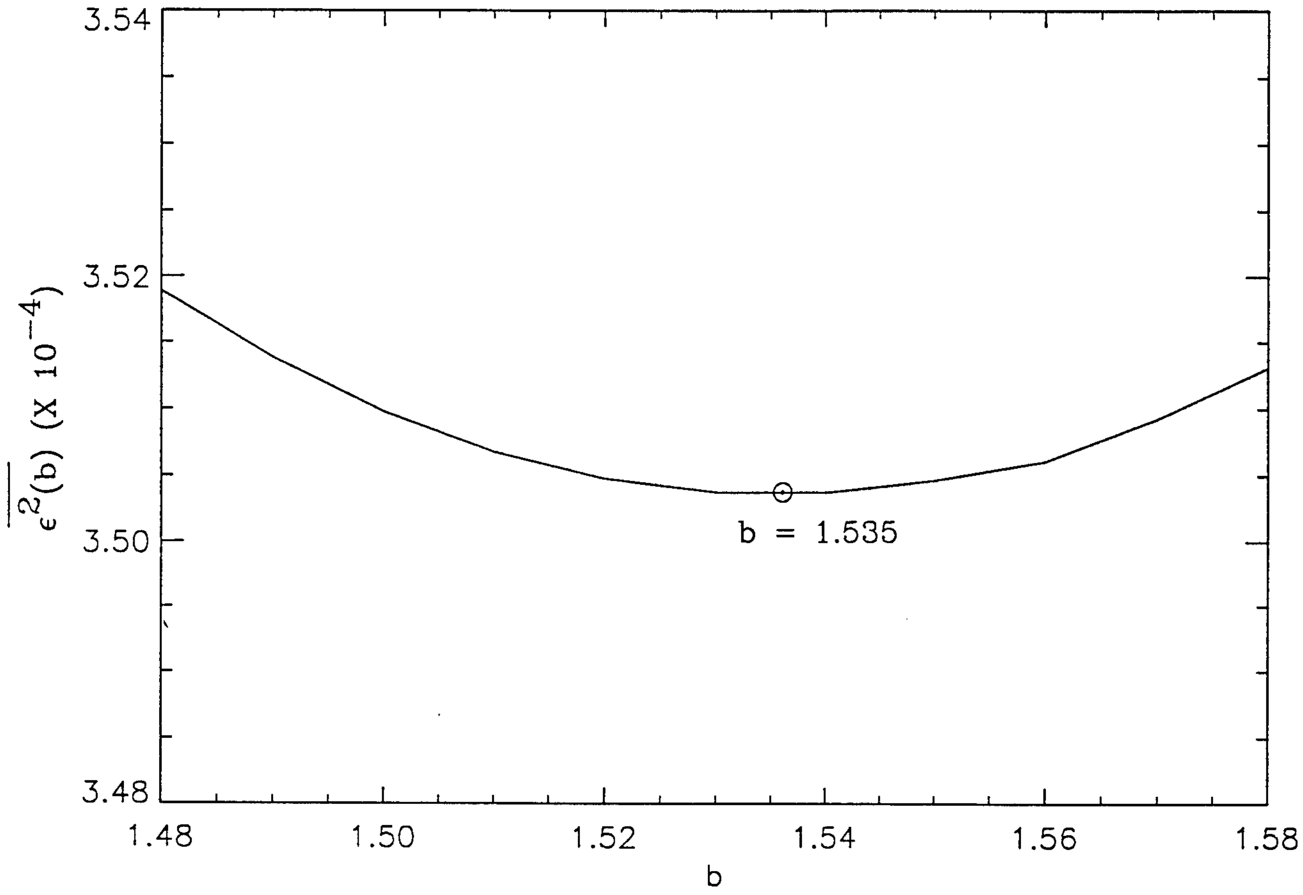


Fig. 8 Mean-squared error found in power spectral density of double-slit aperture as a function of threshold value, b.

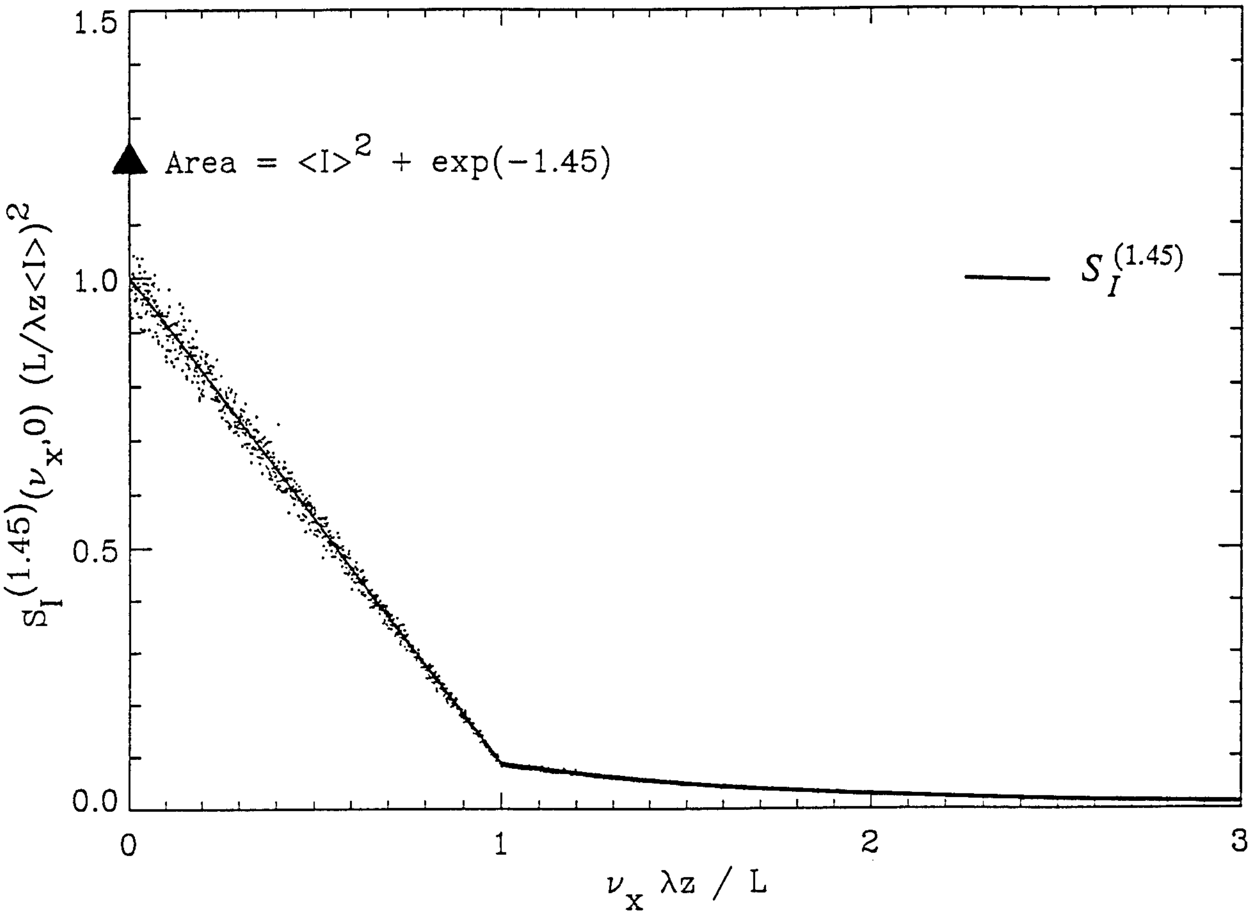


Fig. 9 Power spectral density of simulated laser speckle generated with square aperture and thresholded at b = 1.45. Area refers to the strength of the impulse function.

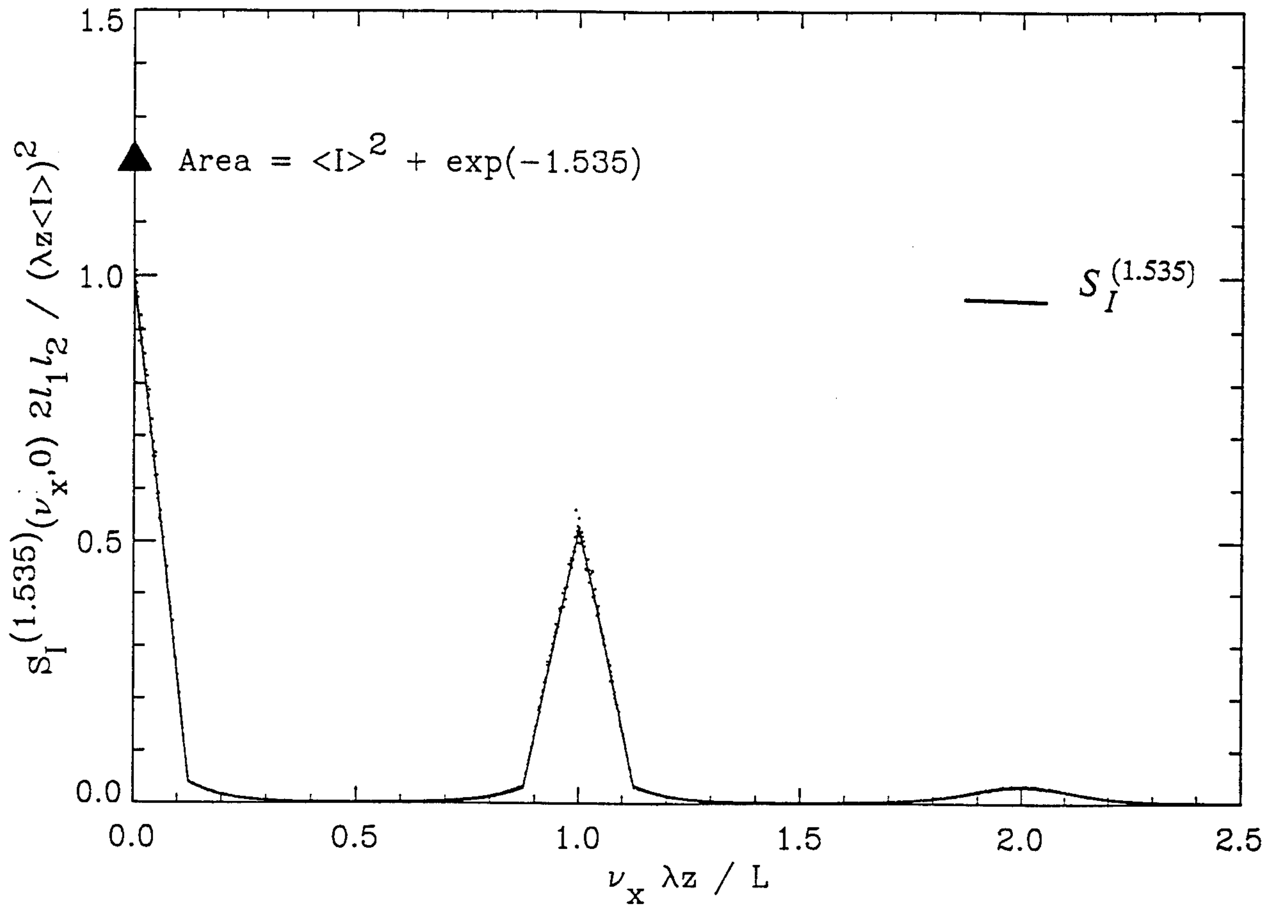


Fig. 10 Power spectral density of simulated laser speckle generated with double-slit aperture and thresholded at b = 1.535. Area refers to strength of impulse function.

Gain-Scheduled Missile Autopilot Design

Jeff S. Shamma
Aerospace Engineering and Engineering Mechanics

Final Report for:
Research Initiation Program
Wright Laboratory

and

The University of Texas at Austin

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

February 17, 1993

Gain-Scheduled Missile Autopilot Design
Final Report
RIP #92–70

Jeff S. Shamma
Department of Aerospace Engineering and Engineering Mechanics
The University of Texas at Austin
Austin, TX 78712

February 17, 1993

Contents

1 Report Summary

This report describes the research undertaken under the Research Initiation Proposal #92–70. The research was concentrated in the areas of 1) gain-scheduled control design and 2) nonlinear and time-varying robust control analysis.

The areas of research are summarized as follows.

1.1 Gain-Scheduled Control Design

Gain-Scheduled Missile Autopilot Design Using Linear Parameter Varying Transformations and μ-Synthesis. This work presents a gain-scheduled design for a missile longitudinal autopilot. The gain-scheduled design is novel in that it *does not* involve linearizations about trim conditions of the missile dynamics. Rather, the missile dynamics are brought to a quasi-linear parameter varying (LPV) form via a state transformation. An LPV system is defined as a linear system whose dynamics depend on an exogenous variable whose values are unknown *a priori* but can be measured upon system operation. In this case, the variable is the angle-of-attack. This is actually an endogenous variable, hence the expression "quasi-LPV". Once in a quasi-LPV form, a robust controller using μ-synthesis is designed to achieve angle-of-attack control via fin deflections. The final design is an inner/outer-loop structure, with angle-of-attack control being the inner-loop, and normal-acceleration control being the outer-loop.

Trajectory Scheduled Missile Autopilot Design. In this work, we present an alternate approach to the design of gain-scheduled controllers. Standard gain-scheduling typically relies on the instantaneous value of the scheduling variable to update the control gains. This approach has the drawbacks of being limited to slow transitions between operating points and requiring a number of possibly tedious point-by-point designs. The present approach is to have the controller gains dynamically evolve according to the scheduling variable's *history*—rather than its instantaneous value. At the cost of increased order of the controller, this approach allows rapid transitions between operating conditions as well as alleviates the need for several point-by-point designs. The approach is demonstrated via a stabilizing control design for a longitudinal missile model.

1.2 Nonlinear and Time-Varying Robust Control Analysis

Fading Memory Feedback Systems and Robust Stability. This work considers fading memory for nonlinear time-varying systems and associated problems of robust stability.

We define two notions of fading memory for stable dynamical systems: uniform and pointwise. We then provide conditions under which stable linear or nonlinear systems exhibit uniform or pointwise fading memory. In particular, we show that (1) all stable discrete-time linear time-varying (LTV) systems have uniform fading-memory, (2) all stable continuous-time LTV systems have pointwise fading-memory, and (3) stable finite-dimensional continuous-time LTV systems have uniform fading-memory.

We then show that a version of the small gain theorem which employs the asymptotic gain of a fading-memory system is necessary for the stable invertibility of certain feedback operators. These results are presented for both continuous-time and discrete-time systems using general ℓ^p or $\mathcal{L}^p$ notions of input/output stability and generalize existing results for ℓ^2 stability. We further investigate fading memory in a closed-loop context. For linear plants, we parameterize all nonlinear controllers which lead to closed-loop pointwise fading memory.

Robust Stability with Time-Varying Structured Uncertainty. We consider the problem of assessing robust stability in the presence of block-diagonally structured time-varying dynamic uncertainty. We show that robust stability holds only if there exist constant scalings which lead to a small gain condition. The notion of stability here is finite-gain stability over finite-energy signals. In sharp contrast to the case of time-invariant dynamic uncertainty, this result is *not* limited by the number uncertainty blocks. These results parallel previous results regarding finite-gain stability over persistent bounded signals.

Nonlinear State Feedback for ℓ^1 Optimal Control. This work considers ℓ^1 optimal control problems with full state feedback. In contrast to $\mathcal{H}^\infty$ optimal control, previous work has shown that linear ℓ^1 optimal controllers can be dynamic and of arbitrarily high order. However, this work shows that continuous memoryless *nonlinear* state feedback performs as well as dynamic linear state feedback. The derivation, which is non-constructive, relies on concepts from viability theory.

2 Research Publications

The following publications acknowledge the support of the Research Initiation Proposal:

1. J.S. Shamma and R. Zhao, "Fading Memory Feedback Systems and Robust Stability," *Automatica,* Special Issue on Robust Control, January 1993, pp. 191–200.

2. J.S. Shamma and J.R. Cloutier, "Gain Scheduled Missile Autopilot Design using Linear Parameter Varying Transformations and μ-synthesis," accepted for publication, *AIAA Journal of Guidance, Control, and Dynamics,* 1992.

3. J.S. Shamma, "Robust Stability with Time-Varying Structured Uncertainty," accepted for publication, *IEEE Transactions on Automatic Control,* 1992.

4. J.S. Shamma, "Nonlinear State Feedback for ℓ^1 Optimal Control," accepted for publication, *Systems & Control Letters,* 1993.

5. J.S. Shamma and J.R. Cloutier, "Trajectory Scheduled Missile Autopilot Design," *Proceedings of the 1st IEEE Conference on Control Applications,* Dayton, Ohio, September 1992.

6. J.S. Shamma, "Robustness Analysis for Time-Varying Systems," *Proceedings of the 31st IEEE Conference on Decision and Control,* Tucson, AZ, December 1992.

7. J.S. Shamma, "ℓ^1 Optimal Control with Nonlinear Full State Feedback," submitted to *IEEE Mediterranean Control Conference,* 1992.

3 Research Summary

In this section, we summarize five of the papers reported in Section 2.

3.1 Gain-Scheduled Missile Autopilot Design Using Linear Parameter Varying Transformations and μ-Synthesis

Future tactical missiles will be required to operate over an expanded flight envelope in order to meet the challenge of highly maneuverable tactical aircraft. In such a scenario, an autopilot derived from linearization about a single flight condition will be unable to achieve suitable performance over all envisioned operating conditions. A particular challenge, however, is that of the missile endgame. This involves the final few seconds before delivery of ordnance. During this phase, a missile autopilot can expect large and rapidly time-varying acceleration commands from the guidance law. In turn, the missile is operating at a high and rapidly changing angle-of-attack.

Traditionally, satisfactory performance across the flight envelope can be attained by gain scheduling local autopilot controllers to yield a global controller. Often the angle-of-attack is used as a scheduling variable. However during the rapid transitions in the missile endgame, a fundamental guideline of gain-scheduling to "schedule on a slow variable" is violated. Given the existing track record of gain-scheduling, any improvement in the gain-scheduling design procedure—especially in the endgame—could have an important impact on future missile autopilot designs.

In this paper, we present a novel approach to gain-scheduled missile autopilot design. The missile control problem under consideration is normal acceleration control of the longitudinal dynamics during the missile endgame. In standard gain-scheduling, the design plants consist of a collection of linearizations about equilibrium conditions indexed by the scheduling variable, in this case the angle-of-attack, α (cf., [25][27]). In the present approach, the design plants also consist of a family of linear plants indexed by the angle-of-attack. A key difference between the present approach and standard gain-scheduling is that this family is *not* the result of linearizations. Rather, it is derived via a state-transformation of the original missile dynamics (i.e., an alternate selection of state variables). Since no linearization is involved, the approach is not limited by the local nature of standard gain-scheduled designs.

Since gain-scheduling generally encounters families of linear plants indexed by a scheduling variable, we shall refer to such a family as a Linear Parameter Varying (LPV) plant. LPV plants *differ* from linear time-varying plants in that the time-variations (i.e., the scheduling variable) is unknown *a priori* but may be measured/estimated upon operation of the feedback system. We shall call such a family quasi-LPV in case the scheduling variable is actually endogenous to the state dynamics (as in the missile problem). In [25][27], it was shown that LPV and quasi-LPV dynamics form the underlying structure of gain-scheduled designs.

The design for the resulting quasi-LPV system is performed via μ-synthesis [4]. Briefly, μ-synthesis exploits the structure of performance requirements and robustness considerations in order to achieve robust performance in a non-conservative manner. Thus, the present approach makes use of gain-scheduling's ability to incorporate modern linear synthesis techniques into a nonlinear design.

Another feature in the present approach is its interpretation of an inner/outer-loop approach to nonlinear control design. In standard gain-scheduling (as well as geometric nonlinear control [19]), one often applies an inner-loop feedback. In gain-scheduling, this feedback is an update of the current trim condition. In geometric nonlinear control, this feedback serves to invert certain system dynamics to yield linear behavior in the modified plant. In either case, unless the inner-loop robustly performs its task, the outer-loop performance and even stability can be destroyed. In other words, any inner/outer-loop approach must be built from the inside out. Reference [25] presents a more detailed discussion of this possibility in the context of standard gain-scheduling.

The present approach also takes an inner/outer-loop approach to the autopilot design. The inner-loop consists of a robust angle-of-attack servo. The reason for the inner-loop is that nonlinear gain-scheduling techniques prefer to directly control the scheduling variable. Such an inner/outer-loop decomposition was also employed in [32], where the reasoning was to avoid non-minimum phase dynamics from the fin deflection to normal acceleration.

The actual regulated variable of interest is the normal acceleration. Thus, the outer-loop serves to generate angle-of-attack commands, α_c, to obtain the desired normal acceleration. A *consequence* of the inner-loop design is that the dynamics from the commanded angle-of-attack, α_c, to the angle-of-attack, α, exhibit a linear behavior *within the bandwidth* of the inner-loop design. Thus, as in geometric nonlinear control, the inner-loop linearizes certain dynamics. However, the approximate linearization stems from the natural *linearizing effect of feedback* (cf., [10]) as opposed to an exact linear geometric condition on the plant state dynamics. Thus, the outer-loop design is essentially a linear design. However, in the design process, it is acknowledged that the linear behavior due to the inner-loop is *approximate* and *bandlimited.*

3.2 Trajectory Scheduled Missile Autopilot Design

Gain-scheduling is a common method of control design for nonlinear plants. Briefly, the main idea is to design several linear controllers based on linearized operating conditions. The control gains are then interpolated according to some scheduling variable which is indicative of the proximity of operating conditions.

Despite its widespread application, current gain-scheduled designs have certain limiting factors. First, since the design is based on a collection of fixed linearizations, the overall design performance is generally limited to slow transitions between operating conditions. Furthermore, the linearization of the dynamics introduces approximations in the design which further limit performance. Another issue is gain-scheduling's need for several point-by-point designs. This can lead to the possibly tedious task of covering the operating range of the plant with a large number of linearized designs—especially in the case of multiple scheduling variables. This also raises the issue of determining how many operating point linearizations are adequate for a particular design. See references [27][28][29] for further discussions on the theoretical properties and limitations of gain-scheduling.

In this paper, we present an alternate approach to gain-scheduling which addresses these issues. The main ideas are as follows. First, extra approximations due to linearizations are limited by expressing the nonlinear plant's dynamics in a form amenable to gain-scheduling via a state *transformation*—rather than linearization. Second, the employed controller gains are not scheduled according to the instantaneous value of the scheduling variable. Rather, the controller gains dynamically evolve according to the trajectory history of scheduling variable. The advantages of trajectory scheduling rather than instantaneous scheduling are twofold. First, it is believed that such an approach assures stability in the presence of arbitrarily fast variations in the scheduling variable. Second, the need to perform several fixed operating point designs is alleviated by allowing the controller gains to dynamically evolve into appropriate values.

3.3 Fading Memory Feedback Systems and Robust Stability

The problem of robust stability analysis (cf., [12] and references therein) is to determine under what conditions a given controller stabilizes a prescribed *family* of possible plants. Typically, this plant family arises from various approximations, simplifications, and limitations in the plant modeling process. One framework for plant family representations is that of unstructured uncertainty. More precisely, the plant family is represented as a nominal plant combined with a norm-bounded perturbation. The theory for robust stability analysis for linear time-invariant systems subject to unstructured uncertainty is well developed (e.g., [7][8][13][15][16][18][21]).

Linear systems typically arise as linearizations of nonlinear systems. Furthermore, adaptive control laws for linear systems are typically nonlinear. Thus, robust stability analysis tools for nonlinear systems are desirable. For nonlinear systems, a standard tool for stability analysis is the small gain theorem [9][24][35]. A limitation of the small gain theorem is that it can often be a conservative sufficient condition for stability. Recent work by the author [26] has shown that the small gain theorem is in fact necessary when considering nonlinear plant families characterized by a norm-bounded perturbation. This work was developed for ℓ^2 (i.e., finite-energy) stability of discrete-time systems. The nonlinear systems considered in [26] are those with *fading memory* (e.g., [6]). Intuitively, a fading memory property means that the current output depends on the recent inputs and not the remote past. Thus, fading-memory is a reasonable assumption for many physical systems.

The objective of this paper is to further develop the analysis of the fading memory systems in the context of robust stability. The results here are presented in an essentially norm-invariant manner. We consider stability over arbitrary $\mathcal{L}_p$ or ℓ_p spaces with $p \in [1, \infty)$. We also consider a form of bounded-input/bounded-output $\mathcal{L}^\infty$ or ℓ^∞ stability with asymptotic decay. Given the norm-invariant nature of the presentation, it is the fading memory property which is isolated and exploited to lead to the desired results.

In this paper, we define two notions of fading memory for stable dynamical systems: uniform and pointwise. In uniform fading memory, the effects of a finite-duration of input eventually vanish depending on the length of the duration only. In pointwise fading memory, the effects of a finite-duration of input also eventually vanish but now depending on the length of the duration and the input itself.

The distinction between uniform and pointwise fading memory turns out to be important. In particular, we will see that all stable discrete-time linear systems exhibit uniform fading memory, all stable continuous-time linear systems exhibit pointwise fading memory, and stable finite-dimensional continuous-time linear systems exhibit uniform fading memory. The pointwise fading memory property also allows us to weaken the conditions under which the small gain theorem is necessary. This leads to a larger class of nonlinear systems for which these results are applicable.

The fading memory property is also considered in a closed-loop context. More precisely, we consider under what conditions a stabilizing compensator leads to a closed-loop system with pointwise fading memory. A key tool in this analysis is the factorization representation for nonlinear operators (cf., [33] and references therein). This allows us to parameterize all nonlinear compensators for linear plants which lead to closed-loop stability with pointwise fading memory. This parameterization takes the form of the familiar linear fractional parametrization (e.g., [17][34]) with the free parameter having fading memory.

3.4 Robust Stability with Time-Varying Structured Uncertainty

This paper addresses the problem of assessing robust stability, i.e., finding conditions under which a feedback system maintains stability in the presence of a prescribed class of modeling errors. Typically, modeling errors are represented as perturbations—either parametric, dynamic, or both—on a nominal model. See [12] and references therein.

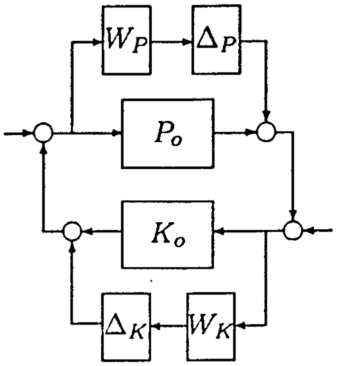


Figure 1: Additive Plant and Controller Uncertainty

The particular paradigm of modeling errors considered here is the so-called "structured dynamic uncertainty" [13][16][20][21]. In this paradigm, modeling errors are represented as normalized norm-bounded dynamical systems occurring at several locations in the feedback loop. For example, Figure 1 depicts a feedback system in which the nominal plant, P_o, and nominal controller, K_o, are both perturbed by additive modeling errors, Δ_P and Δ_K, respectively. Typical assumptions are that the errors satisfy a norm bound such as $\|\Delta_P\| < 1$ and $\|\Delta_K\| < 1$. The weighting blocks W_P and W_K are used to shape and normalize the effects of these errors.

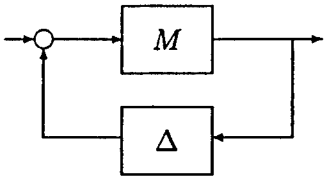


Figure 2: Feedback System after Loop Transformations

Many problems with structured dynamic uncertainty can be transformed to the block diagram Figure 2 (see [16] for a detailed discussion). In this figure, the block M represents a known stable linear closed-loop system composed of the plant, compensator, and various weightings. The "uncertainty" block Δ represents a *block-diagonal* collection of stable linear systems known only to satisfy a given norm bound such as $\|\Delta\| < 1$. The number of block diagonal elements in Δ is often called "the number of uncertainty blocks". In the case of Figure 1, we have

$$M = \begin{pmatrix} W_P K_o (I - K_o P_o)^{-1} & W_P (I - K_o P_o)^{-1} \\ W_K (I - P_o K_o)^{-1} & W_K P_o (I - P_o K_o)^{-1} \end{pmatrix},$$

and

$$\Delta = \begin{pmatrix} \Delta_P & \\ & \Delta_K \end{pmatrix},$$

with two uncertainty blocks.

An immediate sufficient condition for robust stability is the small gain condition $\|M\| \leq 1$ (cf., [9][24][35]). In the case of one uncertainty block—also called unstructured dynamic uncertainty—this condition is also *necessary* for robust stability (modulo some modifications for time-varying systems). That is in case $\|M\| > 1$, there exists an admissible destabilizing perturbation. See [7][8][15] for the case where M is linear time-invariant, [30] for the case where M is linear time-varying, and [26][31] for the case where M is nonlinear with fading memory.

When there are multiple uncertainty blocks, the small gain condition can be an arbitrarily conservative sufficient condition for robust stability. This problem motivated analysis tools such as the structured singular value [13][23]. In particular, one has the following result. In case (1) the underlying notion of stability is finite-gain ℓ^2 stability, (2) M and Δ are linear time-invariant, and (3) the number of uncertainty blocks is three or less, then robust stability holds if and only if

$$\inf_{D \text{ admissible}} \left\| DMD^{-1} \right\| \leq 1.$$

Here an "admissible" D is any diagonal stable minimum phase invertible linear time-invariant system which commutes with with any admissible Δ. Thus robust stability may be tested by searching for "multipliers" (e.g., [9]) which makes the small gain condition non-conservative. In the case of more than three uncertainty blocks, the above scaled small gain condition is once again conservative.

Recent results in [21], however, show that in case (1) the underlying notion of stability is finite-gain ℓ^∞ stability, (2) M is linear time-invariant, and (3) Δ is now linear *time-varying*, then robust stability holds if and only if a scaled small gain condition holds. However, an "admissible" D is now any diagonal gain matrix which commutes with any admissible time-varying Δ. This was later extended to M being linear time-varying [20]. Thus when considering stability over ℓ^∞ with time-varying uncertainty blocks, the equivalence between robust stability and a scaled small gain condition is not restricted by the number of uncertainty blocks.

In this paper, we consider the case of finite-gain stability over ℓ^2 with time-varying uncertainty blocks. In particular, we show that robust stability is again equivalent to a scaled small gain condition. However, in stark contrast to the case of time-invariant uncertainty blocks, this equivalence *is not* restricted by the number of uncertainty blocks. Hence, it seems that it is the time-varying nature of the uncertainty blocks which leads to this equivalence, rather than properties of the underlying signal space.

Related to this work are the results of [5]. There, it is shown that the structured singular value may be computed exactly for certain classes of matrices via a "lifting" operation. This lifting operation may be given the interpretation of enhancing the underlying matrix perturbation structure. The admission of time-varying, rather than time-invariant, uncertainty blocks in the present paper may be viewed as a similar enhancement for dynamical systems (rather than matrices).

3.5 Nonlinear State Feedback for ℓ^1-Optimal Control

This paper investigates the structure of ℓ^1 optimal control problems with full state feedback. The recent paper [11] has shown that even in the case of full state feedback, optimal and near-optimal linear controllers can be dynamic and of arbitrarily high order. This is in contrast to $\mathcal{H}^\infty$ near-optimal control (cf., [14] and references therein) for which full state feedback controllers can be static. This property ultimately reveals an underlying separation structure for $\mathcal{H}^\infty$ optimal control with output feedback. In light of the results of [11] (see also [22]), it seems unlikely that such a separation property holds for linear ℓ^1 optimal controllers.

In this paper, we follow on the work of [11] and consider the utility of *nonlinear* state feedback. We show that continuous nonlinear static state feedback performs as well as dynamic linear state feedback. Thus, the admission of nonlinear feedback removes the necessity of controller dynamics.

The derivation, which is non-constructive, relies on concepts from viability theory [1][2][3]. The main idea is to show that disturbance rejection with a known bounded disturbance set is equivalent to restricting the plant state to evolve in a particular bounded region. In the terminology of viability theory, this bounded region is viable for the closed-loop dynamics with disturbances. Viability theory gives conditions for the existence of state feedback leading to viable trajectories. This feedback is then scaled to assure the desired performance for all disturbances.

References

[1] J.P. Aubin. *Viability Theory.* Birkhäuser, Boston, 1991.

[2] J.P. Aubin and A. Cellina. *Differential Inclusions.* Springer-Verlag, New York, 1984.

[3] J.P. Aubin and H. Frankowvska. *Set-Valued Analysis.* Birkhäuser, Boston, 1990.

[4] G.J. Balas, J.C. Doyle, K. Glover, A. Packard, and R. Smith. *μ Analysis and Synthesis Toolbox User's Guide.* MUSYN, Inc. and the Mathworks, Inc., April 1991.

[5] H. Bercovici, C. Foias, and A. Tannenbaum. Structured interpolation theory. *Operator Theory Advances and Applications*, **47**:195–220, 1990.

[6] S. Boyd and L.O. Chua. Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Transactions on Circuits and Systems*, **CAS–32**:1150–1161, 1985.

[7] M.J. Chen and C.A. Desoer. Necessary and sufficient condition for robust stability of linear distributed feedback systems. *International Journal of Control*, **35**(2):255–267, 1982.

[8] M.A. Dahleh and Y. Ohta. A necessary and sufficient condition for robust BIBO stability. *Systems & Control Letters*, **11**:271–275, 1988.

[9] C.A. Desoer and M. Vidyasagar. *Feedback Systems: Input-Output Properties.* Academic Press, New York, 1975.

[10] C.A. Desoer and Y.-T. Wang. Foundations of feedback theory for nonlinear dynamical systems. *IEEE Transactions on Circuits and Systems*, **CAS–27**(2):104–123, 1980.

[11] I.J. Diaz-Bobillo and M.A. Dahleh. State feedback ℓ^1-optimal controllers can be dynamic. *Systems & Control Letters*, **19**(2), 1992.

[12] P. Dorato, editor. *Robust Control.* IEEE Press, New York, 1987.

[13] J.C. Doyle. Analysis of feedback systems with structured uncertainties. *IEE Proceedings*, Part D, **129**(6):242–250, 1982.

[14] J.C. Doyle, K. Glover, P.P. Khargonekar, and B. Francis. State-space solutions to standard $\mathcal{H}^2$ and $\mathcal{H}^\infty$ control problems. *IEEE Transactions on Automatic Control*, **AC–34**(8):821–830, 1989.

[15] J.C. Doyle and G. Stein. Multivariable feedback design concepts for a classical/modern synthesis. *IEEE Transactions on Automatic Control*, **AC–26**(1):4–16, 1981.

[16] J.C. Doyle, J.E. Wall, and G. Stein. Performance and robustness analysis for structured uncertainty. In *Proceedings of the 21st IEEE Conference on Decision and Control*, pages 629–636, December 1982.

[17] B.A. Francis. *A Course in $\mathcal{H}^\infty$-Optimal Control Theory.* Springer-Verlag, New York, 1987.

[18] T.T. Georgiou and M. Smith. Optimal robustness in the gap metric. *IEEE Transactions on Automatic Control*, **AC–35**(6):673–686, 1990.

[19] A. Isidori. *Nonlinear Control Systems.* Springer-Verlag, Berlin, 2nd edition, 1985.

[20] M.H. Khammash. Necessary and sufficient conditions for the robustness of time-varying systems with applications to sampled-data systems. *IEEE Transactions on Automatic Control*, **AC–38**(1):49–57, 1993.

[21] M.H. Khammash and J.B. Pearson. Performance robustness of discrete-time systems with structured uncertainty. *IEEE Transactions on Automatic Control*, **AC–36**(4):398–412, 1991.

[22] D.G. Meyer. Two properties of ℓ^1-optimal controllers. *IEEE Transactions on Automatic Control*, **AC–33**(9):876–878, 1988.

[23] M.G. Safonov. Stability margins of diagonally perturbed multivariable feedback systems. *IEE Proceedings*, Part D, **129**(6):251–256, 1982.

[24] I.W. Sandberg. An observation concerning the application of the contraction mapping fixed-point theorem and a result concerning the norm-boundedness of solutions of nonlinear functional equations. *Bell Systems Technical Journal*, 44:1809–1812, 1965.

[25] J.S. Shamma. *Analysis and Design of Gain Scheduled Control Systems.* PhD thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, 1988.

[26] J.S. Shamma. The necessity of the small-gain theorem for time-varying and nonlinear systems. *IEEE Transactions on Automatic Control*, AC–36(10):1138–1147, 1991.

[27] J.S. Shamma and M. Athans. Analysis of nonlinear gain scheduled control systems. *IEEE Transactions on Automatic Control*, TAC–35(8):898–907, 1990.

[28] J.S. Shamma and M. Athans. Guaranteed properties of gain scheduled control of linear parameter-varying plants. *Automatica*, **27**(3):898–907, 1991.

[29] J.S. Shamma and M. Athans. Gain scheduling: Potential hazards and possible remedies. *IEEE Control Systems Magazine*, **12**(3):101–107, 1992.

[30] J.S. Shamma and M.A. Dahleh. Time-varying versus time-invariant compensation for rejection of persistent bounded disturbances and robust stabilization. *IEEE Transactions on Automatic Control*, **AC–36**(7):838–847, 1991.

[31] J.S. Shamma and R. Zhao. Fading-memory feedback systems and robust stability. *Automatica*, **29**(11):191–200, 1993.

[32] M. Tahk, M.M. Briggs, and P.K.A. Menon. Applications of plant inversion via state feedback to missile autopilot design. In *Proceedings of the 27th IEEE Conference on Decision and Control*, pages 730–735, Austin, TX, December 1988.

[33] M.S. Verma. Coprime factorizational representations and stability of nonlinear feedback systems. *International Journal of Control*, **48**:897–918, 1988.

[34] D.C. Youla, H.A. Jabr, and J.J. Bongiorno, Jr. Modern Wiener-Hopf design of optimal controllers: Part II. *IEEE Transactions on Automatic Control*, AC–21(3):319–338, 1976.

[35] G. Zames. On th input-output stability of time-varying nonlinear feedback systems. Part I: Conditions using concepts of loop gain, conicity, and positivity. *IEEE Transactions on Automatic Control*, **AC–11**(2):228–238, 1966.

FUZZY AND NEURAL LEARNING: A COMPARISON

Thomas Sudkamp
Associate Professor
Department of Computer Science

Wright State University
Dayton, Ohio 45435

Final Report for:
Research Initiation Program
Wright Laboratories

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

and

Wright State University

October, 1992

FUZZY AND NEURAL LEARNING: A COMPARISON

Thomas Sudkamp
Associate Professor
Department of Computer Science
Wright State University
Dayton, Ohio 45435

Abstract

Fuzzy inference systems and neural networks both provide mathematical systems for approximating continuous real-valued functions. Historically, fuzzy rule bases have been constructed by knowledge acquisition from experts while the weights on neural nets have been learned from data. This paper examines algorithms for constructing fuzzy rules from input-output training data. The antecedents of the rules are determined by a fuzzy decomposition of the input domains. The decomposition localizes the learning process, restricting the influence of a training example to single rule. Fuzzy learning algorithms fill the entries in a fuzzy associative memory using the degree to which the training data matches the rule antecedents. The fuzzy learning algorithms are tested with both precise and noisy training data. Unlike the neural network algorithms, fuzzy learning algorithms require only a single pass through the training set. This produces a computationally efficient method of learning. The effectiveness of the fuzzy learning algorithms is compared with that of a feedforward neural network trained with back-propagation.

1 Introduction

Fuzzy set theory provides a formal system for representing and reasoning with uncertain information. Fuzzy rules encapsulate approximate relationships between the input and the response or, in the terminology of rules, between the antecedent and consequent domains. Historically, fuzzy rule bases have been obtained by knowledge acquisition from experts while the weights on the neural nets have been learned from data. This paper examines algorithms that construct fuzzy rules from input-output training sets. Learning a fuzzy rule base from data begins by decomposing the input and output spaces into fuzzy regions. The regions define the possible antecedents and consequents of the rules. Following the approach of Wang and Mendel, the learning algorithms proceed by determining the entries in a fuzzy associative memory. Each entry represents a localized region in the input space and the value assigned to the entry indicates the response when the input occurs within that region.

Unlike the neural networks, which require multiple passes through the training sets, the fuzzy learning algorithms require only a single pass through the data. This produces a computationally more efficient method of learning. The paper concludes with a comparison of the performance of the fuzzy learning algorithms with backpropagation neural learning. Initial tests have shown that the fuzzy algorithms outperform the neural net algorithms when run with on the same training data.

2 Fuzzy Inference Systems

Fuzzy set theory provides a formalism suitable for representing the imprecise, vague, and ambiguous information. The uncertainty inherent in complex problem domains has made fuzzy representations and inference an increasingly important tool for the analysis of information in expert and database systems, decision analysis, and automatic control systems. Automatic control, in particular, has provided a fertile area for the development of fuzzy inference systems. Throughout this paper, the exposition of fuzzy rule based systems will utilize the terminology of fuzzy control. A thorough introduction to the history and components fuzzy control systems can be found in [6,7].

The input to a fuzzy controller consists of the values of a set of variables that describe the state or configuration of the system being controlled. In fuzzy rule based system depicted in Figure 1, the state variables assume values from the sets U and V. These values may be obtained directly from sensors or from a human operator. When the input is given as precise values, fuzzification produces an interpretation of the input that incorporates the imprecision introduced by the acquisition and measurement of the input.

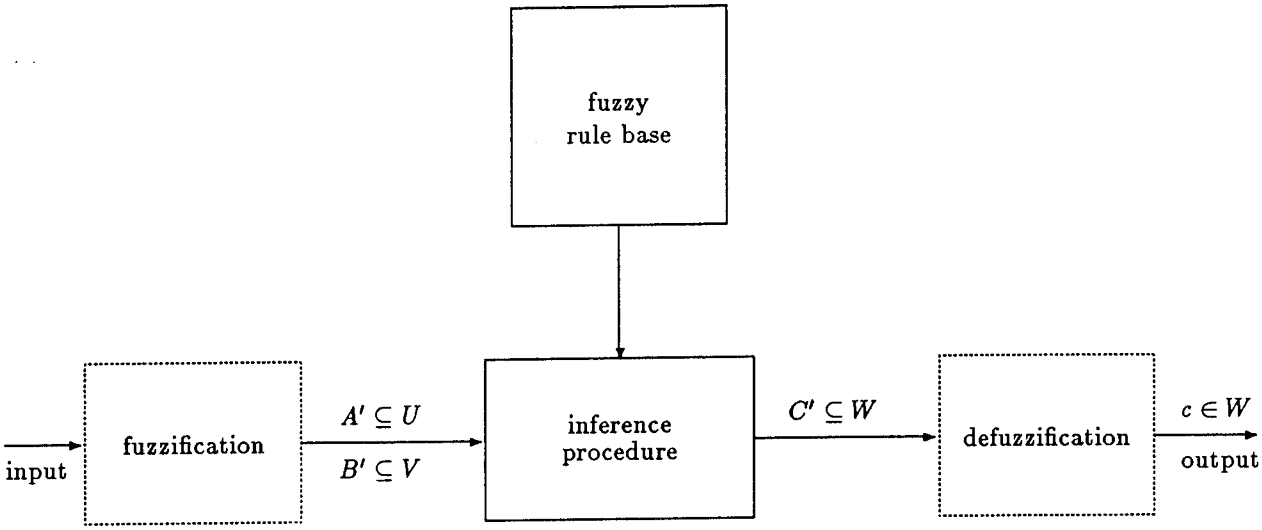


Figure 1: Fuzzy inference system

The rule base contains the information that relates the input conditions to the output responses. Fuzzy rules have the form 'if X is A and Y is B then Z is C' where A and B are fuzzy sets over the input domain U and V and C is a fuzzy set over the output domain W. The essence of fuzzy inference is that of partial matching and similarity. The input A' and B' is compared with the antecendents of the rules in the fuzzy rule base. The compositional rule of inference [12] or a compatibility modification approach [9] utilizes the degree to which the antecedents match the input to produce a fuzzy set C' over the output domain. For situations in which the required response must be a precise action, a defuzzification module transforms the output fuzzy set C' into a single value $c \in W$.

3 Triangular-partition rule bases

This section reviews the properties of fuzzy inference by examining one input and one output systems. A number of conditions are imposed on the decomposition of the input and output universes to produce a standard form for the fuzzy rule base and to facilitate the inference calculations. All the domains under consideration are assumed to be normalized. That is, each will take values from the interval $[-1,1]$. Triangular membership functions are used to subdivide the input and output universes. A fuzzy set A_i defined by a triangular membership function has the form

$$\mu_{A_i(x)} = \begin{cases} (x - a_{i-1})/(a_i - a_{i-1}) & \text{if } a_{i-1} \leq x \leq a_i \\ (-x + a_{i+1})/(a_{i+1} - a_i) & \text{if } a_i \leq x \leq a_{i+1} \\ 0 & \text{otherwise.} \end{cases}$$

The membership values of A_i are nonzero only on the interval (a_{i-1}, a_{i+1}). The point a_i, the unique domain element that has membership value 1 in A_i, will be referred to as the 'midpoint' of A_i. A triangular decomposition of a universe consists of a sequence of triangular fuzzy subsets $A_1, \ldots, A_n$. The leftmost and rightmost fuzzy regions are truncated with the midpoint as the leftmost and rightmost position, respectively. Thus $\mu_{A_1}(-1) = 1$ and $\mu_{A_n}(1) = 1$. Figure 2 shows a triangular decomposition of $[-1,1]$ into five fuzzy regions.

Let $A_1, \ldots, A_n$ be a triangular decomposition of $[-1,1]$ and let a_i be be the midpoint of A_i. We assume that a triangular decomposition forms a fuzzy partition of the underlying universe. That is,

$$\sum_{i=1}^{n} \mu_{A_i}(u) = 1$$

for every $u \in U$. A decomposition of the input and output domains that satisfies the preceding requirements will be called a TP (Triangular Partition) system.

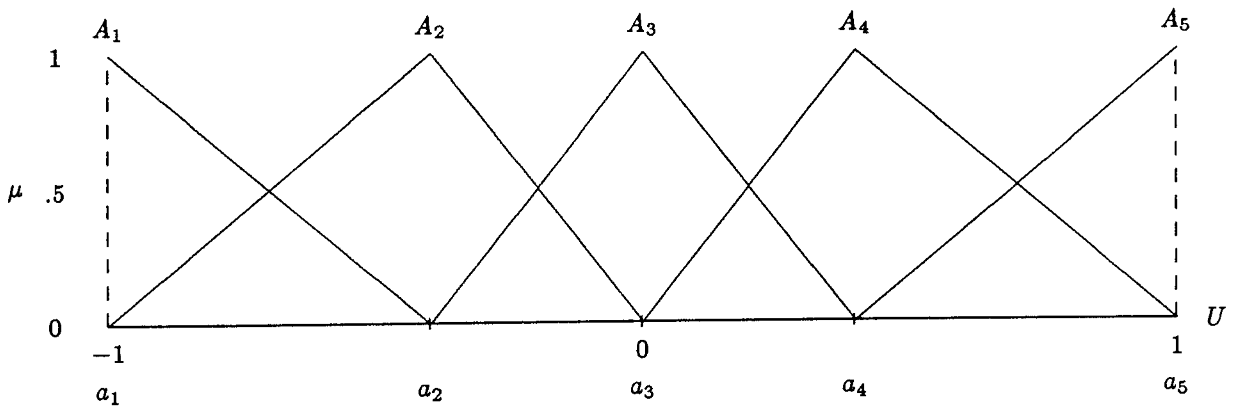


Figure 2: Triangular decomposition

The combination of triangular membership functions and the fuzzy partion requirement imposes strict limitations on the form of the decomposition of the domains. The midpoints of the triangular membership functions divide the input domain into a sequence of intervals $[a_i, a_{i+1}]$. On the region $[a_i, a_{i+1}]$, the only fuzzy sets to assume nonzero membership values are A_i and A_{i+1}. Thus,

$$\mu_{A_i}(x) = 1 - \mu_{A_{i+1}}(x) \tag{1}$$

for every $x \in [a_i, a_{i+1}]$. At a_i, $\mu_{A_i}(a_i) = 1$ and, by the partition condition, $\mu_{A_j}(x) = 0$ for all $j \neq i$. The fuzzy partition also ensures that there is at least one A_j such that $\mu_{A_j}(u) \geq .5$, for every $u \in U$. This property, called .5 completeness [6], guarantees that there is at least one rule whose antecedant significantly matches every possible input.

The input universe U of a one-input one-output TP system is subdivided into triangular fuzzy regions $A_1, \ldots, A_n$ as in Figure 2. The output domain W is decomposed by the fuzzy partition $C_1, \ldots, C_m$. The restrictions on the decomposition make inference in a TP system very straightforward and efficient. As noted above, any input value x has nonzero membership in at most two fuzzy sets over U. Assume that A_i and A_{i+1} are the two fuzzy sets providing nonzero membership for x (Figure 3). The midpoints a_i and a_{i+1} provide sufficient information to determine the line segments that form the membership functions in this region. The equations of lines l_1 and l_2 are

$$\mu_{A_i}(x) = \frac{-x + a_{i+1}}{a_{i+1} - a_i} \tag{2}$$

$$\mu_{A_{i+1}}(x) = \frac{x - a_i}{a_{i+1} - a_i} \tag{3}$$

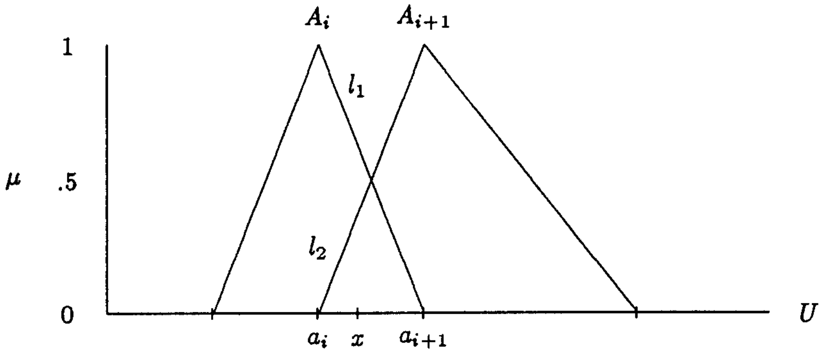


Figure 3: Consecutive regions in TP decomposition

The fuzzy rules associated with regions A_i and A_{i+1} have the form

'if X is A_i then Z is C_r'

and

'if X is A_{i+1} then Z is C_s',

where C_r and C_s are fuzzy regions in the output space. Let c_r and c_s denote the midpoints of C_r and C_s respectively. Using weighted averaging defuzzification, the result z specified by input x is given by

$$z = \frac{\mu_{A_i}(x)c_r + \mu_{A_{i+1}}(x)c_s}{\mu_{A_i}(x) + \mu_{A_{i+1}}(x)}. \tag{4}$$

The membership values $\mu_{A_i}(x)$ and $\mu_{A_{i+1}}(x)$ are obtained directly from Equations 2 and 3. Substituting into (4) and simplifying produces

$$z = \frac{x(c_s - c_r) + a_{i+1}c_r - a_i c_s}{a_{i+1} - a_i}. \tag{5}$$

Equation 5 shows that only two constants are required to determine the appropriate action for any input in an interval $[a_i, a_{i+1}]$. These constants, $(c_s - c_r)/(a_{i+1} - a_i)$ and $(a_{i+1}c_r - a_i c_s)/(a_{i+1} - a_i)$, are completely determined by the rule base and the midpoints of the triangular decomposition of the domains. Thus a TP fuzzy inference system can be represented in tabular form as illustrated in Example 1. Fuzzy rules and inference tables provide equivalent representations of the information required by a TP system. The original fuzzy rule base can be reconstructed from the inference table. The fuzzy sets A_i are obtained from the endpoints the intervals in the table. The consequent of the rule 'if X is A_i' is the region C_r in W whose midpoint is obtained by setting $x = a_i$ in Equation 5.

Example 1: Consider the decomposition of the input universe U into five fuzzy regions with midpoints $a_1 = -1, a_2 = -.4, a_3 = 0, a_4 = .4, a_5 = 1$ (as in Figure 2). Let $c_1 = -1, c_2 = 0, c_3 = 1$ be the midpoints of the decomposition of the output space. A fuzzy rule base of the form

'if X is A_1 then Z is C_1'

'if X is A_2 then Z is C_2'

'if X is A_3 then Z is C_1'

'if X is A_4 then Z is C_2'

'if X is A_5 then Z is C_3'

can be expressed by the table

input interval	$(c_s - c_r)/(a_{i+1} - a_i)$	$(a_{i+1}c_r + a_ic_s)/(a_{i+1} - a_i)$
$[-1, -.4)$	1.67	.67
$[-.4, 0)$	−2.5	1.0
$[0, .4)$	2.5	−1.0
$[.4, 1]$	1.67	.67

With this representation, the evaluation of input and the determination of the output is a simple table lookup followed by a linear evaluation. For example, the action specified by input $x = .2$ is $z = (.2) \cdot (2.5) - 1.0 = -.5$.

The inference process can be made even more efficient by requiring the membership functions to be isoceles triangles with bases of the same width. With this added restriction the midpoints $a_1 = -1, a_2, \ldots, a_n = 1$ divide the input domain into $n-1$ intervals $[a_i, a_{i+1}]$ of the same length. A fuzzy inference system with evenly spaced midpoints will be referred to as a TPE system. Given a TPE system, the amount of computation required for processing input is independent of the number of rules. The tabular information can be stored in a manner that permits direct addressing. Given input x, the row of the table with the appropriate constants is determined by the function $trunc(\frac{2x+2}{n-1}) + 1$. Thus no searching is required to find the appropriate 'rules' in the inference table. The independence of the runtime efficiency on the number of rules has important implications for systems that learn fuzzy rules from data.

It has been shown that, under the constraints of a TPE system, a fuzzy rule base and an inference table are equivalent representations of the information required for fuzzy inference. Another equivalent tabular formulation, the fuzzy associative memory (FAM), provides a representation that is useful in the process of learning fuzzy rules. A FAM is a k-dimensional table where each dimension corresponds to one of the input universes of the rules. The i'th dimension of the table is indexed by the fuzzy sets that comprise the decomposition of the i'th input domain. A 2-dimensional FAM is given in Example 2.

The FAM representation will be modified to produce a numeric inference (NI) table. Rather than labeling the dimensions with the fuzzy set names, the indices will represent the corresponding midpoints of the set. The entries in the table are the midpoints of consequent of the associated rule. Example 2 gives the NI table corresponding to the previously constructed FAM.

Example 2: Consider a fuzzy inference system in which the input comes from domains U and V and the output is in W. Let $A_1, \ldots, A_4$, $B_1, \ldots, B_3$, and $C_1, \ldots, C_5$, be TPE decompositions of U, V, and W respectively. A fuzzy rule base for such a system may be comprised of the rules

'if X is A_1 and Y is B_2 then Z is C_1'
'if X is A_1 and Y is B_3 then Z is C_2'
'if X is A_2 and Y is B_1 then Z is C_1'
'if X is A_2 and Y is B_2 then Z is C_2'
'if X is A_3 and Y is B_2 then Z is C_3'
'if X is A_3 and Y is B_3 then Z is C_4'
'if X is A_4 and Y is B_3 then Z is C_5'

The FAM for this inference system has the form

	B_1	B_2	B_3
A_1		C_1	C_2
A_2	C_1	C_2	
A_3		C_3	C_4
A_4			C_5

The consequent of the rule with antecedent 'if X is A_i and Y is B_j' is entered in the (i, j)'th position.

The NI table constructed from the preceeding FAM is

	-1	0	1
-1		-1	$-.5$
$-.33$	-1	$-.5$	
$.33$		0	$.5$
1			1

Of the four equivalent formulations of the information utilized in fuzzy inference, only one employs rules for knowledge representation. With this in mind, the more generic term *fuzzy inference system* will be used to refer to any of the aforementioned systems.

The popularity and practicality of fuzzy inference derives from the ability to linguistically specify relationships that are too complex or not well enough understood to be described by precise mathematical models. The TPE requirements on the antecedents and consequents may seem too stringent to be amenable to the acquisition of fuzzy rules from experts. Heuristically generated rules often differ in size with smaller regions indicating the areas where precise control is essential and larger regions indicating areas that are not greatly affected by minor changes in the input. If the rules are learned from data, however, the exact form of the membership function need not conform to any intuitive model.

4 Learning Strategies

The impetus behind the development fuzzy control theory was the complexity inherent in most dynamic systems. Fuzzy control employs heuristic, linguistic information to develop the rule base rather than attempting to produce a precise mathematical model or a simulation of the system. Consequently, the knowledge is acquired separately (off-line) and provided to the fuzzy inference system. The recognition of fuzzy inference as a methodology for the universal approximation permits the development of fuzzy rules from data. Rather than interrogating an expert to extract insight into the dynamics of the system, it is possible to record the states of the system and the operator responses and use this data as a training set for learning the appropriate rules.

This section presents the first step in the process of learning the fuzzy rules that comprise a TPE system. The strategy is a modification and extension of the fuzzy associative memory system of Wang and Mendel [10]. As before, a fuzzy inference system with one input and one output will be used to demonstrate the procedure. Following the conventions established in Section 3, the decompositions of the input and output domains are TPE with midpoints $a_1, \dots, a_n$ and $c_1, \dots, c_m$, respectively.

The philosophy behind learning fuzzy rules is that a set of training examples, input-output pairs (x_i, z_i), provide the information for the generation of the rules. A training example with abscissa x_i will affect only the rules whose antecedents assign positive membership values to x_i. With the TPE restrictions, input $x_i \in [a_j, a_{j+1}]$ may have nonzero membership in only A_j and A_{j+1}. The output z_i associated with input x_i

contributes to the determination of the consequent of the rules whose antecedents are 'if X is A_j' and 'if X is A_{j+1}'. Algorithms to learn fuzzy inference rules are comprised of two major components: the specification of the topology and the rule generation or learning algorithm. These topics will be considered in order.

The specification of the topology consists of describing decompositions of the input and output domains. This determines the number of fuzzy rules to be constructed or, equivalently, the size of the FAM or NI table. Since the TPE membership functions are employed, the decompositions are completely determined by the number of regions in the input space (n) and in the output space (m).

The second component required by the learning system is the procedure for generating the rules from the training set. Fuzzy learning is inherently a local process. A training example (x, z) with $x \in [a_j, a_{j+1}]$ is involved only in the determination of the rules with antecedents 'if X is A_j' and 'if X is A_{j+1}'. In terms of the piecewise linear approximating function being constructed, (x, z) contributes solely to the determination of the line segments which have $(a_i, \hat{f}(a_i))$ and $(a_{i+1}, \hat{f}(a_{i+1}))$ as endpoints.

In the learning algorithm of Wang and Mendel, only examples that significantly match the antecedent are used to determine the consequent. Training example (x, z) is considered in the determination of the consequent of 'if X is A_j' only if $\mu_{A_j}(x) \geq .5$. Thus, due to the fuzzy partition property, each example contributes to the generation of only one rule (examples with .5 membership in two contiguous fuzzy regions are assigned to a single region for the learning process). Let T be the set of examples (x, z) that satisfy $\mu_{A_j}(x) \geq .5$. The training instance $(x_i, z_i) \in T$ whose x value has the greatest membership in A_j is used to determine the appropriate output region C_r. This technique will be labeled **L1**.

L1: Let (x_k, z_k) be the training example in T that has the maximal membership in A_j. If more than one example assumes the maximal membership, one is selected arbitrarily. The consequent of the fuzzy rule is the region C_r in the output space in which z_k has maximal membership.

This method of determining the appropriate output region follows the standard fuzzy tradition of using the supremum for specifying the degree to which two fuzzy sets match. The determination of the consequent using a single training example may be adversely affected by erroneous or noisy data in the training set. Inaccuracy in a single training example, say (a_j, z) which has membership 1 in A_j, will be directly reflected in the rule regardless of the accuracy of the remaining examples. To help mitigate the effects of this possibility, the algorithm is modified to incorporate all the training examples in T.

L2: Let

$$w = \frac{\sum_{(x_i, y_i) \in T} \mu_{A_j}(x_i) z_i}{\sum_{(x_i, y_i) \in T} \mu_{A_j}(x_i)} \tag{6}$$

be the sum of the elements in T weighted by their membership degrees. The consequent of the fuzzy rule is defined as the region C_r in the output space in which w has maximal membership.

In **L2**, the property of having each example influence a single rule has been maintained. The summation in the weighted average could easily be extended to include all the examples (x, y) for which $\mu_{A_j}(x) > 0$. This modification would incorporate all the examples in the training set that match the antecedent A_j to any degree.

Example 3: The domain decompositions given in Example 1 are used to illustrate the generation of rules. Let $(-.2, .5)$, $(0, .6)$, and $(.1, .2)$ be the training examples whose x-coordinates fall within the interval $[a_2, a_4]$. The membership values of the abscissas of these points in A_3 are $\mu_{A_3}(-.2) = .5$, $\mu_{A_3}(0) = 1$, and $\mu_{A_3}(.1) = .75$.

Learning technique **L1** uses the single training example $(0, .6)$ to determine the consequent of the rule. C_3 is the region in W in which .6 has maximal membership. Thus **L1** produces 'if X is A_3 then Z is C_3'.

Using the weighted average of the training examples, **L2** selects the consequent that most agrees with

$$\frac{\mu_{A_3}(-.2).5 + \mu_{A_3}(0).6 + \mu_{A_3}(.1).2}{\mu_{A_3}(-.2) + \mu_{A_3}(0) + \mu_{A_3}(.1)} = .411.$$

In this case, the rule 'if X is A_3 then Z is C_2' is produced.

In both **L1** and **L2**, the degree to which the training examples lie within the regions determined by the domain decomposition determines value assigned to the entry in the FAM. Stated in terms of rules, the consequent of the rule is determined by the degree to which the training data matches the antecedent. These learning methods can be easily extended to fuzzy training examples using fuzzy partial matching indices [4,3,2] to determine the degree of match. Throughout this presentation, a training example was defined by a precise value: an ordered pair (x, z) for the one-input one-output case. The extension to fuzzy training examples facilitates the representation of imprecision in the data. A training example would have the form $(\mu_{A'}, \mu_{C'})$ where A' and C' are fuzzy sets over U and W respectively. Linguistically, fuzzy training examples permit data of the form 'if X is near .5 then Z is near .2' to be employed in the learning process.

Another approach to learning fuzzy rules from examples has been proposed by Kosko [5]. Kosko views the construction of a rule base (or FAM) as a search through the space of all fuzzy rule bases over a fixed topology. If the input space is decomposed into n regions and the output into m regions, there are nm possible rules. Any subset of these rules may be considered a fuzzy rule base. A relative frequency approach is used to determine the rules that are selected from the nm possible. For each input region A_i and output region C_j, a count is maintained of the number of training examples (x, z) that fall within the scope of the rule 'if X is A_i then Z is C_j'. That is, the number of examples for which $\mu_{A_i}(x) > 0$ and $\mu_{C_j}(z) > 0$. When the count exceeds a threshold value, the rule 'if X is A_i then Z is C_j' is included in the rule base.

5 Experimental Analysis

The objective of a fuzzy learning algorithm is to use a set of examples to construct an approximation of an unknown real-valued function. To discover the efficacy of the learning techniques, a testing methodology has been developed and exercised. A specific continuous function f is chosen as the target and an approximating function $\hat{f}$ is generated from a set of training examples.

The construction of a training set T of cardinality k begins by randomly selecting k elements from the domain of f. A precise training set for a one-variable function is comprised of k input-output pairs of the form $(x, f(x))$. To test the robustness of learning algorithms, training sets consisting of noisy data are also examined. Such a training set consists of pair $(x, f(x) + e(x))$ where $e(x)$ is an error function. In these experiments, for each point x, the error $e(x)$ is randomly chosen from a uniform distribution over the interval $[-.1, .1]$.

All the target functions considered in this paper have either one or two inputs and single output. That is, $f : [-1, 1] \rightarrow [-1, 1]$ or $f : [-1, 1] \times [-1, 1] \rightarrow [-1, 1]$. The input and output domains are decomposed into an identical number of regions. Consequently, specifying a decomposition of five regions for a two argument function creates a FAM with 25 entries. In this case, the region assigned to an entry will be one of the five in the decomposition of the output domain.

After the training set is processed, the values produced by the approximating function $\hat{f}$ and the target function f are compared on a sample of points evenly distributed over the input domain. For one dimensional functions, the error $|f(x) - \hat{f}(x)|$ is obtained at .01 intervals. The average and maximum error that occur on this set are recorded. The error on two dimensional functions is determined at lattice points separated by distance .01 in each coordinate direction. Tables 1 and 5 give results of the learning algorithms when exercised with precise and noisy training sets, respectively. No results are given for decompositions of 15 and 25 regions for a training set with 25 examples since there is insufficient data to completely generate the rule base.

			L1		**L2**	
function	regions	examples	ave	max	ave	max
			error	error	error	error
$f(x) = x/2$	5	25,100,200	.122	.250	.122	.250
	15	100,200	.037	.071	.037	.071
	25	100,200	.020	.041	.020	.041
$f(x) = x^3$	5	25,100	.044	.125	.044	.125
		200	.044	.125	.089	.500
	15	100	.029	.079	.029	.071
		200	.027	.088	.026	.079
	25	100	.017	.062	.012	.041
		200	.018	.088	.015	.050
$f(x) = \sin(2\pi x)$	5	25	.635	1.251	.568	1.251
		100	.635	1.251	.544	1.000
		200	.615	1.000	.544	1.000
	15	100	.096	.231	.090	.216
		200	.060	.159	.071	.179
	25	100	.043	.167	.032	.088
		200	.026	.092	.031	.083
$f(x,y) = x^2 + y^2 - 1$	5	625	.149	.500	.216	.500
		2500	.082	.250	.216	.500
		5000	.082	.250	.216	.500
	15	2500	.057	.143	.060	.143
		5000	.043	.143	.043	.143
	25	5000	.025	.086	.033	.086

Table 1: Error in learning: precise data

			L1		L2	
function	regions	examples	ave	max	ave	max
			error	error	error	error
$f(x) = x/2$	5	25,100,200	.122	.250	.122	.250
	15	100,200	.056	.139	.056	.139
	25	100,200	.053	.117	.053	.117
$f(x) = x^3$	5	25	.044	.125	.143	.384
		100, 200	.044	.125	.044	.125
	15	100,200	.049	.098	.049	.098
	25	100	.036	.138	.036	.088
		200	.042	.169	.039	.109
$f(x) = \sin(2\pi x)$	5	25	.654	1.251	.604	1.251
		100	.635	1.251	.544	1.000
		200	.615	1.000	.544	1.000
	15	100	.109	.265	.077	.216
		200	.087	.209	.083	.209
	25	100	.076	.250	.053	.134
		200	.066	.544	.047	.134
$f(x, y) = x^2 + y^2 - 1$	5	625	.149	.500	.216	.500
		2500	.082	.250	.216	.500
		5000	.189	.500	.216	.500
	15	2500	.074	.160	.083	.181
		5000	.046	.158	.057	.286
	25	5000	.059	.167	.057	.167

Table 2: Error in learning: noisy data

The function $x/2$ is included in the test set to illustrate the restrictions imposed by the TPE decompositions. Although linear, TPE systems cannot precisely approximate $x/2$ since the approximating functions must satisfy $\hat{f}(a_i) = c_j$ for every rule 'if X is A_i then Z is C_j' in the rule base. The remaining one-variable target functions were selected to demonstrate the ability of the learning algorithms to construct approximations to highly nonlinear target functions. The two-variable function $f(x, y) = x^2 + y^2$ was included to examine the effects of multiple clauses in the antecedents in a fuzzy rule base.

Generally, an increase in the size of the training set is accompanied by a decrease in the error in the approximation. However, increasing the number of examples does not insure a better approximation. This is particularly noticable when the domains decompositions consist of a relatively small number of regions. In Section 4 it was shown that only m^n functions are producable by a TPE system with n input regions and m output regions. When $n = m = 5$, the learning algorithm is limited to selecting one of 3125 possible approximations. This lack of flexibility often negates the effects of an increase in the number of training examples. The approximations of $x/2$ illustrate this property: the increase from 100 to 200 examples fails to improve the approximation, the optimal approximation already having been discovered.

In rare cases, increasing the number of examples may increase the error in the resulting approximation. This occurs when an additional example produces a change in the consequent of a rule. Consider the situation in which additional information causes a change from 'if X is A_i then Z is C_j' to 'if X is A_i then Z is C_k'. The approximation at point a_i then changes from $\hat{f}(a_i) = c_j$ to $\hat{f}(a_i) = c_k$. This change may reduce the error at the point a_i but increase the average error over the line segments beginning and ending at $\hat{f}(a_i)$ due to the shape of f over the corresponding intervals. This is illustrated by the approximations of x^3 (see Table 1). For a decomposition consisting of 25 regions, the approximations constructed by both **L1** and **L2** with 200 examples have greater error than those constructed from 100.

A comparison of the two algorithms shows that **L2** generally outperforms the **L1** on one-variable targets. That is, both the average and maximal error in the approximations produced by **L2** are less than those produced by **L1**. This relationship was also exhibited by tests on an extended set of target functions. As expected, the addition of noise to the input data increases the average error and the maximal error for both learning algorithms. It does not, however, seem to alter the relative performances of the two techniques.

There are anomalous situations in which the noisy data produced better approximations than the corresponding precise data. This can be seen in the approximation of $\sin(2\pi x)$ with 15 regions, 100 examples, and the **L2** learning algorithm. This occurs when the noise fortuitously relocates one of the endpoints of an approximating line segment to more closely match the target function over the entire length of the segment.

6 Fuzzy and neural learning

In this section, the results of the fuzzy algorithms are compared with the approximating functions constructed by feedforward neural networks. Fuzzy learning and neural network learning have different underlying philosophies. Ideally, a neural net will attempt to 'memorize' the data. That is, when the learning converges there should be no the error on the training set. With fuzzy learning, it is expected that the data is approximate and no such exact match on the training data is anticipated or desired.

The topology of a feedforward neural network is defined by the number of layers of nodes and their connections. The network model used in this comparison has a single hidden layer with total connectivity between layers. The number of input nodes is determined by the number or arguments in the target function f. The output layer has a single node that provides the result of the approximation. In the networks examined, the hidden layer consists of 5, 11, and 25 nodes. The neural net is trained using backpropagation with a baseline learning rate of .7. The initial weights are randomly selected from the range $[-.5, .5]$. To help mitigate the effects of the random selection of the initial configuration, each test has been run with several different sets of initial weights.

As has been noted previously, the fuzzy learning algorithms make a single pass through the training set while backpropagation may require multiple passes (epochs) to converge. The network approximations in Table 3 are constructed using one, 50 and 2000 epochs. The single epoch results are included to provide a direct comparison with the single pass of the fuzzy algorithms.

For the smooth one-variable functions, the neural net approach produced approximating functions comparable to the fuzzy algorithms. Fifty epochs were required to generate approximations to the linear function $x/2$ of equal accuracy as the fuzzy approximations. For the cubic function, 2000 epochs were required to produce solutions of similar quality. When the target function was highly nonlinear, the fuzzy algorithms produced approximations that were more accurate than the best neural net generated functions.

The difference between fuzzy learning and neural learning may be attributable to the global nature of neural net learning as opposed to the local nature of fuzzy learning. In fuzzy learning, the influence of a training example is limited to the regions in the domain decomposition within which the input lies. The process of adjusting weights in backpropagation learning considers a training example as global information: all weights within the network may be affected.

The preceding comparisons are not meant to imply any definitive relationship between neural and fuzzy learning algorithms. There are a variety of neural architectures and algorithms, some of which may produce

			1 epoch		50 epochs		2000 epochs	
function	hidden	examples	ave	max	ave	max	ave	max
	nodes		error	error	error	error	error	error
$x/2$	5	25	.322	.758	.018	.040	.011	.024
		100	.241	.529	.012	.043	.009	.030
		200	.190	.424	.009	.030	.006	.021
	11	25	.288	.689	.013	.031	.009	.024
		100	.220	.517	.012	.046	.009	.029
		200	.118	.283	.008	.034	.007	.021
	25	25	.411	.865	.008	.043	.008	.032
		100	.165	.362	.010	.040	.009	.029
		200	.073	.185	.008	.029	.007	.020
x^3	5	25	.447	1.340	.174	.439	.026	.133
		100	.308	1.101	.140	.546	.017	.086
		200	.209	.888	.129	.540	.016	.088
	11	25	.456	1.335	.175	.376	.026	.146
		100	.353	1.132	.152	.557	.016	.088
		200	.153	.687	.144	.555	.015	.085
	25	25	.670	1.606	.165	.396	.034	.168
		100	.347	1.079	.168	.594	.016	.089
		200	.147	.581	.143	.532	.015	.088
$\sin(2\pi x)$	5	25	.652	1.317	.561	1.191	.121	.719
		100	.666	1.380	.581	1.179	.117	.635
		200	.627	1.291	.566	1.098	.129	.625
	11	25	.653	1.314	.560	1.163	.156	.695
		100	.716	1.531	.583	1.224	.085	.212
		200	.625	1.268	.581	1.194	.072	.152
	25	25	.766	1.637	.560	1.144	.519	.153
		100	.826	1.766	.584	1.246	.122	.703
		200	.592	1.171	.587	1.257	.072	.204
$x^2 + y^2 - 1$	5	25	.454	1.021	.398	1.255	.073	.763
		100	.394	1.534	.353	1.545	.055	.666
		200	.386	1.525	.322	1.437	.044	.395
	11	25	.477	1.056	.387	1.405	.070	.489
		100	.404	1.573	.349	1.564	.056	.584
		200	.400	1.585	.270	1.240	.038	.358
	25	25	.414	1.112	.412	1.355	.065	.559
		100	.424	1.639	.335	1.586	.052	.653
		200	.440	1.700	.206	1.263	.036	.322

Table 3: Neural network learning: precise data

results superior to those obtained by a simple one hidden layer feedforward network (see, for example, [1,11,8]). This standard neural architecture has been chosen only to provide a simple comparison between the techniques.

7 Conclusions

Fuzzy learning algorithms provide an efficient single pass method for producing approximating functions from training data. Fuzzy algorithms are local: the decomposition of the input domains into a fuzzy partition produces a set of overlapping regions. The approximation over a region is determined solely by the examples within that region. Restricting the influence of a training example to a local region focuses the information contained in the example. Locality allows fuzzy learning algorithms to produce accurate approximations with a single pass of the training set. An advantage of fuzzy rule based inference is the intuitive linguistic form of fuzzy rules. Fuzzy inference provides the opportunity to combine the two techniques for determining rules, acquiring rules by elicitation from experts and learning from data. Further research is necessary to develop methods for incorporating both types of information in the development of rule bases. In particular, this requires the ability to merge rules defined over different domain decompositions.

References

[1] I. Aleksander and H. Morton. *An Introduction to Neural Computing.* Chapman and Hall, New York, 1990.

[2] V. Cross and T. Sudkamp. Compatibility and aggregation in fuzzy evidential reasoning. In *Proceedings of the 1991 IEEE International Conference on Systems, Man, and Cybernetics*, pages 153–158, 1991.

[3] V. Cross and T. Sudkamp. Compatibility measures and aggregation operators for fuzzy evidential reasoning. In *Proceedings of the North American Fuzzy Information Processing Society*, pages 13–17, 1991.

[4] D. Dubois and H. Prade. On several representations of an uncertain body of evidence. In M.M. Gupta and E. Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 167–181, North Holland, 1982.

[5] B. Kosko. *Neural Networks and Fuzzy Systems: A dynamical systems approach to machine intelligence.* Prentice Hall, Englewood Cliffs, NJ, 1992.

[6] C. C. Lee. Fuzzy logic in control systems: Part I. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(2):404–418, 1990.

[7] C. C. Lee. Fuzzy logic in control systems: Part II. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(2):419–435, 1990.

[8] Yoh-Han Pao. *Adaptive Pattern Recognition and Neural Networks.* Addison-Wesley, Reading, MA, 1989.

[9] I. B. Turksen. Fuzzy second generation expert system design for IE/OR/MS. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, pages 779–786, 1992.

[10] L. Wang and J. M. Mendel. *Generating Fuzzy Rules form Numerical Data, with Applications.* Technical Report USC-SIPI-169, Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089, 1991.

[11] P. D. Wasserman. *Neural Computing: Theory and Practice.* Van Nostrand Reinhold, New York, 1989.

[12] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3:28–44, 1973.

1992 USAF-RDL FACULTY RESEARCH PROGRAM/
GRADUATE STUDENT RESEARCH PROGRAM

Sponsored by the
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Conducted by the
Research & Development Laboratories

R. I. P. FINAL REPORT #140

PHOTOREFLECTANCE FROM GaAs/AlGaAs STRUCTURES

Prepared by: Michael Sydor
Academic Rank: Professor
Department and University: Physics Dept.
Univ. of Mn. Duluth
Research Location: WRL/MLPO
Wright-Patterson AFB
Dayton Ohio 45433
USAF Researcher: William Mitchel

Date: 20 Dec. 1992
Contract No: F49620-90-C-0053

SIMULTANEOUS ELECTROREFLECTANCE/PHOTOREFLECTANCE

by
Michael Sydor
Physics Dept.
University of Minnesota, Duluth
RIP Final Report #140

ABSTRACT

We have developed a new Electroreflectance/Photoreflectance technique (ER/PR) for studies of the internal electric field in layered electronic materials. The technique can be used to determine distribution of applied potentials and the direction of the electric field within layered electronic materials. It can also separate out the effects of the internal electric field modulation from the extraneous modulation in Photoreflectance due to laser generation of free charge carriers. The technique provides the essential data for verification of numerical models of Photoreflectance from layered electronic structures because it allows for isolation of Photoreflectance signatures from various layers and interfaces within the structures. Although we concentrate here on the presentation of the new technique, we outline some important differences between Electroreflectance and Photoreflectance modulation mechanisms, and apply the results to verification of previously reported numerical model results.

INTRODUCTION:

Photoreflectance (PR) and Electroreflectance (ER) are two commonly used techniques for investigation of electronic materials. The techniques measure the changes in reflectance (ΔR) produced by modulating sample's built-in internal electric field $\mathcal{E}$. In PR the modulation is accomplished by illuminating the sample with a variable highly absorbed secondary light source, usually a chopped laser beam, which generates free charge carriers within the sample. The free charge carriers reduce the magnitude of the built-in $\mathcal{E}$ during the laser-on portion of the chopping cycle, and thereby modulate the sample's reflectance.[1] For $|\mathcal{E}| < 10^4$ V/cm, the normalized change in sample's reflectance, ($\Delta R/R$), between the unperturbed and the laser-illuminated condition yields a narrow, ~30 meV, third derivative-like response at the sample's critical point transition energies. When the response is narrow, it can be used in a very accurate determinations of the band-gap energy of semiconductors.[2-5] For higher electric fields, $|\mathcal{E}| > 10^4$ V/cm, $\Delta R/R$ has a broader oscillatory behavior which can be used to determine the magnitude of $\mathcal{E}$.[2-5] However, regardless of the breadth of the response or the magnitude of $\mathcal{E}$, PR data alone is insufficient to determine the direction of $\mathcal{E}$ within a sample because $\Delta R/R$ depends on $|\mathcal{E}|$ and $|\Delta\mathcal{E}|$.[2,3]

The ER technique is very similar to the PR technique.[4] In ER there is no laser modulation, instead the modulation of $\mathcal{E}$ is provided by an external a.c. voltage applied at a semitransparent electrode on the sample surface. $\Delta R/R$ response in ER is often identical to the PR response because both responses depend on $|\mathcal{E}|$ and $|\Delta\mathcal{E}|$.[4] The advantage of using PR over the ER is the fact that PR requires no electrodes which may alter the nature of the built-in potential at the sample surface. On the other hand, PR entails laser generation of free charge carriers which can modulate not only the internal electric field, but may also introduce extraneous effects such as electric field nonuniformity and charge carrier trapping effects which are poorly understood.[6,7] As a result, the most recent improvements in modulated reflectance techniques have concentrated on the development of a Contactless Electroreflectance (CER) which avoids the use of electrodes and the use of laser modulation.[6,8] However, in analysis of modulated reflectance from layered structures it is often useful to know the direction of the internal electric field and to understand the effects of modulation with depth.[9] None of the above techniques taken separately can provide such information, and are thus rather limited in sample characterization and in numerical model verification. On the other hand, the use of simultaneous modulations can provide useful additional information. For instance, PR modulation depends on the laser penetration depth, a property which can be used in Differential Photoreflectance to preferentially suppress and identify the PR from various layers of the sample.[10,11] Another useful property of PR is the fact that unlike ER, PR generally vanishes for the layers with zero electric field. Consequently PR can be used as a null detector in studies of the distribution of applied potentials in layered structures. A third important difference between ER and PR was revealed in our current investigations. We have found that unlike PR, ER modulation becomes ineffective in thin layers of material whose Fermi level is pinned at its interfaces. How the ER and PR modulations differ and complement each other is in general a complex question and much remains to be done on the unraveling of the true nature of either modulation mechanism. We present here an example on how a simultaneous combination of the ER and PR modulations can provide information on the direction of $\mathcal{E}$, and how it can separate the effect of the electric field modulation from the effects due to electron trapping at interfaces. The new technique allows us to compare in detail the differences between ER and PR signals and examine the conditions when these differences are significant. It also helps us understand some experimental and numerical model results which were previously held suspect.

DISCUSSION OF METHOD:

The schematic for the ER/PR apparatus is shown in Fig. 1. The crucial aspect of ER/PR is the precise referencing of the ER modulation relative to the phase of the PR signal whose amplitude is maximized on the lock-in amplifier.

The principle of operation of ER/PR can be best understood by considering a specific case of a single layer of material with an internal electric field $\mathcal{E}$ subjected to the ER modulation. Fig. 2 illustrates such a case with conduction band bending in a single layer under the influence of a positive or a negative ER modulation which changes the ambient internal electric field from $\mathcal{E}$ to $\mathcal{E}'$ or $\mathcal{E}''$ respectively.

In examining Fig. 2, it is important to note that the effect of the positive and negative ER modulation would reverse if the direction of $\mathcal{E}$ and band bending were reversed. It is also important to note that unlike ER, the PR modulation of $\mathcal{E}$ is always toward the flat band condition, independent of the direction of $\mathcal{E}$ because laser injection of free carriers tends to short $|\mathcal{E}|$. As a result, a simultaneous application of ER and PR modulations produces a combined modulation, $|\Delta\mathcal{E}|$, which depends on the polarity of the applied ER modulation field relative to the direction of $\mathcal{E}$ within the sample. When simultaneous ER and PR modulations have an opposing effect on $\mathcal{E}$, for instance if an increase in $\mathcal{E}$ produced by a positive applied ER field is counterbalanced by a simultaneous PR modulation which tends to short $\mathcal{E}$, the combined modulation of $\mathcal{E}$ will be small, and the relative amplitude of $\Delta R/R$ in ER/PR will be small compared with the ER or the PR response taken separately. Thus, since the direction of the applied ER modulation field is always known, the relative amplitude of $\Delta R/R$ in ER/PR can be used to determine the direction of $\mathcal{E}$ relative to the direction of the known applied ER modulation field.

For opposing simultaneous ER/PR modulations, we expect the amplitude of $\Delta R/R$ in ER/PR from any given layer to be smaller than its amplitude in ER or the PR spectrum taken separately. The phase of the ER/PR response will have the phase of either the ER or the PR depending on which modulation was dominant when applied simultaneously. In the case of reinforcing simultaneous modulations, the amplitude of $\Delta R/R$ in ER/PR will be greater than the amplitude of $\Delta R/R$ in the separate PR or ER spectrum.

RESULTS FOR GaAs/AlGaAs HEMT STRUCTURE:

As an example, we consider the ER/PR spectra from a High Electron Mobility Transistor structure (HEMT). The structure has an important practical application and displays field reversal and electron trapping effects which are useful in demonstrating the versatility of the ER/PR technique. The general sample structure and its conduction band in the region of the two-dimensional electron gas (2DEG) is shown in Fig. 3. We demonstrate the various cases of ER/PR and the effect of field reversal in Fig. 4.

The first two traces in Fig. 4 show the separate ER responses, one for a positive, +ER, and one for a negative, -ER, modulation. The third trace shows the separate PR response. In all the traces of Fig. 4, the GaAs signature at ~1.4 eV and the AlGaAs signature at ~1.9 eV come from the modulation of the 2DEG region of the HEMT.[12,13] The response from layers which precede the 2DEG well, (doped AlGaAs/doped GaAs cap), is much broader and is negligible in this case.[10,11,13] (for instance, a piece of the sample with no metal electrode provides the same PR).

In Fig. 4, the trace labeled -ER/PR shows a very low $\Delta R/R$ resulting from simultaneous application of -ER and PR modulations. The simultaneous modulations have opposing effects on $\mathcal{E}$, indicating that $\mathcal{E}$ in the 2DEG region of this sample is negative, pointing toward the illuminated surface. In effect, the negative ER modulation field tends to increase the negative internal field $\mathcal{E}$ while the simultaneous PR laser illumination tends to reduce the magnitude of the negative $\mathcal{E}$, with the net result that the perturbed electric field $\mathcal{E}''$ differs little from the unperturbed $\mathcal{E}$, giving a small $|\Delta\mathcal{E}|$ and a small GaAs and AlGaAs signatures in the -ER/PR trace of Fig. 4. In the case of reinforcing ER/PR modulations, a +ER modulation reduces the magnitude of the negative $\mathcal{E}$, and the simultaneous laser illumination reduces the magnitude of $\mathcal{E}$ even further, giving together a large combined perturbation $|\Delta\mathcal{E}|$ and a large resulting +ER/PR signal as shown in Fig. 4.

The effect of field reversal is exhibited by the last trace in Fig. 4. When all experimental conditions are unchanged, except for the d.c. bias, we observe at -0.1 V bias an inversion of the +ER response, as shown in Fig. 4 by the (+ER -0.1V) trace.[8] Since the GaAs and AlGaAs signatures are known to come from the modulation of the 2DEG well region of the HEMT,[12-14] the reversal of $\mathcal{E}$ at -0.1 V suggests a major change in the concentration of the 2DEG in our sample as a function of the applied d.c. bias. This suspected formation of 2DEG at -0.1 V d.c. bias appears further confirmed at low temperature. At 80° K and -0.1 V bias, the ER modulation of the AlGaAs becomes ineffective, as though the Fermi level in the 2DEG well and the Fermi level at the doped AlGaAs conduction band on the opposite side of the AlGaAs barrier were pinned together. The AlGaAs signature at 80° K, -0.1 V bias vanishes in ER while the GaAs signature becomes small and narrow, as shown in Fig. 5. On the other hand, the laser modulation of the 2DEG region, at 80° K and -0.1 V d.c. bias, remains quite effective. Both PR signatures are well defined at -0.1 V bias and have amplitudes comparable to the PR and the ER signatures from an unbiased HEMT at 80° K, as shown in Fig.5.

To examine the possibility of using ER/PR in isolating the extraneous differences between ER and PR modulations,[14] we consider the case when modulation of $\mathcal{E}$ in a HEMT is equivalent in ER and PR. This condition is best accomplished when ER and PR modulations both drive $\mathcal{E}$ toward the flat band condition, but when

ER and PR modulations occur on the successive halves of the chopper cycle, so that the lock-in amplifier compares R at a constant value of $\mathcal{E}$ over the entire cycle. The individual ER and PR traces for this condition are shown in Fig. 6. The extraneous difference between ER and PR modulation is shown by the ER/PR trace in Fig. 6 and, at a lower laser intensity, in Fig 7. Note a small residual signal in the GaAs band-edge region, and note the total elimination of the AlGaAs signal at 1.9 eV and Quantum Wells signal at ~1.6 eV, which come from the layers lying on either side of the 2DEG well. The residual extraneous signal in Fig. 6 and 7 has an appreciable amplitude in the 'kink' region of the GaAs signature,[12,14] and can not be eliminated in ER/PR at -0.1 V bias. Thus, the PR response at the GaAs band-edge appears to be a composite of two modulations, modulation of $\mathcal{E}$ and an extraneous PR modulation. The suggestion of two PR modulations is further confirmed by a superposition-like effect in PR demonstrated in Fig. 8, where at a negative d.c. bias, the PR signature diminishes, shifts 10 meV toward the higher energy and exhibits an apparent phase reversal relative to the phase of the AlGaAs signature.

Fig. 9. demonstrates the repeatability of our results. The spectra in Fig. 9 show ostensibly the same general results as the previous sample. The sample in Fig. 9 has the same structure as the preceding sample but it has a 100Å undoped AlGaAs spacer versus the 50Å AlGaAs spacer for the sample discussed above. Notice in Fig. 9 a well defined extraneous ER/PR signal 0-30 meV above the GaAs band-edge. This signal appears to be responsible for the kink in the GaAs PR signature.[10] In Fig. 9, the kink in the GaAs PR signature appears only at -0.1 V bias. It is absent in the PR at zero bias, and it is also absent in the ER at -0.1 V bias as shown in Fig. 9, i.e. the kink appears to be the result of the PR modulation of this sample at -0.1 V d.c. bias. However, two other signals appear prominent in the ER and the PR spectra of Fig. 5-9, the low level oscillations extending beyond the kink region, and the signal at ~1.6 eV which comes from the isolating Quantum Wells. The leading energy of the oscillations overlaps with the kink region, where the oscillations actually produce an inflection in the GaAs signature in Fig. 5 and 6, similar to a kink. However the oscillations do not appear to be an integral part of the extraneous PR signal. The oscillations have a distinctly different period (as can be seen in the ER/PR shown in Fig. 7 and 9), and behave differently in ER/PR depending on the sample and the d.c. bias. Notice that the oscillations and the Quantum Well signal behave in unison. Both are prominent in the ER and the PR of Fig. 6, and both disappear or diminish in the ER/PR of Fig. 6, and 7, where they are much smaller than the extraneous ER/PR signal. On the other hand, their behavior is different in the ER/PR of Fig. 9, where both the oscillations and the Quantum Well signal become enhanced and are comparable in amplitude to the extraneous PR signal, as shown in Fig. 9. This is further confirmed in the low temperature data of Fig. 10. Thus, we believe that the oscillations are associated with modulation of ε deeper in the GaAs channel layer, and are not an integral part of the extraneous PR signal. Uncannily, the oscillations also appear in some of our numerical simulations of the reflectance from the 2DEG region, as by curve 2 in Fig. 11. We believe the low level oscillations in the model are fortuitous, see the 1991

Summer Research Report by J. R. Engholm and M. Sydor for further details on the experimental and numerical work done for this project.

Although we did not use ER/PR in an extensive study of the distribution of applied potentials, we thought it interesting to point out the possibility of such application. In the studies presented here, we examined the disappearance of PR signatures as a function of the d.c. bias. It was interesting to observe the flat band condition for the entire HEMT structure because such condition can provide a measure of the AlGaAs/GaAs band-gap offset. In the above samples, the flat band condition for the entire HEMT is achieved at ~300 mV d.c. bias, as shown in the fourth trace of Fig. 6. The 300 mV bias is close to the expected AlGaAs/GaAs conduction band offset, indicating that the applied potential across a HEMT, (in the absence of 2DEG), distributes itself mainly across the AlGaAs/GaAs junction.

CONCLUSIONS:

The results of numerical simulations done as a part of this project showed a disturbing result, in that the sharp Photoreflectance signature from 2DEG observed sometimes experimentally, was obtained numerically only at very low modulations of the 2DEG. The isolation of the signature and its behavior with intensity as shown in Fig. 10 confirms that the numerical result is correct. The extraneous signal at the GaAs band edge becomes most pronounced at low modulations, it is thus often obscured in ordinary PR at the usual modulations exceeding several $\mu W/cm^2$.

ACKNOWLEDGEMENTS:

Research sponsored by the Air Force Office of Scientific Research/AFSC, United States Air Force. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon. We wish to thank the Air Force Systems Command and the Air Force Office of Scientific Research, and MLPO of the Wright Laboratory, Dayton, Ohio for partial sponsorship of this research. We thank Texas Advanced Technology and Advanced Research Programs for partial support in terms of samples. The Research Development Corporation, Culver City, CA also deserves a mention for their help with the administrative aspects of the U. S. A. F. Summer Research program.

REFERENCES

1. E. Y. Wang , W. A. Albers, and C. E. Bleil in *Proc. Int. Conf. on ll-Vl Semiconductor Compounds Providence, R. I.* (1967).

2. D.E. Aspnes , Phys. Rev. 147, 554 (1966); 153, 972 (1967); Aspnes, D.E. and A.A. Studna, Phys. Rev. B 7, 4605 (1973).

3. R. N. Bhattacharya, H. Shen, P. Parayanthal, Fred F. Pollak, T. Coutts , and H. Aharoni, Phys. Rev. B 37, 4044 (1988).

4. R. Glosser and N. Bottka SPIE vol 794, 88, (1987).

5. N. Bottka, D.K. Gaskill, R.S. Sillmon, R. Henry, R. Glosser, J. Electron. Materials 17, 161 (1988).

6. S. L. Mioc, P. M. Raccah and J. W. Garland, 296 / SPIE Vol. 1678, (1992).

7. B. V. Shanabrook, O. J. Glembocki, and W. T. Beard, Phys. Rev. B 35, 2540, (1987)

8. X. Yin, Xinxin Guo, and Fred F. Pollak, G.D. Pettit, J.M. Woodall, T. P. Chin, and C.W. Tu, Appl. Phys Lett. 60 (11), 16 March, (1992).

9. H. M. Hecht, Phys. Rev. B. 41, 7918, (1990-I).

10. M. Sydor and Ali Badakhshan, J. Appl. Phys. 70, 2322, (1991).

11. M. Sydor , Ali Badakhshan, and J. R. Engholm Appl. Phys. Lett. 58, 948 (1991).

12. O. J. Glembocki, B. V. Shanabrook, N. Bottka, W. T. Beard, and J. Comas, SPIE 524, 86 (1985).

13. M. Sydor , Neal Jahren, W.C. Mitchel, W.V. Lampert, T.W. Haas, M.Y. Yen, S.M. Mudare, and D.H. Tomich, J. Appl. Phys. 67 (12), 7423 (1990).

14. E. S. Snow, O. J. Glembocki, and B. V. Shanabrook, Phys. Rev. B 38, 483 (1988)

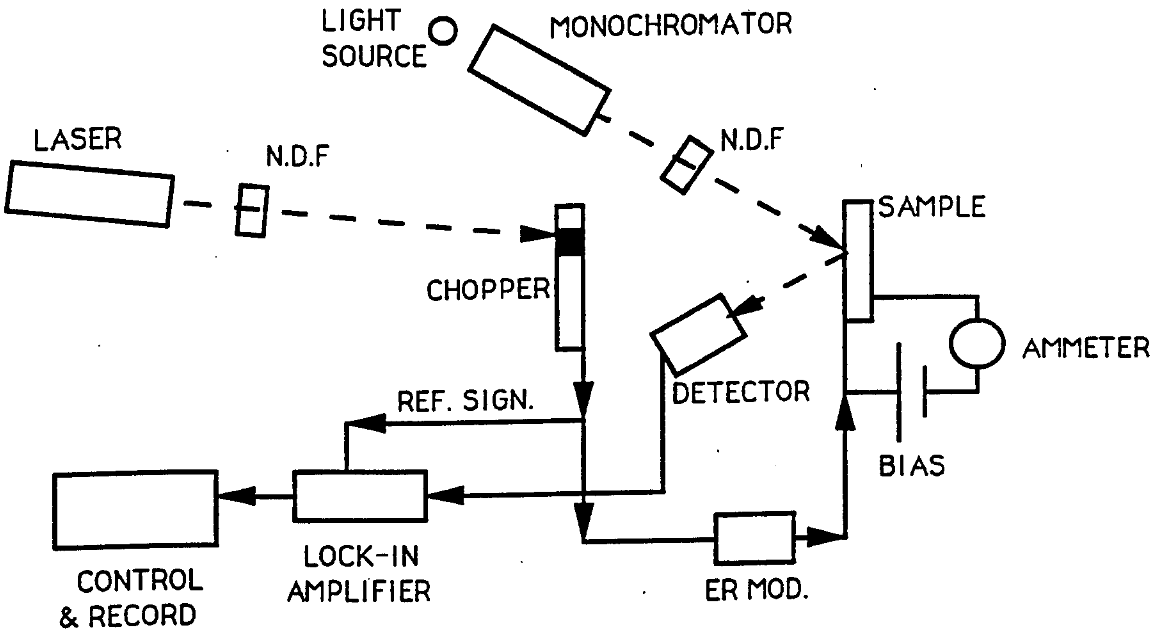


Fig. 1. Schematic of the ER/PR apparatus.

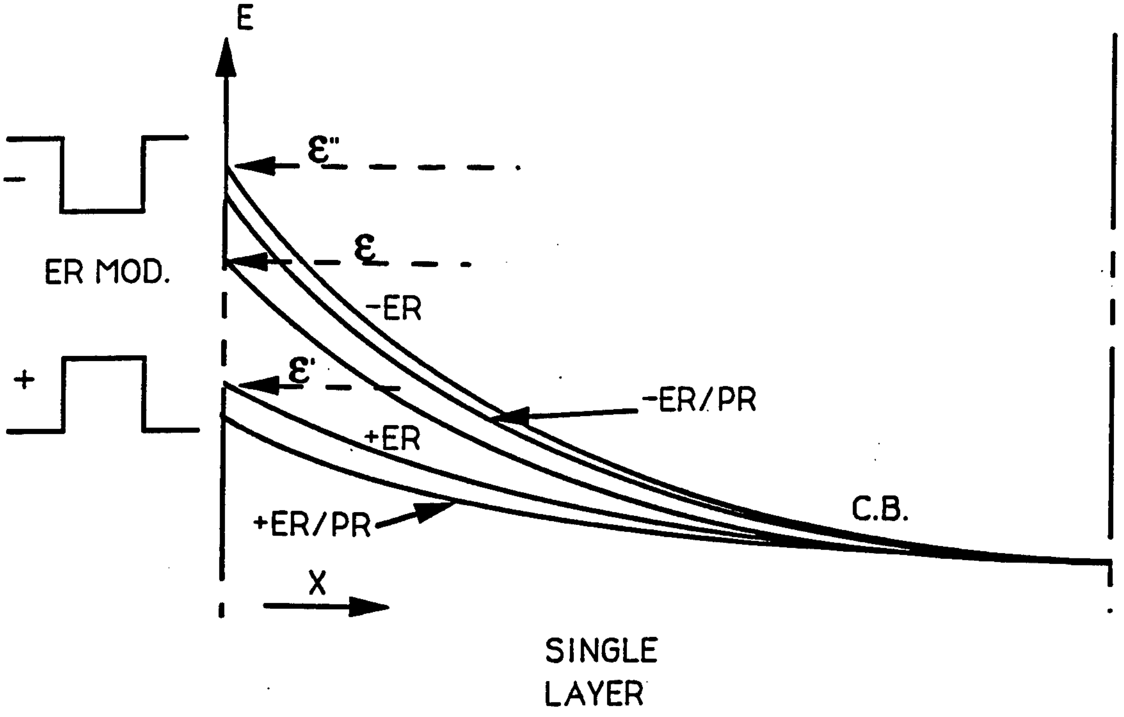


Fig. 2. An example of ER, and ER/PR modulations of a single layer.

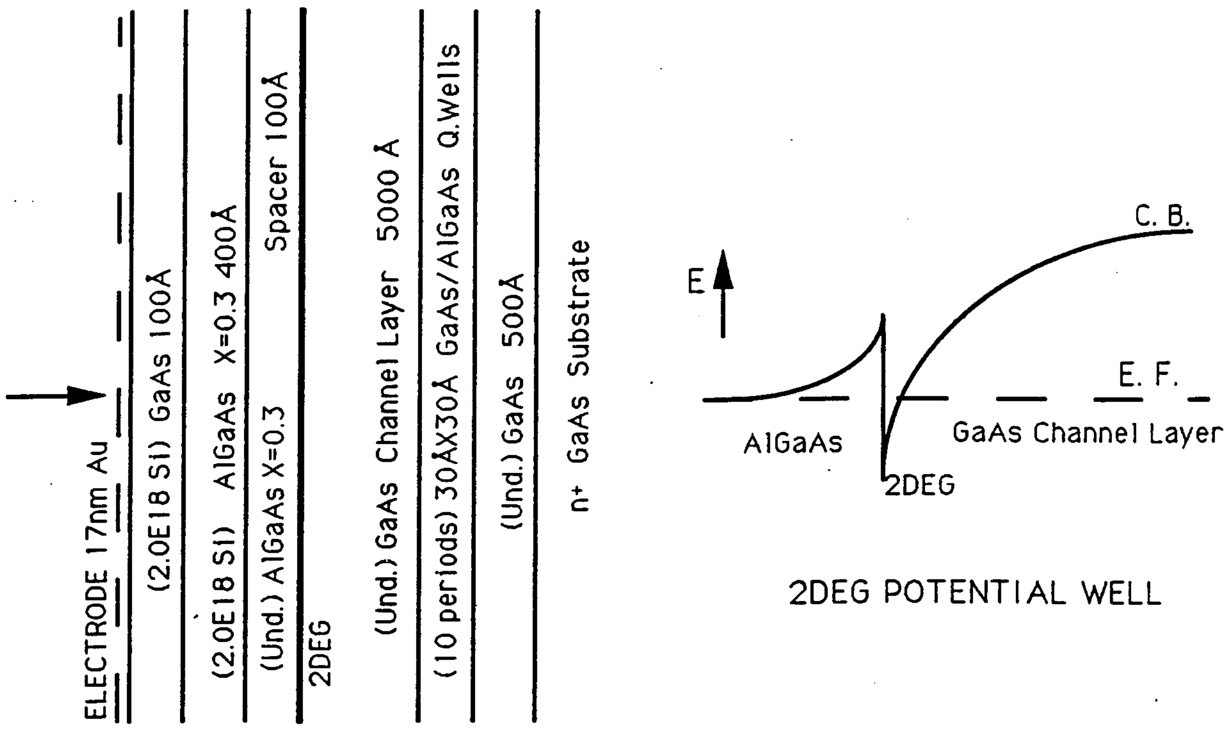


Fig. 3. Sample structure and conduction band bending in the 2DEG region of a HEMT. The thickness of the undoped AlGaAs spacer varied, samples with 50Å, 100Å, and 400Å, were used in this study. Samples with ohmic electrodes and no electrode were also examined in consideration of the electrode effects.

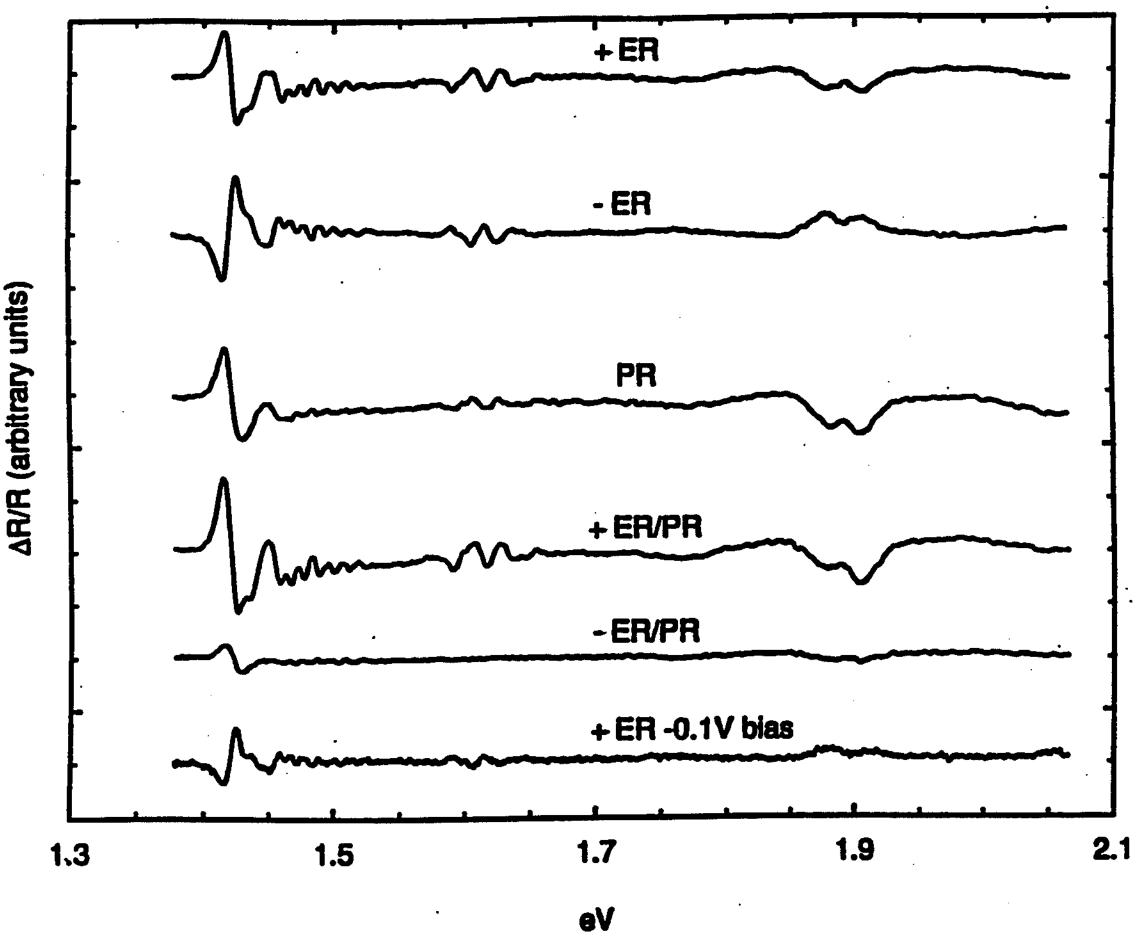


Fig.4. Demonstration of simultaneous ER/PR modulation of a HEMT. The first three traces show the individual responses: +ER, -ER, and PR. The fourth trace, (+ER/PR), represents simultaneous +ER and PR modulations applied on the same half of the modulation cycle. +ER and PR modulations aid each other; both reduce the magnitude of a negative $\mathcal{E}$, giving rise to large combined modulation and large GaAs and AlGaAs signatures at ~1.42 and ~1.9 eV respectively, as shown by the +ER/PR trace. On the other hand, -ER and PR modulations have an opposing effect on a negative $\mathcal{E}$, thus the amplitude of the signatures in -ER/PR is small. The last trace labeled, +ER -0.1 V, shows the reversal of +ER, and thereby reversal of $\mathcal{E}$, at -0.1 V d.c. bias.

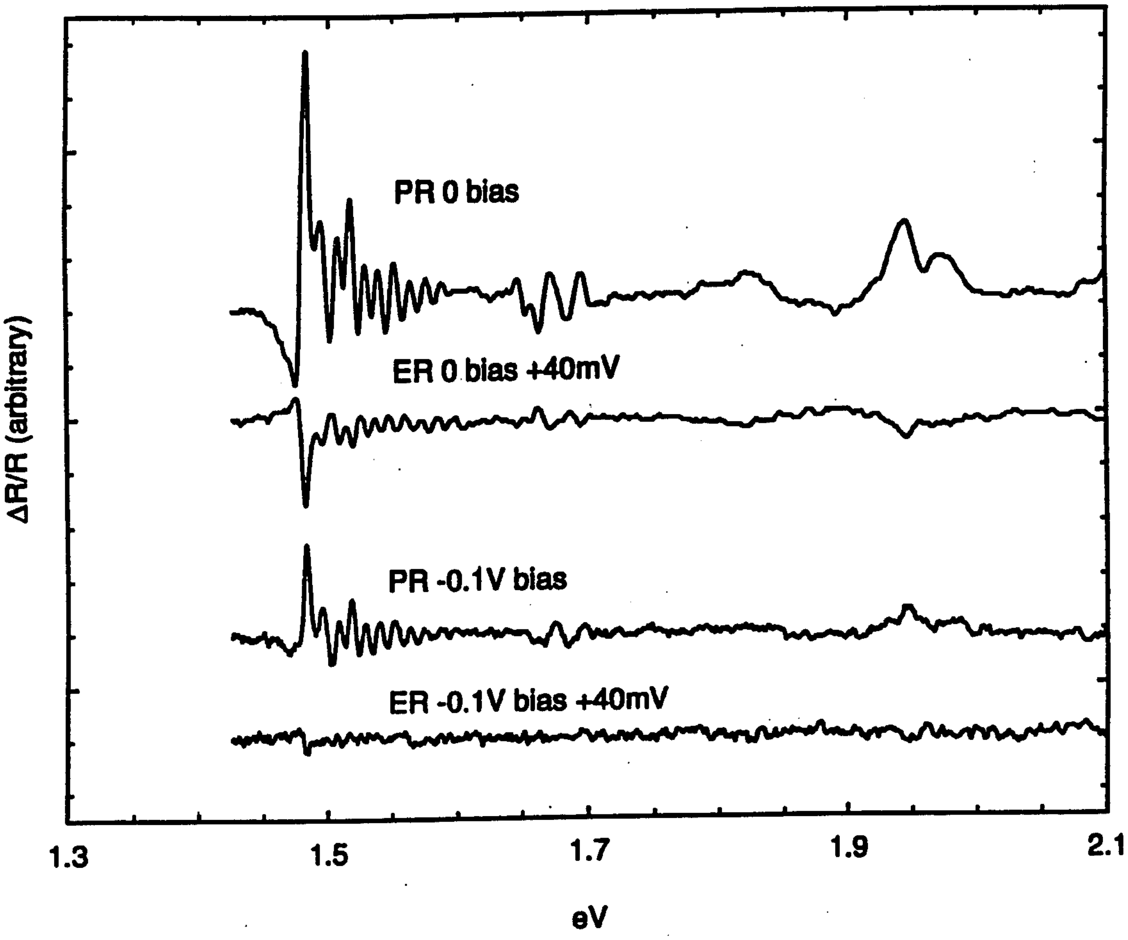


Fig. 5. At 80° K and -0.1 V d.c. bias, the ER modulation of AlGaAs becomes ineffective as shown by the bottom trace, the AlGaAs signature disappears but a low level GaAs ER signature remains. On the other hand, the PR modulation is quite effective at 80° K, both at zero bias and at -0.1 V bias, as demonstrated by the two PR traces.

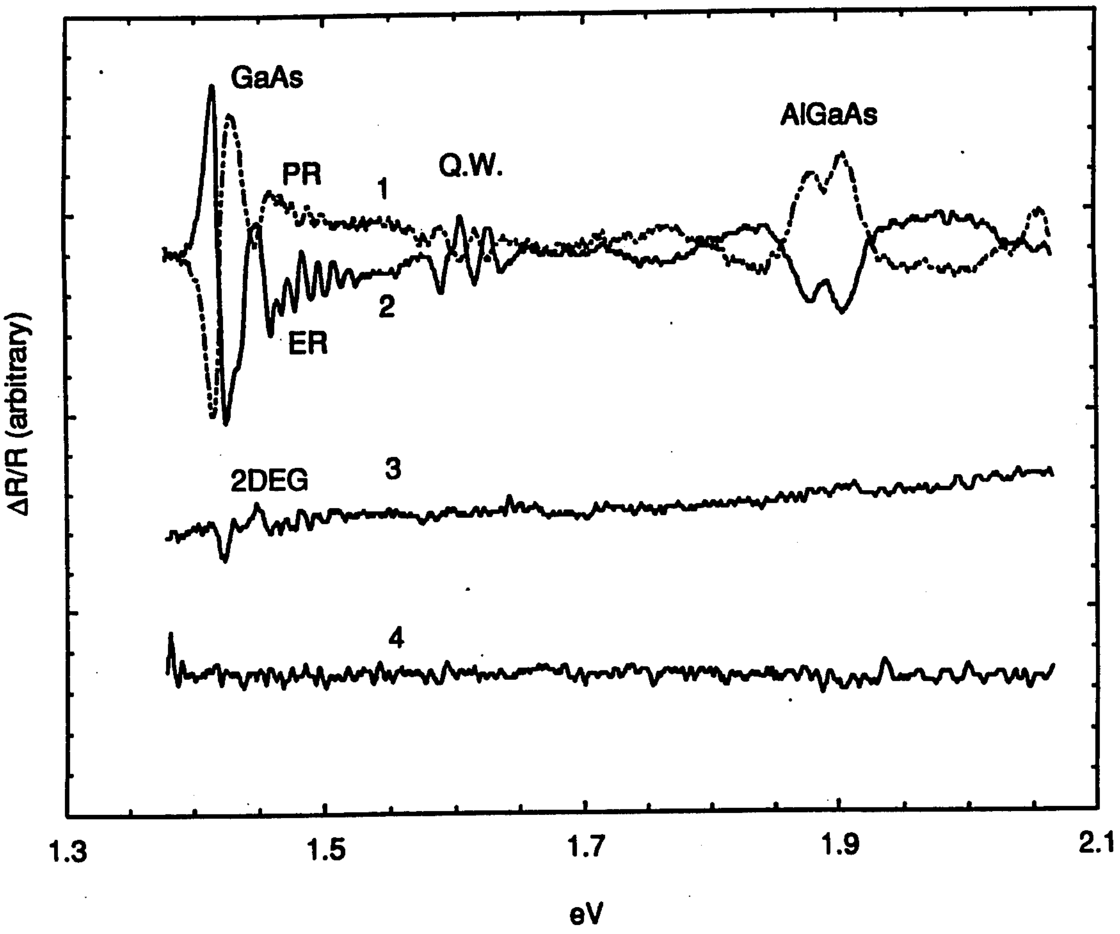


Fig. 6. Isolation of the residual PR signal due to extraneous modulation effects. Trace 1 shows a 60 μW/cm^2 PR. Trace 2 shows an equivalent +ER. Trace 3 shows simultaneous ER/PR when the equivalent +ER and PR modulations occur on the adjacent halves of the modulation cycle. No signal should appear if +ER and PR modulations were identical, i.e. affecting only the magnitude of $\mathcal{E}$ in the 2DEG region. A small but persistent residual signal ~10-30 meV above the band gap remains. Trace 4 shows the PR for flat band condition throughout the HEMT when a 0.3 V d.c. bias is applied across the sample.

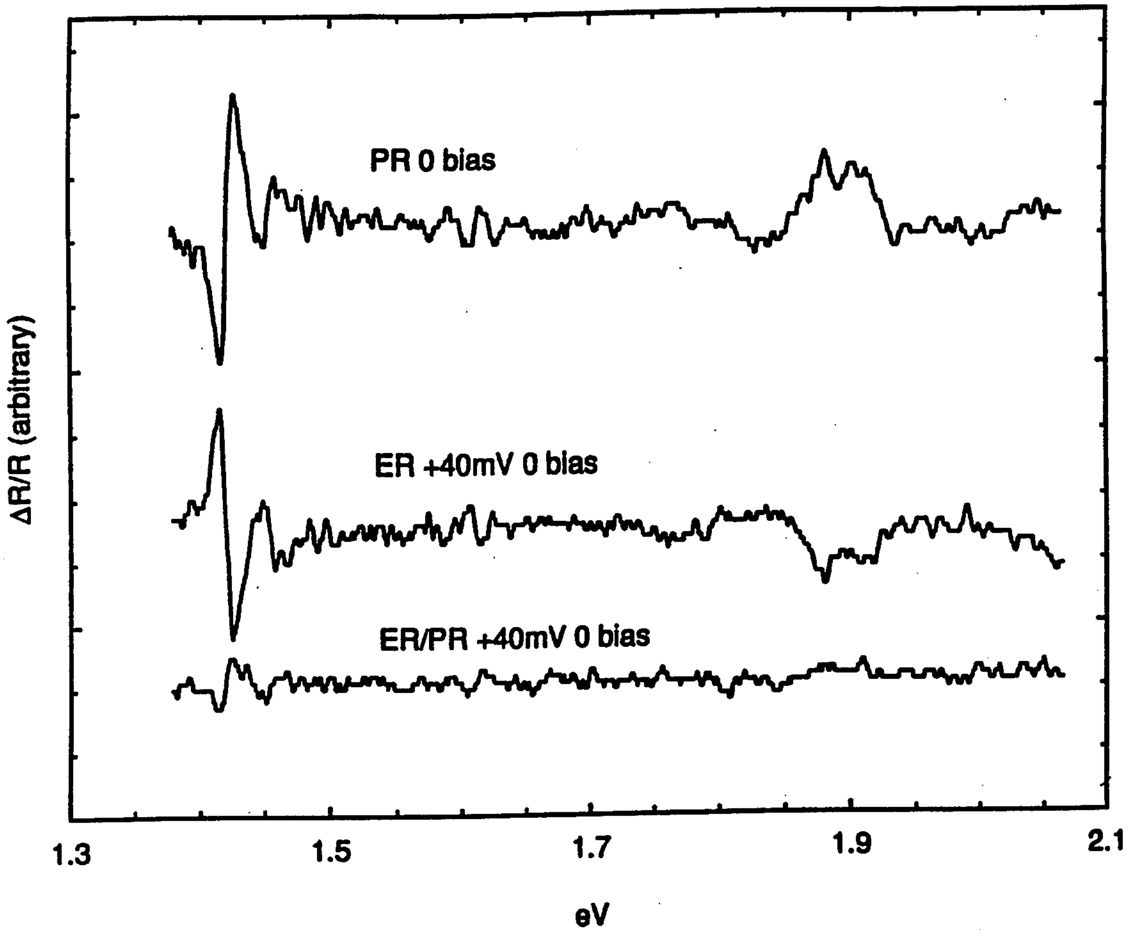


Fig. 7. Isolation of the extraneous PR signal using 14 μW/cm^2 laser intensity shows the same ER/PR result as that shown in Fig. 6. The distorted shape of the extraneous peak at ~1.45 eV is nearly identical in both figures.

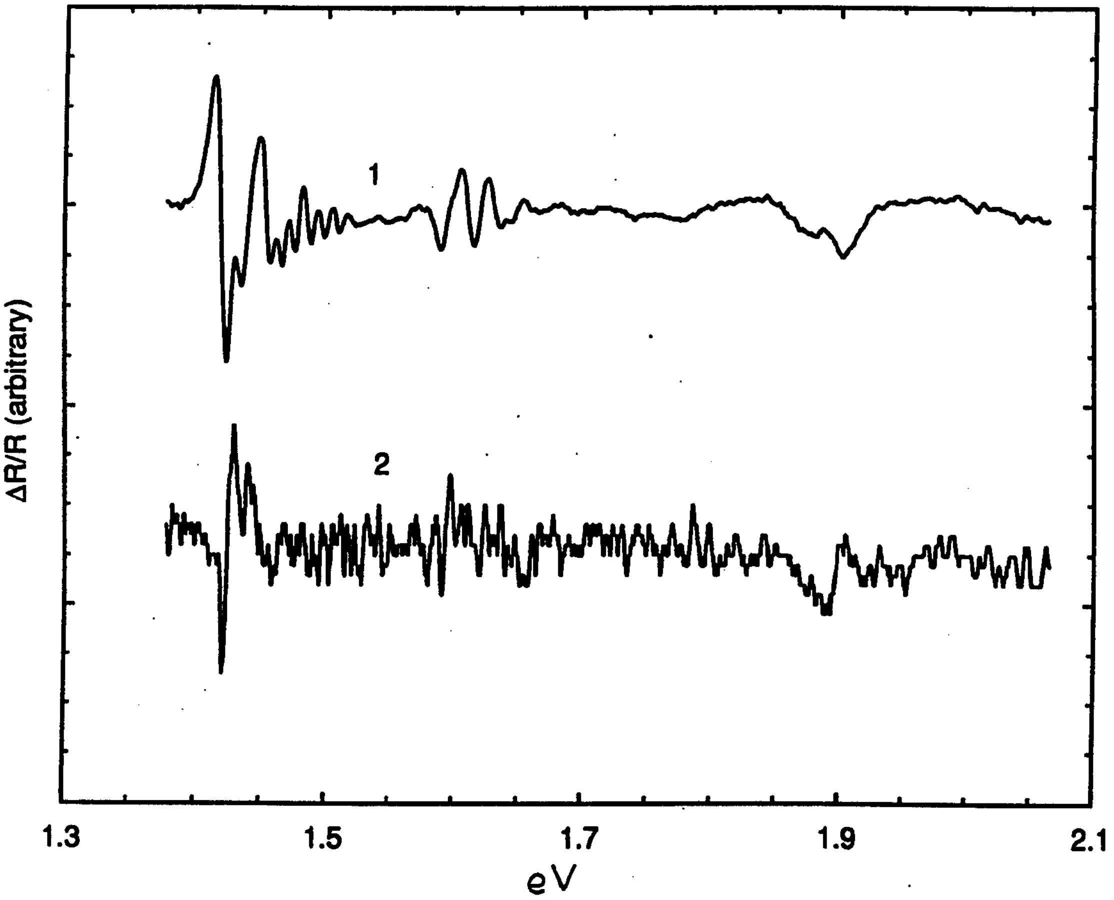


Fig. 8. Evidence of a superposition of two modulation effects in the GaAs PR signature. Trace 1 shows a typical PR at zero bias. Trace 2 shows the PR for a negatively biased HEMT. Comparison of the traces suggests that the GaAs band-edge signal is a composite of two PR modulations. The lower trace shows a 10 meV shift to the higher energy, exhibits an accentuation of the double peaked or kinked structure, and shows an apparent phase reversal of the PR signature relative to the phase of the AlGaAs signature at 1.9 eV. The shape and the energy of the GaAs signature in trace 2 is very much like the shape and energy of the residual ER/PR GaAs signal shown in Fig. 7.

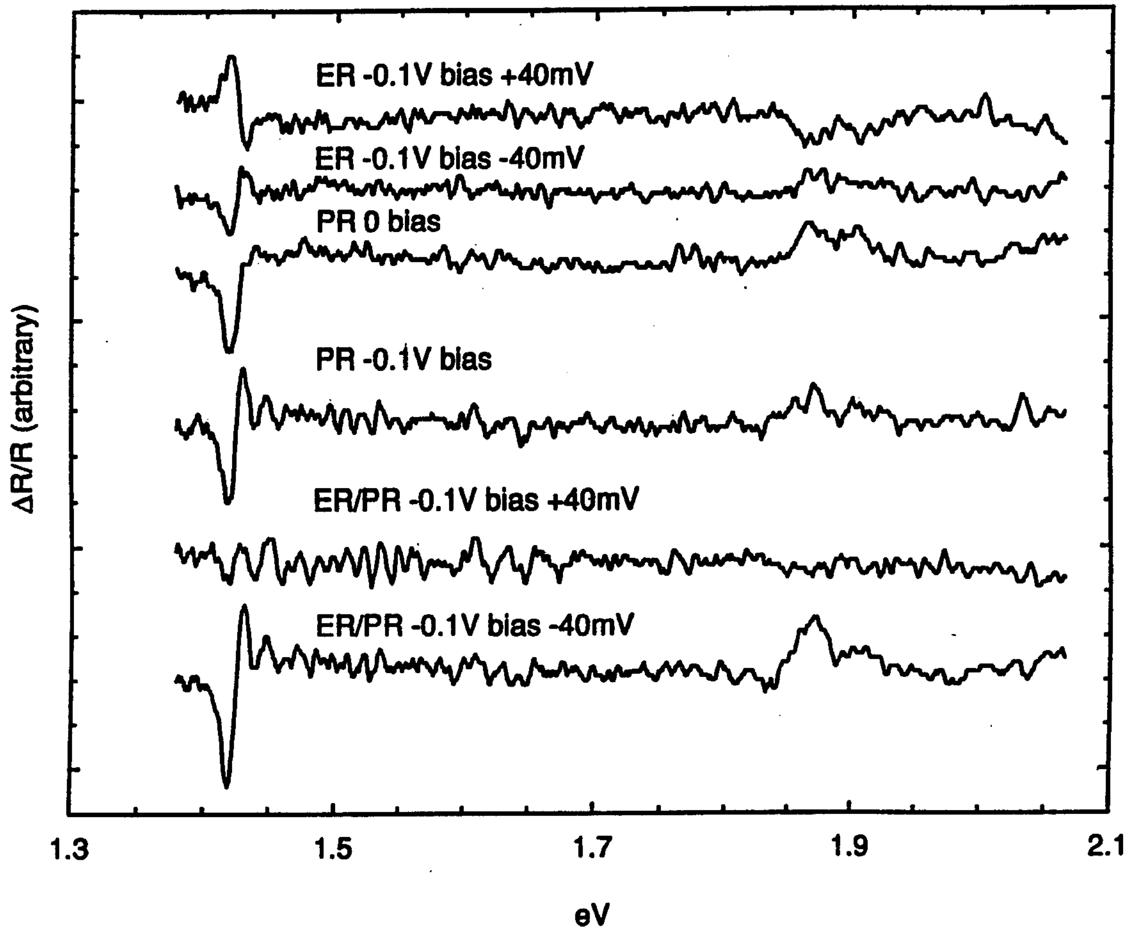


Fig. 9. ER, PR, and ER/PR data at -0.1 V d.c. bias for a HEMT with 100Å undoped AlGaAs spacer (twice that in Fig. 4-8). The ER and PR modulations occur on the alternate halves of the chopping cycle. The PR at zero d.c. bias and the ER at -0.1 V bias show no kinked structure in the GaAs signature. A clear kinked GaAs signature appears in the PR at -0.1 V bias. The residual signal in the kink region is isolated in the ER/PR trace at -0.1 V bias. In this sample, the oscillatory signal beyond the 0-30 meV kink region is enhanced in the ER/PR. It was diminished in the ER/PR in Fig 6 and 7. The extraneous PR signal is often obscured by the onset of the oscillatory signal which appears to come from field modulation deeper in the channel layer.

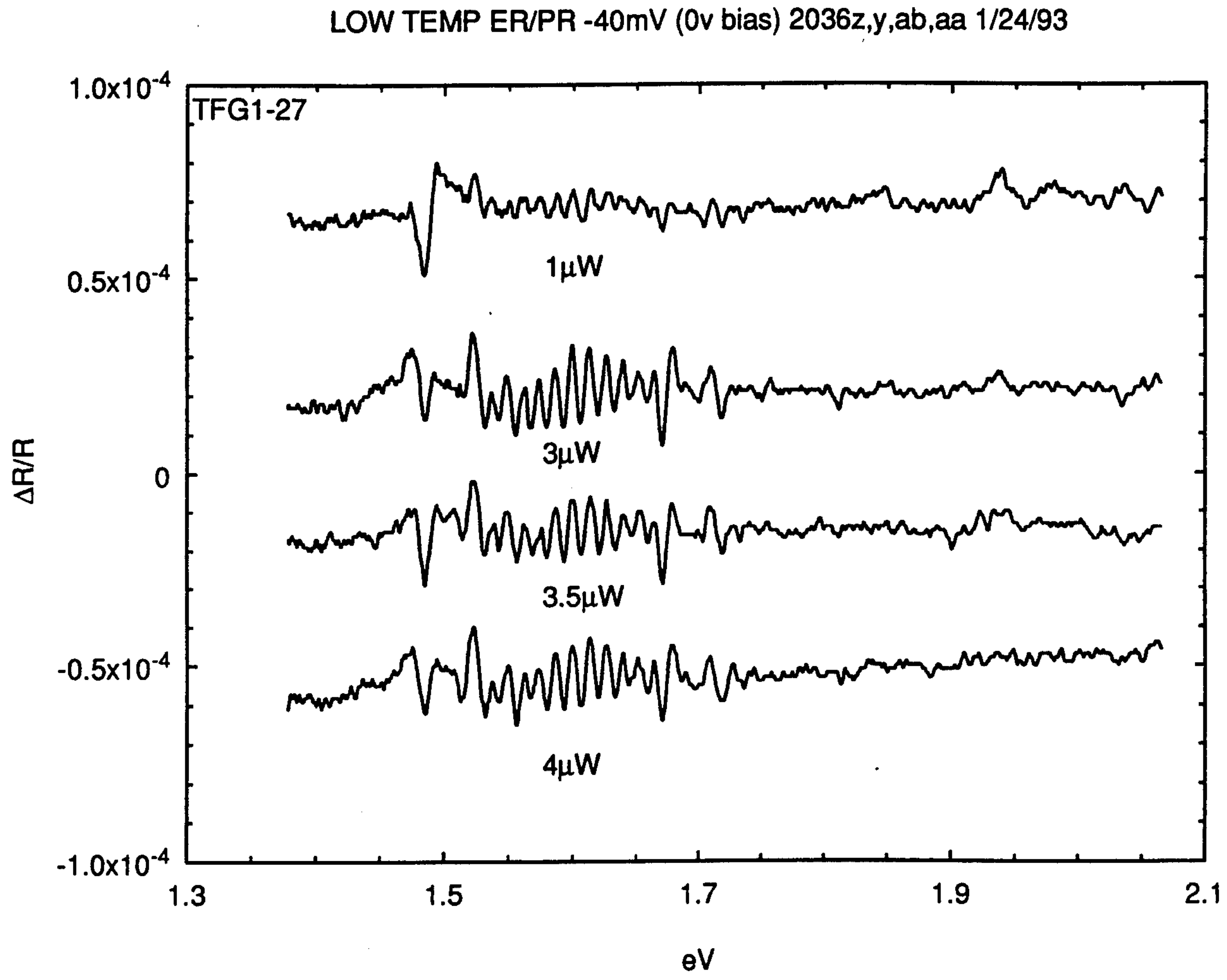


Fig. 10 Behavior of the extraneous Photoreflectance with laser intensity, at 80 K.

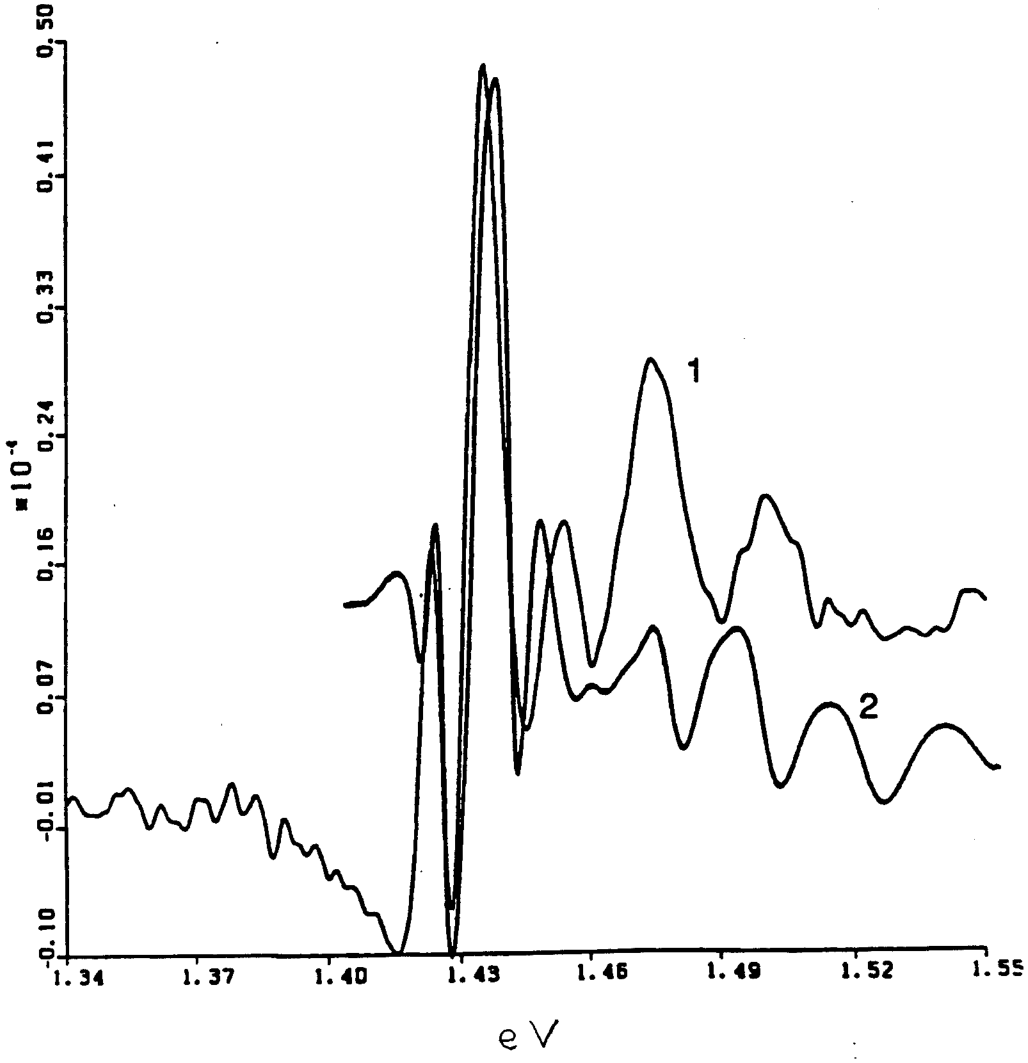


Fig. 11 Comparison of 2DEG PR, curve 1, with numerical model results, curve 2.

1991-1992 USAF - RDL SUMMER FACULTY RESEARCH PROGRAM

GRADUATE STUDENT RESEARCH PROGRAM

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by the

Research and Development Laboratories

FINAL REPORT

MEASUREMENTS OF DROPLET VELOCITY AND SIZE DISTRIBUTIONS

Prepared by:	Richard S. Tankin
Academic Rank:	Professor
Department and	Mechanical Engineering Department
University:	Northwestern University
Research Location:	WL/POSF
	Wright Patterson Air Force Base
	Dayton, Ohio 454336563
USAF Researcher:	Thomas Jackson
Date:	14 January 1993
Contract No:	F49620-90-C-0076

Acknowledgements

I wish to thank WRDC/POSF at Wright Patterson Air Force Base and the Air Force Office of Scientific Research for the sponsorship of this research. I also want to thank Research and Development Laboratories for their efficient handling of all administrative aspects of the program.

This summer's research was beneficial to me and has led to joint research activities with Wright Patterson through AFOSR. Having worked before with Dr. W. M. Roquemore and Dr. T. Jackson, it was no surprise to find them very cooperative and encouraging. The experimental work could not have been accomplished without the expertise and experience of Dr. G. Switzer. The ease and readiness with which Jeff Stutrud handled complex computer problems was invaluable. I found the staff at Wright Patterson to be friendly, stimulating, and helpful.

ABSTRACT

Maximum entropy model used to predict the size and velocity distributions was extended from a two velocity component model to a three velocity component model. The model was also extended to include thermal energy effects - heat transfer to the liquid sheet. The major difference between the two component model and the three component model (neglecting thermal energy effects) is the occurrence of a bi-modal size distribution under certain flow conditions (source terms). This lead to the motivation of this research effort - to determine whether a bi-modal size distribution exists experimentally. It was found that when the pressure atomizer (Allison nozzle) operates at low flow rates a weak bi-modal size distribution is present. Although the maximum entropy model agrees reasonably well with the measured distribution shape, a more pronounced experimental bi-modal distribution was desired. Such a distribution is present in the breakup of a cylindrical liquid jet (Rayleigh problem). Attempts were made to measure size distribtion using the Phase Doppler Particle Ananlyzer without success. Measurements were made using a high speed video camera (Spin Physics).

I.Introduction

Sprays play an important role in many engineering applications; for example, in combustion of liquid fuels, agricultural applications, painting, direct injection condensers, cooling, etc. In all of these applications droplet size and velocity distributions are important parameters in addition to the spray cone angle and droplet penetration. Until recently a main interest of many researchers has been the effect of various parameters on the droplet size distributions. Various techniques have been used to determine the drop size distribution - photographic, optical, collection devices, etc. Each of these techniques have very limited capabilities. Until recently, the most advanced instrument for obtaining a size distribution (integrated over the optical path length) was the Malvern instrument. To determine a radial distribution required the use of Abel inversion; however dense sprays, or asymmetry can complicate this technique. Velocity distributions were rarely considered because it was almost impossible to measure such distributions. Two situations have recently arisen - a vast improvement in the instrumentation needed to measure drop size and velocity distributions and a new approach for predicting droplet size, velocity, and now temperature distributions.

With a Phase Doppler Particle Analyzer (PDPA), it is possible to measure both the local droplet size and velocity distributions. The PDPA is a single point, scattering technique, measuring each droplet as it passes through a small probe volume. A two-component system can measure droplet velocities in two directions. Applications of this instrument have been adequately covered in the literature.

A very brief review of the principle of maximum entropy will be presented because it is not widely discussed in the atomization literature. The concept of information entropy was developed by Claude Shannon [1], and Jaynes [2] who later extended this concept into the now well-known method of maximum entropy formalism. This formalism can be applied to problems which involve probability, i.e., where insufficient information is available to obtain exact solutions. Tribus [3] used the principle of this formalism in thermodynamics and showed that the concepts of heat and temperature in thermodynamics could be defined through the formalism of maximum entropy. This maximum entropy formalism allows one to determine the probability distribution functions for complex systems in physics, chemistry, biology, as well as in many other disciplines by measuring relatively few average (macroscopic) quantities. Application of the maximum entropy is confined in this study to liquid sprays in order to predict the droplet size and velocity distributions. Applications of the maximum entropy principle to atomization problems have been discussed by several researchers - Kelly [4], Sellens and Brzustowski [5,6], Sellens [7,8], Li and Tankin [9,10,11,12], and Chin et al [13] Chin and Tankin [14].

The following problems that we have recently been completed (but not submitted for publication): (1) Comparison between theory and experiments for a pressure atomizer were a bi-modal size distribution occurs; (2) Experimental measurements of the drops formed from a cylindrical jet (Rayleigh problem). At the flow rate selected, satelite drops appear yield a very strong bi-modal size distribution.

II-1. Comparison Between Theory and Experiments for a Pressure Atomizer

In the problem formulation, the control volume consists of the liquid sheet region which extends from the nozzle exit to the breakup plane (see Figure 1).The equations are developed in reference14. The nozzle used in these experiments was an Allison hollow cone, non-swirl spray nozzle having a 90° spray angle. A spray of water issues into a quiescent, saturated air environment at 295°K. The water flow rate is $2.75\times10^{-6}m^3/s$ (2.5 gallons/hr) at a pressure drop across the nozzle assembly of 0.27 MPa (40 psig). This low flow rate was used because the bi-modal size distribution only appeared at low flow rates. The water exiting the nozzle forms a liquid sheet, hollow cone in shape, which breaks up into ligaments and droplets. The ligaments themselves breakup into droplets further downstream. It is desirable to measure the spray as close as possible to the sheet break up region — but downstream of the ligament region. Measurements beyond this point will be influenced by local gas aerodynamics and complicate the comparison with calculations. In addition, if measurements are made far downstream, a large diameter measuring plane will require measurements at many radial positions to account for all the droplets from the spray.

Photograph of the spray (Fig. 2) indicates that sheet breakup occurs at approximately 7.5 mm from the nozzle face. The laser beams that appear in Fig. 2 are located 10 mm from the nozzle exit. At 10 mm, droplets have formed. It should be mentioned that several such photographs were taken and the location of the sheet breakup plane oscillates. This was seen more clearly from video tapes taken of the spray. Based on the photographs and video tapes, the measurement plane was set at 10 mm from the nozzle discharge plane, which is about 2.5 mm beyond the mean position of the sheet break up.

Droplets are sized with an Aerometrics, Inc. two-color, four-beam PDPA. This instrument measures the radius of curvature of droplets passing through the probe volume. A complete discussion of the theory of operation can be found in Bachalo and Houser [15]; a discussion of the operational constraints in practical environments can be found in Jackson [16]. The instrument is configured like a standard laser Doppler anemometer with each beam pair measuring one component of the droplet velocity. At each measurement station 5,000 droplets are measured. Collection times for these many samples are of the order of seconds. The nominal probe volume size is 0.002 mm^3.

Comparison between experiment and calculation demands that the measured values of droplet size and velocity near sheet break up region be representative of the behavior of the entire spray. Several constraints inherent in the PDPA must be addressed. First, measured droplets must be spherical. This requires that the measurement station be sufficiently far from the sheet break up region so that droplets are not oscillating. Second, the PDPA has a dead time of 16 microseconds associated with each measured droplet. Another droplet entering the probe volume during this dead time will not be measured and may prevent the measurement of the first droplet. Thus, in dense spray regions, typical of the sheet break up area, all droplets may not be counted. If the rejection is completely random (not based on a particular droplet size or velocity class), the measurements are still suitable for this purpose. Third, the cross sectional area of the optical probe volume is small compared to the cross sectional area around the sheet break up. Data collected by the PDPA must be extrapolated to the entire spray area at the measurement plane located 10 mm from the nozzle.

The first two considerations influence the percentage of valid signals versus the total attempts. That is, the PDPA attempts to process all Doppler signals. It performs checks on the quality of each signal and rejects those which exceed certain limits. In these experiments valid signals range from about 50% to 75% of the signals collected. Thus, to collect 5,000 valid signals at a challenging location in the spray it was necessary to make as many as 10,000 attempts. The third consideration is one of spray symmetry and of specifying the probe volume cross section. For the symmetry evaluation two orthogonal, full diameter traverses were made. The axis of the spray was vertical for all measurements. An x traverse through the spray yielded measurements of droplet size, and axial and radial droplet velocities; a y traverse yielded measurements of droplet size, and axial and tangential droplet velocities. Measurements were made at radial intervals of 0.5 mm. Figure 3 shows the axial velocity measurements from the x and y traverses. Figure 4 shows the Sauter mean diameter profiles associated with the x and y traverses. The velocity and droplet size profiles along two orthogonal traverses through the spray are in reasonable agreement, indicating the spray has good symmetry about its axis. Thus, size and velocity distributions obtained from the point measurements can be integrated over the ring area associated with each measurement point to yield a total spray measurement. This experimental technique is adequately described in [12].

After initial spray symmetry checks, data were collected to evaluate the calculations. A detailed radial profile was taken at the 10 mm location where measurements were made at 21 radial points from the spray center to its edge. Five thousand valid samples were taken at each location. In looking at the data in Fig. 3, a plot of mean axial velocity profiles, it was observed that there is a rather strong negative (upward) axial velocity in the central portion of the spray (see, for example, Fig. 5 that shows data taken at 3.0 mm from the spray centerline). These negative mean axial velocities extend over about 15 mm of the spray. Since the droplets in the central region of the spray are relatively small (see Fig. 4, which is a plot the Sauter mean diameter profiles), they more or less follow the air flow.

The axial velocity of the air in a hollow cone spray close to the sheet region is positive (downward) because of the drag of the liquid sheet on the surrounding air. To balance this mass flux of air leaving the interior of the hollow cone spray, there must be a flow of air upward (negative axial velocity) in the central portion of the hollow cone spray. This negative mean velocity extends from the spray centerline to a radial position of about 8 mm. At 9 mm from the centerline of the spray, the axial velocity data show a bi-modal distribution. This is shown in Fig. 6. We believe this is due to fluctuations of the sheet region — both spatially and temporally. All the other velocity profiles (at radii ≠ 9mm) have one peak (mono-modal). The data taken at radii less than 9 mm have peaks with negative velocities; those at greater radii have peaks with positive velocities. The air flow (caused by the liquid sheet) carries many of the small droplets some distance downstream, then inward toward the axis of the spray, upward by the vortical structures formed in the air that is interior of the hollow spray cone, and finally downward again in a region close to the interior of the liquid sheet. It is this boundary layer due to the liquid sheet that drives the air recirculating zone process. This flow pattern was deduced in an earlier paper (Figure 3 in reference 17) where negative droplet velocities were observed along the centerline of a hollow cone spray. Therefore, if one were to count the droplets from the centerline of the spray outward to the outer edge of the spray, the small droplets in the interior (radius ≤ 9mm).would be counted three times. Figure 7 shows the normalized axial velocity and Sauter mean diameter profiles. There is a strong spatial correlation between the small droplets and the negative velocities. To avoid this problem of multiple counting of the small droplets, the region analyzed was restricted to radii that are ≥ 9.5 mm. For these measurements the acceptance ratio (validations/attempts) is around 60%; the corrected counts (which appear in the droplet size distribution ouput from the PDPA) raise this ratio (corrected counts/attempts) to about 100 %.

A droplet size distribution is constructed from the individual point measurements, weighting each measurement by their time of collection and the ratio of their optical probe area to the ring area represented at that location. This experimentally determined droplet size distribution will be compared to that calculated by the maximum entropy spray model.

Droplet size is normalized by the mass mean diameter, D_{30}, which was determined from measurements to be 81.43 microns. The resulting experimentally determined probability size distribution is shown in Fig. 8. It should be noted that a bi-modal size distribution appears. One peak occurs at $\overline{D} \approx 0.2$ and the second peak at $\overline{D} \approx 1.3$. A physical explanation for the origin of these peaks is as follows: There are two main sources for the droplets in this spray — the droplets that form from the ligaments and those that form from the thin sheet of liquid that lies between the ligaments at breakup. The ligaments breakup via Rayleigh's capillary instability and form the larger droplets associated with those centered around $\overline{D} \approx 1.3$ (in Fig. 8). The droplets from the thin sheet are associated with those centered around $\overline{D} \approx 0.2$.

In Figure 8 is plotted the numerically predicted size distribution. It is seen that the size distribution obtained using maximum entropy principle agrees with the measured values. This has been written up in detail in a paper that will be submitted for publication.

II-2 Rayleigh Problem

It is well known that satellite drops usually appear in the breakup of a cylindrical jet of liquid that discharges in quiescent air. At first, attempts were made to measure the droplet size distribution using a PDPA setup at WPAFB. However this was not possible for the following reasons:

1. The PDPA instrument manufactured by Aerometrics limits the size of the maximum diameter of the drops to about 300microns. Thus one has to limit the size of the orifice to about 50 microns. Observing the jet from a 50 micron hole with a microscope, satellite drops were seen - but infrequently. The satellite drops have a smaller velocity (larger drag) than the large drops. Thus, the satellite drops coalesce with the large drops downstream - about three or four drop intervals . This means that the PDPA must detect the satellite drop when it first appears (within a couple drop intervals of the breakup point); otherwise it isn't detected. When the satellite droplets are near coalescence with the large droplets, the PDPA probe volume is too large to detect the satellite drops.

2. The breakup region of the cylindrical jet oscillates - thus, at times the cylindrical jet extends beyond the measuring volume (breakup region is downstream); at other times the break up region occurs too far upstream of the measuring volume (satellite drops have already coalesced).

At Northwestern University, we set up a nozzle consisting of a sapphire orifice - 0.030" diameter. We used a Spin Physics video camera to take pictures at 1500 frames per second (see Figure 9). From these pictures we were able to measure the droplet sizes and velocities. In measuring the drop sizes, a circle is fitted to the data directly on the monitor. The software gives the diameter of the circle which is assumed to be the diameter of a shperical drop. With the large drops this is an approximation because the large drops are not spherical in shape as can be seen in Figure 9. An experimentally determined size distribution plot is shown in Figure 10. We are in the process of analyzing these data using the maximum entropy principle.

III. References

1. Shannon, C. E., *Bell System Technical Journal* 27: 379-423 (1948).

2. Jaynes, E. T., *Phy. Rev.* 106: 620-630, 108: 171-190 (1957).

3. Tribus, Myron, *Thermostatics and Thermodynamics*, D. Van Nostrand Company Inc., Princeton, New Jersey, 1961.

4. Kelly, A. J., *J. of Applied Physics*, 47: 5264-5271 (1976).

5. Sellens, R. W., and Brzustowski, T. A., *Atomization and Spray Technology* 1: 89-102 (1985).

6. Sellens, R. W., and Brzustowski, T. A., *Combust. Flame* 65: 273-279 (1986).

7. Sellens, R. W., *Second Symposium on Liquid Particle Size Measurement Techniques*, ASTM, Atlanta, Georgia, 1988.

8. Sellens, R. W., *Part. Part. Syst. Charact.* 6: 17-27 (1989).

9. Li, Xianguo, and Tankin, R. S., *Combust. Sci. and Tech.* 56: 65-76 (1987).

10. Li, Xianguo, and Tankin, R. S., *Combust. Sci. and Tech.* 60: 345-357 (1988).

11. Li, Xianguo, Ph. D Thesis, Northwestern University, Evanston, Illinois, (1989).

12. Li, Xianguo, L. P. Chin, R.S. Tankin, T. Jackson, J. Stutrud, and G. Switzer, *Combustion and Flame*, 86: 73-89 (1991).

13. Chin, L.P., LaRose, P., Tankin, R.S., Jackson, T., Stutrud, J., and Switzer, G., *Physics of Fluids* A 3(8): 1897-1906 (1991).

14. Chin, L.P. and Tankin, R.S., *Winter Annual Meeting of ASME*, (1991).

15. Bachalo, W. D. and Houser, M. J., "Phase/Doppler Spray Analyzer for Simultaneous Measurement of Drop Size and Velocity Distributions," *Optical Engineering*, Vol. 23, pp. 583-590. (1984)

16. Jackson, T. A., "Liquid Particle Size Measurement Techniques, "ASTM *STP 1083* (E. D. Hirleman, W. D. Bachalo, and P. G. Felton, Eds.), ASTM, Philadelphia, Vol. 2, p. 151. (1990)

17. Lee, S. Y. and Tankin, R. S., "A Study of Liquid Spray (Water) in a Non-Condensable Environment," *Int. J. Heat and Mass Transfer*, Vol. 27, pp. 351-361. (1984)

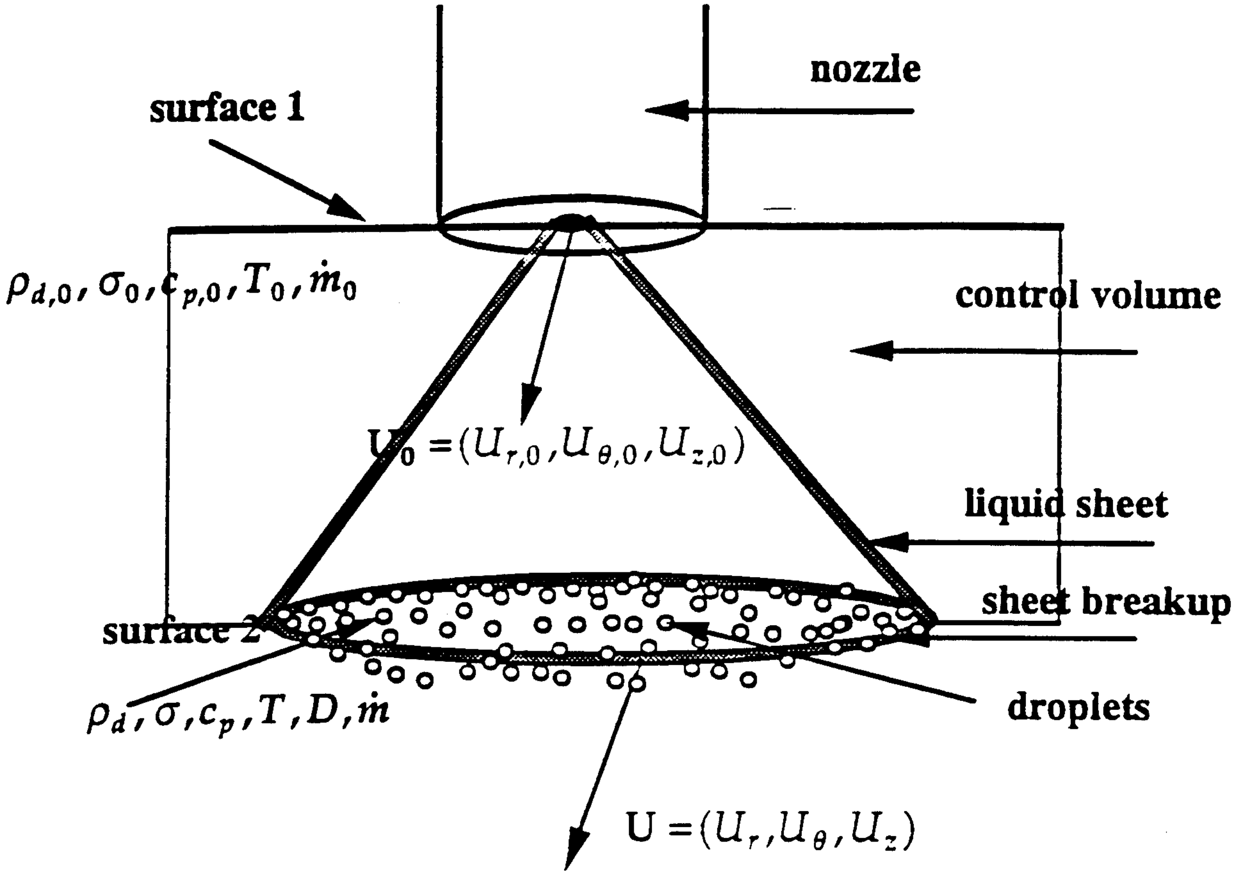

nozzle
surface 1
$\rho_{d,0}, \sigma_0, c_{p,0}, T_0, \dot{m}_0$
control volume
$U_0 = (U_{r,0}, U_{\theta,0}, U_{z,0})$
liquid sheet
sheet breakup
surface 2
$\rho_d, \sigma, c_p, T, D, \dot{m}$
droplets
$U = (U_r, U_\theta, U_z)$

Figure 1 Schematic drawing of control volume for a spray nozzle

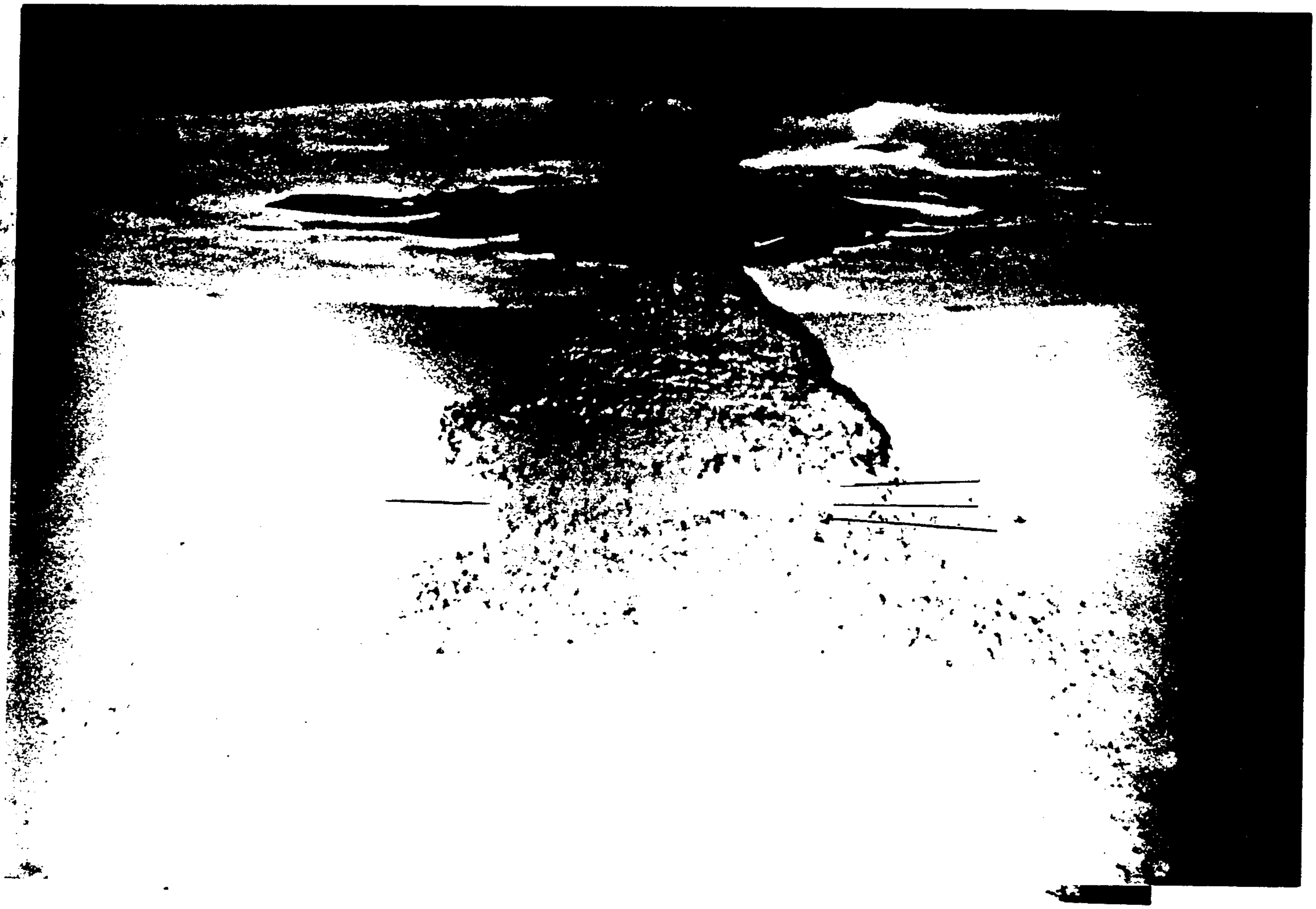


Figure 2 Photograph showing the spray from a pressurized nozzle

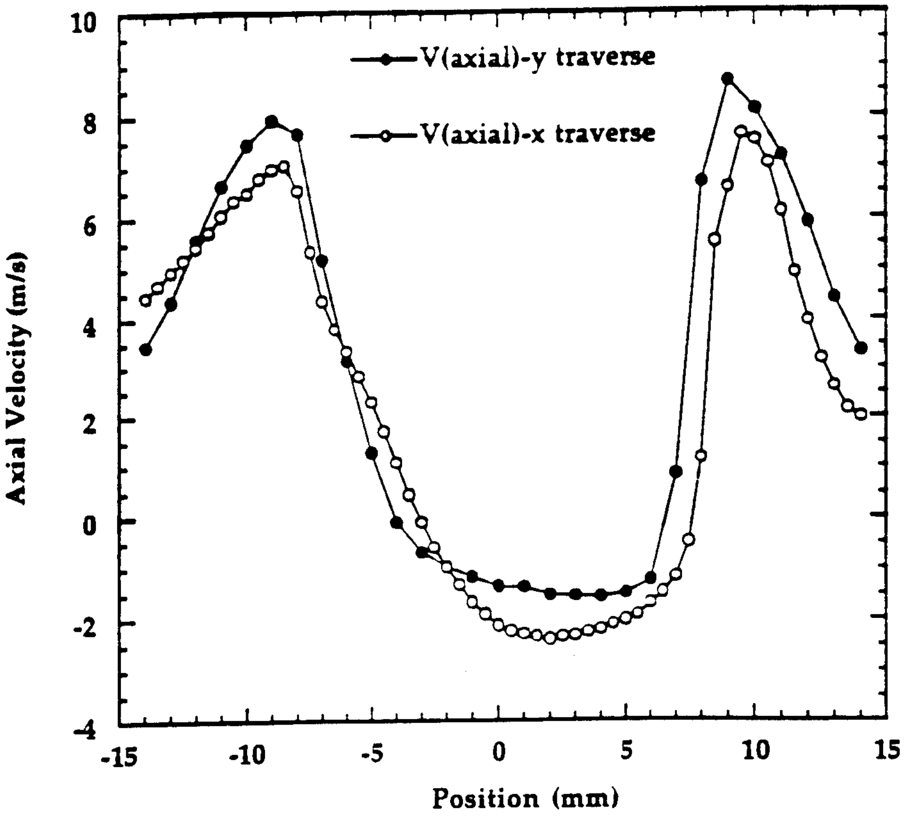


Figure 3 Experimentally measured axial droplet velocities along x and y traverses

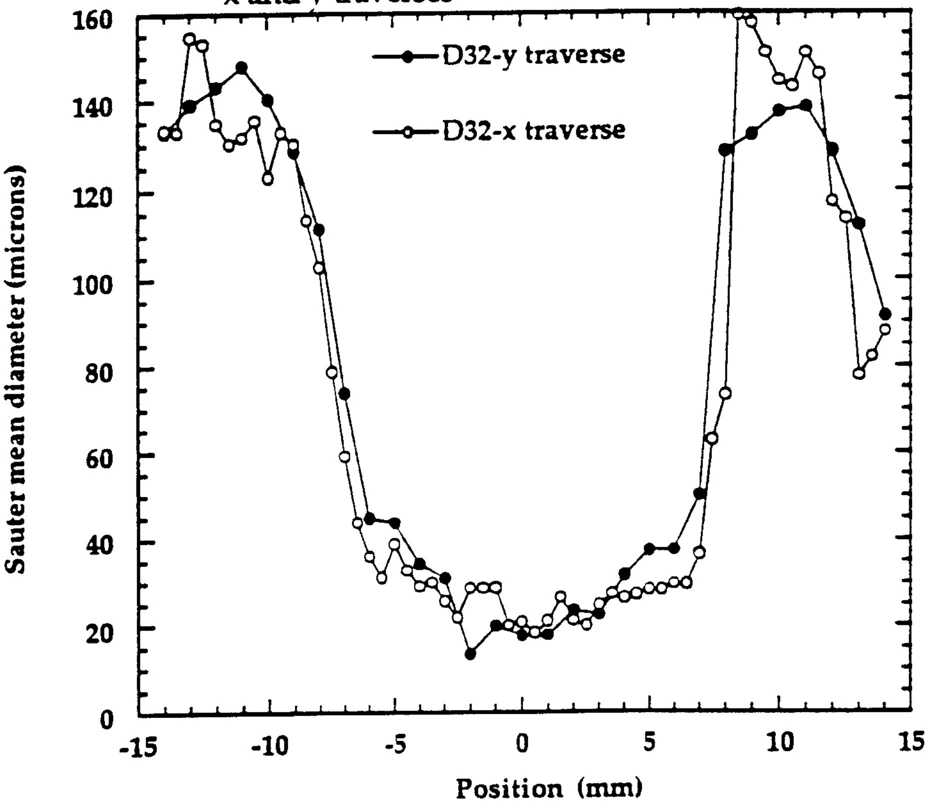


Figure 4 Experimentally measured Sauter mean diameter along x and y traverses

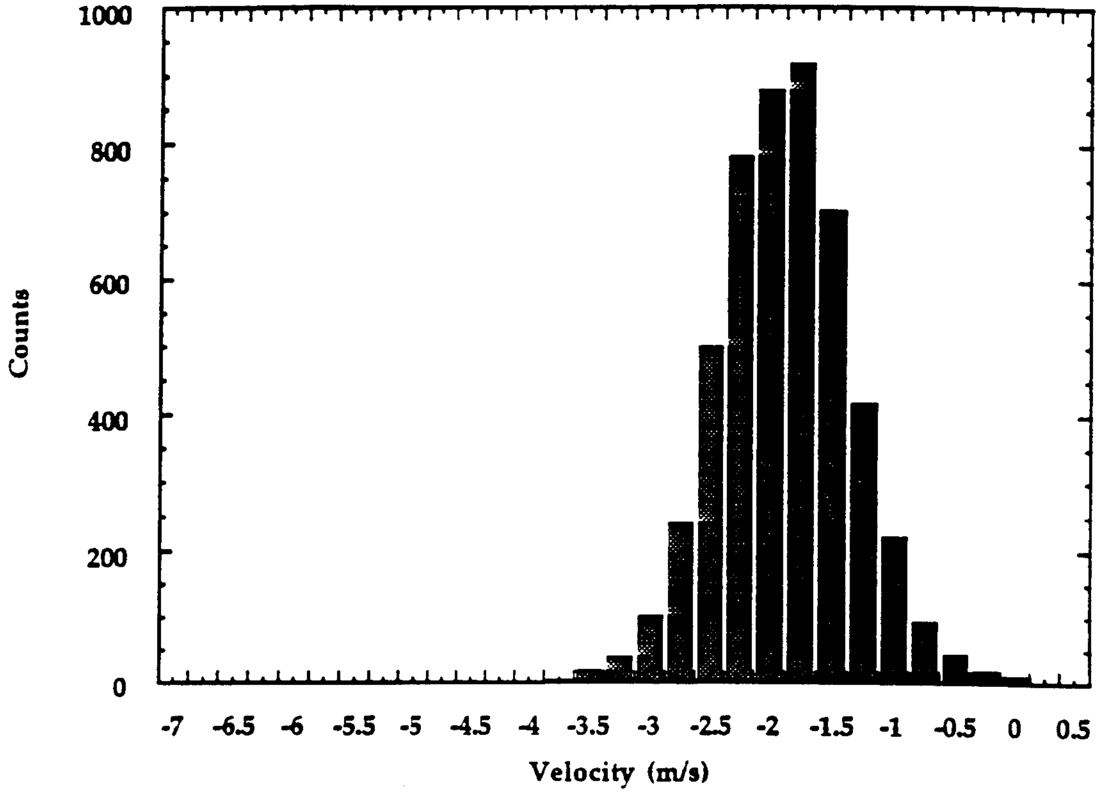


Figure 5 Axial droplet velocity distribution measured at 3 mm from the spray centerline

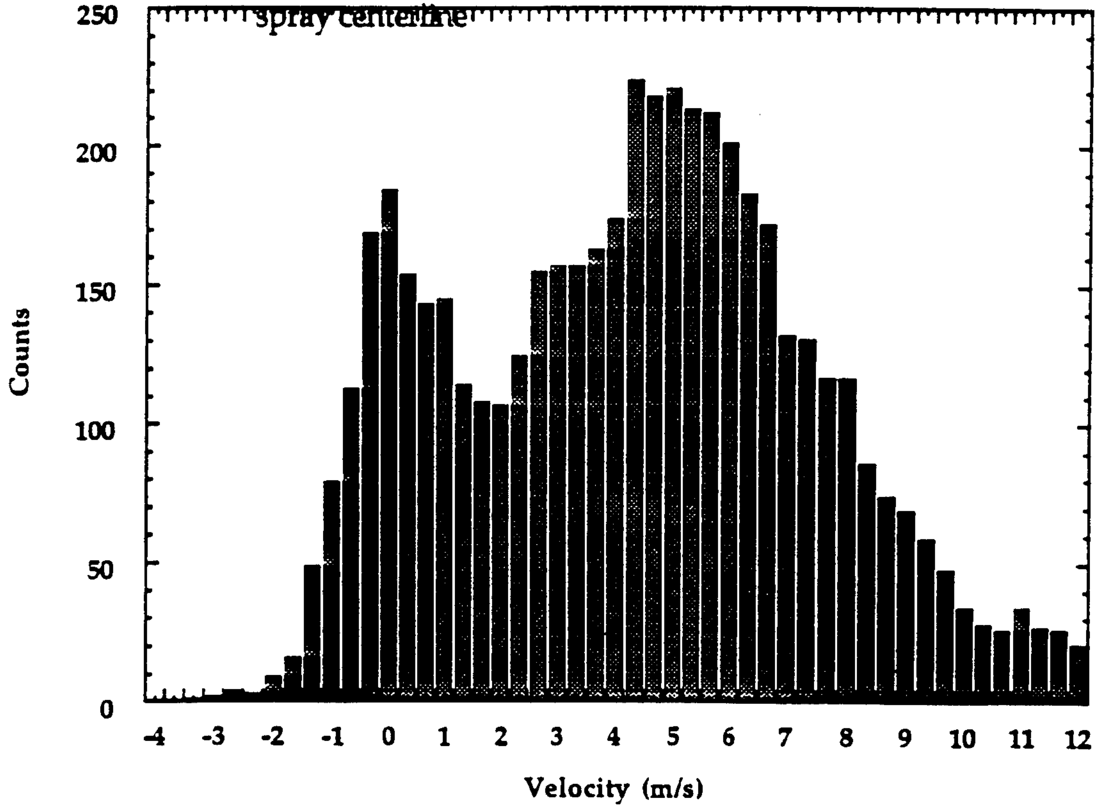


Figure 6 Axial droplet velocity distribution measured at 9 mm from the spray centerline

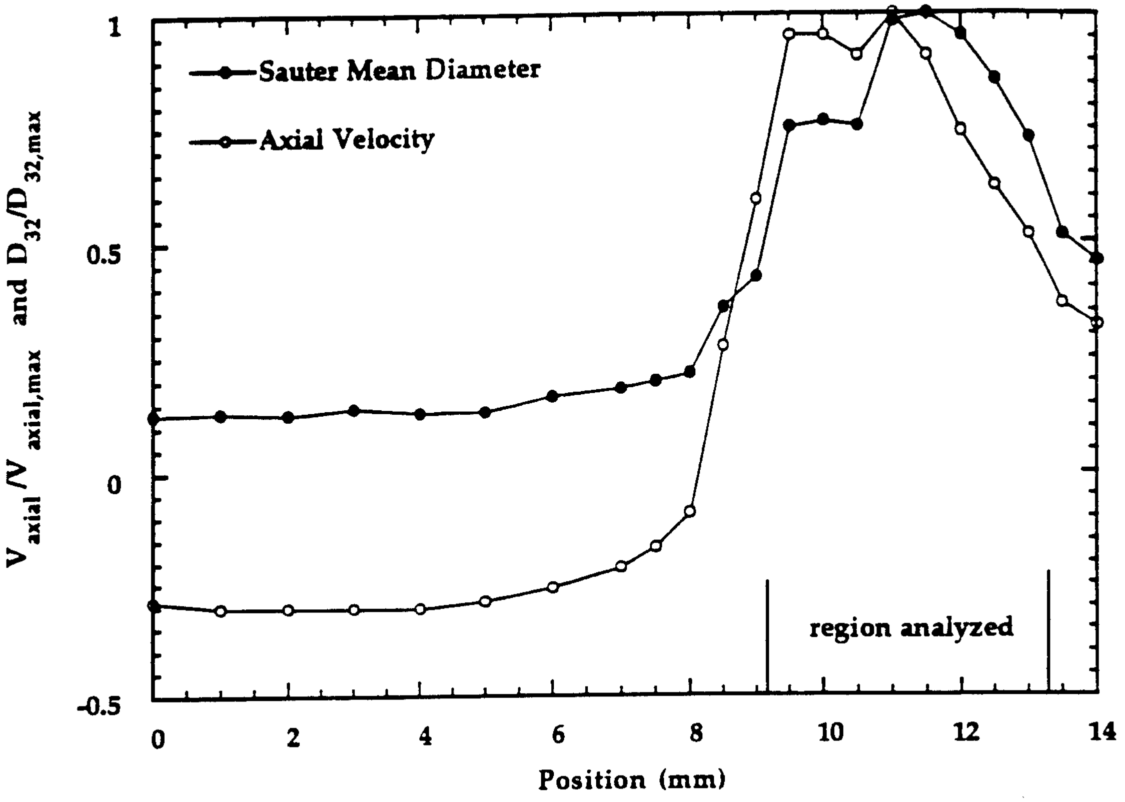


Figure 7 Experimentally measured mean axial velocity and Sauter mean diameter profiles

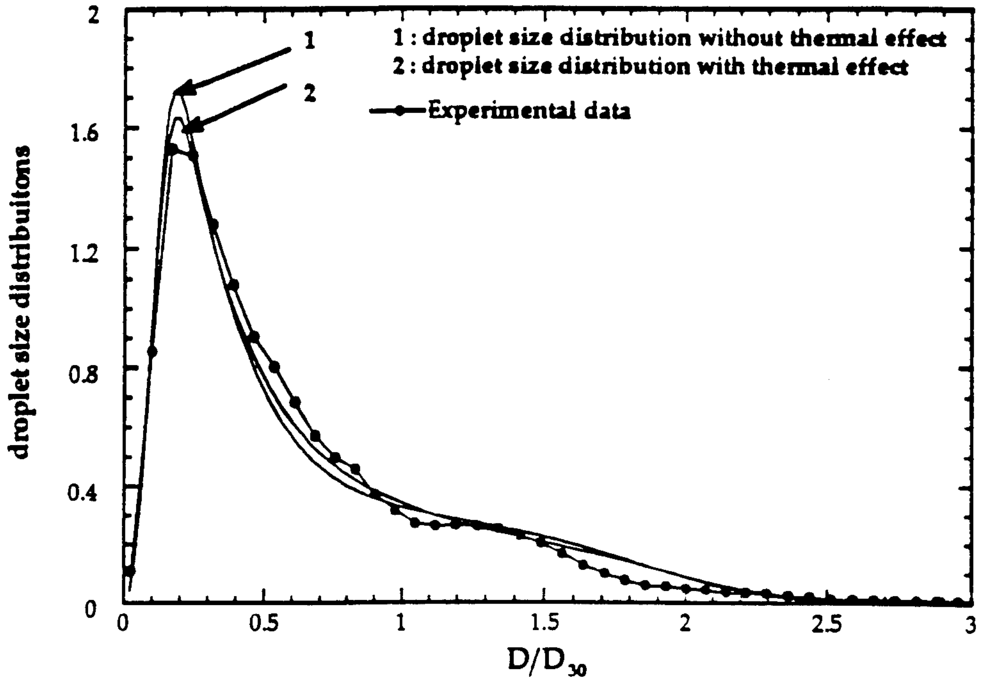


Figure 8. Comparison between calculated and measured droplet size distribuitons

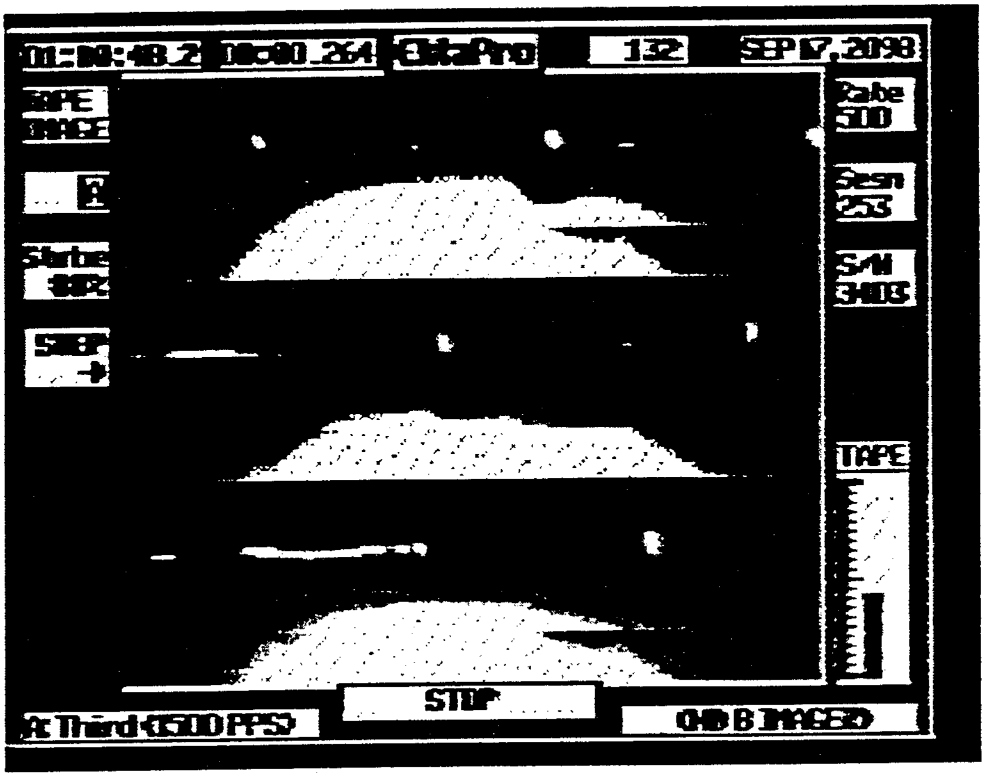


Figure 9. Photograph of data taken using Spin Physics video camera. Flow is from left to right.

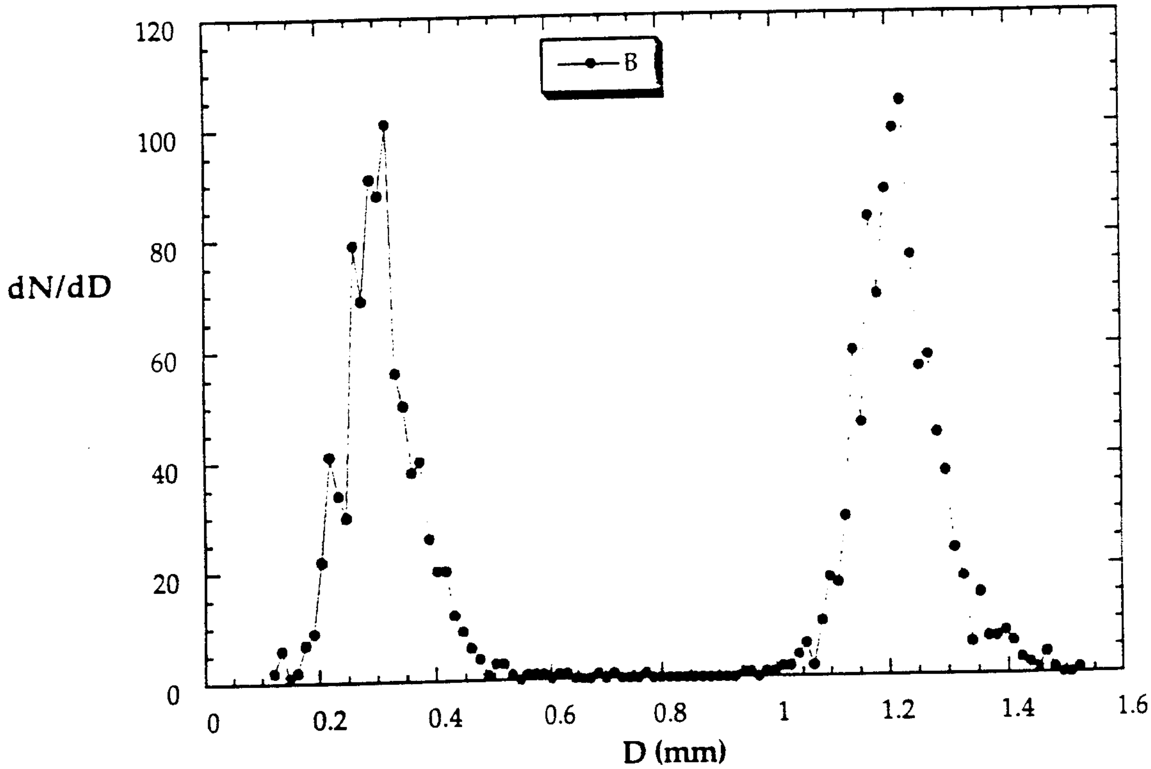


Figure 10. Size distribution plot of data taken with Spin Physics video camera.

Research Report
Research Initiation Program

OXIDATION AND REDUCTION OF CARBONOUS IMPEDIMENTS TO THE LOW TEMPERATURE GROWTH OF DIAMOND FILMS

submitted to

Research & Development Laboratories
Director for AFOSR Summer Research Program
5800 Uplander Way
Culver City, California 90230-6608

by

Fred V. Wells, Associate Professor
Campus Box 8439
Idaho State University
Pocatello, Idaho 83209
(208) 236-2674

January 27, 1993

Contents

Overview

The progress of the proposed research has not been as rapidly as the author would have liked. This has been due to two unforeseen difficulties. The method of constructing an electrode that is thermally and electrically isolated that had been planned was met with limited success, and much of our time during the summer months was spent exploring cost effective ways of to maintain a ceramic to metal vacuum seal at elevated temperatures. While the vacuum would hold initially, failure resulted upon a few cycles. The principal source of the problem is heating an electrode that is four inches in diameter to temperatures above $600^o C$. Typically, diamond films are prepared on small surface areas on the order of 5 mm × 5 mm where localized heating is more readily achieved. While we are currently interested in exploring methods of reducing unwanted deposits, the long range goal is to produce diamond films at temperatures lower than the 400-600°C range. Our group has equipment that approximates industrial scale since the laboratory was equipped with funds from Gould Electronics Semiconductor Division. Furthermore, the methods that we have developed to measure the rate of oxidation and reduction reactions in the plasma environment are dependent upon this scale of equipment. The current design of the electrode uses a pressurized metal o-ring seal from Helicoflex Company as the metal to ceramic seal. These electrodes are able to maintain the vacuum seal after several cycles of raising the temperature to 600°C. The metal o-rings are typically used from cryogenic temperatures to 760°C. It is unfortunate that the author became aware of the pressurized metal o-ring late in the summer.

The second impediment to progress was the failure of our mass spectrometer which prevented progress during the final four months of the formal project time period. Unfortunately, we are a small group and do not have any backup equipment. It is to be emphasized, however, that all materials necessary to complete the proposed measurements are on hand, and these measurements are expected be completed by July 30, 1993 if no other equipment malfunctions occur. The results of the research will be submitted for publication in a reviewed journal (Journal of Applied Physics), and a copy submitted to Research and Development Laboratories.

Electrode Design

The current electrode is mounted by extending the electrode stem through a 1.050" diameter hole in a four inch diameter ceramic (Al_2O_3) disc. This stem is then drawn tight against a metal o-ring by screwing a nut against opposite side of the ceramic disc. A vacuum seal is maintained by an inert gas pressure filled (600 psi) metal (Alloy X750) o-ring obtained from Helicoflex Company, Columbia, South Carolina (part No. U-6312-01431PFB). The ceramic disc is then mounted on a cooled flange that is attached to the plasma chamber. The electrode base is welded to the stem and the cooling channel. Schematics showing the relation of the electrode components and each component is included in Appendix I.

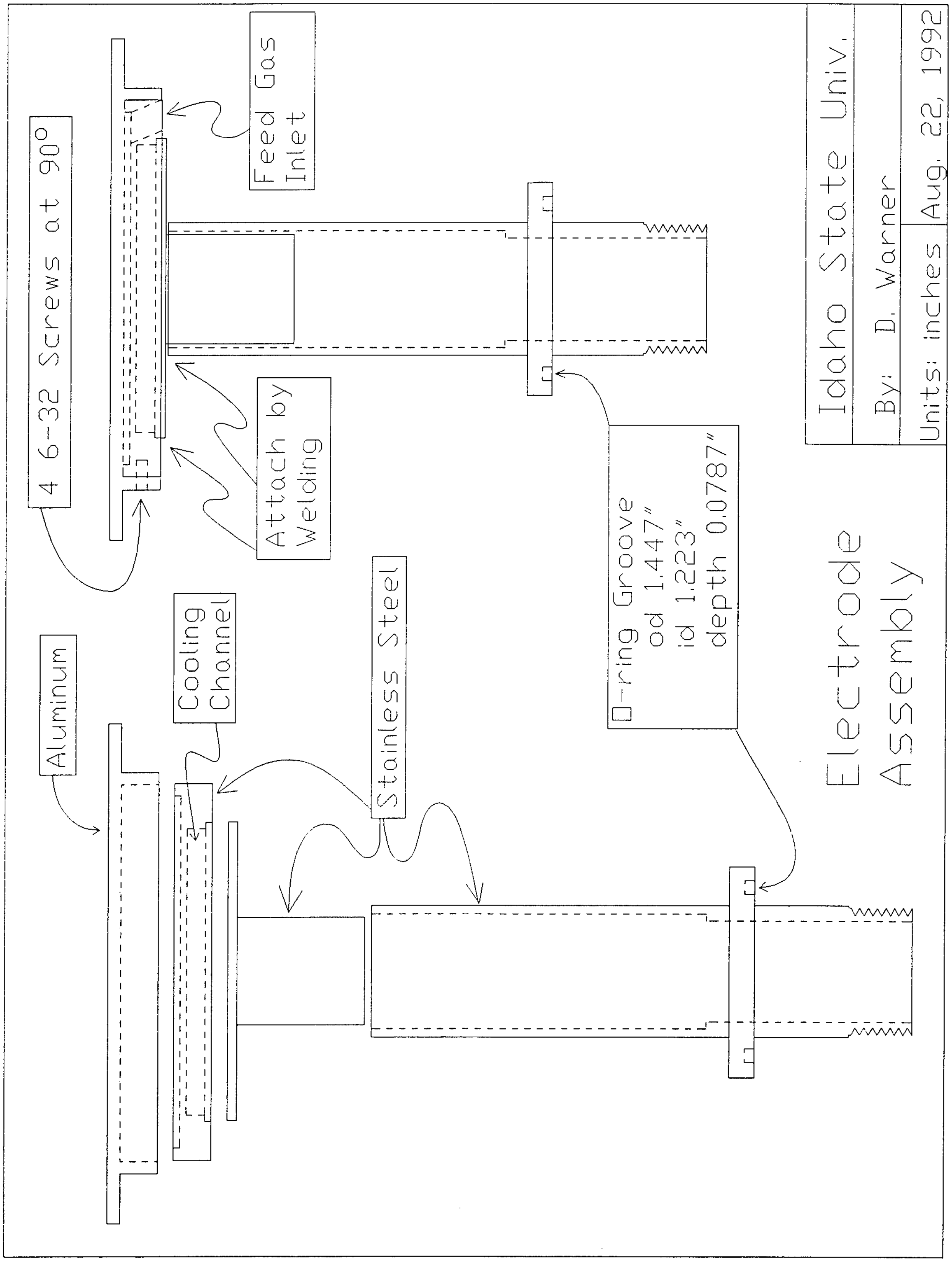
While we had initially constructed bodies using a castable ceramic, the inability to maintain adequate vacuum seals after heating, led us to return to constructing the electrode mostly out of metal parts. A thin wall stainless steel stem was used to attempt to establish a thermal gradient between the surface of the electrode and the lower part of the electrode stem which is attached to the ceramic disc. This is intended to minimize heat loss at higher temperatures.

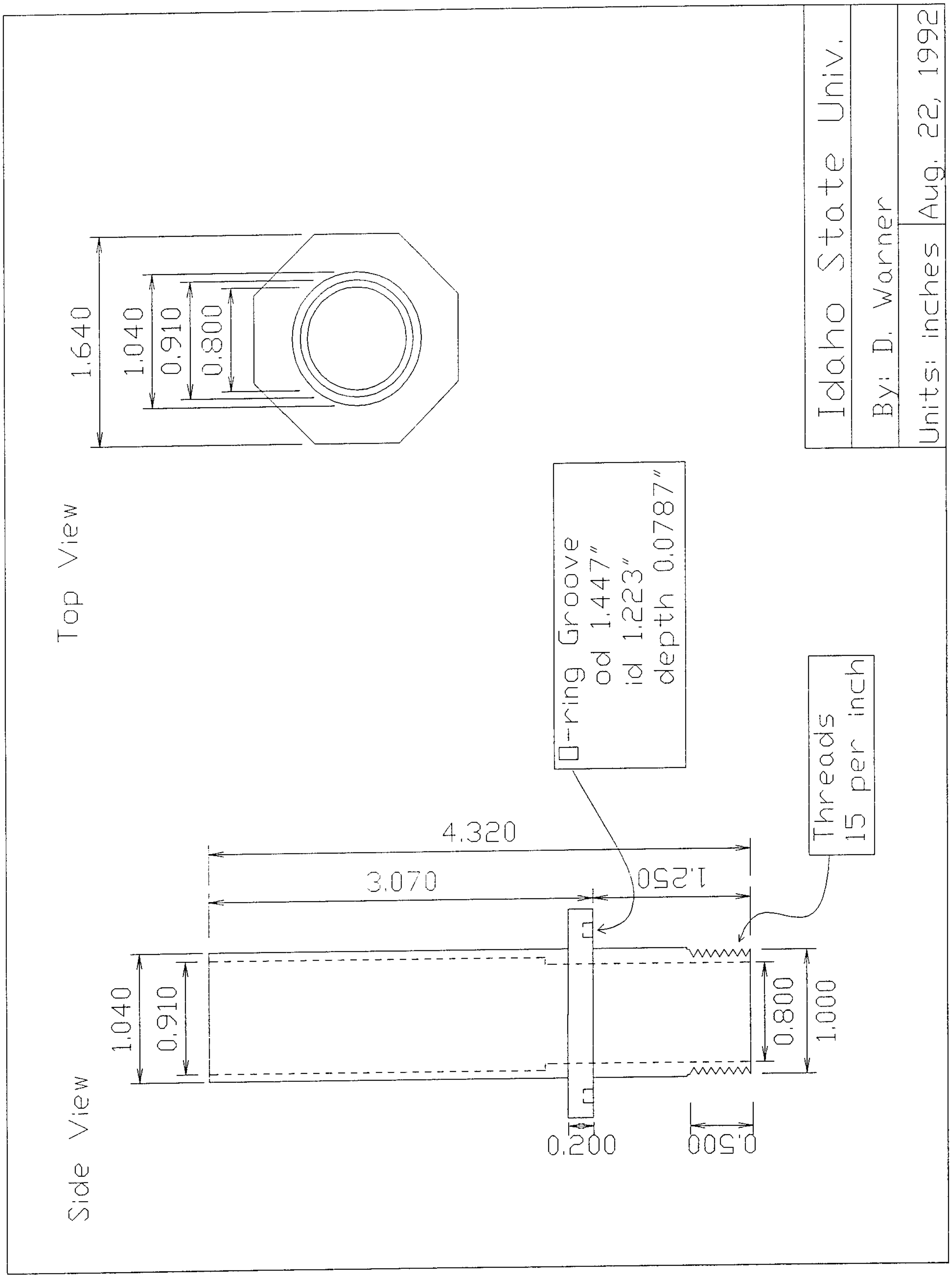
The schematics in appendix I are for the electrode through which the gas is fed. The other electrode is identical, but without the provisions for feeding the gas. A 0.250 inch alumni tube is cemented (Omega, High Temperature Cement) into the gas feed inlet of the cooling channel.

The electrode base has four 0.250 inch holes. Two opposite holes go through the base to provide access to the cooling channel. Alumni tube are cemented into these holes and extend out to the opening in the stem. The electrodes are cooled by circulating air through the cooling channel. The remaining holes house a 150 watt cartridge heater (Omega, CIR 1014/120), and a platinum resistance thermometer (Omega, 1Pt100-K2015) mounted in a 0.250 inch holder.

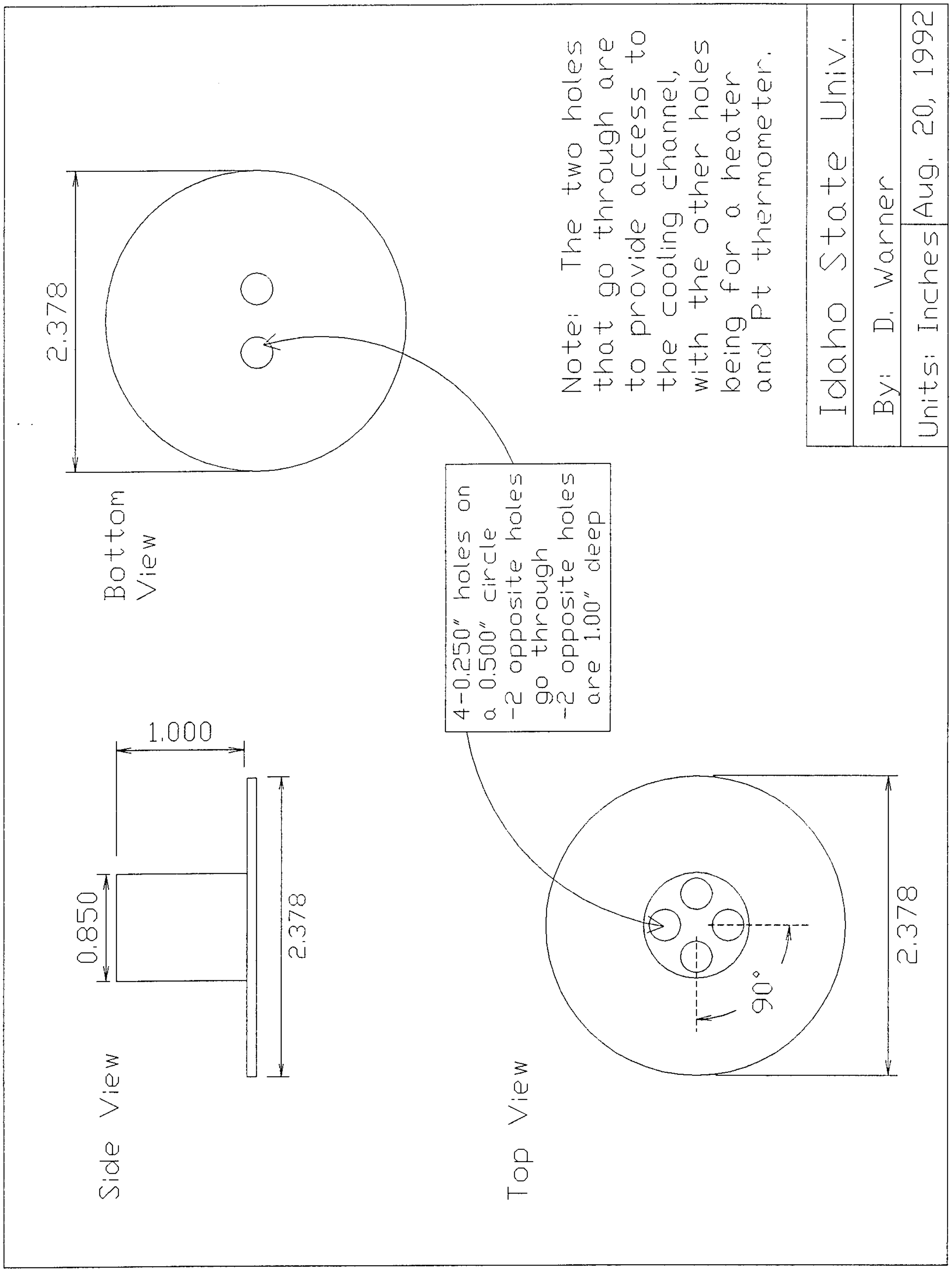
Appendix I.

Schematics of Electrode Components


Electrode Assembly
Idaho State Univ.
By: D. Warner
Units: inches
Aug. 22, 1992
4 6-32 Screws at 90°
Feed Gas Inlet
Attach by Welding
□-ring Groove
od 1.447"
id 1.223"
depth 0.0787"
Aluminum
Cooling Channel
Stainless Steel



Electrode Stem



4-0.250" holes on a 0.500" circle
-2 opposite holes go through
-2 opposite holes are 1.00" deep

Note: The two holes that go through are to provide access to the cooling channel, with the other holes being for a heater and Pt thermometer.

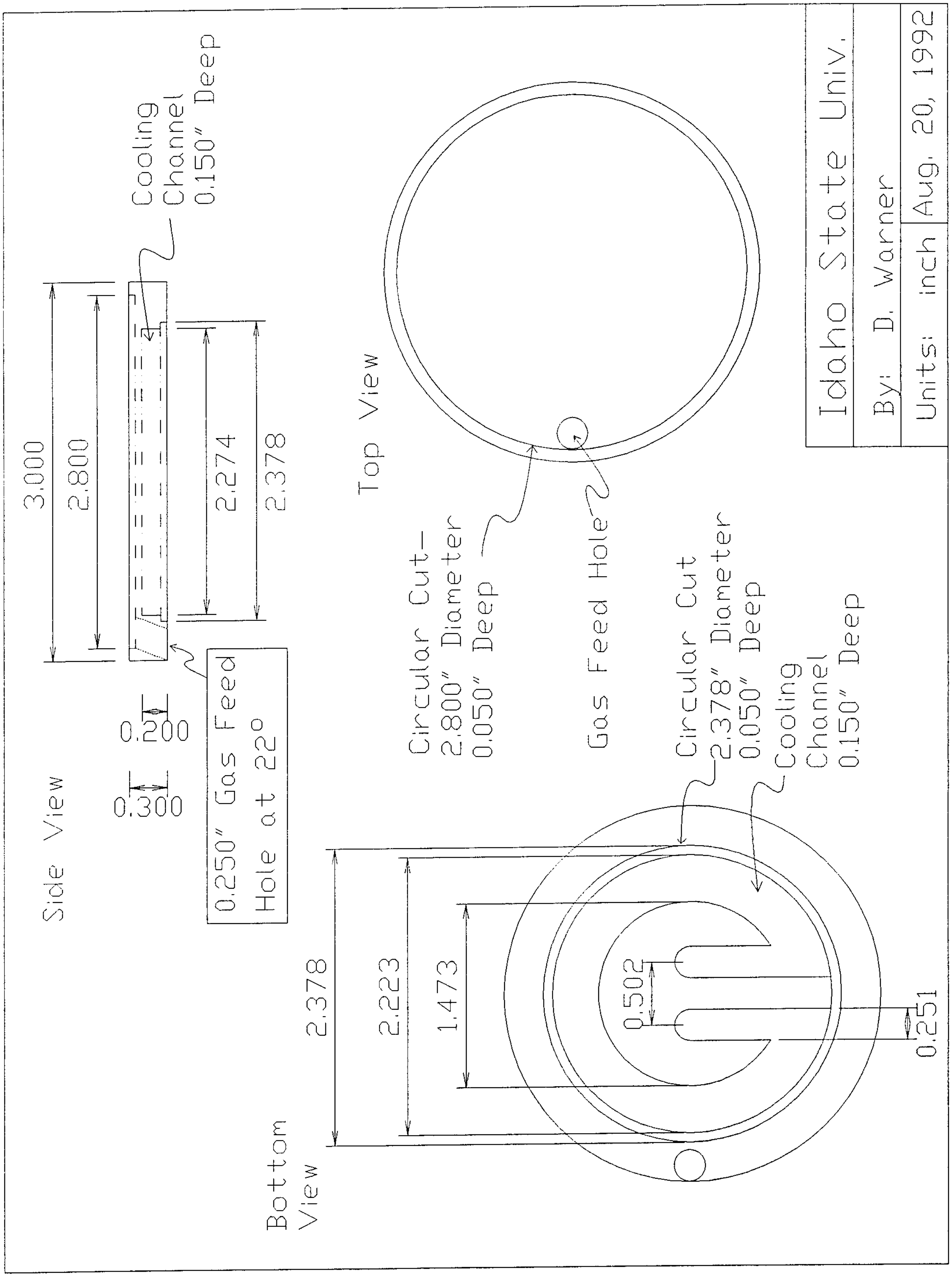
Idaho State Univ.

By: D. Warner

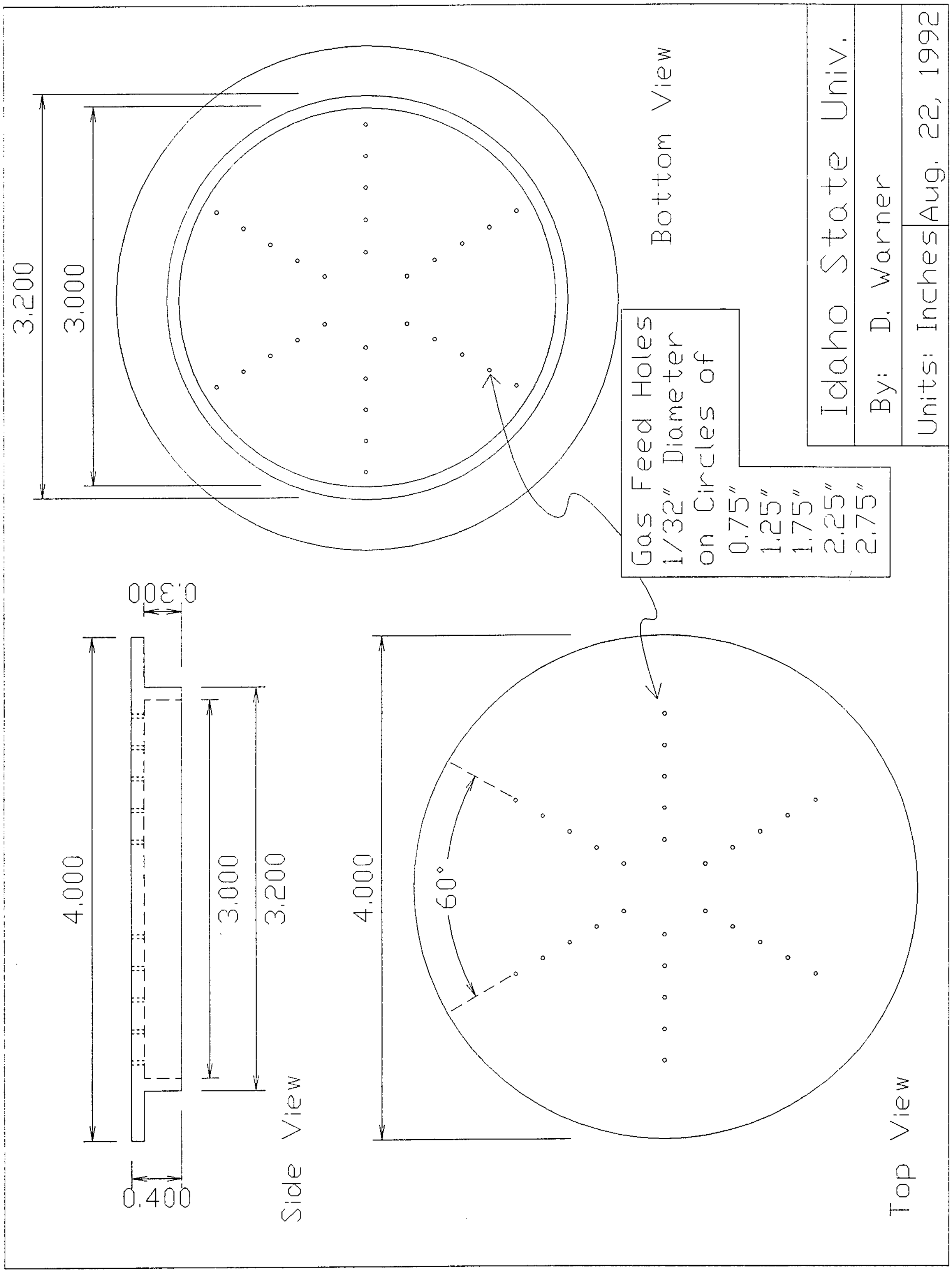
Units: Inches

Aug. 20, 1992

Electrode Base



Cooling Channel



Electrode Surface

SUPERPLASTIC $Y_3Al_5O_{12}$ (YAG)

Jeff Wolfenstine
Assistant Professor
Department of Mechanical and Aerospace Engineering

University of California, Irvine
Department of Mechanical and Aerospace Engineering
Irvine, CA 92717-3975

Final Report for:
Research Initiation Program
Materials Laboratory at Wright-Patterson AFB

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

December 1992

SUPERPLASTIC $Y_3Al_5O_{12}$ (YAG)

Jeff Wolfenstine
Assistant Professor
Department of Mechanical and Aerospace Engineering
University of California, Irvine

Abstract

The elevated temperature mechanical behavior of polycrystalline two-phase Al_2O_3-$Y_3Al_5O_{12}$ (matrix) materials, 50 vol.% Al_2O_3-50 vol.% $Y_3Al_5O_{12}$ [50YAG] and 25 vol.% Al_2O_3-75 vol.% $Y_3Al_5O_{12}$ [75YAG] was investigated. Both the 50YAG and 75YAG materials exhibited a brief normal primary creep transient (<1 to 2%), stress exponent close to unity ($n=1.2\pm0.1$) and an activation energy for creep of approximately 590 kJ/mole. It is postulated that the creep behavior of both the 50YAG and 75YAG materials is controlled by a Nabarro-Herring diffusional creep mechanism rate-controlled by either yttrium or aluminum lattice diffusion in the the majority $Y_3Al_5O_{12}$ phase. The creep rate of the two-phase Al_2O_3-$Y_3Al_5O_{12}$ materials is a strong function of the volume fraction of the Al_2O_3 second phase. As the amount of Al_2O_3 increases the creep rate of the two-phase Al_2O_3-$Y_3Al_5O_{12}$ materials increases. The two-phase Al_2O_3-$Y_3Al_5O_{12}$ materials were not superplastic as a result of the "large" grain size ($L\approx2$-3 μm) of the $Y_3Al_5O_{12}$ matrix and "low" theoretical density (98%).

SUPERPLASTIC $Y_3Al_5O_{12}$ (YAG)

Jeff Wolfenstine

I. INTRODUCTION

It is known that polycrystalline $Y_3Al_5O_{12}$ (YAG) is under consideration as a potential matrix material in ceramic composites that are be used in air at high-temperatures (T>1200°C) [1,2]. In order to form YAG based composites into the required shapes with a minimum amount of subsequent machining and joining it is desirable to achieve superplasticity in these materials. Superplasticity usually requires a fine grain size matrix. Typically 1 µm or less for ceramic materials. In order to keep the grain size of the matrix from growing at the superplastic forming temperature the presence of a second phase is usually required. There are many requirements that a second phase must meet in order to be utilized. For example, no reaction should occur between the second phase and the matrix. In addition, for the case of ceramics it is desirable that the thermal expansion coefficients of the matrix and second phase be similar as to prevent grain boundary separation upon cooling. A potential second phase oxide which can added to the $Y_3Al_5O_{12}$ (matrix) which meets these requirements is Al_2O_3 (alumina).

It is the purpose of this report to present some preliminary results on the elevated temperature mechanical behavior of polycrystalline two-phase Al_2O_3-$Y_3Al_5O_{12}$ (matrix) materials with emphasis on achieving superplasticity in these materials and to compare the results to polycrystalline single phase $Y_3Al_5O_{12}$. The shape of the creep curve, variation of the stress exponent and activation energy for creep as a function of volume fraction Al_2O_3 will be presented.

II. EXPERIMENTAL PROCEDURE

(1) Materials

The Al_2O_3-$Y_3Al_5O_{12}$ system was chosen for the present study. Two different compositions were selected: (1) 50 vol.% Al_2O_3-50 vol.% $Y_3Al_5O_{12}$ [50YAG] and 25 vol.% Al_2O_3-75 vol.% $Y_3Al_5O_{12}$ [75YAG]. Powders of the two compositions were obtained from Ceralox Corporation (Tucson, AZ). The cation impurities were less than 10 ppm except for Si and Na which were less than 20 ppm in both compositions. Both materials had a mean particle size of about 0.5 µm. X-ray diffraction of the powders revealed them to be two-phase (Al_2O_3 and $Y_3Al_5O_{12}$) with a volume fraction of approximately 0.78 $Y_3Al_5O_{12}$ for the 75YAG material and 0.47 $Y_3Al_5O_{12}$ for the 50YAG material. The powders were heated for 2 hours at 200°C to remove any absorbed gases and excess water that remained after chemical processing.

The powders were uniaxially cold-pressed at room temperature into right circular cylinders. The cylinders were placed on alumina spacers and sintered at 1600°C in air until both materials achieved a density of greater than 98% of the theoretical density. Densities were measured using the Archimedes method with water as the liquid. For the 50YAG material sintering times of greater than 72 hours were required whereas, the 75YAG material required sintering times of at least 120 hours. After sintering the specimens were polished so that the ends were parallel to each other. The ratio of the height to width of the samples was approximately 2:1 with a typical specimen having a diameter of approximately 2 mm and a height of about 4 mm.

(2) Creep Testing

The samples were deformed under uniaxial compression in a dead-load apparatus. All the creep tests were preformed at a single stress and temperature. The creep experiments were conducted under conditions of approximately constant true stress. To ensure constant true stress, load adjustments were made after 0.01 true strain intervals. The magnitude of the load adjustment was based on the assumption that the specimen volume remained constant throughout the experiment and the deformation was homogeneous. Prior to testing the specimens were annealed in air at the testing temperature for 1 to 2 hours. Polished Al_2O_3 platens separated the sample from the SiC pistons. It was observed that when SiC platens were used with both the 75YAG and 50YAG materials that a reaction occurred between the sample and the platens. Creep tests were carried out in air at temperatures between 1350 to 1400°C under true stresses between 1 to 20 MPa. Samples were typically deformed to true strains between to 0.10 to 0.15. After completion of creep testing samples were cooled under load.

(3) Microstructural Observations

Microstructures of both the 75YAG and 50YAG materials prior to and after testing were examined using scanning electron microscopy (SEM). The samples were polished using conventional techniques and thermal etched at 1500°C for 30 minutes to 4 hours. No grain growth due to thermal etching was observed. The linear intercept method was used to determine the grain size.

III. RESULTS

(1) As-Sintered Microstructure

The microstructure of the 50YAG and 75YAG materials after sintering is shown in Fig. 1. Figure 1 A is a secondary electron image of the 50YAG whereas, Fig. 1B is of the 75 YAG material. Figure 2 is a high magnification image of only the $Y_3Al_5O_{12}$ matrix in the 75YAG material. The $Y_3Al_5O_{12}$ matrix in the 50YAG material exhibited a similar structure to that shown in Fig. 2. From Figs. 1 and 2 several important points are noted. First, the Al_2O_3 (dark) phase is fairly uniformly distributed throughout the $Y_3Al_5O_{12}$ (light) matrix in both the 75YAG and 50YAG materials. Second, the grain size of the Al_2O_3 phase is larger than that for the $Y_3Al_5O_{12}$ matrix in both materials. Third, the Al_2O_3 grain size in the 50YAG material is greater than in the 75YAG material. The average linear intercept grain size, L, of the Al_2O_3 phase in the 75YAG material is about 5 μm. In the 50YAG material the average linear intercept grain size of the Al_2O_3 phase is about 15 μm. Fourth, both materials are highly dense with the porosity located primarily at grain boundary junctions. Fifth, the $Y_3Al_5O_{12}$ matrix exhibits a equiaxed grain structure in both materials. Sixth, the linear intercept grain size of the $Y_3Al_5O_{12}$ matrix in the 75YAG material ($L \approx 2.8$ μm) is smaller than that observed in the 50YAG material ($L \approx 3.2$ μm). Image analysis of backscattered images of the 50YAG and 75YAG materials yielded a volume fraction of $Y_3Al_5O_{12}$ equal to 0.48 and 0.79, respectively. These results are in good agreement with the volume fraction of $Y_3Al_5O_{12}$ determined using x-ray diffraction.

(2) Deformed Microstructures

Examination of the deformed microstructures in both the 75 YAG and 50 YAG materials revealed that no dynamic grain growth occurred as a result of the deformation. The grains remained fairly equiaxed after deformation at all temperatures of testing. However, the micrographs revealed evidence of cavitation after testing. The amount of cavitation as a result of the deformation was estimated from micrographs of the deformed samples and Archimedes method to be between 2 to 4 %.

(3) Shape of Creep Curves

A representative creep curve for the 50YAG and 75 YAG materials is shown in Fig. 3. Figure 3A is for the 50YAG material. Figure 3B is for the 75YAG material. The data in Fig. 3 are plotted as logarithm of strain rate versus true strain. An examination of Fig. 3 reveals two important results.

First, both creep curves exhibit a normal primary creep period, during which the creep rate decreases with strain. This is followed by a steady-state period where the creep rate remains essentially constant with strain. Second, in both materials a primary creep strain of 0.01 to 0.02 is observed prior to the onset of steady-state behavior. The length of the primary creep stage was always less than 2% for the entire stress and temperature range investigated for both materials.

(4) Stress Dependence of the Steady-State Creep Rate

The dependence of the steady-state creep rate on the applied stress for the 50YAG and 75YAG materials at four temperatures (1350, 1363, 1375 and 1400°C) is shown in Fig. 4. Figure 4A is for the 50YAG material. Figure 4B is for the 75YAG material. The data in Fig. 4 are plotted as logarithm of steady-state strain rate, $\dot{\varepsilon}$, versus applied stress, σ, on a logarithmic scale. The slope of the curves yields the stress exponent, n, according to the relation, $\dot{\varepsilon}=k\sigma^n$, where k is a constant incorporating temperature and microstructural dependencies. The value of the stress exponent for both materials is about the same, 1.2±0.1, and essentially independent of temperature. The value of the stress exponent obtained for both the 50YAG and 75 YAG materials is in excellent agreement with the value of 1.2 exhibited by single phase $Y_3Al_5O_{12}$ of similar grain size ($L \approx 3.0$ µm) over a temperature between 1400 to 1610°C [3,4].

(5) Activation Energy for Creep

The activation energy for creep, Q, was determined for both materials from a plot of logarithm $\dot{\varepsilon}$ versus inverse absolute temperature, 1/T, at constant stress and grain size. Figure 5 is such a plot for both the 50 YAG and 75 YAG materials at two different stresses. Figure 5A is for the 50YAG material. Figure 5B is for the 75YAG material. The activation energy for creep determined from the slope of the straight lines in Fig. 5 for the 50YAG and 75YAG materials is approximately 570 kJ/mole for the 50YAG material and 610 kJ/mole for the 75YAG material. These values are in good agreement with each other and with the value of 584 kJ/mole obtained for single phase polycrystalline $Y_3Al_5O_{12}$ in the n=1.2 region [3,4].

IV. DISCUSSION

The observations that the type of creep curve and length of the primary creep stage, stress exponent and the activation energy for creep for the 50YAG and 75YAG materials are similar suggests that these materials have the same rate-controlling deformation mechanism. The value of the stress exponent close to unity (n=1.2±0.1) and short primary creep stage (<1-2%) suggests that the deformation takes place by a diffusional creep mechanism. It is likely that the creep behavior of the two-phase Al_2O_3-$Y_3Al_5O_{12}$ materials is controlled by a Nabarro-Herring diffusional creep mechanism rate-controlled by lattice diffusion of the slowest species (yttrium or aluminum or oxygen) in the majority $Y_3Al_5O_{12}$ phase. This suggestion is based on the following observations: i) The activation energy for creep for the two-phase Al_2O_3-$Y_3Al_5O_{12}$ materials (≈590 kJ/mole) is in good agreement with the value exhibited by polycrystalline single phase $Y_3Al_5O_{12}$ in the n=1.2 region (584 kJ/mole) where it was postulated that Nabarro-Herring diffusional creep mechanism was the dominant deformation mechanism [3,4]. ii) The activation energy for creep of the two-phase Al_2O_3-$Y_3Al_5O_{12}$ materials is higher than that for lattice diffusion of aluminum in Al_2O_3 (476 kJ/mole) [5]. It is generally accepted that Nabarro-Herring diffusional creep of Al_2O_3 of grain size L≈3 μm is rate-controlled by lattice diffusion of the aluminum cation [6]. iii) The activation energy for creep for the two-phase Al_2O_3-$Y_3Al_5O_{12}$ and single $Y_3Al_5O_{12}$ materials is in agreement with the value for single crystalline $Y_3Al_5O_{12}$ (650-700 kJ/mole) in the power-law region where dislocation creep is controlled by lattice diffusion of the the slowest moving species [7]. The fact that the activation energy for creep for the two-phase Al_2O_3-$Y_3Al_5O_{12}$ and single phase $Y_3Al_5O_{12}$ materials is in close agreement with that exhibited by single crystalline $Y_3Al_5O_{12}$ combined with the observation that the activation energy for oxygen lattice diffusion in single crystalline $Y_3Al_5O_{12}$ is about 325 kJ/mole [8] suggests that the Nabarro-Herring diffusional creep of the two-phase Al_2O_3-$Y_3Al_5O_{12}$ materials is rate-limited by lattice diffusion of the cations, either yttrium or aluminum, in the majority $Y_3Al_5O_{12}$ phase. However, since no data is available for lattice diffusion of either yttrium or aluminum in $Y_3Al_5O_{12}$ it is impossible to determine which of cations is rate-controlling the deformation behavior of the two-phase Al_2O_3-$Y_3Al_5O_{12}$ materials.

The effect of volume fraction of Al_2O_3 on the creep behavior of $Y_3Al_5O_{12}$ is shown in Fig. 6. Creep data at 1400°C for the 50YAG, 75YAG and single phase YAG materials are shown. The data for single phase YAG were extrapolated from higher stresses using n=1.4 [3,4]. The grain size of the YAG matrix phase for each material is listed in the figure. From Fig. 6 it is observed that as the volume fraction of Al_2O_3 increases the creep rate of the two-phase mixture increases. For example, at an applied stress equal to 10 MPa the creep rate of the 75YAG material is seven times faster than that for single phase YAG, whereas the creep rate of the 50YAG material is almost thirty times faster than that for single phase YAG. The creep rate of the 50YAG material is three times faster than the 75YAG material over the entire stress range. Addition of the second phase Al_2O_3 to the $Y_3Al_5O_{12}$ matrix has caused the creep resistance to decrease. As can been observed from Fig. 6 the grain size of the $Y_3Al_5O_{12}$ matrix for the three materials is slightly different. Thus, it is possible that the difference in creep resistance for these materials is not only a result of the difference in volume fraction of Al_2O_3 but also due to the different $Y_3Al_5O_{12}$ matrix grain sizes.

It was previously suggested that the creep behavior of the 50YAG, 75YAG and single phase YAG materials is a result of a Nabarro-Herring diffusional creep mechanism. For the case of a Nabarro-Herring diffusional creep mechanism the creep rate should vary inversely with the grain size squared [9,10]. Thus, it is possible to eliminate the effect of grain size on the creep rate for the data shown in Fig. 6 by normalizing the creep rate with respect to the grain size squared. Figure 7 is a plot of the normalized creep rate, $\dot{\varepsilon}L^2$, versus stress for the data shown in Fig. 6. Examination of Fig. 7 confirms the results of Fig. 6, that indeed the addition of the second phase Al_2O_3 to the $Y_3Al_5O_{12}$ matrix has caused the creep resistance of the material to decrease. This result may seem at first surprising since Al_2O_3 has a slightly higher melting point, T_m, (T_m≈2055°C) compared to $Y_3Al_5O_{12}$ (T_m≈1950°C) and thus, the addition of a higher melting phase to the matrix in the case where no reaction between the phases occurs would cause the creep rate of the mixture to decrease. However, the lattice diffusivity of the rate-controlling species, aluminum, in Al_2O_3, is higher than the predicted lattice diffusivity of either yttrium or aluminum in the majority $Y_3Al_5O_{12}$ phase (based on creep data [3]). Thus, the addition of Al_2O_3 with a faster diffusing rate-controlling species than in $Y_3Al_5O_{12}$ would cause the creep rate of the mixture to increase with addition of Al_2O_3 to $Y_3Al_5O_{12}$ as observed in Figs. 6 and 7. Finally, it should be noted that the difference in creep rate between the 50 YAG and 75 YAG materials is mainly a result of the volume fraction difference and not a result of the grain size difference of the Al_2O_3 phase in each material. If a grain size normalization is applied to the Al_2O_3 phase (L≈5 μm for the 50 YAG and L≈15 μm for the 75 YAG) in each material assuming that the Al_2O_3 phase deforms by a Nabarro-Herring diffusional creep mechanism this would cause the creep rate of the 50YAG material relative to that for the 75YAG material to be enhanced by an additional factor of about 1.5 over that shown in Fig. 6. This normalization leads to an even greater effect of volume fraction of the Al_2O_3 phase on the creep rate of the Al_2O_3-$Y_3Al_5O_{12}$ mixtures.

As a result of the "large" grain sizes (L≈2-3 μm) of the YAG matrix phase for both the 50YAG and 75YAG materials and "low" theoretical density (98%) maximum strains to failure of only about 0.20 to 0.30 were achieved before severe cavitation and microcracking occurred. Consequently, it was not possible to achieve superplasticity in these materials. Attempts to reduce the grain size to below 1 μm and obtain higher density samples (relative densities >99%) are currently underway. Preliminary results using a reactive hot-pressing method have shown that YAG matrix grain sizes on the order of ≈1 μm can be obtained however, the relative density is less than 96%.

V. CONCLUSIONS

(1) The existence of a brief normal primary creep transient (<1 to 2%) prior to steady-state behavior and a stress exponent close to unity (n=1.2±0.1) for both two-phase materials, 50 vol.% Al_2O_3-50 vol.% $Y_3Al_5O_{12}$ [50YAG] and 25 vol.% Al_2O_3-75 vol.% $Y_3Al_5O_{12}$ [75YAG], suggests that the deformation of both these materials occurs by a diffusional creep mechanism.

(2) The activation energy for creep for both the 50YAG (570 kJ/mole) and the 75YAG (610 kJ/mole) materials is similar and in good agreement with that obtained for polycrystalline single phase YAG and single crystalline YAG.

(3) It is postulated that the creep behavior of both the 50YAG and 75YAG materials is controlled by a Nabarro-Herring diffusional creep mechanism rate-controlled by either yttrium or aluminum lattice diffusion in the the majority YAG phase.

(4) The creep rate of both the 50YAG and 75YAG materials is a strong function of the volume

fraction of second phase Al_2O_3. As the amount of Al_2O_3 increases the creep rate of the two-phase Al_2O_3-$Y_3Al_5O_{12}$ materials increases.

(5) The two-phase Al_2O_3-$Y_3Al_5O_{12}$ materials were not superplastic as a result of the “large” grain size ($L \approx 2$-3 μm) of the $Y_3Al_5O_{12}$ matrix and “low” theoretical density (98%).

REFERENCES

1. K. Keller, T. Mah and T. A. Parthasarathy, Ceram. Eng. Sci. Proc., 11 [7-8], 1122 (1990).
2. T. Mah, T. A. Parthasarathy and L. Matson, Ceram. Eng. Sci. Proc., 11 [9-10], 1617 (1990)
3. T. A. Parthasarathy, T. Mah and K. Keller, Ceram. Eng. Sci. Proc., 12 [9-10], 1767 (1991).
4. T. A. Parthasarathy, T. Mah and K. Keller, J. Am. Ceram. Soc., 75 [7], 1756 (1992).
5. A. E. Paladino and W. D. Kingery, J. Chem. Phys., 37, 957 (1962).
6. W. R. Cannon and T. G. Langdon, J. Mater. Sci., 23, 1 (1988).
7. G. S. Gorman, Ceram. Eng. Sci. Proc., 12 [9-10], 1745 (1991).
8. H. Handea, Y. Miyazawa and S. Shirasaki, J. Cryst. Growth, 68, 581 (1984).
9. F. R. N. Nabarro, Rep. Conf. on the Strength of Solids, Physical Soc., London, 1948, p.75.
10. C. Herring, J. Appl. Phys., 21, 437 (1950).

LIST OF FIGURES

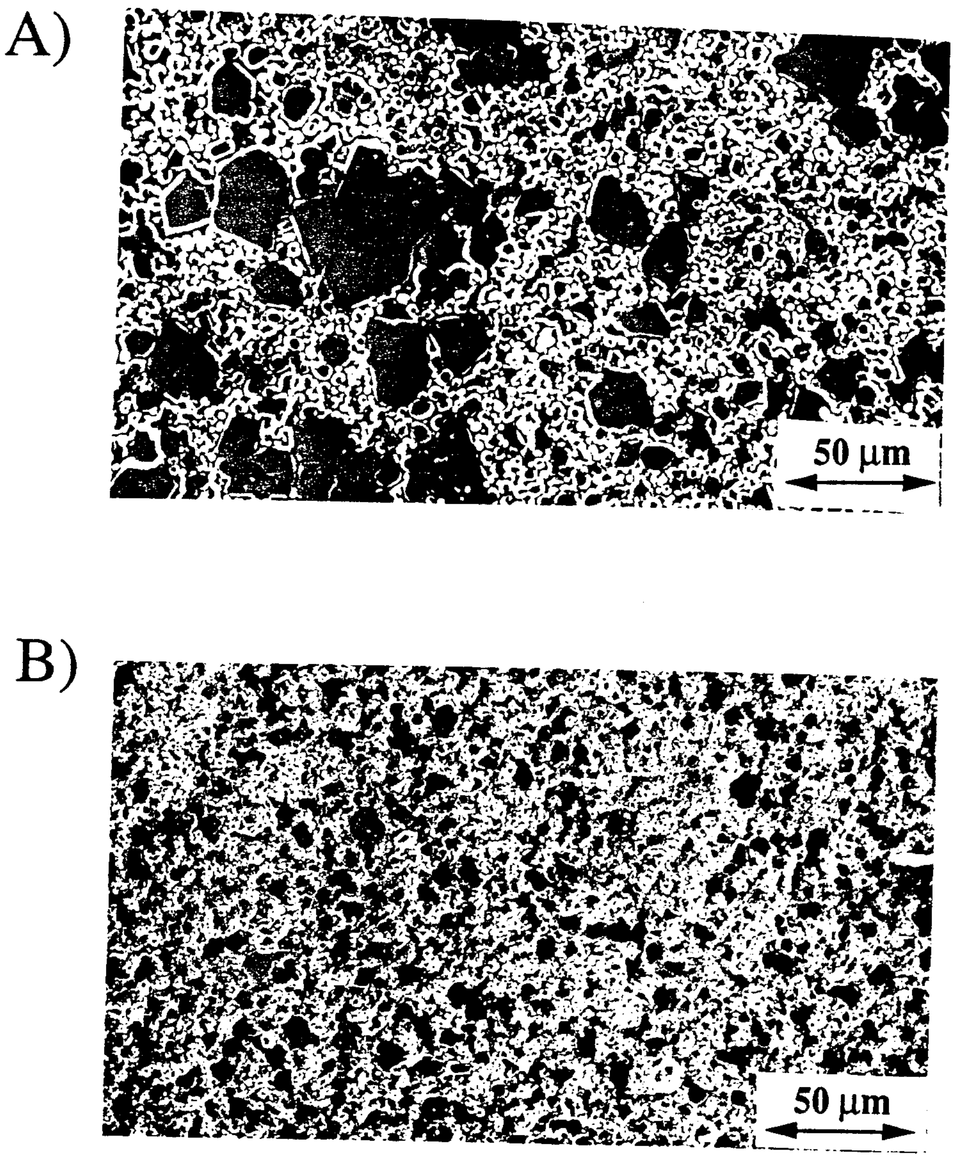


FIG. 1. Microstructure of the as-sintered A) 50YAG and B) 75YAG materials.

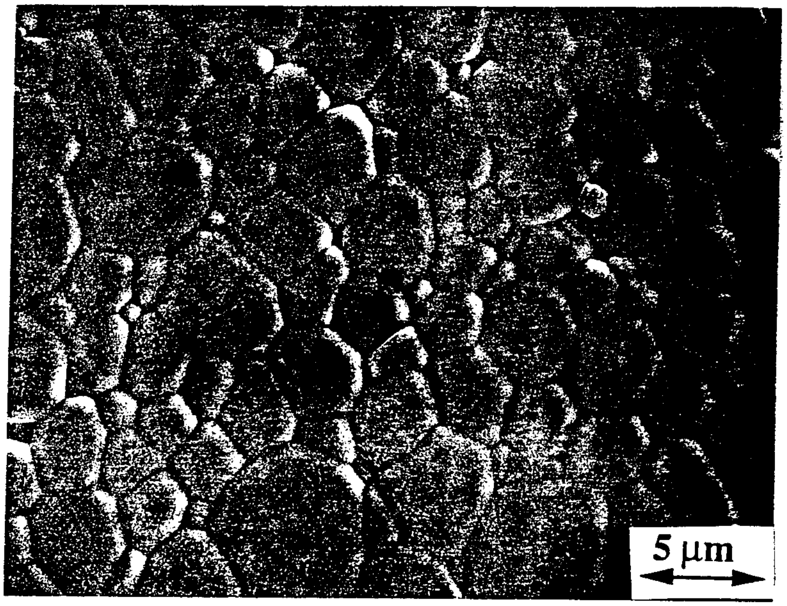


FIG. 2. Microstructure of only the $Y_3Al_5O_{12}$ matrix in the 75YAG material.

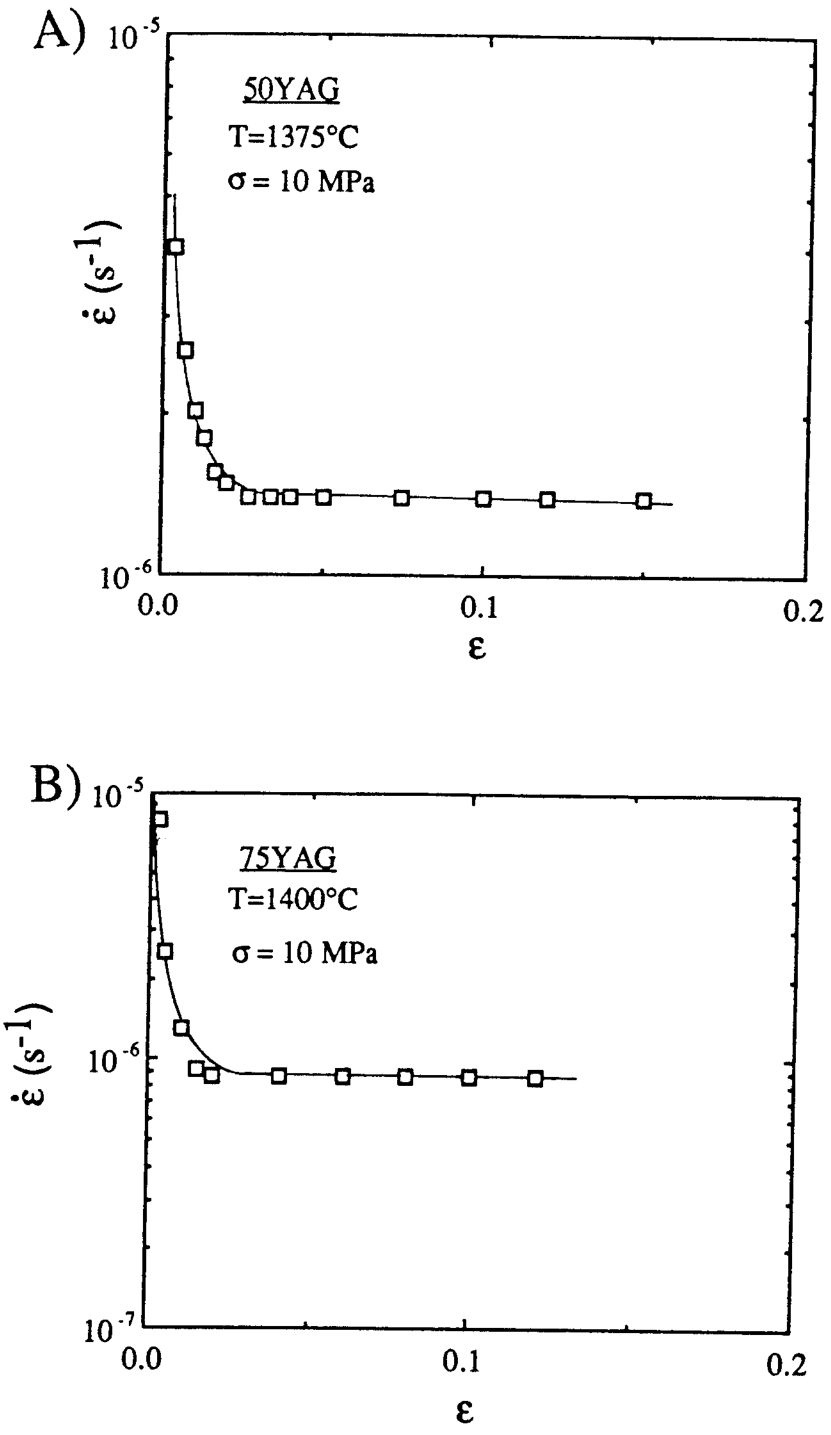


FIG. 3. Typical creep curve for the A) 50YAG and B) 75YAG materials.

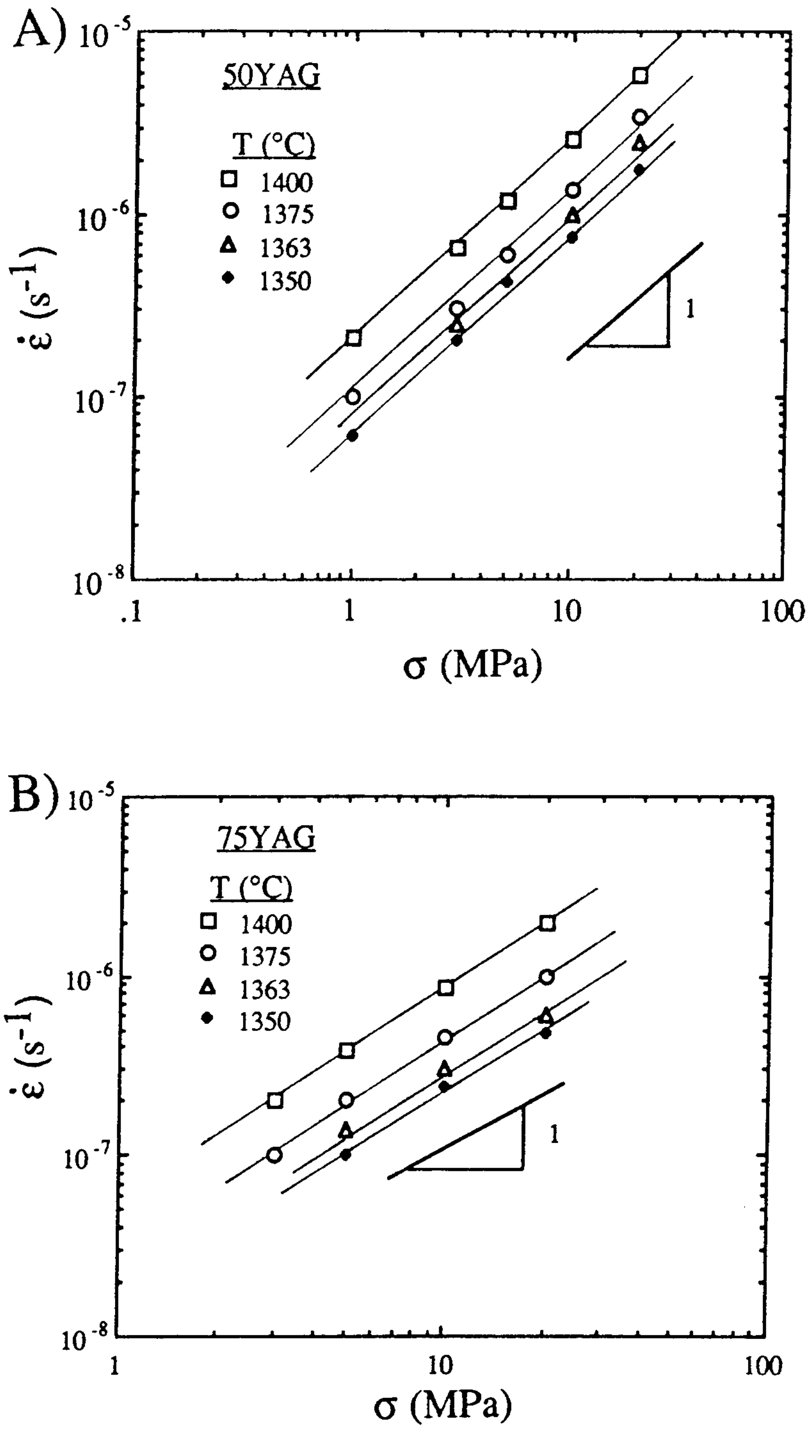


FIG. 4. Steady-state creep rate versus stress for the A) 50YAG and B) 75YAG materials.

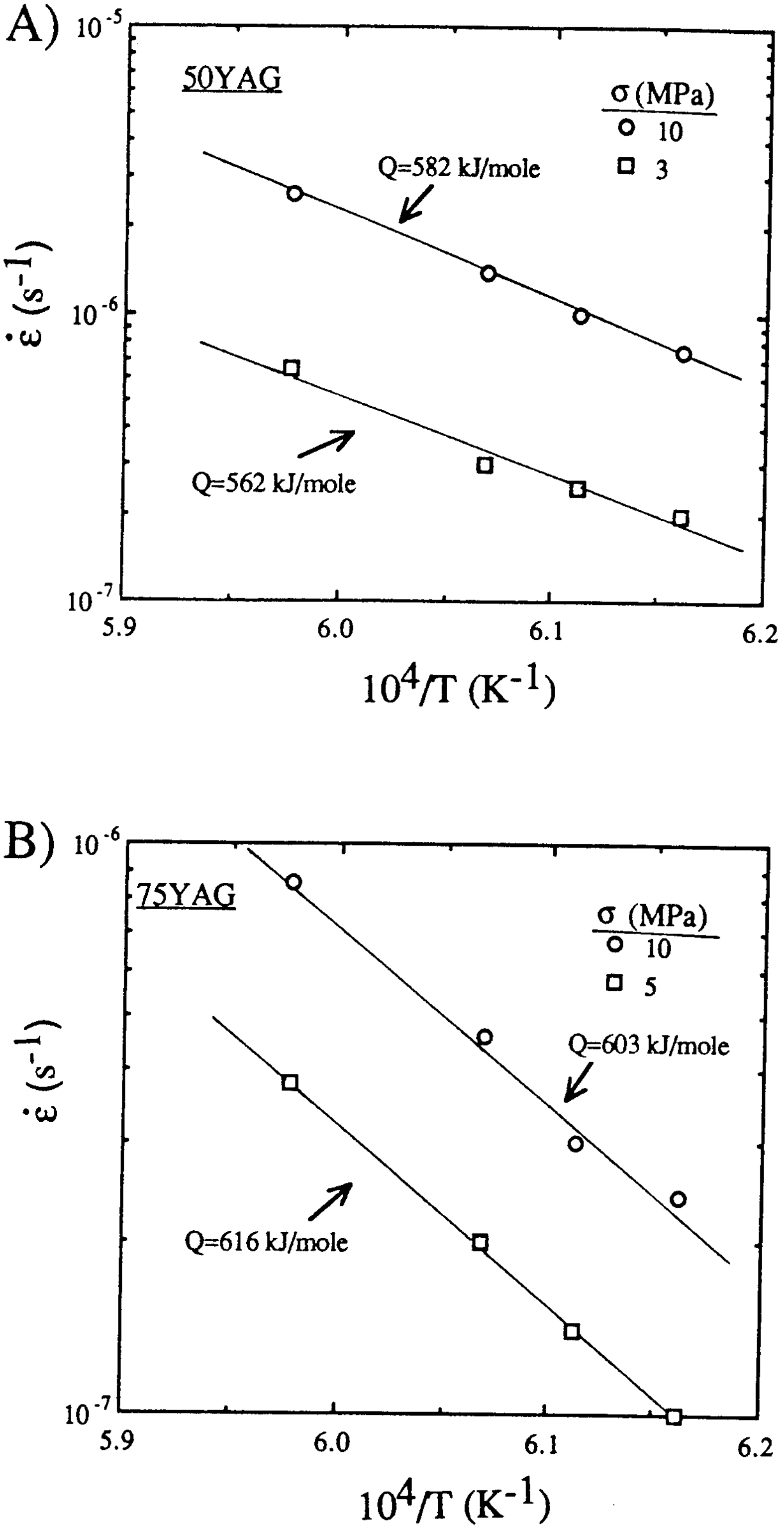


FIG. 5. Steady-state creep rate versus reciprocal absolute temperature for the A) 50YAG and B) 75YAG materials.

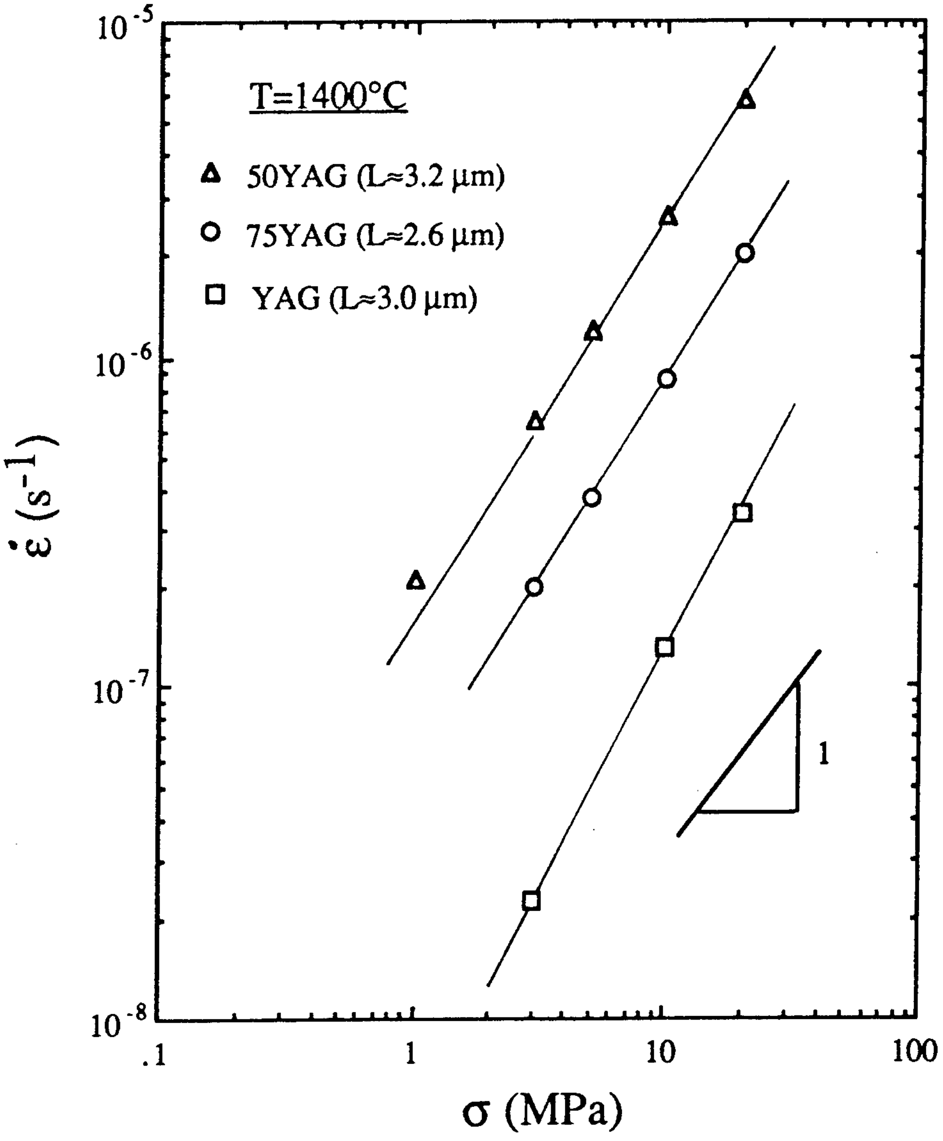


FIG. 6. Steady-state creep rate versus stress for the 50YAG, 75YAG and single phase YAG materials.

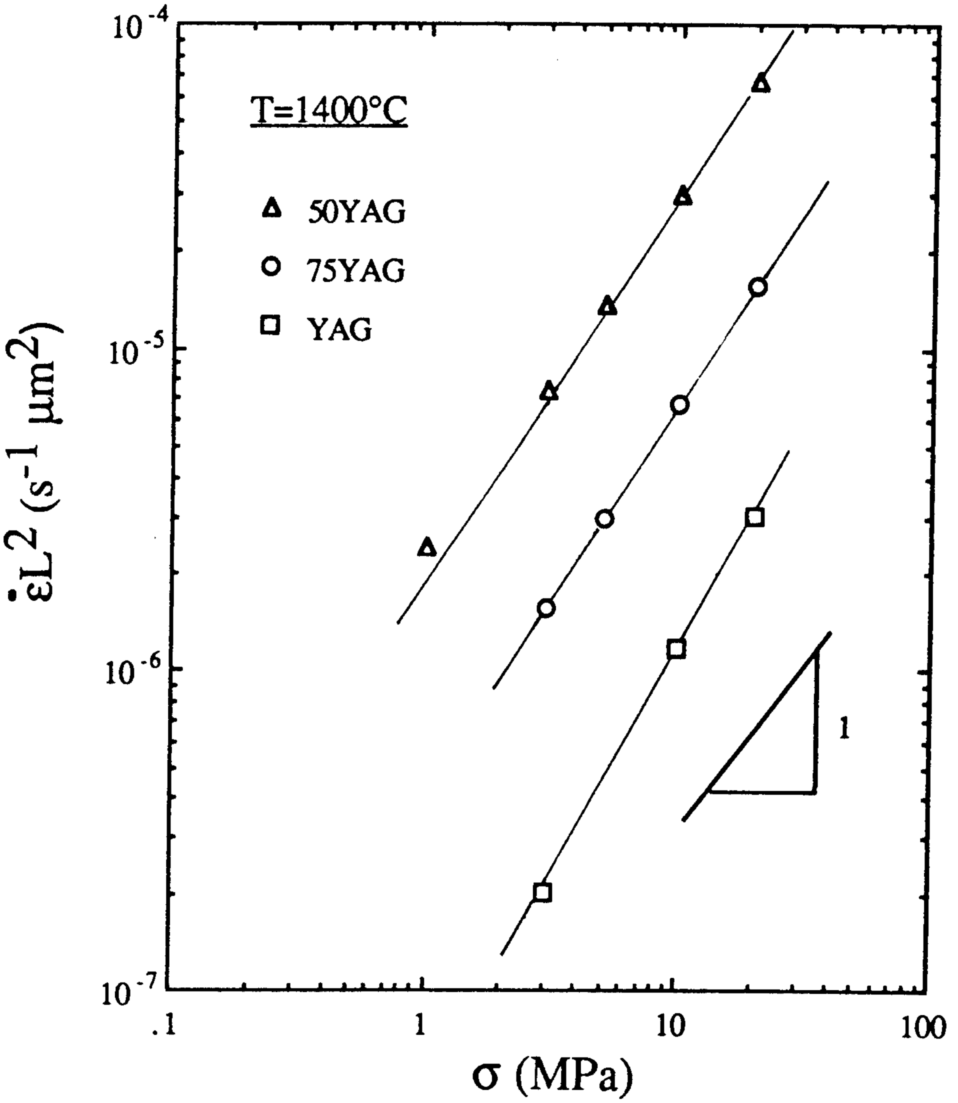


FIG. 7. Grain size normalized steady-state creep rate versus stress for the 50YAG, 75YAG and single phase YAG materials.

PATTERN THEORY: EXTENSIONS AND APPLICATIONS

James S. Wolper
Assistant Professor
Department of Mathematics

Idaho State University
Pocatello, ID 83209-8085

Final Report for:
Research Initiation Program
USAF Wright Laboratories

Research Sponsored by:
Air Force Office of Scientific Research
Bolling AFB, Washington, DC
and
Idaho State University

December, 1992

PATTERN THEORY: EXTENSIONS AND APPLICATIONS

James S. Wolper
Assistant Professor
Department of Mathematics
Idaho State University

Abstract

Pattern Theory is an ongoing effort of the USAF Avionics Laboratory to develop an engineering theory of algorithm design. The primary tool is Decomposed Function Cardinality (DFC). Decomposed Function Cardinality is a measure of the complexity of a binary function $f\colon \{0,1\}^n \rightarrow \{0,1\}$; briefly, DFC($f$) is the number of bits that must be specified in a decomposition of f. DFC is a new basic technique in theoretical computer science.

DFC shows promise in many areas of theory and applications, but it is rather difficult to compute as $n \rightarrow \infty$. This report describes research into the use of PT in computer vision. It also describes research into more efficient methods for computing DFC. One method shows that DFC has polynomial growth with n in families of functions subject to recurrence relations. Another method uses an Error Correcting Code, developed for this effort, as a more efficient preprocessing step in estimating DFC. These codes are of independent interest in algebraic geometry and representation theory.

I. Introduction

Pattern Theory views a computation as a binary function of n variables, that is, as $f : \{0,1\}^n \rightarrow \{0,1\}$. The *cardinality* of such a function is the number of input–output pairs needed to describe it completely. Ordinarily, this cardinality is 2^n, that is, the function value must be specified for each of the 2^n possible inputs. But some functions have the property that they *decompose* into a composition of functions of fewer variables. For example, there might be vectors x_1, x_2, x_3 and x_4 of variables and appropriately–domained functions F, ϕ and g such that $f(x_1, x_2, x_3, x_4) = F(\phi(g(x_1, x_2), x_3), x_4)$. (This is only one of many possible forms for a composition.) In many cases, the combined cardinalities of the component functions (in the example, F, g, and ϕ) is less than the cardinality of f. The *Decomposed Function Cardinality* of f is the minimum of the cardinalities of its decompositions.

Decomposed Function Cardinality (DFC) is a measure of the complexity of a binary function $f: \{0,1\}^n \rightarrow \{0,1\}$; briefly, DFC($f$) is the number of bits that must be specified in a decomposition of f. This is similar to the notion of *Kolmogorov complexity*, which is (informally) the minimum length of a program to compute the function; DFC is more concrete. DFC is also of interest because exhaustive experimentation indicates that "real-world" functions tend to have *low* DFC, while random functions have *high* DFC. Thus, low DFC indicates that there is some kind of "pattern" in the function. For more details consult the Pattern Theory Technical Report [RNTG].

DFC shows promise in many areas of theory and applications (see, eg, [W1]), but it is rather difficult to compute as $n \rightarrow \infty$; even $n = 10$ is difficult. Thus we have embarked on a program to estimate DFC effectively.

This report describes two kinds of results on estimating DFC. These results are primarily of theoretical interest for the moment. However, lurking in the background is a potential application strategy, namely, computing or estimating DFC of a function by comparing it either to a small class of functions whose DFCs are known (at least asymptotically), or to the functions represented by codewords.

The report also discusses some new applications of PT to improve a previously-developed computer vision system. Other similar applications are under consideration by researchers at Wright Laboratories (see [BRNA]).

II. Results

This effort has produced three different types of result. First, more potential *applications* of Pattern Theory have been identified. However, applications depend on the ability to calculate (or estimate) DFC. Thus, there are results about *behavior of DFC in families* and results about *relations between DFC and coding theory*.

IIa. Applications

Most of the newly envisioned applications of PT are in computer vision. PT could be used to enhance the computer vision system described in [W2] (which was partially developed with support from a previous RIP grant) in several ways. First, as discussed in

[W1], PT can be used for segmentation (this is similar to the use of Kolmogorov complexity for segmentation described in [T]).

More significantly, the novel aspect of the system described in [W2] is the use of *entropy*, which is associated with information content, to choose features of interest (high entropy means low information). The features are responses from a network of Gabor filters responding at various scales and orientations. Entropy depends on the global distribution of filter responses and is therefore subject to sampling variability, especially with an unknown image. It is now proposed to use DFC instead of entropy to make this choice. DFC has the advantage of being intrinsic to the feature vector; in a sense, DFC searches for *intrinsic* interest, while entropy searches for *relative* interest. Note that DFC and entropy are inversely related, ie, entropy chooses the features which are unpatterned, while DFC chooses features which are patterned.

Here is an encoding scheme which would make calculation of such DFCs feasible. Suppose that the filter has d directions at each scale, where d is usually a power of 2. Filter responses are quantized as integers $0, 1, \ldots, M$. where $M \leq q$. Encode a response of i as a string of 2^i zeroes and 2^{M-i} ones. For example, if $d = 4$ and $M = 3$ then the feature vector $(7, 1, 3, 2)$ would be encoded as $0^{128}0^2 1^{126}0^8 1^{120}0^4 1^{124}$, where 0^n indicates a string of n consecutive 0s.

The DFC of the string of length 2^M consisting of 2^i zeroes followed by 2^{M-1} ones is estimated using the algorithm behind the Ada Function Decomposition program described in [RNTG]. The string represents a function of M variables. These variables are partitioned into two sets, the function is represented in tabular form (where the variables in one partition determine the row to which an entry is assigned, and those in the other determine the column), and the number of distinct columns is counted. In this case, the partition to use is $\{x_1\}$ (row variable) and $\{x_2, \ldots, x_M\}$ (column variables). The table has 2 rows, corresponding to function values where $x_1 = 0$ or where $x_1 = 1$. The first row consists of 2^i zeroes, followed by all ones; the second row is all ones. The column multiplicity is two. If $M = i$ then the columns are all the same, so the value of the function is the value of function is x_1, so it has DFC $= 2$. As i decreases the number of relevant variables increases, so the DFC is 2^{M-i+1}.

When we concatenate d of these words we introduce $\log_2(d)$ new variables to indicate which subword is relevant, and add the DFCs of the subwords. Thus

PROPOSITION. The DFC of a word of length 2^M consisting of 2^i zeroes followed by 2^{M-1} ones is $\leq \log_2(d) + \sum_j 2^{M-i_j+1}$.

If all of the i_j are distinct then the DFC tends toward the higher end of this inequality, but if there is a pattern then there are interactions among the variables which reduce DFC. However, if there is a pattern then the information content tends to be higher.

IIb. Families

The results of this effort concerning growth of DFC in families of functions are discussed in the paper "Polynomial Growth of DFC" which will be submitted for publication. A copy of that paper is enclosed as Appendix A.

IIc. Codes

The plan of research for this effort proposed to relate DFC to the mathematical problem of sphere packing. In this problem, one seeks to arrange spheres of equal radius in n-space in a manner that occupies as much of the space as possible. The sphere packing problem is very closely related to the construction of error correcting codes (ECCs), an area whose greatest advances have been made using algebraic geometry. The book [TV] is a detailed treatise on these subjects (a review of this book was written during the research period, and has been submitted to *SIAM Review*).

The original thought had been to derive a sphere packing from a binary function, and to derive results relating the efficiency of this packing to the function's DFC. After reading [TV], however, it seemed more sensible to relate the DFC to certain ECCs, which is an equivalent problem. However, none of the codes in the tables in [MS], a standard reference of ECCs, was sufficiently large. Thus an effort was made to construct a new family of large codes. The results of that effort are contained in the paper "Schubert Varieties and Linear Codes", which will be submitted for publication. This paper describes the construction of a new family of ECCs based on new results in pure mathematics. A copy of the paper is enclosed as Appendix B.

These results led to a new scheme for estimating DFC. The function is treated as a received word, and standard decoding algorithms are used to find the nearest codeword. DFC of the codewords will have been precomputed. The entry-by-entry difference between the function and the codeword should have a relatively small number of nonzero entries, so its DFC is easy to compute. By "Asymptotic Semi-Linearity" ([W1]), the DFC of the function is estimated by the DFC of the codeword plus the DFC of the difference. Similarly, since these are so-called linear codes in a vector space over the field with 2 elements, DFC need only be calculated for the basis vectors, and asymptotic semi-linearity can be used to estimate DFC for the remaining codewords.

Thus, the codewords become kinds of "signposts" (Ross's phrase) in the space of binary functions, enabling a more rapid estimation of DFC.

III. Conclusions

Pattern Theory (PT) continues to show promise as a theoretical and applied tool. Pattern Theory provides a uniform approach to problems in complexity theory, signal processing, image processing, data compression, and coding. DFC is a simple but powerful concept.

This effort has used Pattern Theory as a tool for traditional engineering applications, as a method for obtaining the kind of results on complexity associated with computer science, and the attempt to simplify its computability has inspired results of independent interest in coding theory and pure mathematics. General methods for estimating the growth of DFC were developed and explored. A new method for estimating DFC was developed and a substantial theoretical problem concerning its implementation was solved. Potential applications of PT to computer vision were explored.

Many avenues remain to be explored, however. The FERD (Function Estimation by

Recomposing Decompositions) effort described in [RNTG] shows the use of PT in a neural–network like situation, although with vastly superior performance for some problems. To this author's knowledge, no other effort has been made to compare PT with other methods of parallel computation. Nor has any effort been made to use parallel computation to calculate DFCs, although the problem has a natural embedding onto a hypercube architecture such as that of the Connection Machine (TM). As these machines become more widely available, PT should expand to take advantage of their capabalities and to *add* to these capabilities.

REFERENCES

[BRNA] Breen, Michael, T. Ross, M. Noviskey, and M. Axtell, "Pattern Theoretic Image Restoration", to be presented at SPIE Symposium on Electronic Imaging, Jan 1993.

[MS] MacWilliams, F. J., and N. J. A. Sloane. *The theory of error–correcting codes*. North–Holland Mathematical Library, v. 16 (1978).

[RNTG] Ross, Timothy, M. Noviskey, D. Gadd, and T. Taylor. "Pattern Theory: An Engineering Paradigm for Algorithm Design", WL-TR-91-1060.

[T] Telgarsky, R., "Target Extraction via String Matching", presented at University of Colorado–Colorado Springs Conference on Computer Vision, May, 1992.

[TV] Tsfasman, M. A., and S. G. Vlăduţ. *Algebraic–Geometric codes*. Mathematics and its Applications, Soviet Series, 58. Kluwer Academic Publishers, Dordrecht, 1991.

[W1] Wolper, James, "Aspects of Pattern Theory", Final Report, 1992 Summer Faculty Research Project.

[W2] Wolper, James, "A Gabor–transform based system for Image Recognition", presented at University of Colorado–Colorado Springs Conference on Computer Vision, May, 1992.

Appendix A

Polynomial growth of DFC

Jim Wolper
Idaho State University
wolperj@howland.isu.edu

11 March 1992

Research partially supported by AFOSR under the RIP program

I. Introduction

Decomposed Function Cardinality (DFC) is a measure of the complexity of a binary function $f\colon \{0,1\}^n \to \{0,1\}$; briefly, DFC($f$) is the number of bits that must be specified in a decomposition of f. For more details consult the Pattern Theory Technical Report [RNTG].

II. Recursion and DFC

The key to estimating DFC asymptotically is recursion (or, more generally, recurrence). Informally, we try to compute DFC(f_0) using a "divide–and–conquer" strategy. To be precise, suppose we are given a collection $\mathbf{C}$ of $n+1$ functions of a single variable, ie, $\mathbf{C} = \{f_0, f_1, f_2, \ldots, f_n\}$. (Generally the f_i will be DFC(g_i) for some function of interest g_i; the variable for the f_i is the number of variables for the g_i.) A *recurrence* in $\mathbf{C}$ is a function F of n variables such that $f_0(m) = F(f_1(m_1), f_2(m_2), \ldots, f_n(m_n))$ with all $m_j < m$.

For convenience, define DFC$_n(f)$ to be the DFC when f has n variables.

We do not exclude the possibility that f_0 is one of the f_i; but it is important to note that if f_0 does appear as an argument to F then it has fewer variables. This means that one can use induction to determine a bound for DFC(f_0).

The following lemma is an obvious consequence of the definition of DFC. It generalizes "asymptotic semilinearity of DFC" [W].

LEMMA. Suppose F is a recurrence in $\mathbf{C}$, ie, that

$$f_0(m) = F(f_1(m_1), f_2(m_2), \ldots, f_n(m_n))$$

with all $m_j < m$, and suppose that $f_0 \neq f_i$ for $i > 0$. Then

$$\mathrm{DFC}(f_0) \leq \mathrm{DFC}(F) + \sum_{i=1}^{n} \mathrm{DFC}(f_i).$$

QED

Recall that a function $f(x)$ is $O(g(x))$ is there exists a constant C such that $|f(x)| \leq C|g(x)|$ for $x \gg 0$.

COROLLARY. If F is a recurrence for f_0 then

$$\mathrm{DFC}(f_0) = O((\mathrm{DFC}(F)) \cdot \max\{\mathrm{DFC}(f_1), \mathrm{DFC}(f_2), \ldots, \mathrm{DFC}(f_n)\}).$$

Proof. This follows from standard facts about O growth. Notice that if f_0 is one of the f_i then the DFC which appears is based on a smaller n. **QED**

THEOREM. If F is a recurrence for f_0 and if F and f_i ($i > 0$) have polynomial growth then so does f_0.

Proof. If f_0 does not appear in the recurrence F then the result is obvious. If not, then we must use induction on the number of variables in f_0. So suppose that $\mathrm{DFC}_m(f_0) = O(m^k)$ for $m < n$. Then the contribution of the m–variable f_0 to the "max" in the corollary is less than n^k. However, by the corollary, it is the max which determines the final growth rate. Hence the growth is polynomial. **QED**

III. Using recursion and recurrence

The results in section II indicate the following strategy for estimating DFC of some function f: namely, try to find a recurrence for f using functions of known asymptotic DFC. this section illustrates several examples of the strategy. Part of the current PT effort is to build up a library of functions with known asymptotic DFC.

One class of such results was obtained during the 1992 Summer Faculty Research Project at Wright Labs [W]. Consider the predicate $E_k(w)$, which is true if the binary string w has exactly k ones. This can be, in turn, constructed from the predicate $F_k(W)$, which is true if w contains k or more ones, since $E_k(w) = F_k(w)$ AND NOT $F_{k+1}(w)$. The predicate F_k can be computed from *MAJGATE* (ie, the predicate which is true when more than half of the inputs are one) by adding enough ones to the input; if $r > n - 2k$ then $k + r > \frac{n+r}{2}$, so r ones suffice.

Clearly if F_k has polynomial DFC then so does E_k, since E_k can be decomposed into two F_ks with some $O(1)$ overhead. Similarly the overhead in constructing F_k from *MAJGATE* is $O(n)$, so the latter has polynomial DFC if the former does.

There is a recurrence for *MAJGATE* involving F_ks and some small overhead. Group the input in 3s; from each group of 3 there are two signals. "2 or more ones" or "1 or 3 ones". The DFC so far is $16\frac{n}{3}$. The "2 or more ones" form the input to a function A which is $F_{n/4}$; by induction A has polynomial DFC. These also form the input to the function $B = F_{n/6}$, which also has polynomial DFC by induction. Similarly, the "1 or 3 ones" form the input to a function C which is *MAJGATE* for $\frac{n}{3}$, which, again by induction, has polynomial DFC. B and C are inputs to an AND gate ($O(1)$) whose output is D. with $O(1)$ effect on DFC. D and A are ORed, with $O(1)$ effect again.

Notice that A is 1 when $\frac{n}{4}$ of the groups have at least 2 ones, ie, when there are at least $\frac{n}{2}$ ones. B is 1 when $\frac{n}{6}$ of the groups have at least 2. C is one when $\frac{n}{6}$ have at least one 1. D is (B AND C); NOT D is (NOT B OR NOT C). If NOT B. then fewer than $n/6$ of the 3 variable packets have 2 or more ones, which means that it is impossible to have $n/2$ ones. If NOT C, then fewer than $n/6$ of the 3 variable packets have 3 ones, so again it is impossible to have $n/2$ ones. Thus either D or A guarantees that there are at least $n/2$ ones. We have proven

THEOREM. The predicates $E_k =$"exactly k ones", $F_k =$ "k or more 1s". and MAJGATE have polynomial growth of DFC. **QED**

Notice that the proof above depended on a genuine recurrence relation for the functions, not just a recursion.

Another class of functions with polynomial growth of DFC are the *interval acceptors*. An interval acceptor f has the property that its value is 1 if the input string has the form $0^a 1^b 0^{n-a-b}$ for some $a > 0$. A recurrence among these functions can be built up with $O(1)$ overhead; thus, there is an easy inductive proof that DFC has polynomial growth. The recurrence is constructed as follows: suppose the input string is $x_1, x_2, \ldots, x_n$. Then consider the following table of values of x_1, x_2, and $f(x_2, \ldots, x_n)$.

x_1	x_2	$f(x_2, \ldots, x_n)$	$f(x_1, \ldots, x_n)$	remarks
0	0	0	0	
0	0	1	1	$00^a 1^b 0^{n-a-b-1}$
0	1	0	0	
0	1	1	1	$01^b 0^{n-1-b}$
1	0	0	0	
1	0	1	?	$10^a 1^b 0^{n-1-a-b}$ OR 10^{n-1}
1	1	0	0	
1	1	1	1	$11^b 0^{n-1-b}$

Note that when $f(x_2, \ldots, x_n) = 0$ then $f(x_1, \ldots, x_n) = 0$ as well.

The case $(1, 0, \ldots)$ is ambiguous when the $n-1$ variable predicate is 1. This is resolved by considering the value of x_3, at $O(1)$ cost. Thus

THEOREM. DFC for interval acceptor functions has polynomial growth. **QED**

Note that this Theorem fits the "Pattern Theory Philosophy" quite well. This philosophy (as expressed in the Technical Report [RNTG]) is that random functions generally have high DFC, while 'naturally occuring' functions generally have low DFC. For each a and b, the probability of an interval $0^a 1^b 0^{n-a-b}$ is

$$\frac{1}{2^a}\frac{1}{2^b}\frac{1}{2^{n-a-b}} = \frac{1}{2^n}.$$

Thus, the probability that a string is an interval is $O(\frac{n^2}{2^n})$.

Appendix B.

Linear codes and Schubert Varieties

James S. Wolper
Department of Mathematics
Idaho State University
Pocatello, ID 83209
wolperj@howland.isu.edu

10 August 1992

Abstract: A family of linear "Algebraic–Geometric" codes is constructed from line bundles on Schubert varieties and analyzed using techniques from Representation Theory.

AMS(MOS) 1991 Subject Classification (Primary) 94B27 (Secondary) 14M99. 20G10.

Research partially supported by the AFOSR.

I. Introduction.

Recent work in the constructions of Error–Correcting Codes has exploited techniques from algebraic geometry, especially from the geometry of varieties over a finite field; [TV] contains a complete description of this work. Most of the codes constructed from varieties are of the *Reed–Muller* or *Goppa* type. In these, one starts with a variety X over $GF(q)$, the field with q elements, and chooses a subset $P \subset X$ and a line bundle (locally free sheaf of rank 1) $L \to X$. There is a linear transformation $H^0(X, L) \to V$, where V is a $GF(q)$ vector space of dimension $|P|$, essentially defined by "evaluating" all of the sections in $H^0(X, L)$ (Čech cohomology) at points of P. The codewords are the image of $H^0(X. L)$ under this map. (The value of a section is not well–defined, so a choice must be made in evaluation.)

Most of the codes so constructed have used a curve X; this is natural since so much is known about the Picard group (= group of line bundles) of a curve. Comparatively little work has been done with varieties of higher dimension, notable examples being the work of Chakravarti [C] and Ryan and Ryan [RR]. In the current work, a family of codes is constructed from Schubert subvarieties of flag manifolds G/B, where G is a semisimple algebraic group defined over $GF(2)$ (ie, a Chevalley group) and B is a Borel subgroup. Line bundles over G/B correspond to irreducible representations of G. Each such a representation can, by the Bott–Borel–Weil Theorem ([D]), be constructed as $H^0(G/B. L_\chi)$ where L_χ is the line bundle associated to some character χ of B. The dimension of $H^0(G/B. L_\chi)$ (ie, the degree of the representation) can be determined from the Weyl Character Formula, and many of the properties of these codes can be deduced from the combinatorial machinery of representation theory.

Here is an outline of the rest of this paper. Section II contains a brief résumé of results about algebraic groups: all of this material is standard. Section III discusses Schubert varieties and is also standard. In section IV, the codes are constructed, and various parameters are estimated. It works out that the most interesting codes are constructed from fundamental representations of $SL(n)$, and codes so constructed are discussed in section V. The final section describes some unanswered questions.

This work was partially inspired by the necessity of having linear codes in vector spaces of large dimension. Such codes are part of a scheme to make certain calculations in computational complexity feasible. The calculations involve a complexity measure called Decomposed Function Cardinality.

II. Algebraic Groups.

In this section, k is a field of any characteristic. The material here can be found in [Bo].

A *Linear Algebraic Group G* over k is a (reduced, irreducible) variety over k which is also a group; the group operations are k–morphisms; and there exists an embedding $G \to GL(n, k)$ for some n. A torus T is a linear algebraic group which is isomorphic to a product of $GL(1,k)$s. Suppose T is a *maximal* torus in G (such T exist for dimension reasons). A subgroup $B \subset G$ is a *Borel* subgroup if it is maximal among the connected, solvable subgroups. It is a fundamental fact that all Borel subgroups are conjugate. The quotient space G/B is the *flag variety*. It is smooth. The *Weyl group* is $W = N(T)/T$, where $N(T)$ is the normalizer of T in G. If G is semisimple. W is finite.

When $G = SL(n,k)$ or $GL(n,k)$. take T to be the subgroup of diagonal matrices. and B to be the subgroup of upper triangular matrices. The Weyl group is isomorphic to the symmetric group S_n of permutations of $\{1,\ldots,n\}$. Here is an explanation for the name: let $G = SL(n,k)$, and let V be a vector space of dimension n over k. Then a *flag* in V is a sequence of subspaces $V_1 \subset V_2 \subset \ldots \subset V_n = V$, where $\dim(V_i) = i$. G acts transitively on the space of all flags in V with isotropy B, so G/B can be identified with the space of flags.

The representation theory of G is intricate and similar to the representation theory of Lie groups; it was worked out by Chevalley and others ([Ch]). The book [H] is an excellent introduction. It is easiest to describe in terms of the Lie algebra $\underline{g}$ of G. which is, as a vector space, the tangent space to G at the identity element; the exponential mapping exp:$\underline{g} \to G$ enables one to go back-and-forth between them. Let $\underline{h}$ be a Cartan subalgebra of $\underline{g}$, that is, the Lie algebra of a maximal torus. Let $\rho : \underline{g} \to GL(n,k)$ be a representation on a vector space V of dimension n. A *weight* for ρ is function $\alpha : \underline{h} \to k$ such that $\{v \in V : \rho(h).v = \alpha(h).v \text{ for all } h \in \underline{h}\}$ is non–empty. It is a theorem that every irreducible representation of $\underline{g}$ (and hence of G) has a unique "highest weight" (highest with respect to a well-defined order on the space of characters). Similarly, given a so–called dominant, integral character of T, its derivative is a weight and is, in fact, the highest weight of some representation of $\underline{g}$. The dimension of such a representation can be determined by the *Weyl character formula*; rather than discuss this in its full generality. we will discuss it in an *ad hoc* manner as needed.

Each $\underline{g}$ comes equipped with a representation on $\underline{g}$, called the *adjoint* representation. There is a non–commutative product, the *bracket* $[a,b]$ on $\underline{g}$, and the adjoint representation is defined by $\mathrm{ad}(x)(y) = [x,y]$. A weight of the adjoint representation is called a *root*. The classification of G is based on Dynkin diagrams constructed from the roots: a node of the diagram corresponds to a "simple"root. and nodes are connected by edges (with multiplicity) depending on the angle (with respect to an appropriate inner product) between the corresponding subspaces in $\underline{h} \otimes \mathbf{R}$. Any dominant integral weight is a sum of roots with positive integer coefficients.

The most important example, both in the general theory and in this work. is the case of $SL(n)$. Then T and B are defined as above, and G/B is the space of full flags. The Lie algebra $\underline{g}$ is the set of $n \times n$ matrices with the bracket $[x,y] = xy - yx$. and $\underline{h}$ is the subalgebra of diagonal matrices. Define $\epsilon_i : \underline{h} \to k$ as the homomorphism taking an element of $\underline{h}$ to its i^{th} diagonal entry. The roots $\alpha_i = \epsilon_i - \epsilon_{i+1}$ form a basis for the root system. The i^{th} *fundamental representation* has highest weight $\epsilon_1 + \ldots + \epsilon_i$. Its exponential takes a diagonal matrix in T to the product of its first i entries (which is nonzero). Such representations are realized on $k^n \wedge \ldots \wedge k^n$ (i factors); the dimension of the representation space is then $\binom{n}{i}$, as can also be determined from the Weyl Character Formula. A weight (for T) is dominant if it has the form $t_1^{m_1}t_2^{m_2}\dots t_n^{m_n}$ where t_i is the i^{th} diagonal entry of T and the m_i are integers with $m_1 > m_2 > \dots > m_n$.

In the general case, every character χ of T is a dominant weight, and thus corresponds to a unique (up to isomorphism) representation with highest weight χ; it is a fact that this representation extends to B. There are several ways to construct such representations: we will construct them as the vector space of sections of a certain line bundle on G/B. Define an equivalence relation $\equiv$ on the product $G/B \times k$ by $(xbB, k) \equiv (gB, \chi(b)k)$, where $b \in B$. The quotient is a line bundle L_χ over G/B. The cohomology $H^0(G/B, L_\chi)$ is represented by functions $f : G \rightarrow k$ satisfying $f(xb) = \chi(b)^{-1}f(x)$. G acts on this cohomology group, thus defining arepresentation. (Strictly speaking, when using cohomology, one should consider G defined over the algebraic closure of k, and then restrict to the fixed points of the Frobenius mapping.)

The irreducible representations of G are given the structure of a semigroup under the *Young product*; this semigroup is generated by the fundamental representations. The Young product of the irreducible representations with highest weights χ_1 and χ_2 has highest weight $\chi_1\chi_2$, and the representation space is realized by taking the products of sections in $H^0(G/B, L_{\chi_1})$ and $H^0(G/B, L_{\chi_2})$

A basis for $H^0(G/B, L_{\omega_i})$, where ω_i is the character of the i^{th} fundamental representation of $SL(n)$, is constructed as follows. First, there is a "distinguished" section f_0.

PROPOSITION. Define $f_0 : G/B \rightarrow k$ by $f_0(g) :=$ upper left $i \times i$ minor of g. Then f_0 is a section of ω_i.

PROOF. A simple computation. **QED**

PROPOSITION. Suppose w is an element of the Weyl group of G and f is in $H^0(G/B, L_\chi)$. Define $f_w(g) := f(wg)$. Then f_w is in $H^0(G/B, L_\chi)$.

PROOF. Another straightforward computation. **QED**

PROPOSITION. The functions f_w constructed from f_0 above span $H^0(G/B, L_{\omega_i})$ when $G = SL(n)$.

PROOF. The Weyl group acts on sections in the indicated manner. The isotropy of f_0 is the direct product of the subgroup that fixes $\{i+1, \dots, n\}$ and the subgroup that fixes $\{1, \dots, i\}$. This has $i!(n-i)!$ elements, so the size of the orbit is $\frac{n!}{i!(n-i)!}$, which is the dimension of $H^0(G/B, L_{\omega_i})$. The distinct f_w are linearly independent by inspection. **QED**

III. Schubert varieties.

There is a decomposition of G into disjoint double cosets BwB, where $w \in W$: this is called the Bruhat decomposition, and it naturally leads to a decomposition of G/B into a disjoint union of affine spaces. Let C_w denote the image of BwB in G/B, and let X_w be its closure. X_w is called a *Schubert variety*. It is a fundamental fact that X_w is a disjoint union of C_y, $y \in W$. This defines a partial order (the *Bruhat order*) on the set of Schubert varieties in G/B (and hence on the Weyl group) by $X_1 \leq X_2 \iff X_1 \subset \overline{X}_2$. Proctor ([P]) has defined isomorphic partial orders on integer sequences.

In the example of $SL(n)$, the Weyl group is isomorphic to S_n, and the integer sequences are permutations of $\{1, \ldots, n\}$. The order is generated by the "moves" of exchanging s_i and s_j when $s_i < s_j$ and $i < j$. The dimension of a Schubert cell is the length of the associated permutation.

It is easy to count the number of points in a Schubert variety X_w when k is a finite field, because X_w is a disjoint union of affine spaces. Thus, a cell of dimension i has $|k|^i$ points. To find the cardinality of a variety, add up the cardinalities of its constituent cells.

A line bundle L over G/B restricts to a line bundle over a Schubert variety $X_w \subset G/B$. This is intuitive when X_w is non–singular, but many Schubert varieties are singular, and the theory becomes less intuitive (recall that a line bundle in this context is really a locally free sheaf of rank 1). The paper [W] discusses a combinatorial algorithm for deciding if X_w is singular when $G = SL(n)$.

IV. Codes from Schubert varieties.

Assume from now on that k is a finite field. Choose an algebraic group G over k, a Borel subgroup B, and character χ of a maximal torus $T \subset B$. The characer χ extends to a character of B. Then we have a line bundle L_χ over G/B. Let m =dim $(H^0(G/B, L_\chi))$, and let $f_1, \ldots, f_m$ be a basis for $H^0(G/B, L_\chi)$. Choose a Schubert variety X_w and let $N = |X_w|$. Order the points of X_w in some arbitrary manner $x_1, \ldots, x_N$, and choose arbitrary representatives for the x_i (recall that the x_i are cosets in G/B). Then define the linear code $C(k, G, B, \chi, w)$ as follows.

DEFINITION. Let V be a k–vector space of dimension N, and define a linear transformation $C : H^0(G/B, L_\chi) \to V$ by $C(f) = (f(x_1), \ldots, f(x_N))$. The code is the image of C.

This is an $[N, \leq m, d]$ code over k, where d must be determined. If the map C is *injective* then the code has parameters $[N, m, d]$. By [Ch, Exposé 15], the i^{th} fundamental representation has a section whose zero–locus is a certain codimension one Schubert variety X_w. (This is illustrated in the example of Section V.) The map $C(k, G, B, \chi, w)$ then fails to be injective. Also, if $y < w$ in the Bruhat order, the map $C(k, G, B, \chi, y)$ fails to be injective. However, this particular failure of injectivity has no effect on the asymptotic results below.

PROPOSITION. If $\dim(X_w) = r$ then $N = O(2^r)$.

PROOF. The difference $X_w - C_w$ is a union of cells of dimension smaller than r, and X_w has a unique cell of dimension r. **QED**

Recall that the *rate* of an $[N, m, d]$ code is m/N.

PROPOSITION. There exists a family of codes $C(k, G, B, \chi, w)$ whose rate is asymptotically 1.

PROOF. Let $G = SL(n, k)$, B be the subgroup of upper triangular matrices, and let χ be the character of the i^{th} fundamental representation. Then $m = \binom{n}{i} = O(n^i)$. Choose any w with length $O(i)$. **QED**

Of course, by choosing smaller X_w, the rate can be lowered.

Estimating d, the minimum weight of the code, is more complicated. By definition, the weight of the codeword corresponding to the image of a section f is $N - |\{x_i : f(x_i) = 0\}|$.

PROPOSITION. Let ω_i be the minimal weight of the code $C(k, G, B, \chi_i, w)$, $i = 1, 2$, and let ω_{12} be the minimal weight of the code constructed from the Young product of χ_1 and χ_2. Then $\omega_{12} \leq \min(\omega_1, \omega_2)$.

PROOF. The zero-locus of a product of two sections contains the zero-loci of the factors. **QED**

Thus, we focus on the fundamental representations. First, consider the i^{th} fundamental representation of $SL(n)$. Recall that a basis for $H^0(G/B, \chi)$ is given by the determinant of the upper-left $i \times i$ minor of row permutations of g, the argument. If i is small, it is more likely that the minor has determinant zero; conversely, if i is large, it is less likely that the minor has determinant zero. Thus, the codes with the highest minimal weight correspond to larger values of i. When $i = n$ the determinant is non-zero, and the line bundle is trivial.

Recall that the *relative minimum distance* of an $[N, m, d]$ code is d/N. The argument above (with the example of $G = SL(n, k)$) shows:

PROPOSITION. There exists a family of codes $C(k, G, B, \chi, w)$ whose relative minimum distance is asymptotically 1. **QED**

V. Examples.

In this section we work out examples of codes constructed from a fundamental representation of $SL(3, k)$ where $k = GF(2)$. These codes exceed the Gilbert Varshamov bound. The constructions generalize quite easily.

The Weyl group for $SL(3, k)$ is the symmetric group S_3, which has 6 elements; hence, there are 6 Schubert cells. These will be represented by the permutations of abc. Here is a set of coset representatives; these can be interpreted as flags by taking V_i to be the span of columns $1, \dots, i$.

$$acb\text{: } \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\}$$

$$acb\text{: } \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \right\}$$

$$bac\text{: } \left\{ \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\}$$

$$bca\text{: } \left\{ \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \right\}$$

$$cab\text{: } \left\{ \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \right\}$$

$$cba\text{: } \left\{ \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$
$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \right\}$$

Let ω_2 be the second fundamental representation. Then $H^0(G/B, L_{\omega_2})$ has dimension $\binom{3}{2} = 3$, and a basis is given by $f_0, f_{(23)}, f_{(13)}$, where f_0 is given by the upper left 2×2 determinant of the matrix, and f_w is defined in section II.

Consider $X = bca$, which has 9 points. The code $C(k, SL(3), B, \omega_2, bca)$ is a $[9, 3, 4]_2$ code, as can be determined by direct computation. This is illustrated below, where the points of the Schubert variety X_w are named $w_1, w_2, \dots$ based on the order above. Thus, eg, bca_2 is $\begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

x_i	$f_0(x_i)$	$f_{(23)}(x_i)$	$f_{(13)}(x_i)$
abc_1	1	0	0
acb_1	0	1	0
acb_2	1	1	0

bac_1	1	0	0
bac_2	1	0	0
bca_1	0	0	1
bca_2	0	1	1
bca_3	1	0	1
bca_4	1	1	1

Also, notice the values on cab, where $f_{(13)}(x_i)$ vanishes:

x_i	$f_0(x_i)$	$f_{(23)}(x_i)$	$f_{(13)}(x_i)$
cab_1	1	1	0
cab_2	0	1	0
cab_3	0	1	0
cab_4	1	1	0

Thus, $C(k, SL(3), B, \omega_2, cab)$ is a $[9, 2, 6]_2$ code

VI. Further questions.

The example in V and its generalizations are amenable to computer analysis. The asymptotic results of IV indicate that much higher rate codes can be constructed. Another way to increase the rate is to use a Young product of characters; however, this tends to lead to a lower minimum distance.

Since the codes constructed here are linear codes, standard decoding algorithms apply. In particular, Reed–Muller codes are easy to decode in hardware; see [MS].

The construction here can be generalized as follows: let P be a parabolic subgroup, that is, a subgroup containing a Borel subgroup. Then a representation of G with character χ determines a vector bundle over G/P. By the Bott–Borel–Weil theorem, exactly one of the cohomology groups $H^i(G/P, \chi)$ is non-zero, and G acts on this group. It is not clear whether better codes can be constructed in this manner.

REFERENCES

[Bo] Borel, Armand, *Linear Algebraic Groups.* Graduate Texts in Mathematics, Number 126. NY: Springer–Verlag.

[C] Chakravarti, M., "The generalized Goppa codes and related designs from Hermitian surfaces ...", in *Coding Theory and Applications*, Lecture Notes in Computer Science, volume 311. NY: Springer–Verlag.

[Ch] Chevalley, Claude, *Seminaire sur la classification des groupes de Lie algébriques.* (Mimeographed notes) Paris (1956 – 8).

[D] Demazure, Michel, "Une démonstration algébrique d'un théorème de Bott". *inventiones mathematicae* **5**(1968), 349 – 356.

[H] Humphreys, James E., *Introduction to Lie algebras and representation theory*. Graduate Texts in Mathematics, Number 9. NY: Springer–Verlag.

[MS] MacWilliams. F. J., and N. J. A. Sloane, *The theory of error–correcting codes*. North-Holland Mathematical Library, v. 16 (1978).

[P] Proctor. R., "Classical Bruhat orders and lexicographic shellability", *Journal of Algebra* **77** (1982), 104 – 126.

[RR] Ryan, Charles T. and Kevin M. Ryan, "An application of geometry to the calculation of weight enumerators", *Congr. Numer.* **67**(1988), 77–89.

[TV] Tsfasman, M. A., and S. G. Vlăduţ. *Algebraic–Geometric codes*. Mathematics and its Applications, Soviet Series, 58. Kluwer Academic Publishers, Dordrecht, 1991.

[W] Wolper. J., "A combinatorial approach to the singularities of Schubert varieties", *Advances in Math.* **76** (1989) 184 – 193.

THE IDENTIFICATION OF NONLINEARITY IN STRUCTURAL SYSTEMS

Lawrence D. Zavodney
Assistant Professor
Department of Engineering Mechanics
Ohio State University
155 W. Woodruff Ave.
Columbus, OH 43210

currently
Associate Professor and Chairman
Department of Engineering
Cedarville College
P.O. Box 601
Cedarville, OH 45314-0601

Final Report for:
Research Initiation Program
Wright Laboratory

Sponsored by:
Air Force Office of Scientific Research
Wright Patterson Air Force Base, Ohio

and

Ohio State University

December 1992

THE IDENTIFICATION OF NONLINEARITY IN STRUCTURAL SYSTEMS

Lawrence D. Zavodney*
Assistant Professor
Department of Engineering Mechanics
Ohio State University

Abstract

The identification of cubic nonlinearity in structural systems exhibiting the jump phenomenon was studied analytically and experimentally. An axially-tensioned clamped-clampd (ATCC) elastic steel beam (27.25" long x 0.502" wide x 0.0156" thick) was analyzed; the governing partial differential equation containing the leading nonlinear term caused by mid-plane stretching was used as the mathematical model. This equation was discretized by Galerkin's Method. Several shape functions were used to approximate the mode shape; they included the linear string, a linear clamped-clamped (CC) beam, and a linear ATCC beam. The temporal equation governing the modulations of the shape function was solved by the Method of Multple Scales; two different detuning parameters were used. The identification technique utilized information obtained from the jump-down points during a frequency sweep up and, by matching it to that predicted by the perturbation solution, an estimate for the cubic coefficient was obtained. The results of the various analytical methods were compared to experiments performed on the ATCC beam mounted on an 1100-lb force electrodynamic shaker. It was shown that the technique works very well (less than 3% error). It was also shown that using the easily obtained string shape function as a shape function for the ATCC beam gave extremely good results (i.e., less than 3% error) whereas when the CC beam solution was used, it resulted in a 24% error. The technique is easy to use and will lend itself to implementation on a computer or in a routine modal analysis when cubic nonlinear terms are present.

*Currently Associate Professor and Chairman, Department of Engineering, Cedarville College, Cedarville, Ohio 45314.

The Identification of Nonlinearity in Structural Systems

Dr. Lawrence D. Zavodney

1. Introduction

An important branch of modern engineering is the analysis and prediction of the dynamic behaivior of structures and systems. When structures move back and forth about an equilibrium postition they are vibrating. Aerospace structures are especially susceptible to vibration because they must be lightweight. The most important aspect of vibration is the phenomenon of resonance, which is manifested by very large motions at certain frequencies. Large vibration can render a structure useless for its intended purpose or cause catastrophic failure.

Linear systems exhibit a resonance for each natural frequency of the system. Nonlinear systems can exhibit all of the resonances in linear systems plus additional resonances and behaviors that are unique to nonlinear systems. In addition to the conventional external resonances, systems may also be excited parametrically--which may cause additional (typically subharmonic) resonances. Hence, using a mathematical model that contains the approproate nonlinear and parametric terms is essential if the correct behavior is to be predicted. If predicting resonance is of concern, then the dominant nonlinear terms must be identified both qualitatively and quantitatively.

For this reason, the problem of understanding nonlinear systems and identifying dynamic systems has been receiving increasing attention because of the importance being given to the accurate prediction of the response of structures to various loading environments [1-8]. The problem of identifying nonlinear systems is further compounded because commercial modal analyzers on the market today (and the software that supports them) are capable of identifying only linear systems [9]. In some cases, the system may behave linearly at some frequencies, and nonlinearly at other frequencies [10-12].

All structures exhibit nonlinear dynamic behavior to some degree; for some it is quite small and, for these cases, linear models are adequate. For other structures the nonlinear behavior is appreciable and, in some cases, the nonlinear behavior dominates the dynamic response. Nonlinear behavior may be due to material properties (nonlinear constitutive law), geometric asymmetry, midplane strecthing (especially during large amplitude vibration), nonlinear damping, or any combination of these and other sources. In a majority of these cases, the nonlinearity causes a deviation from linear behavior, but for others, it introduces new and unique phenomena that have no counterparts in linear theory [1-3]; some examples include subharmonic, superharmonic, combination, and internal resonances, jump phenomenon, saturation phenomenon, self-excited oscillations, bifurcations, and nonexistence of periodic oscillations (chaos). An example of a common nonlinear behavior is shown in Figure 1.1. This figure shows the frequency response of a clamped-clamped flat plate subject to swept-sine excitation. The frequency sweep up and down reveals a typical hardening nonlinearity giving rise to the nonlinear jump phenomenon.

In general, a linear mathematical model used to describe the response of a structure is the least precise model because it can always be improved by including nonlinear terms to do three things: (1) to increase the accuracy of the predicted response, (2) to extend the useful range of a solution, e.g., for larger displacements, and (3) to explain or predict new phenomena that have no counterparts in linear theory.

Current research on the identificatifon of nonlinear systems is focusing largely on the nonlinearity that is characterized by small but finite deviation from linear behavior. Hence, there is a need to develop techniques that are capable of identifying appropriate nonlinear terms in systems that cannot be characterized by small deviation from linear behavior. Methods that presuppose linearity cannot identify behavior unique to nonlinear systems. Furthermore, the identification scheme should also be able to quantitatively estimate these terms. Because most nonlinear terms are quite small when compared to the linear terms, they often pose problems when it comes to obtaining good measurements. When brute force techniques are used to estimate all system parameters simultaneoulsy, it is possible to lose the nonlinear effects in the noise. In some cases the smallness of the nonlinear terms is somewhat deceiving because their effect on the response can be quite pronounced.

The objective of this research effort was to develop an easy-to-use method that would identify quantitatively one of the most common nonlinear behaviors--the jump phenomon caused by a cubic nonlinearity. One of the goals of the project was to implement the proceedure experimentally to determine its feasability in the laboratory or in the field.

1.1 Discussion of the Method

The basic idea motivating the research is that nonlinear systems sometimes exhibit unique behavior. When it does, it is necessary to modify the mathematical model to account for the nonlinear behavior. If, for example, the system exhibits a jump when the frequency is swept up or down, then a quadratic or cubic term must be included in the restoring force. One way to determine the form of the nonlinear term is to examine the nature of the harmonic content. If the system exhibits a third harmonic in the forced response, then one would conclude that a cubic term is present; likewise, if a second harmonic is present in the response, then one would conclude that a quadratic term is present. Typically, if a quadratic or cubic term is present, some distortion to the waveform will be present, but it may be difficult to quantify it. However, if the amplitude of motion was increased during a sinusoidal sweep up and down, it may be possible to cause a jump to occur. If a jump occurs, and one can measure the amplitude and frequency at which the jump occurs, then it is possible, in theory, to choose a coefficient that would cause a jump at the same frequency and with the same amplitude. In other words, if the system to be identified already exhibits a jump, we use the bifurcation information (jump down) to identify the cubic term. If the system does not exhibit a jump, then we increase the amplitude of the excitation to cause a jump in the system response for the express purpose of identification. In other words, we accentuate the qualitative nonlinear behavior for the purpose of identifying quantitatively the nonlinearity.

The basic solution of the Duffing Equation has been known for some. time now; recently newer perturbation solutions have appeared [15,16]. The basic feature of the cubic nonlinearity is the backbone curve; for linear systems it is perfectly vertical whereas for nonlinear systems it is bent--to the right for hardening systems and to the left for softening systems. In either case, the response follows the backbone curve. The jump in the response occurs when a vertical tangent occurs in the frequency response curve. An implicit assumption in this method is that the jump point is very close to the backbone curve. Hence, the method essentially fits jump-point data to a suitable backbone curve--which uniquely defines the coefficient of the nonlinear term. What remains is to determine the best method of developing the backbone curve.

The first method, called "Method I" uses Multiple Scales with a detuning parameter that measures the deviation from the linear natural frequency [15]. From the analysis, one can relate the jump points and frequency at which the jumps occur to the coefficient of the cubic term. By generating several jump points, a curve fit to the data will yield a line whose slope is directly related to the cubic coefficient. Method II also uses Multiple Scales but uses a different detuning parameter and yields results that, although are more involved and complicated, more accurately predict the jump points [16]. Details are given in the next chapter.

2. Derivation of the Non-Linear Beam Equation

In this chapter we summarize the derivation of the beam equation. By assuming planar transverse motion with no coupling to the out-of-plane and torsional modes, the transverse displacement of a beam element is given by $v(x,t)$. By further restricting the analysis to a clamped-clamped beam, we can assume the slopes (i.e. the rotation of a beam element) to be everywhere small. Euler-Bernoulli beam theory gives moment-curvature relationship (using a small slope approximation) as

$$M = EIv_{xx} \qquad (2.1)$$

where E is the modulus of elasticity and I is the area moment of inertia about the longitudinal axis. The subscripts following $v(x,t)$ represent partial derivatives with respect to the listed variables. Cutting the beam at s, as shown in Figure 2.1, and considering the contributions to the moment from the remainder of the beam to the right, we have

$$M = M_1 + M_2, \tag{2.2}$$

where M_1 is the contribution due to the transverse inertial force, and M_2 is the contribution due to the reaction forces at the end of the beam. The contributions due to inertia are given by

$$M_1 = -\int_s^L \rho A(v_{tt} + y_{tt})\xi d\xi, \tag{2.3}$$

where ρ is the mass density of the beam and $y(t)$ represents the transverse motion of the supports. It is assumed that both supports move simultaneously. Considering the reaction forces shown in Figure 2.1, the contribution to the internal moment may be expressed as

$$M_2 = M_B - R_B(L - x) + (P + p)v, \tag{2.4}$$

where M_B and R_B are the reactions at the clamped end, P is the applied static axial load, and p is the axial load arising from midplane stretching in the beam. The axial load p can be evaluated based upon the longitudinal strain in the beam, and is given by

$$p = \frac{AE}{L}\int_0^L \varepsilon_x dx, \tag{2.5}$$

where ε_x is the axial component caused by a transverse displacement and does not include the axial static strain caused by the initial tension. Assuming small slopes, we may approximate the longitudinal strain as

$$\varepsilon_x = \frac{\partial u}{\partial x} \approx \frac{ds}{dx} - 1. \tag{2.6}$$

A Taylor series approximation to ds/dx, truncated at the second-order term, yields

$$\varepsilon_x \approx \frac{1}{2}\left(\frac{\partial v}{\partial x}\right)^2. \tag{2.7}$$

Substituting into (2.5) and (2.6) yields

$$p \approx \frac{1}{2}\frac{AE}{L}\int_0^L \left(\frac{\partial v}{\partial x}\right)^2 dx, \tag{2.8}$$

and combining with (2.1), the moment equation can now be expressed as

$$EIv_{xx} = -\int_s^L \rho A\left(v_{tt} + y_{tt}\right)\xi\, d\xi + M_B + R_B(s - L) + (P + p)v. \tag{2.9}$$

Taking derivatives twice with respect to x and noting that the reaction forces are constant with respect to x, the equation of motion for the beam becomes

$$EIv_{xxxx} + \rho A\left(v_{tt} + y_{tt}\right) - Pv_{xx} - \frac{AE}{2L}v_{xx}\int_0^L v_x^2 dx = 0. \tag{2.10}$$

An approximate solution to this nonlinear equation will be obtained using a variational method to discretize the PDE. The Galerkin method uses, as a weighting function, one of the system eigenfunctions. In this case the eigenfunction is unknown. Suitable approximations will be sought from three sources: the linear string, the linear clamped-clamped beam, and the linear clamped-clamped beam with an axial load. The objective here is to quantify the error incurred by using easier-to-obtain shape functions.

3. Exact Solution of the Linear Loaded Beam Equation for Clamped-Clamped Boundary Conditions

3.1 Linear String

The problem of the linear string has been formulated and solved (e.g., [17]). The governing equation for the linear string is

$$\frac{\partial^2 v}{\partial x^2} = \frac{\rho}{T}\frac{\partial^2 v}{\partial t^2}, \tag{3.1}$$

where T is the tension in the string and ρ is the mass per unit length of the string. The general solution to the linear string of length L fixed at both ends, is given by

$$v(x,t) = \sum_{n=1}^{\infty}\left(C_n Sin(\omega_n t) + D_n Cos(\omega_n t)\right) Sin\,\frac{n\pi x}{L} \tag{3.2}$$

where

$$\frac{\omega_n L}{c} = \omega_n L\sqrt{\frac{\rho}{T}} = n\pi, \qquad n = 1,2,3,\ldots \tag{3.3}$$

and each value of *n* corresponds to the nth mode of vibration. The first mode of vibration, for example, would have a mode shape given by

$$\Phi(x) = Sin\left(\pi\frac{x}{L}\right). \tag{3.4}$$

The deflection and the first three derivatives are shown in Figure 3.1. Note that this solution does not satisfy the boundary conditions for the clamped-clamped beam and using this shape function to generate an approximate solution for the nonlinear beam equation will contain some error. However, for the beam under consideration, the string mode shape does represent a close approximation to the actual shape away from the fixed ends.

3.2 Linear Clamped-Clamped Beam

The problem of the linear clamped-clamped beam has been formulated (e.g., [17]). The governing differential equation for transverse deflection is given by

$$\frac{\partial^2}{\partial x^2}\left(EI\,\frac{\partial^2 v}{\partial x^2}\right) = -\rho\,A\,\frac{\partial^2 v}{\partial t^2}, \tag{3.5}$$

where *E* is the modulus of elasticity for the beam and *I* is the area moment of inertia of the beam cross section. For free vibration, the transverse displacement is expressed by

$$v(x,t) = \Phi(x)Sin(\omega t - \theta), \tag{3.6}$$

and for constant *E* and *I*, (3.5) reduces to

$$EI\,\frac{d^4\Phi}{dx^4} - \rho\omega^2\Phi = 0. \tag{3.7}$$

The general solution for this differential equation is given by

$$\Phi(x) = ACosh(\beta x) + BSinh(\beta x) + CCos(\beta x) + DSin(\beta x), \tag{3.8}$$

where $\beta = \rho\omega^2/EI$, and $A, B, C,$ and D are constants determined by boundary conditions. The boundary conditions for the clamped-clamped beam of length L are given by

$$\Phi(0) = \Phi(L) = 0 \tag{3.9}$$

$$\frac{d\Phi}{dx}(0) = \frac{d\Phi}{dx}(L) = 0. \tag{3.10}$$

Applying the boundary conditions at $x = 0$ to (3.8) the shape function becomes

$$\Phi(x) = A\big(Cosh(\beta x) - Cos(\beta x)\big) + B\big(Sinh(\beta x) - Sin(\beta x)\big). \tag{3.11}$$

The boundary conditions at $x = L$, when applied to (3.11), generate a characteristic equation given by

$$Cos(\beta L)Cosh(\beta L) = 1, \tag{3.12}$$

from which β may be determined. The relationship between A and B is then determined from the matrix coefficients, and is given by

$$A = -B\frac{\big(Sinh(\beta L) - Sin(\beta L)\big)}{\big(Cosh(\beta L) - Cos(\beta L)\big)}. \tag{3.13}$$

The equation for the first mode shape is given by (3.11) and (3.13) using the lowest root of (3.12) which is $(\beta_1 L)^2 = 22.373$. Deflection, slope, bending and shear diagrams for the clamped-clamped beam are shown in Figure 3.2.

3.3 Axially Loaded Linear Clamped-Clamped Beam

The equation governing the linear free vibration of an axially loaded clamped-clamped beam is given by

$$\rho A\frac{\partial^2 v}{\partial t^2} + EI\frac{\partial^4 v}{\partial x^4} - P\frac{\partial^2 v}{\partial x^2} = 0 \tag{3.14}$$

where ρ and E are material parameters, A and I are functions of the beam cross-section and P is the static tensile load on the beam. A solution assumed to be of the form

$$v(x,t) = \Phi(x)Sin(\omega t - \theta) \tag{3.15}$$

where the function $\Phi(x)$ is the shape function and $Sin(\omega t-\theta)$ is the temporal response. When substituted into the governing partial differential equation (3.14) we obtain

$$EI\Phi^{iv} - P\Phi^{ii} - \omega^2\rho A\,\Phi = 0\,. \tag{3.16}$$

The general solution to (3.16) is

$$\Phi(x) = C_1 Sin(\lambda_1 x) + C_2 Cos(\lambda_1 x) + C_3 Sinh(\lambda_2 x) + C_4 Cosh(\lambda_2 x) \tag{3.17}$$

where the C_1, C_2, C_3, C_4, λ_1 and λ_2 are constants determined by application of the boundary conditions. Substituting the shape function into the governing differential equation yields equations for λ_1 and λ_2, given by

$$\lambda_1 = \sqrt{\frac{-P}{2EI} + \sqrt{\left(\frac{P}{2EI}\right)^2 + \frac{\omega^2\rho A}{EI}}} \quad \text{and} \tag{3.18}$$

$$\lambda_2 = \sqrt{\frac{P}{2EI} + \sqrt{\left(\frac{P}{2EI}\right)^2 + \frac{\omega^2\rho A}{EI}}}\,. \tag{3.19}$$

The boundary conditions for a clamped-clamped beam of length L are given by

$$\begin{aligned} \Phi(0) &= 0 \\ \Phi(L) &= 0 \\ \Phi'(0) &= 0 \\ \Phi'(L) &= 0. \end{aligned} \tag{3.19}$$

The resulting shape function is given by

$$\Phi(x) = C_1\left(Sin(\lambda_1 x) - \frac{\lambda_1}{\lambda_2} Sinh(\lambda_2 x)\right) + C_2\left(Cos(\lambda_1 x) - Cosh(\lambda_2 x)\right) \tag{3.20}$$

where the coefficients C_1 and C_2 are related by

$$C_1 = C_2 \frac{Cosh(\lambda_2 L) - Cos(\lambda_1 L)}{Sin(\lambda_1 L) - \frac{\lambda_1}{\lambda_2} Sinh(\lambda_2 L)}\,. \tag{3.21}$$

The characteristic equation for the beam problem is derived from the determinant of the coefficient matrix for C_1 and C_2 and is given by

$$\left(Sin(\lambda_1 L) - \frac{\lambda_1}{\lambda_2} Sinh(\lambda_2 L)\right)(\lambda_1 Sin(\lambda_1 L) + \lambda_2 Sinh(\lambda_2 L)) + \\ (\lambda_1 Cos(\lambda_1 L) - \lambda_1 Cosh(\lambda_2 L))(Cos(\lambda_1 L) - Cosh(\lambda_2 L)) = 0 \tag{3.22}$$

where λ_1 and λ_2 are functions of ω. This equation may be solved for the natural frequencies of the beam. Note that (3.17), (3.18), (3.21) and (3.22), when viewed as functions of ω, are extremely sensitive to truncation errors. Numerical calculations of the λ_i (and hence the natural frequencies) must be made to a very high degree of precision (e.g. 15 digits) in order to avoid numerical instabilities in the mode shape calculations. The material and geometric properties for the beam under consideration are given in Table 3.1. Figure 3.3 shows the deflection, slope, shear, and moment for the first-mode response.

4. Solution of the Nonlinear Problem

Galerkin's procedure was used to discretize the partial differential equation and was applied to the nonlinear beam equation (2.10), shown again below,

$$\rho A \frac{\partial^2 v}{\partial t^2} + EI \frac{\partial^4 v}{\partial x^4} - P \frac{\partial^2 v}{\partial x^2} - \frac{EA}{2L} \frac{\partial^2 v}{\partial x^2} \int_0^L \left(\frac{\partial v}{\partial x}\right)^2 dx = -\rho A \frac{\partial^2 y}{\partial t^2}. \tag{2.10}$$

Assuming a separation of variables solution

$$v(x,t) = \sum_{i=1}^{n} q_i(t) \Phi_i(x), \tag{4.1}$$

where $\Phi_i(x)$ is the spatial shape function of the ith mode, and $q_i(t)$ is the time modulation function, and applying Galerkin's procedure for a single mode to (2.10), we obtain

$$\ddot{q}\rho A\int_0^L \Phi^2 dx + q\left(EI\int_0^L \Phi^{IV}\Phi dx - P\int_0^L \Phi^{II}\Phi dx\right) + \\ q^3\left(-\frac{EA}{2L}\int_0^L \left(\Phi^I\right)^2 dx\int_0^L \Phi^{II}\Phi dx\right) = -\rho A\frac{\partial^2 Y}{\partial t^2}\int_0^L \Phi dx \tag{4.2}$$

where the subscript has been dropped. This equation can be rewritten as

$$\ddot{q} + \omega_1^2 q + Kq^3 = F(t), \tag{4.3}$$

where the constant coefficients and $F(t)$ are

$$\omega_1^2 = \frac{EI\int_0^L \Phi^{IV}\Phi dx + P\int_0^L \Phi^{II}\Phi dx}{C} \tag{4.4}$$

$$K = \frac{-\frac{EA}{2L}\int_0^L \left(\Phi^I\right)^2 dx\int_0^L \Phi^{II}\Phi dx}{C} \tag{4.5}$$

$$F(t) = \frac{-\rho A\frac{\partial^2 Y}{\partial t^2}\int_0^L \Phi dx}{C} \tag{4.6}$$

where

$$C = \rho A\int_0^L \Phi^2 dx \tag{4.7}$$

and Φ^I represents the first derivative of the shape function with respect to x, Φ^{II} represents the second derivative with respect to x, and so on.

4.1 Approximation of the Nonlinear Solution Using the Linear String Shape Function

The shape function generated for the first mode of the linear string is applied to the analysis, and is given by

$$\Phi(x) = Sin\left(\frac{\pi x}{L}\right). \tag{4.8}$$

Applying this to equations (4.4) and (4.5) results in the coefficients for the temporal equation in algebraic form leading to

$$\omega_1^2 = \frac{\pi^2}{\rho A L^4}\left(EI\pi^2 + L^2 P\right), \tag{4.9}$$

and

$$K = \frac{E\pi^4}{4\rho L^4}. \tag{4.10}$$

4.2 Approximation of the Nonlinear Solution Using the Linear Clamped-Clamped Beam Shape Function

The shape function for the first mode of the linear clamped-clamped beam is given by

$$\Phi(x) = A\left(Cos(\beta x) - Cosh(\beta x)\right) + \left(Sin(\beta x) - Sinh(\beta x)\right) \tag{4.11}$$

where

$$A = \frac{Sinh(\beta L) - Sin(\beta L)}{Cos(\beta L) - Cosh(\beta L)}. \tag{4.12}$$

Numerical evaluation of the equations (4.4) and (4.5) generates the coefficients for the temporal equation (4.3). For the first mode, $\beta_1 L$ = 4.730041, and yields for the temporal equation

$$\omega_1^2 = \frac{0.0354242}{\rho A}\left(0.0256267 EI + 0.467696 P\right) \tag{4.13}$$

$$K = \frac{0.000142178 AE}{\rho A (1.61643021)^2}, \tag{4.14}$$

where A is the cross-sectional beam area, E is the modulus of elasticity, and P is the axially applied static tensile load.

4.3 Approximation of the Nonlinear Solution Using the Linear Axially Loaded Clamped-Clamped Beam Shape Function

The shape function for the first mode of the axially loaded clamped-clamped beam is given by

$$\Phi(x) = C_1\left(Sin(\lambda_1 x) - \frac{\lambda_1}{\lambda_2} Sinh(\lambda_2 x)\right) + C_2\left(Cos(\lambda_1 x) - Cosh(\lambda_2 x)\right) \tag{4.15}$$

where the coefficients C_1 and C_2 are related by

$$C_1 = C_2 \frac{Cosh(\lambda_2 L) - Cos(\lambda_1 L)}{Sin(\lambda_1 L) - \frac{\lambda_1}{\lambda_2} Sinh(\lambda_2 L)}, \tag{4.16}$$

and, λ_1 and λ_2 are functions of the material parameters and axial loading applied to the beam, given by

$$\lambda_1 = \sqrt{\frac{-P}{2EI} + \sqrt{\left(\frac{P}{2EI}\right)^2 + \frac{\omega^2 \rho A}{EI}}} \tag{4.17}$$

$$\lambda_2 = \sqrt{\frac{P}{2EI} + \sqrt{\left(\frac{P}{2EI}\right)^2 + \frac{\omega^2 \rho A}{EI}}}\ . \tag{4.18}$$

Numerical evaluation of equations (4.4) and (4.5) using the first-mode shape function yields the coefficients for the temporal modulation equation of the first mode. Since the shape function is dependent on embedded material parameters, the coefficients are not expressible in forms similar to equations (4.13) and (4.14). Consequently, numerical evaluation of the shape function must be performed for each set of material and loading parameters.

5. Perturbation Solution of the Nonlinear Temporal Modulation Problem

5.1 Method I: Perturbation Solution Using Linear Detuning

The temporal equation may be expressed, with the addition of a linear viscous damping term, by

$$\ddot{q} + \omega_1^2 q + 2\varepsilon\mu\dot{q} + \varepsilon\alpha q^3 = F(t) = \varepsilon F_0 Cos(\Omega t), \tag{5.1}$$

where $\varepsilon\alpha$ corresponds to K in the temporal equation (4.3). Following Nayfeh and Mook[15], the excitation frequency is expressed as a linear deviation from the natural frequency in the form

$$\Omega = \omega_1 + \varepsilon\sigma \tag{5.2}$$

where σ is a detuning parameter and ε is a small dimensionless parameter which is proportional to the amplitude of the response. The temporal variable q is expressed as a function of various time scales and ε given by

$$q(t,\varepsilon) = q_0(T_0,T_1) + \varepsilon q_1(T_0,T_1) + \ldots \tag{5.3}$$

where $T_0 = t$ and $T_1 = \varepsilon t$. Using (5.2), the excitation term in equation (5.1) is expressed as

$$F(t) = \varepsilon F_0 Cos(\omega_1 T_0 + \sigma T_1), \tag{5.4}$$

where F_0 is a constant. The time derivatives become

$$\frac{d}{dt} = D_0 + \varepsilon D_1 + \varepsilon^2 D_2 + \ldots \tag{5.5}$$

$$\frac{d^2}{dt^2} = D_0^2 + 2\varepsilon D_0 D_1 + \varepsilon^2 (2 D_0 D_2 + D_1^2) + \ldots \tag{5.6}$$

where

$$D_n = \frac{\partial}{\partial T_n}. \tag{5.7}$$

Substituting equations (5.3), and (5.4)-(5.6) into equation (5.1) and setting the terms with like powers of ε to zero, we obtain

$$D_0^2 u_0 + \omega_1^2 u_0 = 0 \tag{5.8}$$

and

$$D_0^2 u_1 + \omega_1^2 u_1 = -2 D_0 D_1 u_0 - 2\mu D_0 u_0 - \alpha u_0^3 + F_0 Cos(\omega_1 T_0 + \sigma T_1). \tag{5.9}$$

The solution of (5.8) may be written in terms of complex exponential functions as

$$u_0 = A(T_1) e^{i\omega_1 T_0} + \overline{A}(T_1) e^{-i\omega_1 T_0} \tag{5.10}$$

where $A(T_1)$ is as yet undetermined and the over bar indicates the complex conjugate. Equation (5.10) is substituted into equation (5.9), resulting in

$$\begin{aligned} D_0^2 u_1 + \omega_1^2 u_1 = &-\left(2i\omega_1\left(A' + \mu A\right) + 3\alpha A^2 \overline{A}\right) e^{i\omega_1 T_0} \\ &- \alpha A^3 e^{3i\omega_1 T_0} + \left(F_0/2\right) e^{i(\omega_1 T_0 + \sigma T_1)} + cc, \end{aligned} \tag{5.11}$$

where cc indicates the complex conjugate of the preceding terms. The secular terms of equation (5.11) are eliminated by satisfying

$$2i\omega_1\left(A' + \mu A\right) + 3\alpha A^2 \overline{A} - \left(F_0/2\right) e^{i\sigma T_1} = 0. \tag{5.12}$$

Expressing $A(T_1)$ in complex exponential form

$$A=(a/2)e^{i\beta}, \tag{5.13}$$

where a and β are functions of T_1, and substituting into equation (5.12), we obtain

$$a' = -\mu a + \frac{F_0}{2\omega_1} Sin(\sigma T_1 - \beta) \tag{5.14}$$

and

$$a\beta' = \frac{3\alpha}{8\omega_1} a^3 - \frac{F_0}{2\omega_1} Cos(\sigma T_1 - \beta). \tag{5.15}$$

Equations (5.14) and (5.15) govern the first-order approximate solution to the temporal equation; $q(t)$ may be expressed approximately as

$$q = aCos(\omega_1 t + \beta) + O(\varepsilon). \tag{5.16}$$

Equations (5.14) and (5.15) are nonautonomous; they can be transformed into autonomous equations with

$$\gamma = \sigma T_1 - \beta. \tag{5.17}$$

The resulting amplitude-modulation and phase-modulation equations are given by

$$a' = -\mu a + \frac{F_0}{2\omega_1} Sin\gamma, \tag{5.18}$$

and

$$a\gamma' = \sigma a - \frac{3\alpha}{8\omega_1} a^3 + \frac{F_0}{2\omega_1} Cos\gamma. \tag{5.19}$$

Steady-state oscillations occur when $a' = \gamma' = 0$ and correspond to the fixed points of the right-hand sides. Squaring and summing (5.18) and (5.19) yields the frequency response function given by

$$\left[\mu^2 + \left(\sigma - \frac{3\alpha}{8\omega_1} a^2\right)^2\right] a^2 = \frac{F_0^2}{4\omega_1^2}. \tag{5.20}$$

Solving for σ yields

$$\sigma = \frac{3\alpha}{8\omega_1} a^2 \pm \left(\frac{F_0^2}{4\omega_1^2 a^2} - \mu^2\right)^{\frac{1}{2}}. \tag{5.21}$$

This curve is the typical hardening (or softening) response that exhibits the jump phenomenon. For a jump down to occur in the response amplitude, the radical in equation (5.21) must vanish; the resulting equation for the jump point is given as

$$\sigma = \frac{3\alpha}{8\omega_1} a^2 . \tag{5.22}$$

This relationship may be used to identify α from experimental frequency and amplitude data for a jump point. Similarly, from the radicand in equation (5.21), the damping coefficient can be identified as

$$\mu = \frac{F_0}{2\omega_1 a}, \tag{5.23}$$

where F and a are positive numbers.

5.2 Method II: Perturbation Solution Using Quadratic Detuning

A modified perturbation solution of the temporal modulation equation was developed by Burton and Rahman [16] using the Method of Multiple Scales. Using a slightly modified version of the temporal equation, given by

$$\ddot{q} + \omega_1^2 q + 2\varepsilon\mu\,\dot{q} + \varepsilon q^3 = F(t) = \varepsilon F_0 Cos(\Omega t), \tag{5.24}$$

the time scale is transformed with

$$T = \Omega t . \tag{5.25}$$

Substituting equation (5.25) into (5.24) results in

$$\Omega^2\ddot{q} + \omega_1^2 q + 2\varepsilon\mu\,\Omega\dot{q} + \varepsilon q^3 = F(t) = \varepsilon F_0 Cos(T), \tag{5.26}$$

where the overdots now refer to differentiation with respect to the nondimensional time T. A new expansion parameter α based upon the postulated existence of a steady-state response amplitude a_0 is defined as

$$\alpha = \frac{\varepsilon a_0^2}{\left(4\omega_1^2 + 3\varepsilon a_0^2\right)} . \tag{5.27}$$

The old expansion parameter ε may be expressed as a function of α as

$$\varepsilon = \frac{\omega_1^2}{a_0^2}\left(\frac{4\alpha}{1-3\alpha}\right). \tag{5.28}$$

In this case, the detuning parameter σ is defined in terms of the frequency squared and measures deviation from the backbone curve; it is given by

$$\Omega^2 = \left(\omega_1^2 + \frac{3\varepsilon a_0^2}{4}\right)(1+\alpha\sigma), \tag{5.29}$$

where the first term in parenthesis is an approximate equation for the backbone curve. Equation (5.29) may be alternatively expressed in terms of the new detuning parameter as

$$\Omega^2 = \omega_1^2\left(\frac{1}{1-3\alpha}\right)(1+\alpha\sigma). \tag{5.30}$$

Substituting (5.30), (5.27), and (5.28) into (5.26) yields an equation in terms of the new detuning parameter and is given by

$$\begin{aligned} &\omega_1^2\left(\frac{1}{1-3\alpha}\right)(1+\alpha\sigma)\ddot{q} + \omega_1^2 q + 2\varepsilon\mu\Omega\dot{q} + \frac{4\omega_1^2\alpha}{a_0^2(1-3\alpha)}q^3 \\ &\quad = \frac{4\omega_1^2\alpha}{a_0^2(1-3\alpha)}F_0 Cos(T) \end{aligned} \tag{5.31}$$

Defining a new damping parameter as

$$\eta = \frac{4\mu\Omega}{a_0^2}, \tag{5.32}$$

equation (5.31) may be simplified to

$$(1+\alpha\sigma)\ddot{q} + (1-3\alpha)q + 2\alpha\eta\dot{q} + \frac{4\alpha}{a_0^2}q^3 = \frac{8\alpha}{a_0^2}F_0 Cos(T). \tag{5.33}$$

The final step in the transformation of the temporal equation is to introduce a normalized temporal variable r defined by

$$r = \frac{q}{a_0}, \tag{5.34}$$

which when substituted into (5.33) yields

$$(1+\alpha\sigma)\ddot{r}+(1-3\alpha)r+2\alpha\eta\dot{r}+4\alpha r^3=\frac{8\alpha}{a_0^3}F_0Cos(T). \tag{5.35}$$

This transformed equation is solved following the method used previously; in this case α is the scaling parameter rather than ε. The resulting equations are analogous to equations (5.8) and (5.9), and are given by

$$D_0^2r_0+r_0=0, \tag{5.36}$$

and

$$D_0^2r_1+r_1=-\sigma D_0^2r_0-2D_1D_0r_0-2\eta D_0r_0-4r_0^3+3r_0+\frac{8F_0}{a_0^3}CosT. \tag{5.37}$$

As before, equation (5.36) has a solution of the form

$$r_0=\frac{a}{2}e^{i(T_0+\beta)}+cc, \tag{5.38}$$

which is substituted into equation (5.37). When the secular terms from the resulting equation are eliminated, two equations are generated and are given by

$$a'=-\eta a-\frac{4F_0}{a_0^3}Sin\beta, \tag{5.39}$$

and

$$a\beta'=-\frac{\sigma a}{2}-\frac{3}{2}a\left(1-a^2\right)-\frac{4F_0}{a_0^3}Cos\beta. \tag{5.40}$$

The steady-state solutions are obtained in the same manner as before, setting the derivatives of a and β to zero, and squaring and summing the equations. This leads to an equation for the detuning parameter which is given by

$$\sigma=\pm2\sqrt{\left(\frac{4F_0}{a_0^3}\right)^2-\eta^2}\,. \tag{5.41}$$

The equation for the frequency response curve may be obtained by substituting (5.41) into equation (5.29), and is given by

$$\Omega^2 = \left(\omega_1^2 + \frac{3\varepsilon a_0^2}{4}\right)\left(1 \pm 2\alpha\sqrt{\left(\frac{4F_0}{a_0^3}\right)^2 - \eta^2}\right). \qquad (5.42)$$

As before, the condition for the occurrence of a jump-down bifurcation requires the vanishing of the radical in equation (5.42), which yields a relationship between the amplitude and frequency of excitation, the amplitude of the response, and the damping coefficient. This relationship is given by

$$\mu = \frac{F_0}{a_0\ \Omega}. \qquad (5.43)$$

At the jump down bifurcation, equation (5.42) becomes

$$\Omega^2 = \left(\omega_1^2 + \frac{3\varepsilon a_0^2}{4}\right), \qquad (5.44)$$

which can be used to identify the cubic coefficient ε in (5.24) from experimental frequency and response amplitude data.

6. Experimental Effort and Results

Experiments were performed to identify the value of the cubic coefficient in equation (4.3). The dominant source of nonlinearity for a beam is mid-plane stretching, which motivated the choice of a clamped-clamped beam for this experiment. Dimensions of the beam are given in Table 3.1. A schematic of the fixture used to support the beam for these experiments is shown in Figure 6.1. A schematic of the instrumentation is shown in Figure 6.2.

Early tests of the support fixture indicated the need for additional reinforcement to minimize even slight motion in the axial direction. Additional stiffness was added with angle iron near the top of the supports with three supporting sections along the span of the beam, as shown in part (b) of Figure 6.1. The interior sides of the clamps were lightly grooved in a direction perpendicular to the axial direction of the test beam to prevent the slipping (which occurred in cases of large beam deflections in preliminary experiments). Four longitudinally aligned strain gages (Micro-Measurements type EA-06-120LZ-120) were applied at one end of the test beam to form a four-arm active bridge. Since the circuit was wired to measure strain due to bending, the location was chosen to avoid the inflection point.

The shaker used in this experiment was an Unholtz-Dickie 1100-lb shaker. Sinusoidal excitation was provided by a Wavetek model 650 variable-phase synthesizer. Spectral response data was collected and analyzed by a Data Physics signal analyzer board using DP420 software installed on an 80486 PC-compatible personal computer. Amplitude response data was also measured on a Hewlett Packard 3582A Spectrum Analyzer, and excitation levels were measured using a Bruel & Kjaer voltmeter, type 2432, which monitored acceleration levels recorded by a Bruel & Kjaer accelerometer, type 4381, attached to a Bruel & Kjaer charge amplifier, type 2635.

An estimate for the modulus of elasticity was obtained experimentally. The gages were configured in a Wheatstone half bridge to measure longitudinal strain. A load-strain curve was generated using weights applied to the beam; the modulus measured was $31.815\text{x}10^6$ psi. A correlation between strain gage output and the midspan displacement of the beam was obtained using a non-contacting displacement transducer.

For the axially-loaded experiments, a tension of 51.56 pounds was applied to the beam. The natural frequency for the first mode, estimated using the spectral response of the beam subjected to a low-level random excitation, was 45.2201 Hz. Frequency sweeps were performed at four levels of excitation, ranging from 2.00 g's to 0.50 g's, and are shown in Figures 6.3 through 6.6. During these sweeps, which included both up and down sweeps, the response amplitudes and frequencies at the jump down points were recorded. The jump down bifurcation points are listed in Table 6.1.

7. Analysis of Experimental Data

It was shown by equation (5.22) that, at the jump down bifurcation points, when the linear detuning from the linear natural frequency is used, the frequency and amplitude are related by

$$\sigma=\frac{\Omega-\omega_1}{\varepsilon}=\frac{3\alpha}{8\omega_1}a^2 . \tag{5.22}$$

Hence, to identify the cubic coefficient α, one performs several experiments at various levels of excitation such that jump-down bifurcations occur. At least two different amplitude levels are required to identify a straight line with slope $3\alpha / 8\omega_1$. When three or more points are available, a least squares linear curve fit to the data, as shown in Figure 7.1, can be performed. The experimental data given in Table 6.1 and plotted in Figure 7.1 yields an estimate for $K = 1.3298\text{x}10^6$ $(\text{rad/sec-in})^2$.

Equation (5.44) shows that, at the jump down bifurcation points, the frequency and amplitude are related by

$$\Omega^2 = \omega_1^2 + \frac{3\varepsilon a^2}{4}. \tag{5.44}$$

Using this equation as the model to which the experimental data is fit yields an estimate for $K = 1.8151\text{x}10^6$ $(\text{rad/sec-in})^2$, where the data is shown in Figure 7.2.

The coefficient for the cubic term estimated from Galerkin's procedure applied to the nonlinear equation of motion was found to be

$$K=\frac{-\frac{EA}{2L}\int_0^L(\Phi^I)^2dx\int_0^L\Phi^{II}\Phi dx}{\rho A\int_0^L\Phi^2dx}. \tag{4.5}$$

When K is evaluated for the experimental clamped-clamped beam using the shape functions for the string, the clamped-clamped beam with no axial load, and the axially loaded linearized clamped-clamped beam, the estimates for the cubic coefficient are $1.8361\text{x}10^6$, $2.2623\text{x}10^6$, and $1.8655\text{x}10^6$ $(\text{rad/sec})^2/\text{in}^2$ respectively. Table 7.1 lists the estimates of the cubic coefficient from the experimental jump-point data. Table 7.2 summarizes the percent error between theoretically and experimentally estimated values.

8. Summary and Conclusion

A technique was developed and tested to identify quantitatively the coefficient of the cubic term; the results show that it works extremely well. Method II yields the most accurate results.

Several strategies were developed and tested in this project. Given the degree of difficulty in obtaining the solution to the clamped-clamped beam with an axial load undergoing vibration, several alternative approaches were considered. The Galerkin procedure required numerical evaluation of complex integrals; since the exact shape function is unknown, suitable approximations were used. For this project, three solutions were considered: the linear string, the linear clamped-clamped beam with no axial load, and the linear clamped-clamped beam with an axial load.

The theoretical analyses showed that the estimated value of the cubic coefficient, obtained from experiments, was quite large; it was between 1.3×10^6 and 1.8×10^6. The theoretical values ranged from 1.8×10^6 to 2.3×10^6. The first observation to be made is that the experimental results were of the same order of magnitude as that predicted by the theory, but were consistently on the low side. The second observation is that the Method II analysis of the experimental data came the closest to the value of the cubic coefficient predicted theoretically. The third observation is that the string shape function gave surprisingly consistent results when compared with the solution obtained from the clamped-clamped beam with an axial load; one would have thought that the clamped-clamped beam with no axial load would have been a better function because the boundary conditions were completely satisfied. That it did not can be explained as follows.

The clamped-clamped beam with no axial load has zero slope at the boundaries. As the axial tension is increased, the inflection point in the transverse displacement moves towards the boundaries. As a result, the shape function becomes more and more like that of a string. For the case considered, i.e., P = 51.56 lbs, the shape function was very similar to the string but obviously, the boundary conditions were not satisfied. However, the discrepancy is largely confined to the region very close to the boundaries. The computation of the coefficients for the temporal equation involves numerical evaluation of integrals of the shape function and its derivatives over the entire length of the beam. Hence, the solution that most closely approximates the shape and whose derivatives most closely approximates those of the C-C beam with an axial load will yield the most accurate result. One can visually compare the results (Figures 3.1-3.3) and note the similarities (realizing to check the scales).

A final comment to be made regarding the very large size of the coefficient; Perturbation Method I assumes that the perturbation parameter is small. In this analysis, we observe that it is not. Perturbation Method II does not assume that the perturbation parameter is small. That both give a similar (thought not identical) result leads one to conclude that the dependant variable is indeed so small that the product of the large coefficient and the small cubic term act to cause the whole term to be small.

The method has the potential to be automated on a computer and could be included in a routine modal analysis. It is fairly easy to use and yields results consistent with the theory and assumptions made regarding the quality of trial functions to describe the spatial response.

9. References

1. Evan-Iwanowski, R. M., *Resonance Oscillations in Mechanical Systems*, Elsvier Scientific Publishing Co., Amsterdam, 1976.

2. Nayfeh, A. H. and Mook, D. T., *Nonlinear Oscillations*, Wiley-Interscience, New York, 1979.

3. Schmidt, G. S. and Tondl, A., *Nonlinear Vibrations*, Akademie-Verlay, Berlin, 1986.

4. Tomlinson, G. R., "Detection, identification and quantification of nonlinearity in modal analysis - a review," *Proceedings of the Fourth International Modal Analysis Conference*, Orlando, FL, 837-843, 1985.

5. Busby, H. R., Nopporn, C. and Singh, R., "Experimental modal analysis of nonlinear systems: a feasibility study," *The Journal of Souns and Vibration*, **180**(3), 415-427, 1986.

6. Zavodney, L. D., "Can the modal analyst afford to be ignorant of nonlinear vibration phenomena?" *Proceedings of the Fifth International Modal Analysis Conference*, London, England, 1987.

7. Natke, H. G., Juang, J.-N. and Gawronski, W., "A brief review on the identification of nonlinear mechanical systems," *Proceedings of the Sixth International Modal Analysis Conference*, Kissimmee, FL, 1569-1574, 1988.

8. Nayfeh, A. H., "Parametric identification of nonlinear dynamic systems," *Computers and Structures*, **20**, 487-493, 1985.

9. Brown, D. L., "Modal analysis - past, present, future," *Proceedings of the First International Modal Analysis Conference*, Orlando, FL, ix-xiii, 1982.

10. Nayfeh, A. H. and Zavodney, L. D., "Experimental observation of amplitude- and phase-modulated responses of the two internally coupled oscillators to a harmonic excitation," *The Journal of Applied Mechanics*, **110**, 706-710, 1988.

11. Zavodney, L. D. and Nayfeh, A. H., "The response of a single-degree-of-freedom system with quadratic and cubic nonlinearities to a fundamental parametric resonance," *The Journal of Sound and Vibration*, 1988, **120**(1), 63-93.

12. Zavodney, L. D. and Nayfeh, A. H., and Sanchez, N. E., "The response of a single-degree-of-freedom system with quadratic and cubic nonlinearities to a principal parametric resonance," *The Journal of Sound and Vibration*, 1989, **129**(3), 417-442.

13. Zavodney, L. D., "The identification of nonlinearity in structural systems: theory, simulation, and experiment," *Applied Mechanics Reviews*, 1991, **44**(11), Part 2, 5295-5303.

14. Zavodney, L. D. and Hollkamp, J. J., "Experimental identification of internally resonant nonlinear systems possessing quadratic nonlinearity," *Proceedings of the 32nd Structures, Structural Dynamics, and Materials Conference*, Baltimore, MD, April 8-10, 1991, 2755-2765.

15. Nayfeh, A. H. and Mook, D. T., *Nonlinear Oscillations*, Wiley-Interscience, New York, 1979, pp 162-170.

16. Burton, T. D., and Rahman, Z., "On the multi-scale analysis of strongly non-linear forced oscillators," *International Journal of Non-Linear Mechanics*, **21**, 135-146, 1986.

17. Thomson, W. T., *Theory of Vibrations with Applications*, Prentice Hall, New Jersey, 1988.

Table 3.1. Material and geometric properties used for the calculation of the linear axially loaded clamped-clamped beam shape function.

Thickness	0.0156 in
Width	0.502 in
Length	27.25 in
Young's Modulus (Experimental)	$31.815\text{x}10^6$ psi
Mass Density	0.2956 lb/in^3

Table 6.1. Jump down bifurcation points located during frequency sweeps.

Excitation Level	Excitation Frequency	Response Amplitude
2.00 g's RMS	82.90 Hz	0.1375 in^2
1.50	80.00	0.1240
1.00	72.90	0.0979
0.50	64.30	0.0656

Table 7.1. Estimates of the cubic coefficient from experimental jump-point data.

Method	Equation	Cubic Coefficient K (rad/sec-in)^2
I	(5.22)	$1.3298\text{x}10^6$
II	(5.44)	$1.8151\text{x}10^6$

Table 7.2. Predictions of the cubic coefficient using shape functions from various linear theories (from equation (4.5).

Shape Function	Cubic Coefficient K	Percent Error Method I	Percent Error Method II
Axially Loaded Linear Clamped-Clamped Beam	1.8655×10^6	40%	2.8%
Clamped-Clamped Beam with No Axial Load	2.2623×10^6	70%	24%
String	1.8361×10^6	38%	1.2%

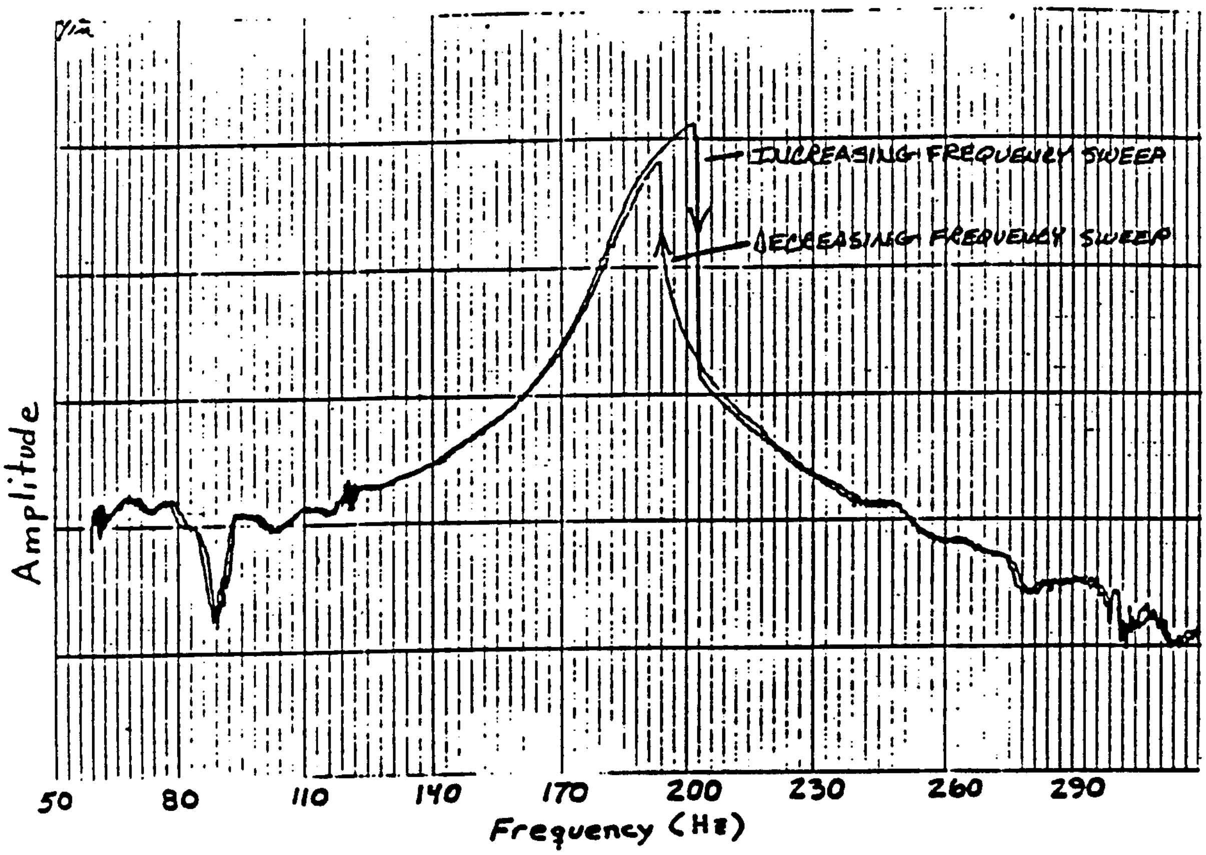


Figure 1.1. A typical frequency response of a rectangular pannel subject to harmonic excitation; the jump phenomena is clearly visibly in the sweep up and down (data provided by the Acoustic and Sonic Fatigue Group in the Flight Dynamics Laboratory at WRDC, Dayton, OH).

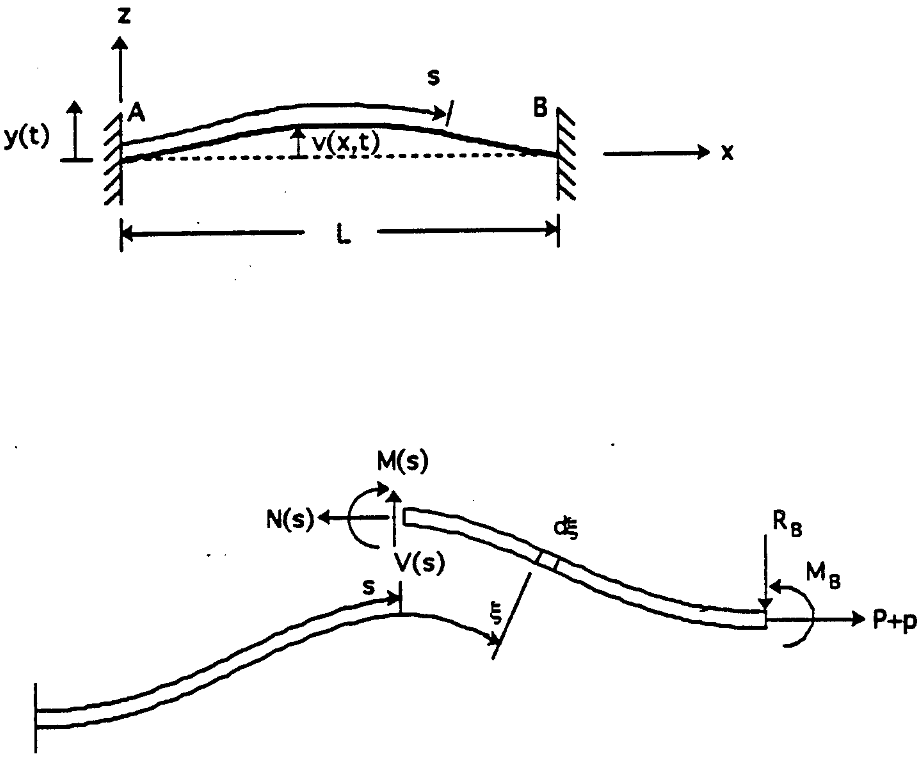


Figure 2.1. Clamped-Clamped beam subjected to support motion.

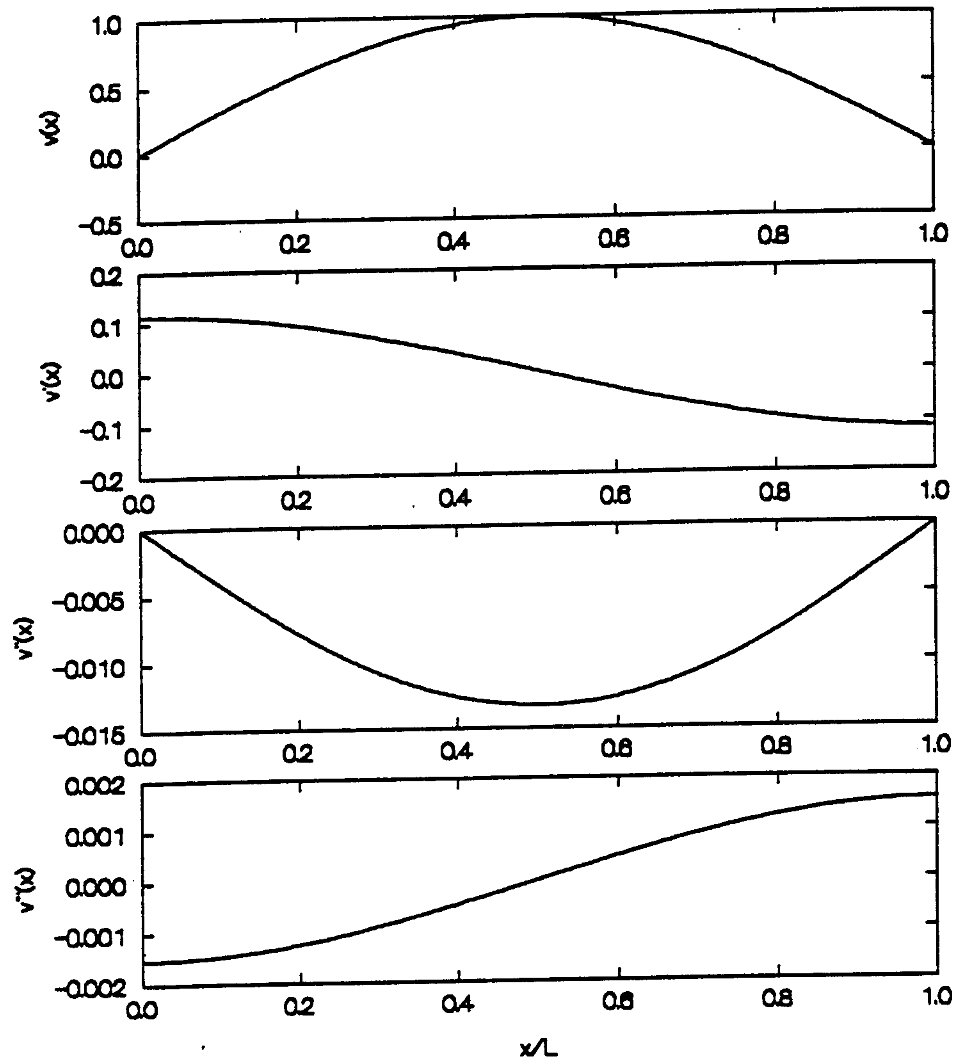


Figure 3.1. Shape function and associated derivatives for the string.

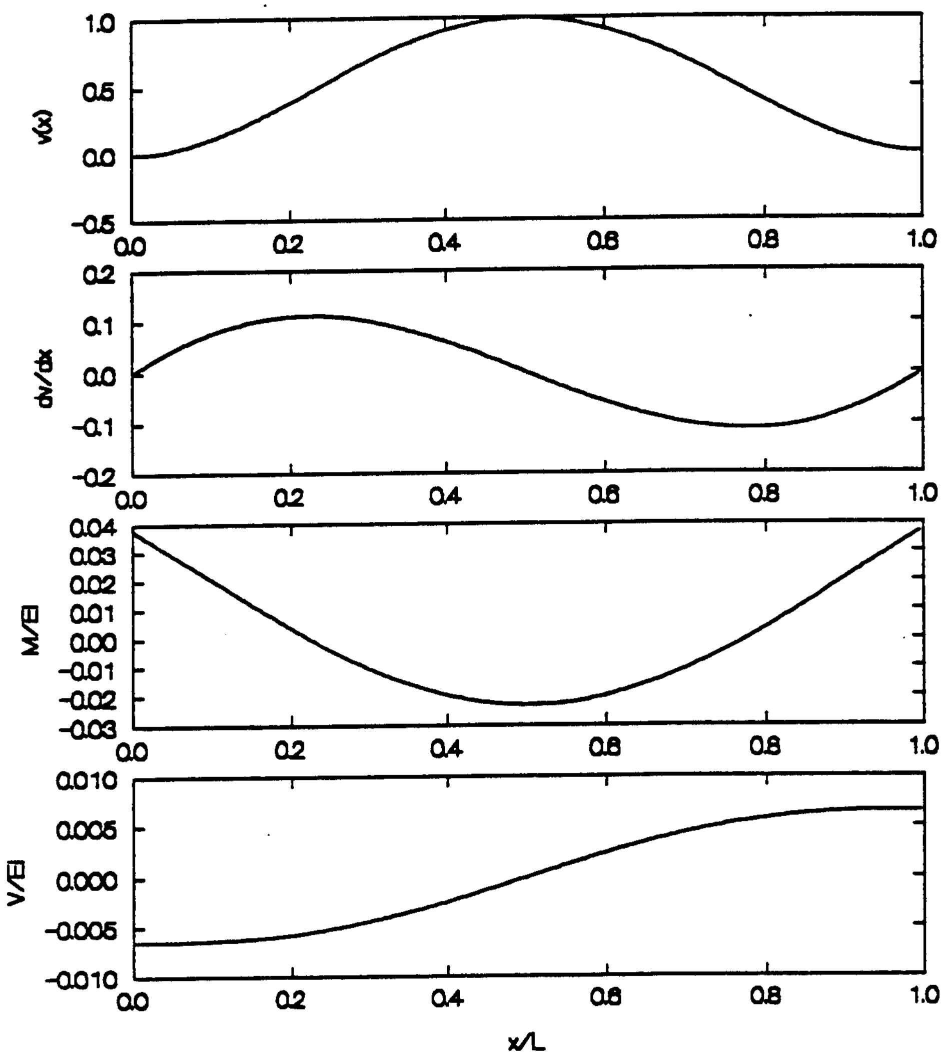


Figure 3.2. Shape function, slope, moment, and shear for the clamped-clamped beam with no axial load.

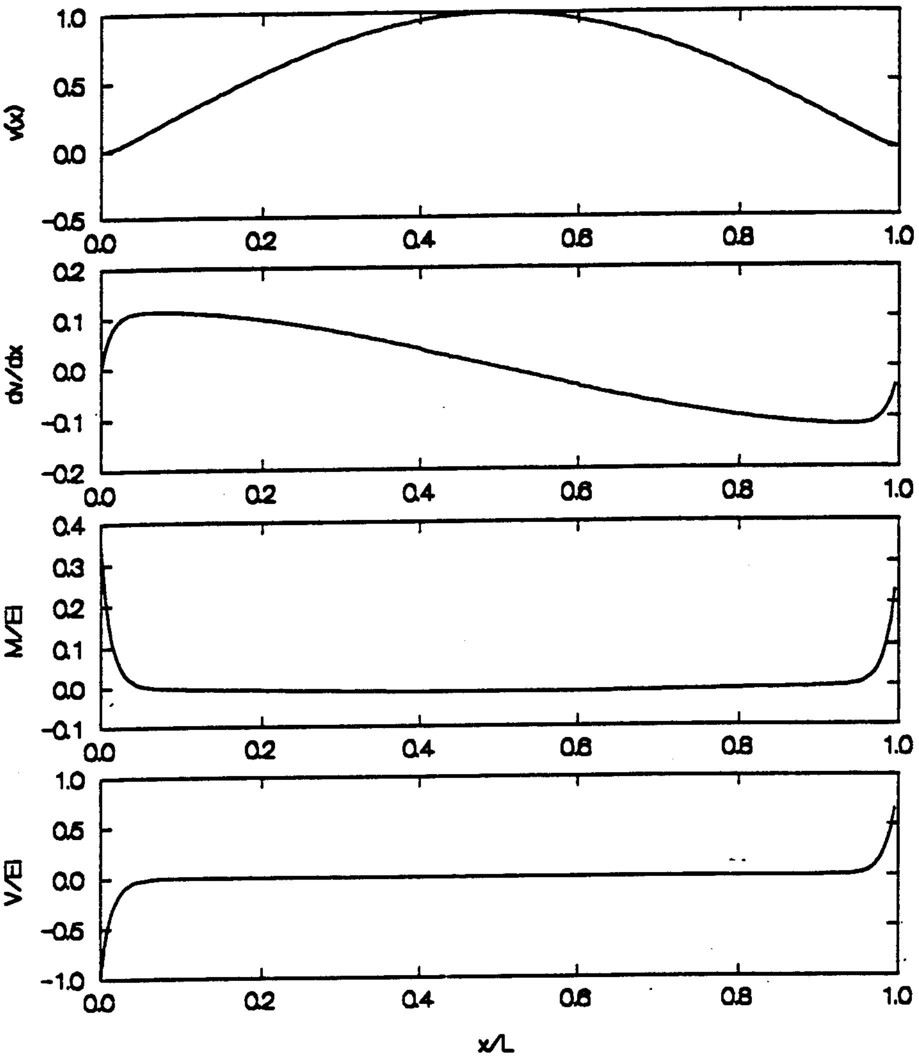


Figure 3.3. Shape function, slope, moment, and shear for the axially loaded clamped-clamped beam using parameters given in Table 3.1.

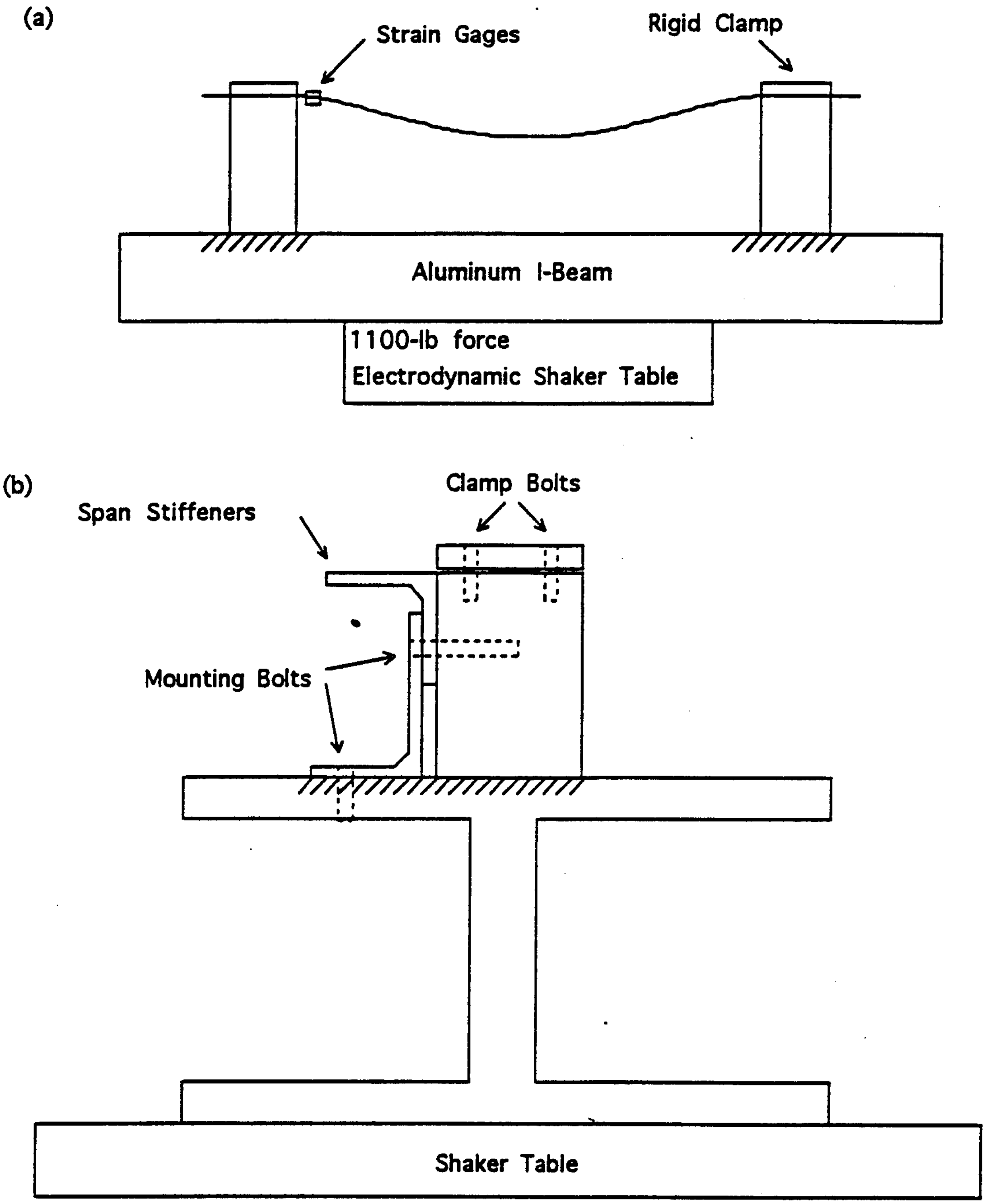


Figure 6.1. Beam mounting fixture as seen from (a) the side, additional support braces not shown, and (b) the end.

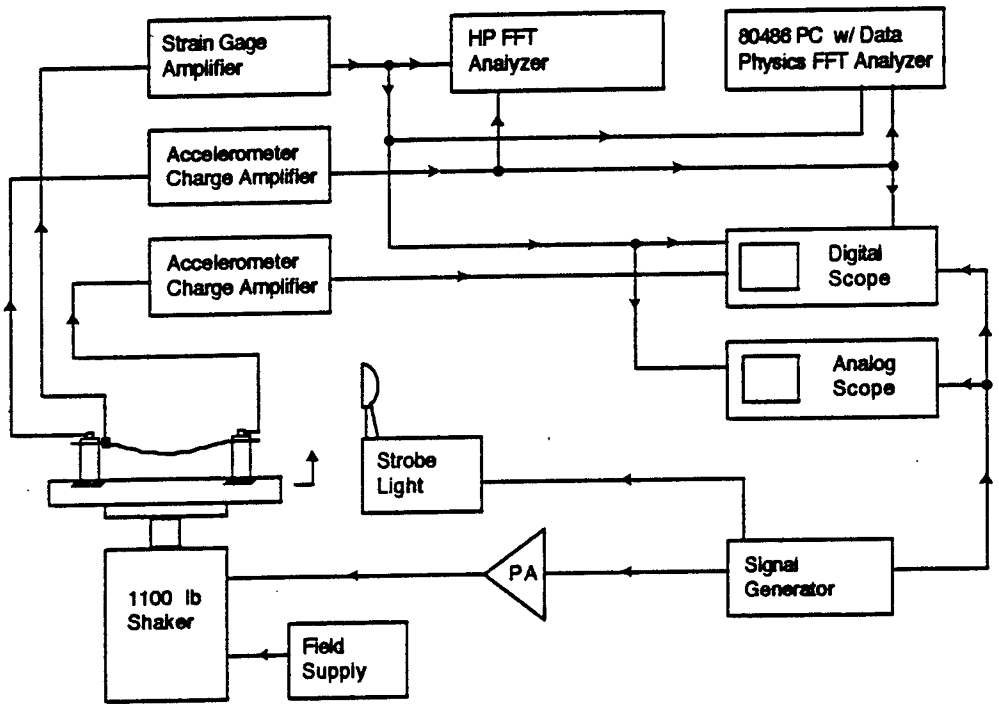


Figure 6.2. Typical instrumentation setup.

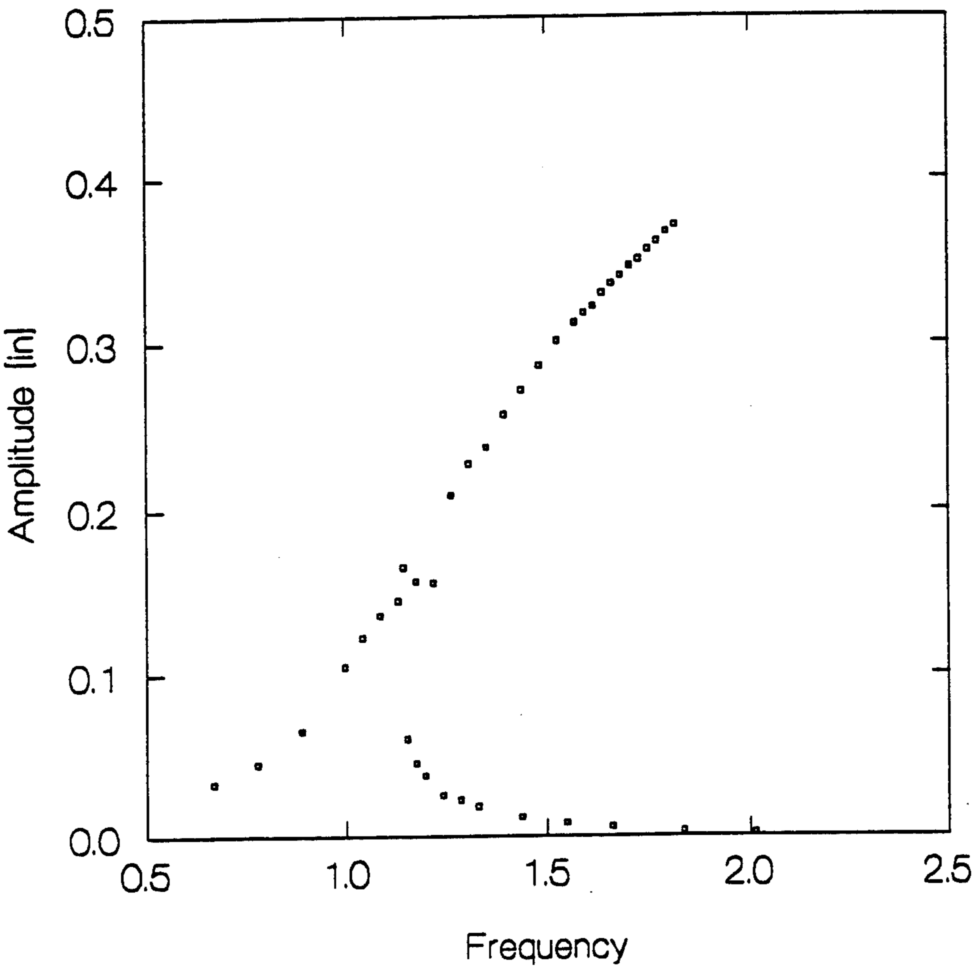


Figure 6.3. Amplitude response, a_o, as a function of nondimensional excitation frequency, Ω/ω_o, for an excitation level of 2.00 g's RMS.

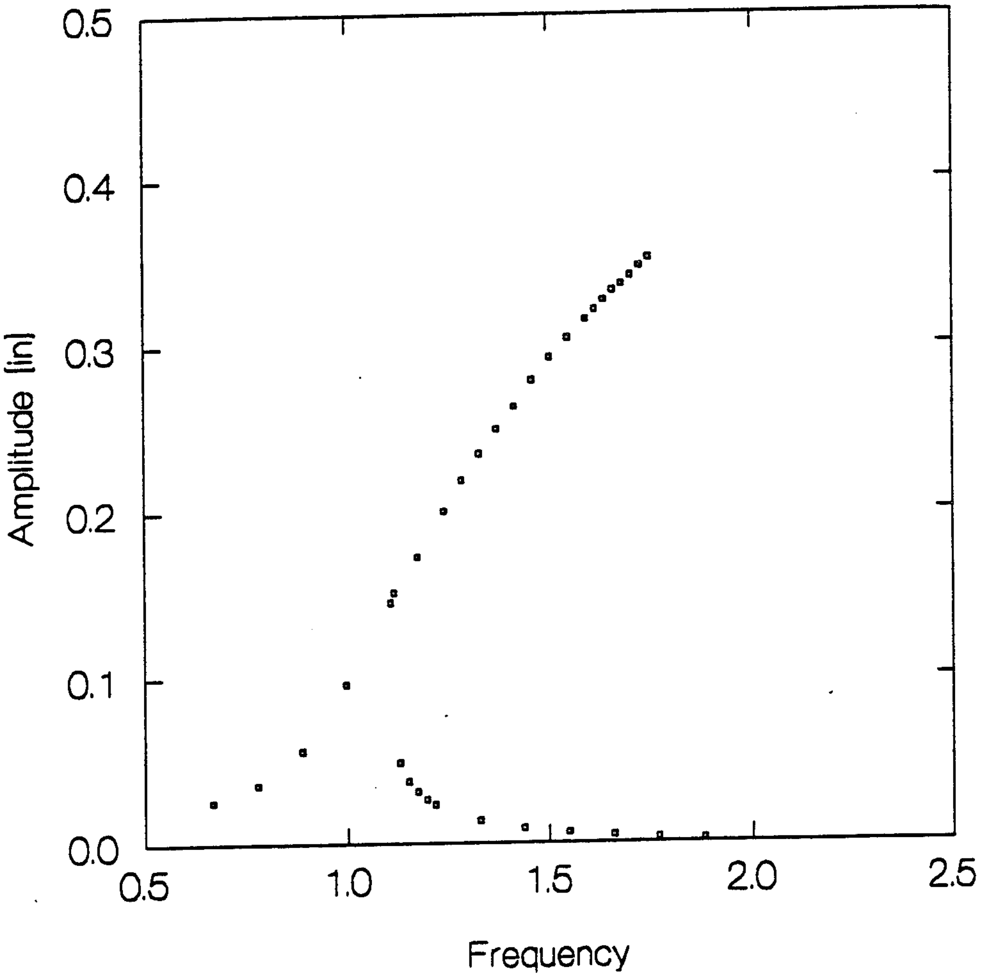


Figure 6.4. Amplitude response, a_o, as a function of nondimensional excitation frequency, Ω/ω_o, for an excitation level of 1.50 g's RMS.

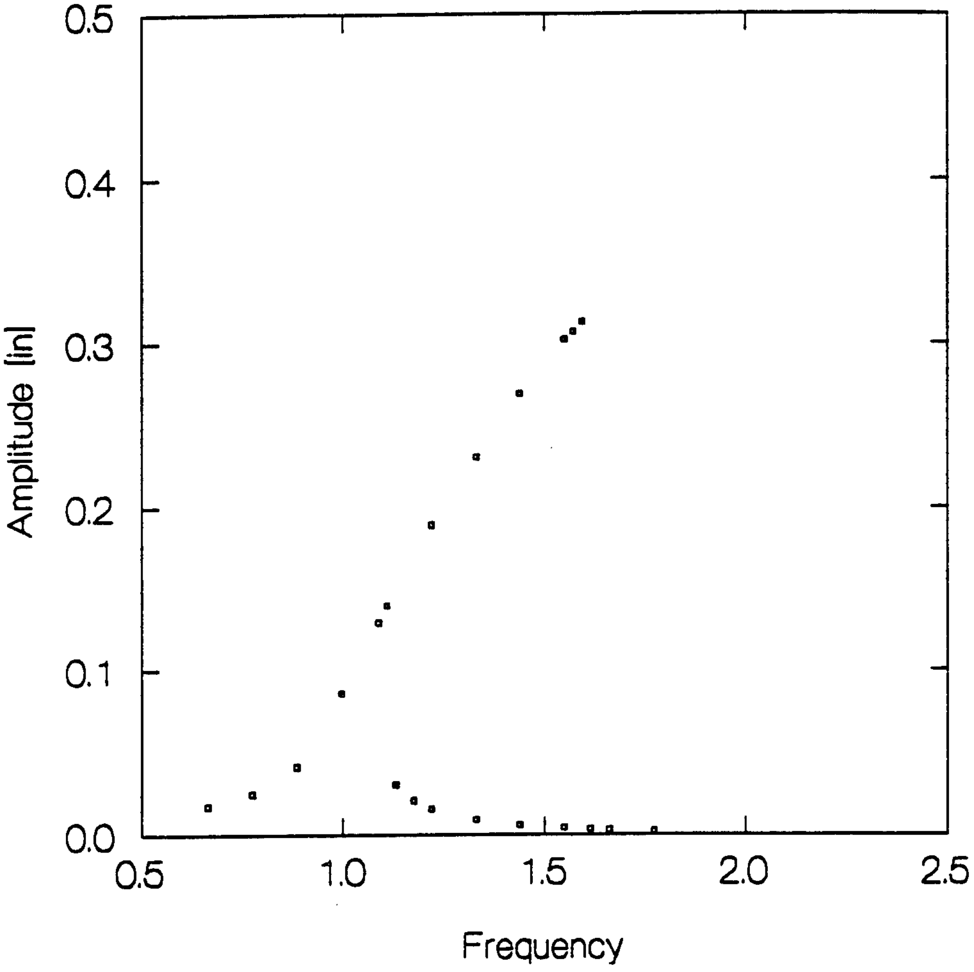


Figure 6.5. Amplitude response, a_o, as a function of nondimensional excitation frequency, Ω/ω_o, for an excitation level of 1.00 g's RMS.

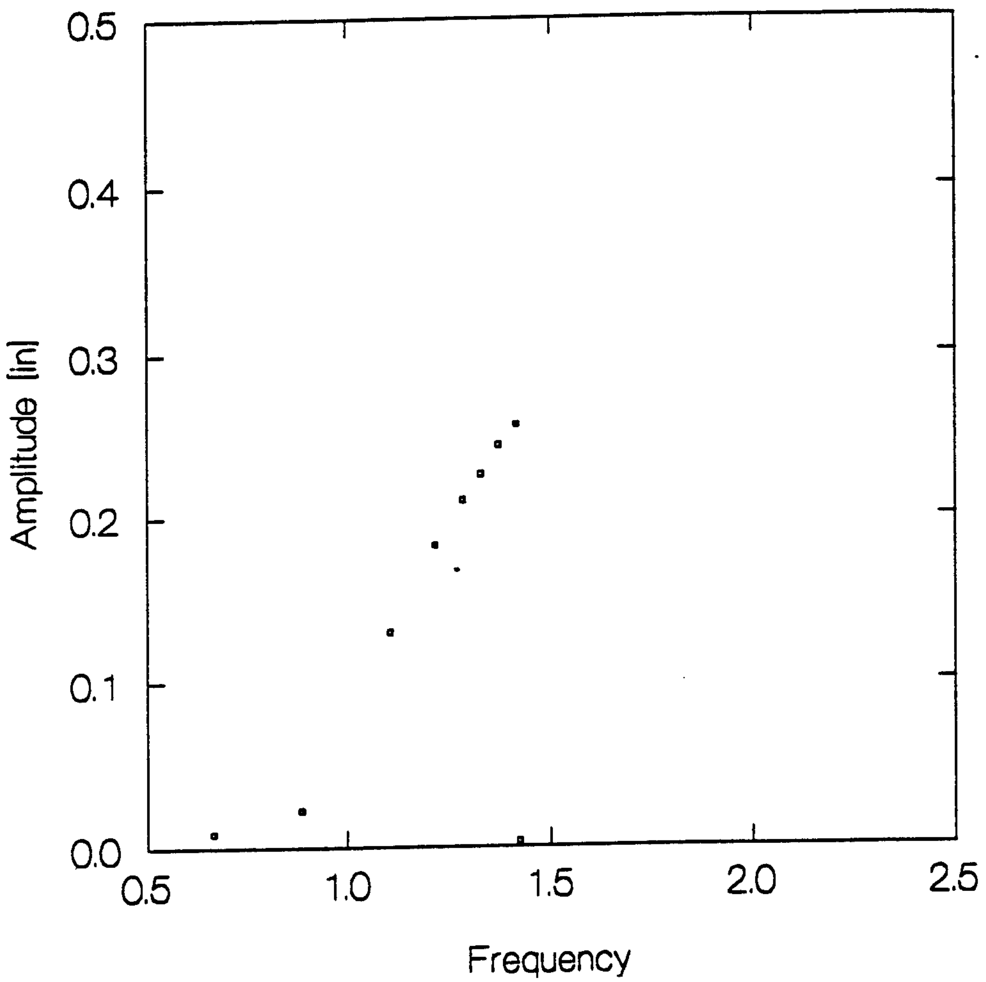


Figure 6.6. Amplitude response, a_o, as a function of nondimensional excitation frequency, Ω/ω_o, for an excitation level of 0.500 g's RMS.

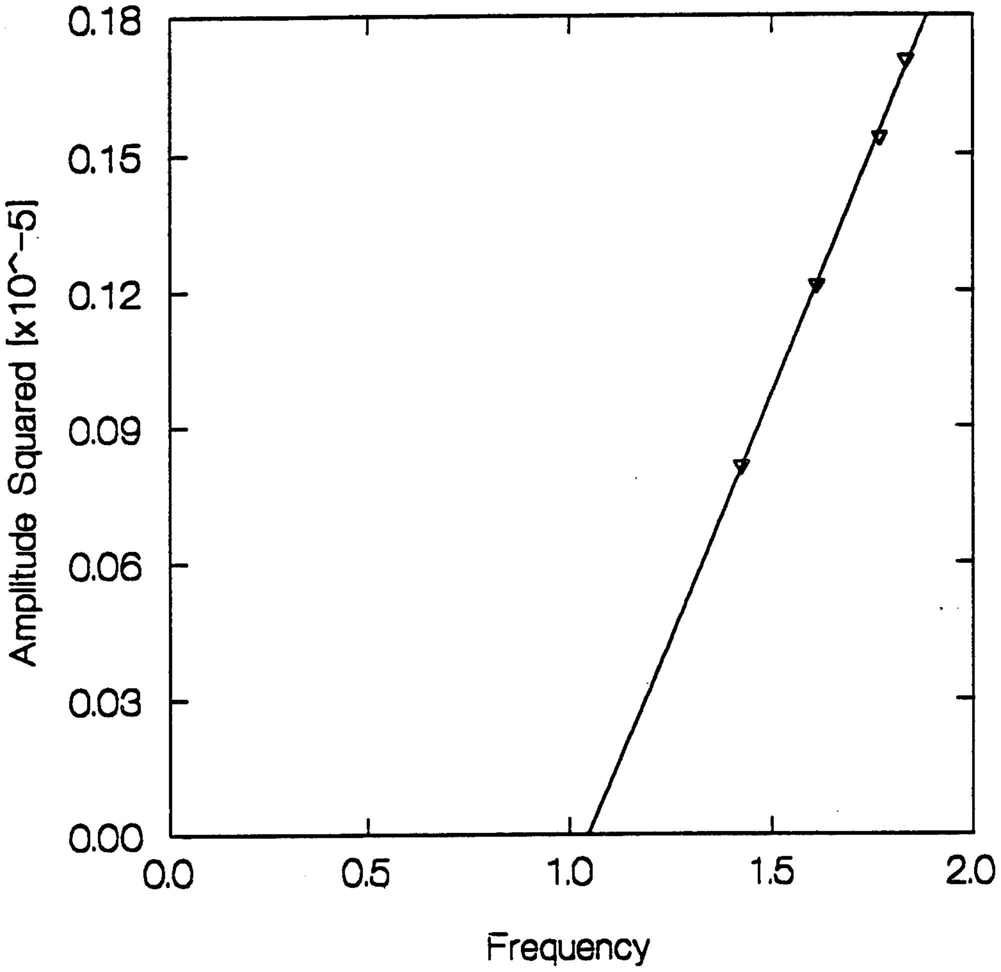


Figure 7.1. Scaled Amplitude Squared $[a^2/\omega_1^2]$ vs. Nondimensionalized Frequency $[\Omega/\omega_1]$.

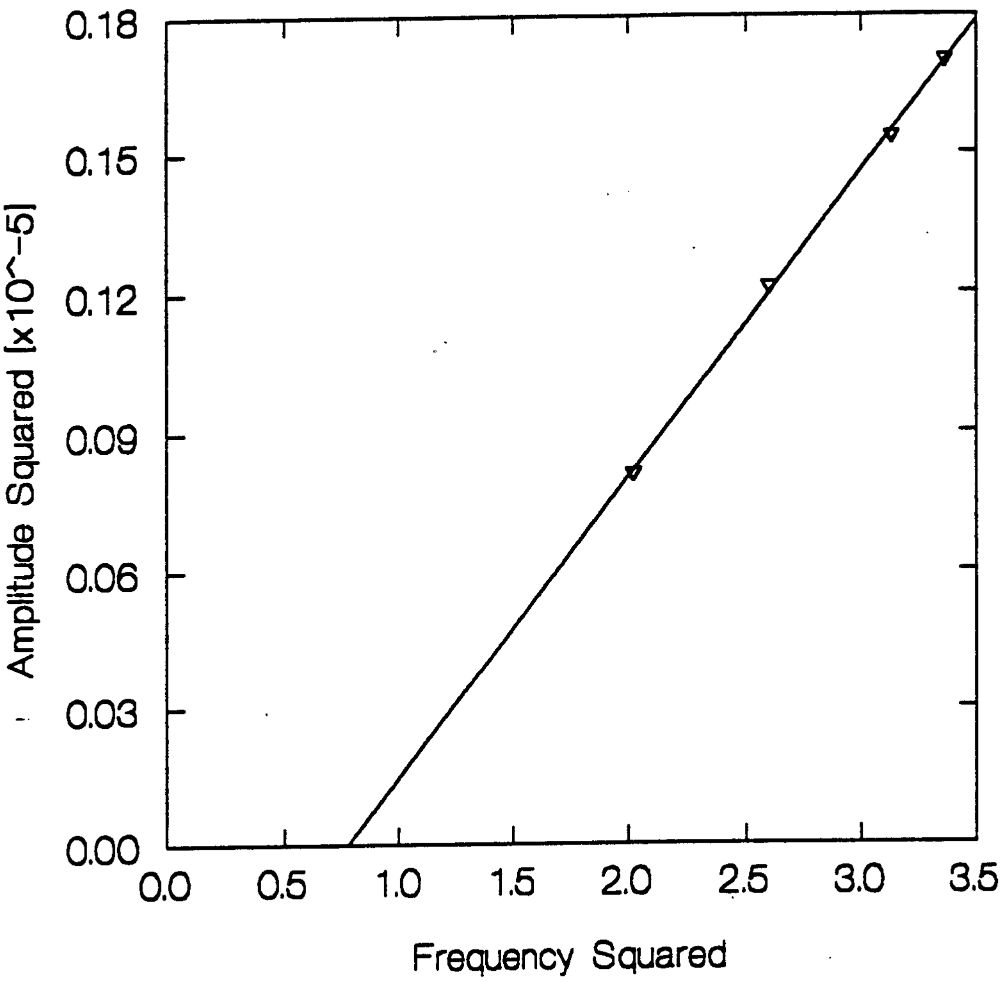


Figure 7.2. Scaled Amplitude Squared $[a^2/\omega_1^2]$ vs. Nondimensionalize Frequency Squared $[\Omega^2/\omega_1^2]$.

A REPORT ON AN INVESTIGATION OF A SENSOR DESIGN FOR STRATA IDENTIFICATION BY AN EARTH PENETRATING DEVICE

Wayne J. Zimmermann
Associate Professor
Department of Mathematics and Computer Science

Texas Woman's University
P.O.Box 22865
Denton, TX 76204

Final Report for:
Research Initiation Program
Eglin AFB-Fuse Data Group

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

and

Texas Woman's University

December, 1992

A REPORT ON AN INVESTIGATION OF A SENSOR DESIGN FOR STRATA IDENTIFICATION BY AN EARTH PENETRATING DEVICE

Wayne J. Zimmermann
Associate Professor
Department of Mathematics and Computer Science
Texas Woman's University

Abstract

This paper reports on the work which addresses two problems associated with strata identification using a penetrating device. The first is concerned with the investigation of the design of a sensor used for strata identification with the requirement that it be a subcomponent of an earth penetrating device. The study approaches the investigation of the various designs by employing various mathematical/computer models. The parameters defining the model can then be used in selecting the shape and composition of the material. The second problem addresses the concept of strata identification under the indicated requirements, that is, by the use of a penetrating device. Thus, the models are used to predict the effect of the strata on the sensor.

AN INVESTIGATION OF A SENSOR DESIGN FOR STRATA IDENTIFICATION FOR USE IN AN EARTH PENETRATING DEVICE

1. Introduction

This paper reports on the work done in addressing the following problem: Consider a cylindrical device which is moving through a layered medium such as the earth. Is it possible to identify the interface between each layer based on the variation of resistive forces? As will be seen this question has meaning only if each layer has a significantly distinct resistive force to the motion of the device. An immediate and related question is: How and what techniques could be used to determine resistive force? A related, but simpler question, is: What equations would describe the expected signal of a sensor moving through a two layer medium, each of infinite thickness? Would the equations define the sensor itself or would they describe an associated electrical output? To allow for simplicity of experimentation, typically the chosen layers could be: air and water. Embedded in these questions is the more complex question: How does the signal relate to the physical activity within the sensor?

An important aspect of this study is that it contains a review of the various design techniques used to develop a single axis sensor having a minimum of ringing associated with the mass-spring configuration.

2. Experimental Procedure -- A General Approach

The purpose of this section is to provide a general description of experimental procedure.

A sensing device is encapsulated within a cavity of a free-fall cylindrical device. The device is supported at a known height. On release the device falls, striking a laminated target consisting of several layers of known materials. The materials are selected according to their hardness. For example, a soft target may consist of four strata: air, balsa, foam and air, where the selection of the indicated target materials is to accommodate the slow velocity of the penetrating device. A hard target may be composed of a number of layers consisting of packed sand, concrete, steel, wood then air. Figure 1(a) illustrates the structure of a typical target. The intent of the experiment is to test the response of various sensors as the device moves through the target. Figure 1(b) indicates a theoretical response while Figure 1(c) represents a current typical sensor response and Figure 1(d) illustrates an acceptable response.

A problem directly related to that of sensor design is the packaging design or mounting design. This problem considers the various methods in which the sensor is to be encapsulated or mounted within the penetrating device. This question presupposes that the sensor is a device which is independent of the penetrating device. Assuming such, we are back to the first question. The question of packaging introduces a number of problems concerned with the method. The various methods may include: one, the sensor is suspended within a cavity filled with spherical glass beads of radius r. To minimize looseness, which is equivalent to increasing the rigidity of the packing material (glass spheres), the packing is done by vibrating the penetrating device. This would support the sensor but would not permit it to move at random. Further, such a packaging configuration would be the cause of information lose due to the deformation characteristics of the packing material. Since the beads are small the quantity used to encapsulate the device would large and hence any modeling of the interaction of the beads, the sensor and the penetrating device would require some type of statistical approach. To reduce the loss of information additional pressure could be applied to the contents of the cavity during vibration, thereby increasing the rigidity in the mounting scheme but does so by deforming the glass beads. But too much pressure would create a state similar to a hard mount, that is, the sensor is mounted directly to an internal surface of the penetrating device.

As with any other system we cannot restrict our investigation simply to the design of the sensor. We must address questions related to signal processing in a real-time domain. Clearly, a fast and accurate method for removing the ringing and thereby generating an acceptable response would reduce the problems related to the sensor's hardware design. The design of such an algorithm could be based on a technique which would remove the sensor's response to the Delta Function from the sensor's response to the target. Three approaches using this approach are currently under investigation: one is based on a minimal least square fit of a linear combination of decayed sinusoidal functions. Another is based on the use of the Fourier Transform and its relatives--Walsh Functions and Wavelets. The third method under consideration is based on the use of a class of neural networks. Regardless of the processing method chosen, the method must be able to identify the transition of the device from one layer to another layer in real-time by correctly interpreting a segment of the sensor's output signal. To provide fast identification the number of sample points must be small and the sampling rate must be high. But if the identification is to be accurate the set of sample points must be sufficiently large. By using a very high speed sampling device we could possibly obtain the larger sampling set. But if the velocity of the device is large then we must either increase the speed of the sampling device or reduce the size of the sample set. Details concerning this dilemma are addressed in a later section.

3. The Mathematical Model of a Cantilever--A Review

Since one of the sensor is base on the use of a cantilever the following mathematical models describing the motion of such a device are reviewed.

A Uniform Cantilever. An age old problem, it is discussed extensively by William Nash, [5], and by Hurty and Rubinstein, [4]. Hurty and Rubinstein base their model on the equation

$$\frac{d^4z}{dx^4} - \frac{m\omega^2}{EI} z = 0 \qquad (1)$$

where ω denotes the natural frequencies, m the mass, E denotes Young's modulus of elasticity, and I denotes the moment of inertia of the cross section. They indicate that this equation holds for a nonuniform beam. For the present we consider the beam to be uniform and the quantities ω, m and EI to be constants. Here the variable x denotes the position of the point from the free end of the beam and z denotes the deflection of the free end. The solution to this fourth-order differential equation takes the general form:

$$z(x) = C_1\cosh\beta x + C_2\sinh\beta x + C_3\cos\beta x + C_4\sin\beta x \qquad (2)$$

where

$$\beta^4 = \frac{m\omega^2}{EI} \tag{3}$$

If we assume the boundary conditions to be:

$$z(0) = \frac{dy}{dx}(0) = 0$$

$$z(L) = \frac{dy}{dx}(L) = 0$$

where L denotes the length of the beam, then the solution takes the form:

$$z(\tfrac{x}{L}) = \frac{1}{2}\,\frac{[\cosh\beta L + \cos\beta L][\sinh\beta L(\frac{x}{L}) - \sin\beta L(\frac{x}{L})]}{\sinh\beta L\ \cos\beta L - \cosh\beta L\ \sin\beta L}$$

$$-\frac{1}{2}\,\frac{[\sinh\beta L + \sin\beta L][\cosh\beta L(\frac{x}{L}) - \cos\beta L(\frac{x}{L})]}{\sinh\beta L\ \cos\beta L - \cosh\beta L\ \sin\beta L} \tag{4}$$

Figure 2 is a plot of deflection z as a function of x according to equation (4). Hence, it defines the curvature of the beam.

We should note that Hurty and Rubinstein used a fourth order linear differential equation to describe the deflection of a uniform cantilevered beam whereas Nash used a second order linear differential equation defined by

$$EI\frac{d^2z}{dx^2} = w(L - x)^2 \tag{5}$$

Clearly this equation is based on amount of curvature induced in the cantilever. Nash solved this equation in general and on applying the boundary conditions defined by the beam's fixed end determined the final form of the deflection curve to be:

$$EIz = -\frac{w}{24}(L - x)^4 - \frac{wL^3}{6}x + \frac{wL^4}{24} \tag{6}$$

Using (6) the maximum deflection is given by

$$EI\ z_{max} = -\frac{wL^4}{8} \tag{7}$$

A Constrained Cantilever. As with the previous model, this is a well defined problem that has been extensively explored. We include a review this configuration since it is used as a basic principle in the design of one possible sensor. Figure 3 illustrates this configuration.

Consider the problem of determining the deflection of a beam subject to a uniform load of w lbs per unit length. If we replace the distributed load by its resultant of wL lbs acting at the midpoint of the length L then on taking moments about fixed point we have

$$\sum M_c = R_1 b - \frac{wL^2}{2} = 0 \quad \text{or} \quad R_1 b = \frac{wL^2}{2} \tag{8}$$

Summing forces vertically

$$\sum F_v = \frac{wL^2}{2b} + R_2 - wL = 0$$

or

$$R_2 = wL - \frac{wL^2}{2b}$$

For $0 < x < a$ the bending moment equation in the left overhanging region is given by $M = -\frac{wx^2}{2}$. Hence, the differential equation of the bent beam in that region is

$$EI\frac{d^2z}{dx^2} + \frac{w}{2}x^2 = 0 \qquad 0 < x < a \tag{10}$$

On solving this equation with appropriate conditions yields

$$EI\ z = -\frac{w}{24}x^4 + C_1\ x + C_2 \qquad (11)$$

The bending moment in the region between the supports, $0 < x < L$, is

$$M = -\frac{z\ x^2}{2} + R_1(x - a)$$

and the differential equation for the beam in this region is

$$EI\frac{d^2z}{dx^2} + \frac{wx^2}{2} - \frac{wL^2}{2b}(x - a) = 0. \quad a < x < L \qquad (12)$$

The general solution to this equation is given by

$$EI\ z = -\frac{w\ x^4}{24} + \frac{wL^2}{12b}(x - a)^3 + C_3x + C_4 \qquad (13)$$

The conditions associated with this configuration are: a) at $x = a$, $z = 0$, b) at $x = a$, $z = 0$, c) at $x = L$, $z = 0$, and d) at $x = a$, the slope given define (11) must equal that defined by (13). Insertion of these conditions yield the solution

$$EI\ z = -\frac{wx^4}{24} + \frac{w(L^4 - a^4)x}{24b} - \frac{wL^2bx}{12} + \frac{wa^4}{24}$$

$$- \frac{w(L^4 - a^4)a}{24b} + \frac{wL^2ab}{12} \qquad (14)$$

for $0 < x < a$. And for $a \leq x < L$

$$EI\ z = -\frac{wx^4}{24} + \frac{w(x^4 - a^4)x}{12b} + \frac{w(L^4 - a^4)x}{24b}$$

$$- \frac{wL^2bx}{12} + \frac{wa^4}{24} - \frac{w(L^4 - a^4)a}{24b} + \frac{wL^2ab}{12}. \qquad (15)$$

Vibrating Cantilever. The above discussions were restricted to describing the shape of the cantilever under a constant force. Here we consider the cantilever as a vibrating mass/spring. If the free end of the beam is displaced an amount z(L,0) on release it will vibrate and the displacement z will be a function of x, the position along the beam, and t, the time. See Figure 4 which illustrates a typical movement.

Kreysizg, [8], indicates that small free vertical vibrations of a uniform cantilever beam are govern by the fourth-order equation

$$\frac{\partial^2 z}{\partial t^2} + c^2 \frac{\partial^4 z}{\partial x^4} = 0 \qquad (16)$$

where $c^2 = EI/\rho A$. Here E denotes Young's modulus of elasticity, I denotes the moment of inertia of the cross section with respect to the y-axis as indicated in Figure 5. The symbol ρ denotes the density, and A the cross sectional area. We note that the Equation 16 is a forth order linear partial differential equation. We say this noting that all nonlinear element, if existing are assumed to be sufficiently small and therefore ignored. Using separation of variables that is z(x,t) = F(x)G(t) it follows that

$$\frac{F^{iv}}{F} = - \frac{G''}{c^2 G} = \beta^4 \text{ (a constant).} \qquad (17)$$

and hence

$$F(x) = A \cos\beta x + B \sin\beta x + C \cosh\beta x + D \sinh\beta x \qquad (18)$$

and

$$G(t) = A \cos(c\beta^2 t) + B \sin(c\beta^2 t) \qquad (19)$$

The resulting solution to the vibrating cantilever is given by $z(x,t) = F(x)\cdot G(t)$.

The above derivation was based on equation (16). If we recall

$$m \frac{\partial^2 z}{\partial t^2} = P(x) \qquad (20)$$

where P(x) is the external load per unit length and we use

$$EI\frac{d^4 W}{dX^2} - T \frac{d^2 W}{dX^2} = P(x) \qquad 0 \leq x \leq L \qquad (21)$$

then using a generalized form of (21) we have

$$m \frac{\partial^2 z}{\partial t^2} = EI \frac{\partial^4 z}{\partial X^2} - T \frac{\partial^2 z}{\partial X^2} \qquad 0 \leq x \leq L \qquad (22)$$

Cole, [3], presents equation (21) without derivation. He simply indicates that it is based on the theory of an elastic beam with tension which supports a given load distribution. Further, he indicates that the equation fits well if the deflection W is small. Clearly, this addresses the problem of nonlinearity. In (22) we replaced W with z since we are using z to represent the deflection. Further, the constants provide in (21) are defined as:

E = constant modulus of elasticity.

I = constant moment of inertia of cross section about the neutral axis.

T = constant external tension.

Applying the technique separation of variables equation (22) takes the form:

$$\frac{EI\ F^{iv}\ -\ T\ F''}{mF} = -\frac{G''}{G} = \beta^4 \tag{23}$$

where β is a constant. This yields two differential equations and associated solutions:

$$\frac{G''}{G} = -\beta^4 \tag{23}$$

which yields

$$G(t) = A\cos(\beta^2 t) + B\sin(\beta^2 t) \tag{24}$$

and

$$\frac{EI\ F^{iv}\ -\ T\ F''}{mF} = \beta^4. \tag{25}$$

To determine the solution to this fourth order linear differential equation we use the auxiliary equation. We find the roots to be

$$R_1 = \sqrt{\frac{T+\sqrt{T^2 + 4\ E\ I\ m\ \beta^4}}{2EI}}$$

$$R_2 = -\sqrt{\frac{T+\sqrt{T^2 + 4\ E\ I\ m\ \beta^4}}{2EI}}$$

$$R_3 = \sqrt{\frac{T-\sqrt{T^2 + 4\ E\ I\ m\ \beta^4}}{2EI}}$$

$$R_4 = -\sqrt{\frac{T-\sqrt{T^2 + 4\ E\ I\ m\ \beta^4}}{2EI}}$$

The roots R_1 and R_2 are real since the term $T^2 + 4\ E\ I\ m\ \beta^4$ is positive. The roots R_3 and R_4 are imaginary since the radicand $T-\sqrt{T^2 + 4\ E\ I\ m\ \beta^4}$ is negative. Hence the solution to (25) is:

$$F(x) = a\mathbf{e}^{R_1 t} + b\mathbf{e}^{R_2 t} + c\mathbf{e}^{R_3 t} + d\mathbf{e}^{R_4 t} \qquad (26)$$

We note that the first two terms are exponential, that is, hyperbolic functions defined over the x-axis. This results from R_1 and R_2 being real. Since R_3 and R_4 are imaginary the last two terms are trigonometric.

Physically, the system is simply a vibrating beam (ruler) clamped to a desk which has had its free end externally displaced then released. As such it will vibrate over time. If we restrict our measurements the free end of the beam then the F component is a constant, F(L), and the solution we are concerned with is:

$$z(L,t) = F(L)\cdot[A\cos(\beta^2 t) + B\sin(\beta^2 t)]. \qquad (27)$$

Clearly equation (27) is simply the harmonic oscillation problem.

Vibrating Cantilever. In the above the cantilever is assume to be enclosed in a motionless environment. In the problem of interest the environment is in motion. Further, the motion is not uniform it does involve acceleration. In Seto's text, [10], a model for a seismic instrument is developed. Figure 6 illustrates the configuration of the sensor's design and indicates that the two variables of interest are z_1 and z_2. The base is attached to the body have an unknown force, due to the resistance the strata imposes on the penetrating device.

The forces acting on the mass are: the spring force,

$$k(\tfrac{1}{2} + \tfrac{1}{2})\ (z_1 - z_2),$$

and the damping force,

$$c(\dot{z}_1 - \dot{z}_2),$$

assuming $z_1 > z_2$. Using $\sum F = ma$, the equation of motion is

$$-k(z_1 - z_2) - c(\dot{z}_1 - \dot{z}_2) = m\ddot{z}_1 \qquad (28)$$

Let z denote the relative motion, i.e., $z = z_1 - z_2$; then

$$z_1 = z - z_2, \quad \dot{z}_1 = \dot{z} - \dot{z}_2 \quad \text{and} \quad \ddot{z}_1 = \ddot{z} - \ddot{z}_2$$

thus (28) takes the form

$$m(\frac{d^2z}{dt^2} + \frac{d^2z_2}{dt^2}) + c\frac{dz}{dt} + k\, z = 0$$

or

$$m\frac{d^2z}{dt^2} + c\frac{dz}{dt} + k\, z = -m\frac{d^2z_2}{dt^2} \qquad (29)$$

If $z_2(t)$ is known then we can solve (29), a second order differential equation defining a forced harmonic oscillator.

Vibrating Constrained Cantilever. Here we make use of equations (18) and (19) because the only load applied to the beam is its reaction to the impact which is equivalent to the assumption that the beam is uniformly loaded. Due to a negative acceleration caused by impact at time t=0 deceleration can be approximated by a step function. The induced force will cause the lever arm of the cantilever to deflect. Inertia will drive the arm beyond the maximum stable deflection given by equation (7). Since the arm is beyond the stable point it will reverse its direction. This motion will carry it through the system's stable deflection point. Thus the cantilever is similar to a vibration spring/mass. To damp the oscillation a STOP is employed. Its function is to shorten the arm's length when the beam is moving in one direction. This increases the frequency but the energy acquired by the STOP dampens the oscillatory motion. The material used to constrain the upward motion should be investigated as to its ability to dampen the beam's vertical movement.

To determine the motion we have approached the problem as follows: let t_{max} denote the time at which the beam has moved from its 'rest' position to maximum deflection and let t_{stop} denote the time for the arm to make contact with the STOP. We can then define $t_1 = t_{max} + t_{stop}$, the time it take for the arm to swing from it's stable position to it maximum deflection and back to the STOP.

Next we can find the t_2, the time it takes the arm to travel from first contact with the STOP to it maximum deflection is the opposite direction then back to the point of non-contact with the STOP. This is similar to finding t_1. We note that if the motion is significantly large the lever will bend around the STOP. Hence we must solve the same equations as used in determining t_1 but with a shorten arm length L'.

We then determine t_3, the travel time for the beam under an unconstrained environment. In this way we determine the time and motion of the beam in its varying state.

To repeat, once the beam has made contact with the STOP it then react with a shorter swing and a higher frequency. This is due directly to the shorten length of the free beam. An effect of this action is to consume kinetic energy within the cantilever and thereby dampen its motion.

4. Design of Experiment

Sensor Design. The current investigations are based to three type of sensors, all of which employ a strain gauge. Hence all measurements of the deformation of a solid body are converted into a voltage level.

Our first approach addressing the problem of sensor design is to make use of a cantilever which would yield an acceptable response. Initially, the design took the form indicated in Figure 3. In this design we see that the beam is not constrained when its movement is downward, but is constrained when the movement is upward. This configuration was selected in that it should help dampen the oscillatory movement of the beam. The principle noted is: A short beam of cross-section C does not oscillate as much, magnitude wise, nor as long as a longer beam having the same cross-section. Further, the material used to build the STOP will absorb some of the energy it is compressible. It is seen that under sufficient resistance to the penetrating device the beam would not be near the contact point with the STOP, hence it would not, for small movement, be constrained. But if the beam's movement is sufficiently large then the STOP's position would have an effect and help damp out the vibration. See Figure 7.

Another sensor being considered is the ENDEVCO 7270A(60K). Eventually the ENDEVCO 7270A(20K) and the ENDEVCO 7270A(6K) will be considered. In addition we are investigating a sensor based on the strain gauge measurements of a small deformable cylinder encapsulated within a cavity of the penetrating device. As before, this configuration will be considered in various environments by varying the materials used to construct the cylinder and by varying the method by which the cylinder is encapsulated within the device. Further the shape of the cylinder could be varied, for example, it could allow for a taper or an inverted taper mounted on a parallelpiped.

General Layout. A penetrating device show in Figure 8 was constructed for the purpose of testing each of the indicated sensors. The sensor is packed in the cavity using various packing material and packing techniques. The first experiments used a spherical glass bead. The sensor was placed in the cavity and additional beads were added thereby enveloping the sensor with beads which are packed to minimize the wasted volume. The procedure used to minimize the void is to pour the beads into the cavity while the device is being vibrated. The device is then supported at a selected height. Initially a height of three feet was used. For greater dropping distances the device is guided along a wire towards its target. Various targets are used and their composition is generally selected from one of the following configurations:

* a block of foam
* a layer of foam with a cavity
* a layer of foam, wood, foam
* a layer of wood
* a layer of packed sand
* slightly cured (green) cement,
* cured cement,

The acquisition system was set to trigger on impact. To guarantee the entire signal would be captured the device would strike a thin layer of wood half a meter above the target to activate the trigger in the acquisition system.

Future experiments are to be repeated for a drops of 6 feet, 20 feet, 50 feet and 150 feet. The target's composition will be modified to accommodate the higher velocities. To provide normality of the penetrating device to the surface of the target at contact a guideline will be used for the larger drops. Further we will use a pair of breaker lights to ascertain the velocity.

Assumptions. In trying to formulate a theoretical model, one based on the acquired data, a number of assumptions will be imposed. They include:

* On striking the target the principal axis of the penetrating device is taken to be normal to the plane of contact surface of the target.
* The only resistive force impose by the target is along the principal axis of the penetrating device.
* The device will continue to move through the target along the line defined by the principal axis.
* The drag due to the frictional contact between the target and the device is zero. Later this will be removed.
* The diameter of the path within the target is larger than the diameter of the penetrating device.
* The material defining each layer of the target is isotropic.
* Catastrophe, the sudden reduction in resistive force due to the blowoff of target material on the exit side of the target is not taking into account.
* For some materials, such as concrete, rock, etc., the reaction to a load is nonlinear. Such materials

will support a load for a time, compressing under the load until the support mechanism break down. At that point the material will shatter and move around the leading edge of the penetrating device.

* The material defining the penetrating device supports a much larger load than the material defining any layer of the target. Hence, the device does not compress significantly with respect to the target material.

5. Review of Mathematical Models

Inverse Method for Evaluating the Sensor's Output. This approach, the Inverse Problem Formulation, is being investigated with the idea that it might yield a fast accurate method for identifying the interface between two distinct strata. To make use of the method requires an accurate model of the phenomena. Since any model describing the event will involve a partial differential equation, such an approach is feasible. The technique consists in selecting the parameters defining the system and generating a solution. The solution is then compared with the data. Once the parameters defining the differential equation are determined the associated solution of the equation is used to describe the phenomena. Regardless of the method employed the goal of any technique is to make a correct identification with a minimum of sample points as the projectile is moving through the media. Hence, our device must sample-process-identify as the projectile moves through the media.

An apparent weakness of this method might appear to be its need to solve an associated equation, but this is not the case. The approach would be to create a small data base which would map the various parameters to the various scenarios. Hence, a fast scheme based on a table look-up would provide the appropriate interpretation. The response would be to simply indicate that the device entered a new strata and that the strata is composed of one of several materials based on hardness, i.e., resistance. Clearly, a device based strictly on resistive forces could not differentiate between to materials or stratas having the same resistive force.

Direct Method for Modeling a Synthetic Sensor. Another approach is to use the data to describe the structure directly by use of the data, hence, rather than determining the differential equation the signal is used to describe the impact forces which are then used to define the strata's composition. This method is referred to as the Forward Problem. In this method we are required to formulate an accurate model described by the current events. To illustrate, a possible model would be able to identify the strata by use of the fact that the signal would indicate a larger current (or voltage) when the device is reacting to a greater resisting force.

In both methods the real problem is noise; noise due to instrumentation and noise due to real event. To begin with, the device is not moving uniformly through the media or the device did not impact the target at a normal thereby resulting in broaching, and the material defining each strata is not isotropic.

Three Smoothing Procedures. In this segment we report that our investigation consisted in testing the feasibility of three methods for processing the sensor's output. The first method is based on a two layered perceptron, a neural network, which is trained moving the device through a high resistant media for a brief period of time. Once the network is trained using an actual system the weights are stored in ROM to be used by other systems. Hence, the production systems will not accommodate learning. The second is to determine a best fit using a linear combination of damped sinusoidal functions. This is done because our claim is that the difference between the acquired data and the data generated by the best fit defines the expected signal as generated by the sensor's response to its environment. A damped sinusoidal is removed from the best fit resulting in a signal which represents the signal defined by those forces due only to impact. Part of the research done here will address the data acquired after the data has been accumulated. The remainder will deal with looking for some a smoothing method appropriate for a small sample set acquired at a high sample rate. This would address the problem of predicting transition of the penetrating device from one layer to another layer in real-time.

6. Sensor Design using Finite Element Method

A Review of FEM. It suffices to note that the formulation of the Finite Element Method used in this study follows that outlined in Baran's text, Finite Analysis on Microcomputers, [1]. In depth details of the process can be found in Bathe and Wilson, [2]. We will omit discussion of this process since we made use of a commercial finite element package, IMAGES-3D by Clestial Software.

FEM Model of a Cantilever. A model for an unconstrained cantilever has been formulated using approximately 100 nodes. It is our intention to increase the node count to 200. Further, we are in the process of implementing a constrained cantilever. To increase speed and accuracy we are currently writing our own FEM model. When all programming is completed this section describes the method used to model a constrained cantilever. The first program will be based on use of a FEM analysis of a cantilever. A second program for modeling the cantilever is based on techniques described by J. M. Haile in his text Molecular Dynamics Simulation, [6], is being written.

FEM Model of a Tapered Cylinder. When all programming is completed and tested this section will describe the method and its results in modeling a small metal cylinder. As with the cantilever the first program will be based on use of a FEM analysis of a cantilever while the second is based on techniques described dynamics simulation, [6]. It should be evident this model is similar to that based on measuring the deformation of the penetrating device itself since the device is cylinder. We will incorporate the actual shape of the cylindrical sensor to allow for taper or inverse taper.

7. Results

Details of data acquisition. For each sensor, which consists of a cantilever with four strain gauges configured as a bridge, we mount the sensor within the cavity of the penetrating device. A voltage is applied to one side of the bridge and an amplified voltage reading is taken from the other side. A RAPID System ADC sampling at 50,000 samples per second is used to capture the output signal for processing. The signal is processed after the experiment has been preformed, hence the problem of identifying interface structures is done by using a sliding subset of the data, thereby emulating realtime processing. Sampling begin at impact by use of a trigger. Current experiments consists of dropping the device from various heights for each given target. Since the dropping distance are not large, a maximum of six feet, the targets selected to date are: styrofoam, sand, dirt and wood.

Review of data in the time domain. The data indicated below show the weakness in the current system. Based on the idea that current over a short time interval correlates to the intensity of the force we call see that it is difficult to determine the intensity of the force. If the system would damp quickly then it should settle to a level proportional to the intensity of the force. Figure 9 and Figure 10 are samples of a typical signal for the situation under investigation.

Review of data in the frequency domain. If we generated the associated frequency domain data based on a short sample of the time domain data we would expect that an increase in the intensity of the force would cause a shift to higher frequencies. Such a system will require that we maintain a log of previous frequency distributions. As the device travels from one media to another the data, as it appears in frequency space, would be compared to the previous data set to see if a frequency shift or spike has occurred. Further, we note that if the time domain data damps quickly then the frequency domain information would reflect these event by recording high frequency data.

Identification Under Time Constraints. Since we are concerned with the problem of using the sensor in a real-time environment, the sensor will be part of an autonomous system which must identify the strata as the device moves through the media. This requires that the classification be based on a minimal number of points obtained during the projectiles transition from one strata to another. Here testing is done on using as few as 10 sample points to as many as 100 samples points. For low velocity projectiles we could use 100 sample points since the device will not travel a great distance. For high velocity projectiles it is necessary to use fewer sample points thereby keeping the distance the projectile has moved during the sampling period to a minimum.

8. Conclusion

In summary this work is far from complete and hence many more experiments are required since the selection of sensor material and sensor design as well as mounting techniques are critical to the effectiveness of the device. Techniques for processing the data are paramount. Formulation of models to aid the designing future devices will play a major role in our continued investigation.

Experimental result have yielded signals in the class indicated by the signal presented in Figures 9 and 10. Future plans include an investigation of various sensors of similar design built around the use of stiffer material. Further, we must somehow implement the theory provided by the seismic sensor if we are to interpret the resulting signal with any significant accuracy.

Lastly, the signal has an embedded sinusoidal signal which we would like to remove. The form of this signal is

$$x = Ate^{-\alpha t}\sin(ft - p) + \sum_{i=1}^{N} A_i e^{-\alpha_i t}\sin(f_i t - p_i)$$

which has nice traits relative to the Fourier Transform. Further, the equation satisfies the ordinary differential equation

$$x'' + [\alpha + \frac{\alpha - 2}{t}]\, x' + [\alpha^2 + f^2 - \frac{2\alpha}{t} + \frac{2}{t^2}]x = 0$$

This equation is similar to Bessel's equation; yet it is significantly different to warrant an further investigation.

References

[1] Baran, Nicholas, Finite Element Analysis on Microcomputers, McGraw-Hill, Inc., New York, NY, 1988

[2] Bathe, K., Wilson, L., Numerical Method in Finite Element Analysis, Prentice Hall, Englewood Cliffs, NJ, 1976

[3] Cole, Julian, Perturbation Methods in Applied Mathematics, Blaisdell Publishing Company, Waltham, MA, 1968

[4] Hurty, Walter, Rubinstein, Moshe, Dynamics of Structures, Prentice Hall, Inc., Englewood, NJ, 1964

[5] Nash, William, Strength of Materials, McGraw-Hill, Inc., New York, NY, 1972

[6] Haile, J. M., Molecular Dynamics Simulation--Elementary Methods, John Wiley & Sons, Inc., New York, 1992

[7] IMAGES-3D, Clestial Software Inc., Berkeley, CA, 1990

[8] Kreyszig, E., Advanced Engineering Mathematics, John Wiley and Sons, Inc., NY, 1964

[9] Romanov, V., Inverse Problems of Mathematical Physics, VNU Science Press, BV, 1987

[10] Seto, William W., Mechanical Vibrations, McGraw-Hill Inc., New York, 1964

[11] Webster, A., Partial Differential Equation of Mathematical Physics, Hafner, NY, 1947

APPENDIX

Miscellany

This section is concerned with the presentation of some relevant facts related to the problem of designing a strata sensor and acquisition system for an earth penetrating device.

First, we should have a general idea of how fast a device is traveling when dropped a distance h in a vacuum on a planet having a gravitational acceleration g. This is a well know formula given by $v = \sqrt{2gh}$ where h denotes the distance of the fall. The initial velocity is assumed to be zero. Based on this formula we can formulate the table:

Table 1. Dropping distance vs velocity

height dropped (feet)	velocity ft/sec
6	19.5
10	25.3
100	80.0
200	113.1

Next, we would like to determine a formula which would yield the average g's incurred in stopping a device within a given distance, say s. The device has traveled a distance h during its fall, hence the velocity is determined by the use of the velocity/distance formula. Now using

$$\int_{\sqrt{2gh}}^{0} v \, dv = - G \int_{0}^{s} dx$$

we have $$gh = Gs.$$

Therefore $$G = \frac{h}{s} \cdot g$$

where h is the distance of fall and s is the stopping distance.

Based on this relationship we now formulate the following table:

Table 2. Number of g's verses height and stopping distance

	stopping distance				
height	200 ft	50 ft	10 ft	1 ft	0.1 ft
6	..	..	..	6	60
10	..	..	1	10	100
10,000	50	200	1,000	10,000	100,000
20,000	100	400	2,000	20,000	200,000

Now to capture 1000 g's for the above we need a sample rate of 24,000 samples per second. Argument: The time to go through 50 feet given the dropping height is 10,000 is: 0.125/50 = 0.0025. If we desired a 10 sample points per foot then we would want a sample point every 0.00025 which equivalently is 4000 samples/sec. The time to go through 10 feet as given above and dropped from 30,000 is: 0.043/1000 = 0.000043. If we wish 10 sample points per foot then we would want a sample point every 0.0000043 which is 24,000 samples/sec

Assuming uniform resistive forces then $t = \frac{v}{\#g\text{'s} \cdot 32}$. For example $t = \frac{800 \text{ f/s}}{50 \cdot 32 \text{ f/s}^2}$ = 0.50 seconds. Hence we can formulate the table below.

Table 3. Stopping times as a function of velocity and g's.

height dropped	velocity	time to stop (body of table)			
(feet)	ft/sec	10g's	50g's	200g's	1000g's
10	20	0.63sec	0.013	0.003	0.0006
10,000	800	2.50sec	0.500	0.120	0.025
20,000	1131	3.53sec	0.710	0.180	0.035
30,000	1385	4.33sec	0.860	0.220	0.043

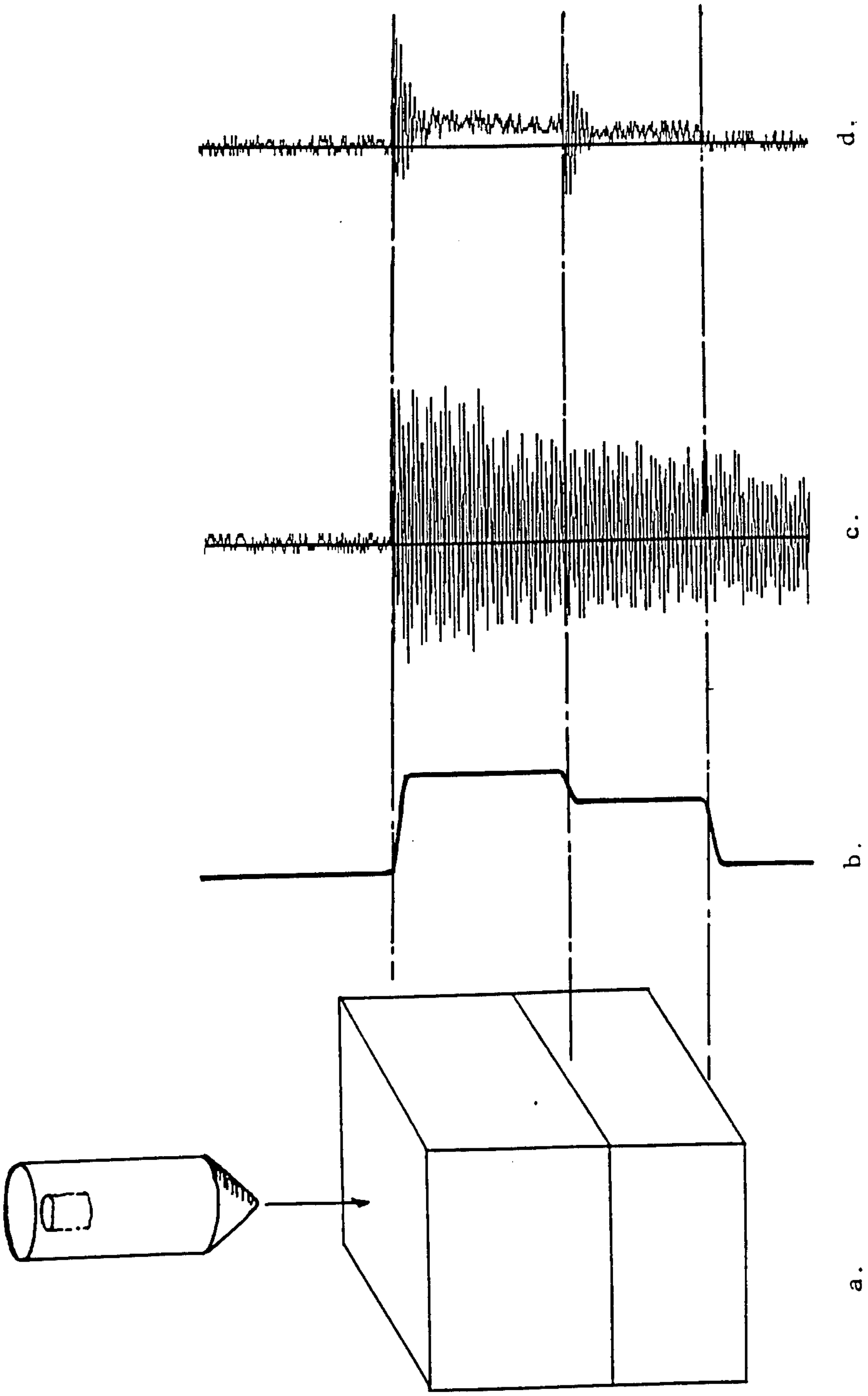


Figure 1. a. A Cross-section of the target, b. A theoretical signal, c. a typical response, and d. a desired response.

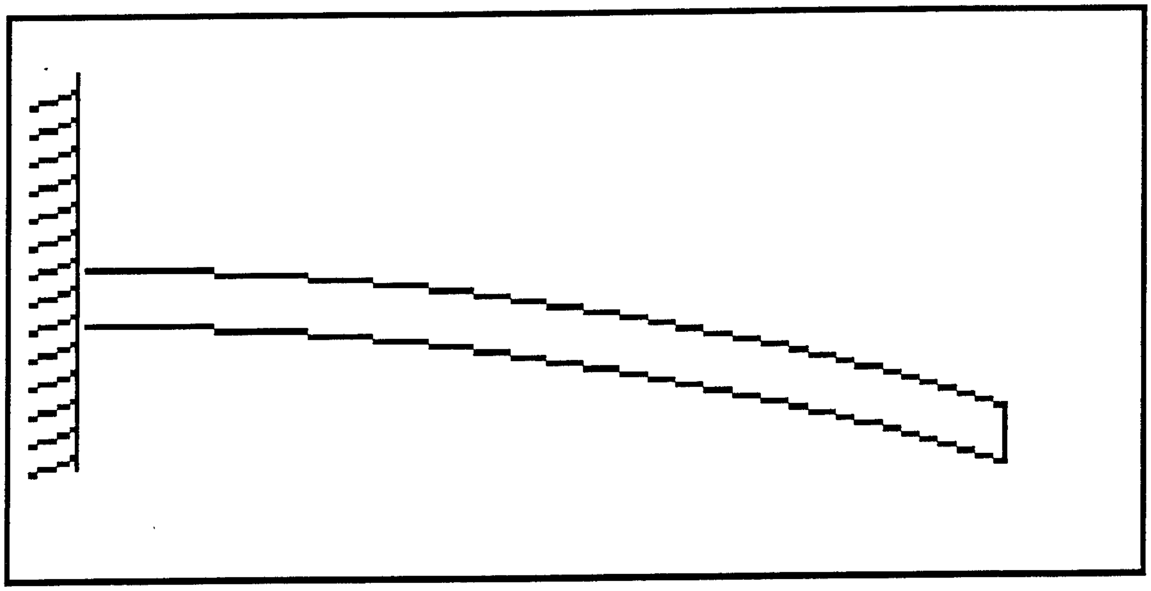


Figure 2. A typical deflection response for a uniform cantilever. The deflection z is a function of the distance, x, along the beam. Defines the shape of the beam.

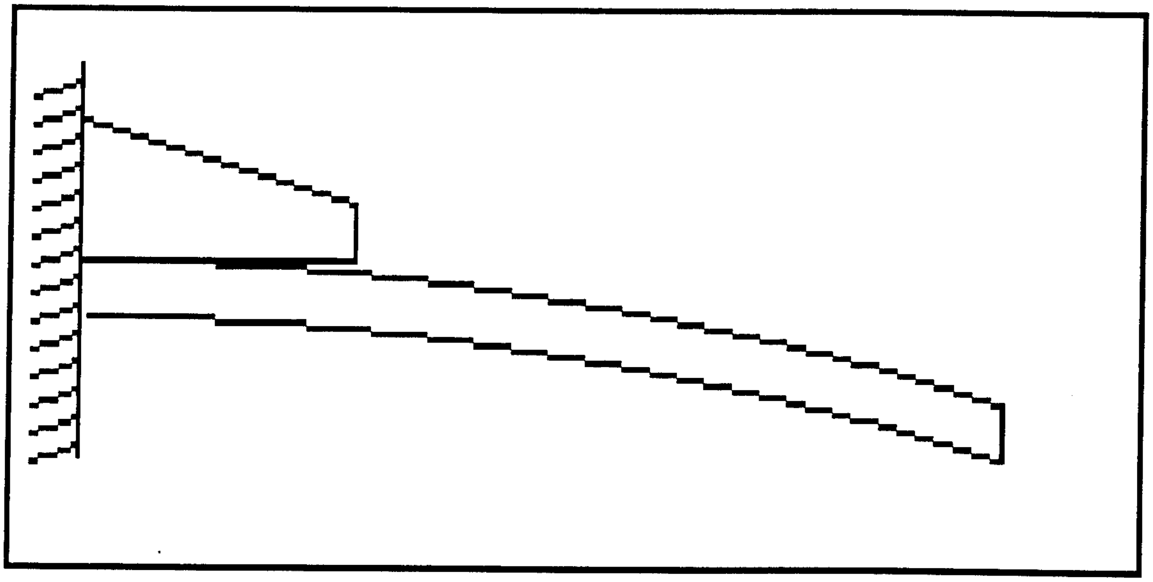


Figure 3. Configuration of a constrained cantilever. The short cantilever is referred to as the STOP. The stop is short and thick to minimize its movement.

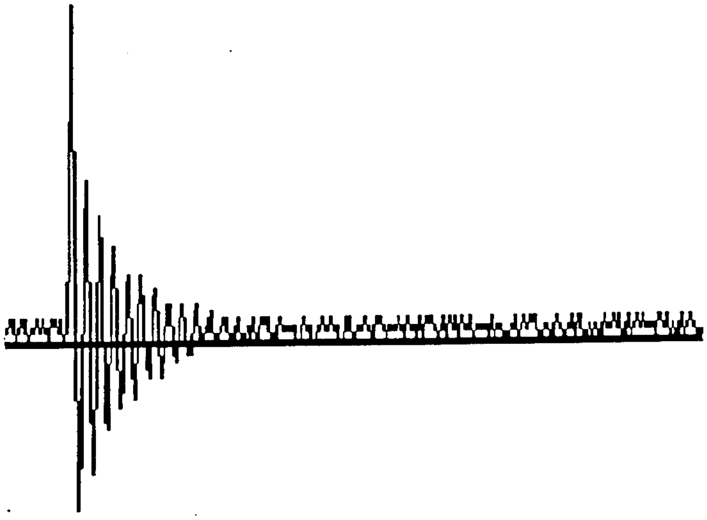


Figure 4. A plot of the deflection of the beam's free end as a function of time.

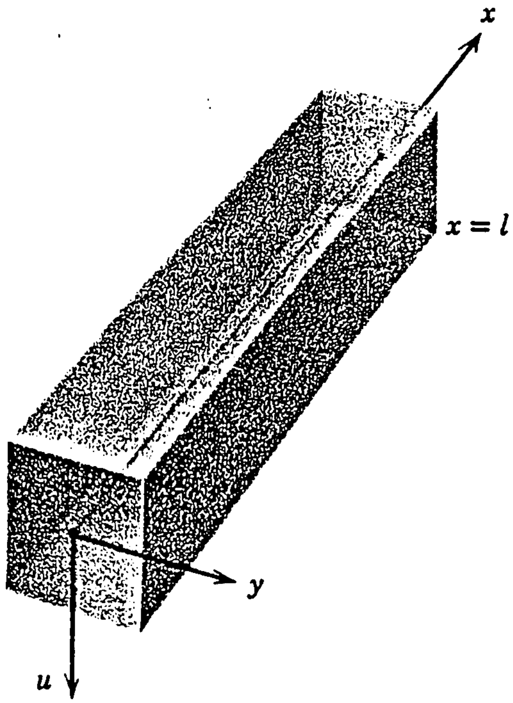


Figure 5. Assigned reference frame for an undeformed beam.

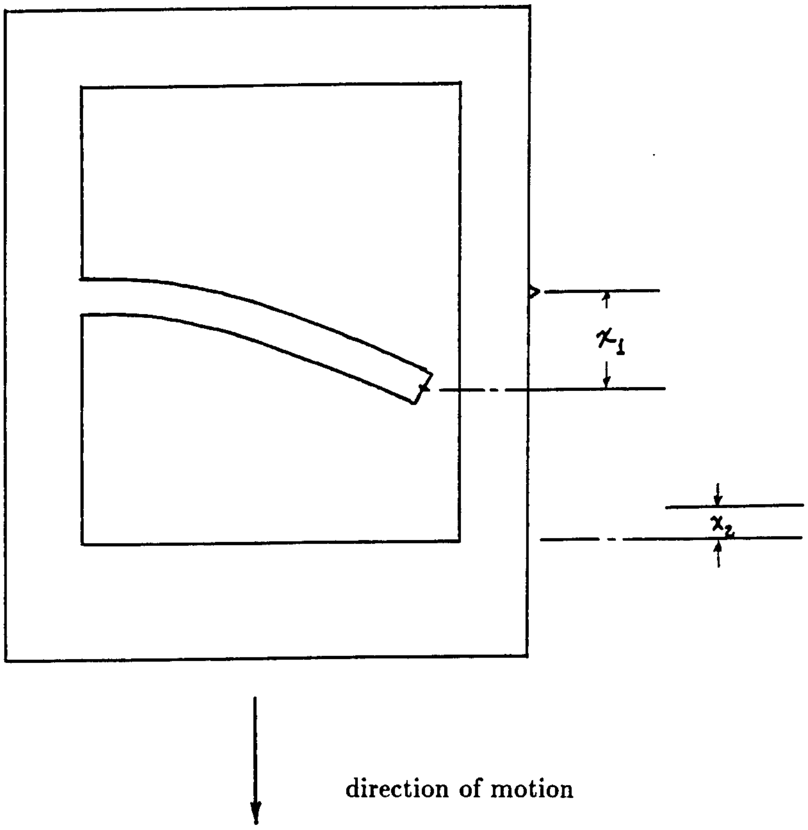


Figure 6. A model for a seismic instrument. The environment is in motion. Here x_1 denotes the movement of the beam and x_2 denotes the movement of the device. The sensor is moving in the indicated direction.

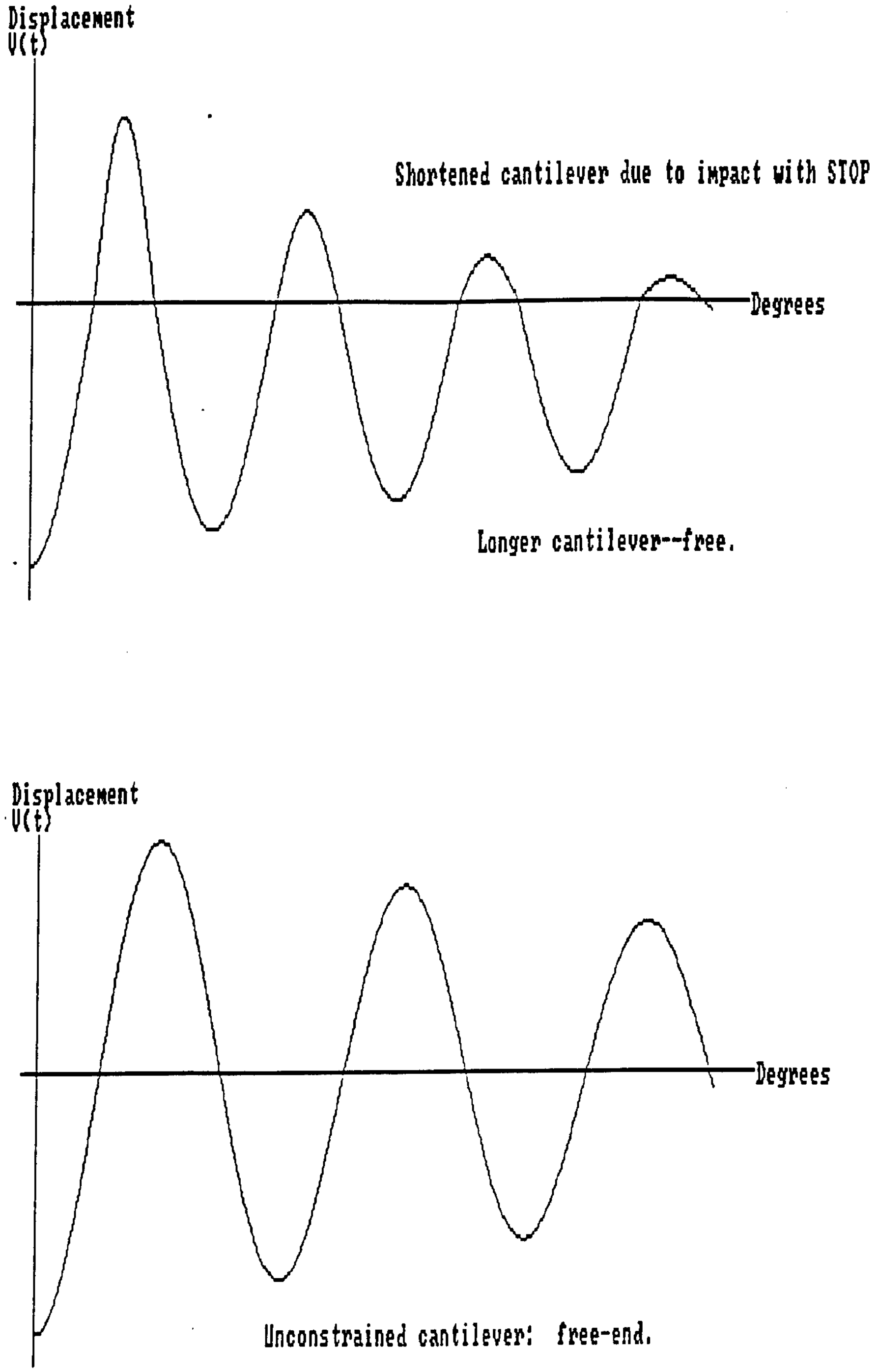


Figure 7. A typical response signal for a constrained beam using a stop. The unconstrained side illustrates lower frequency and a longer swing. The constrained side illustrates the higher frequency and a much shorten swing. The linear segment represents a forced movement of the free end to its selected displacement. The second plot is defined by the motion of an unconstrained cantilever.

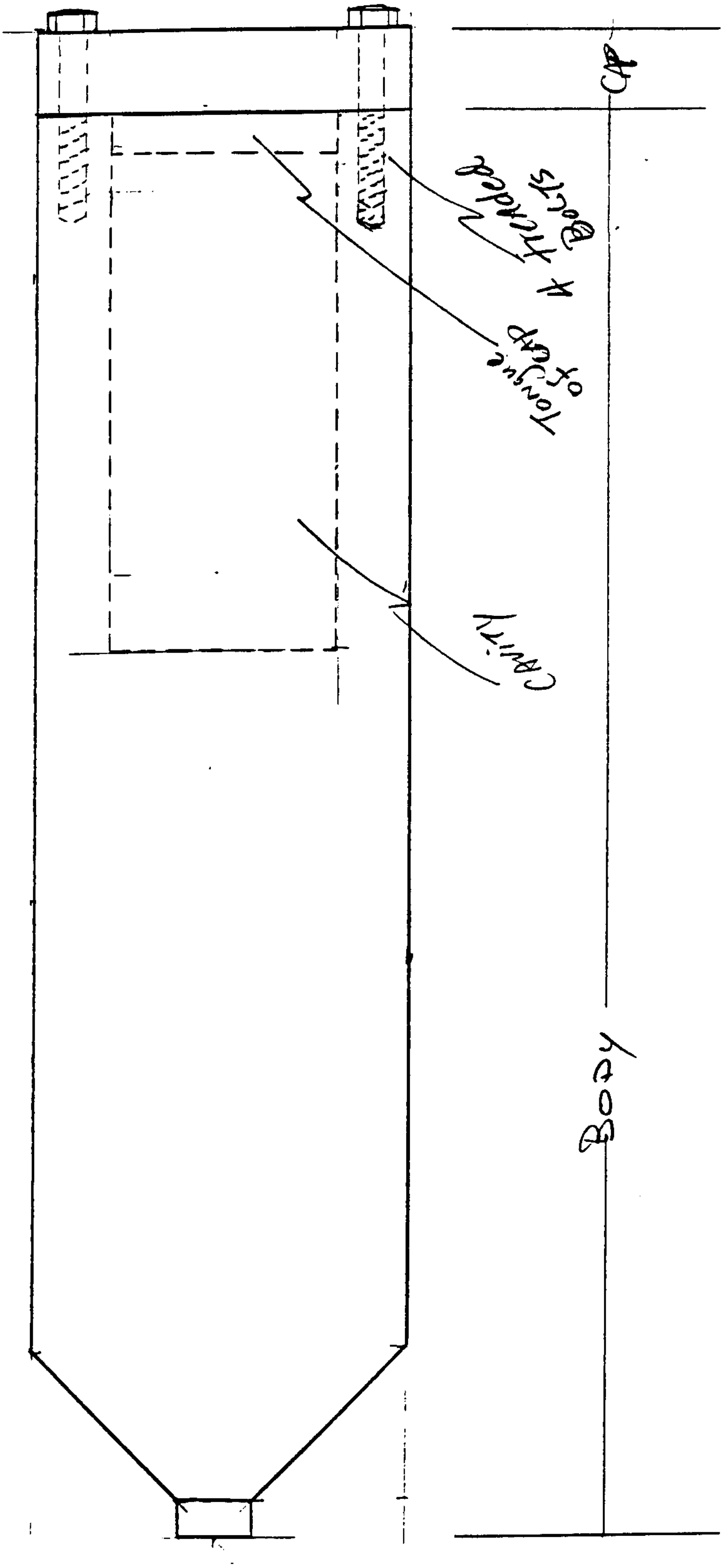


Figure 8. A containment device -- used as a projectile to impact target.

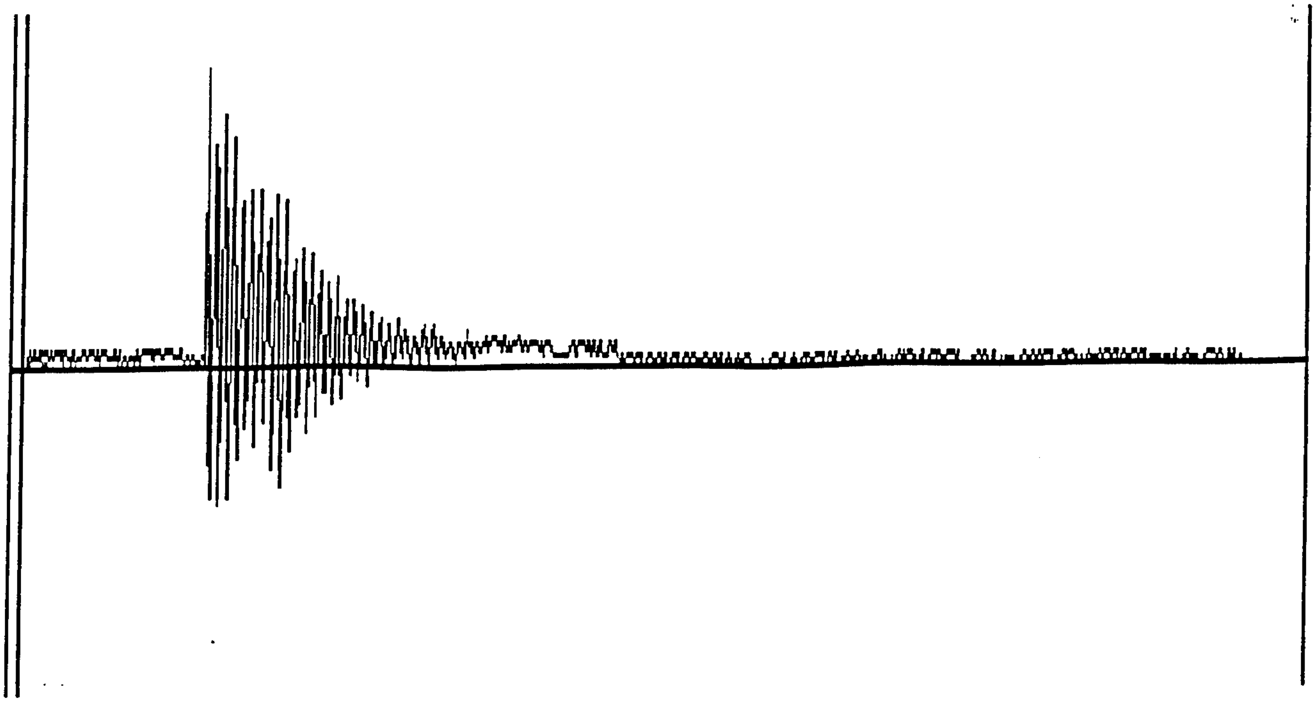


Figure 9. A typical response signal from the impact of the device with a two layered target. Drope height is six feet. Plot is based on using every other sample point.

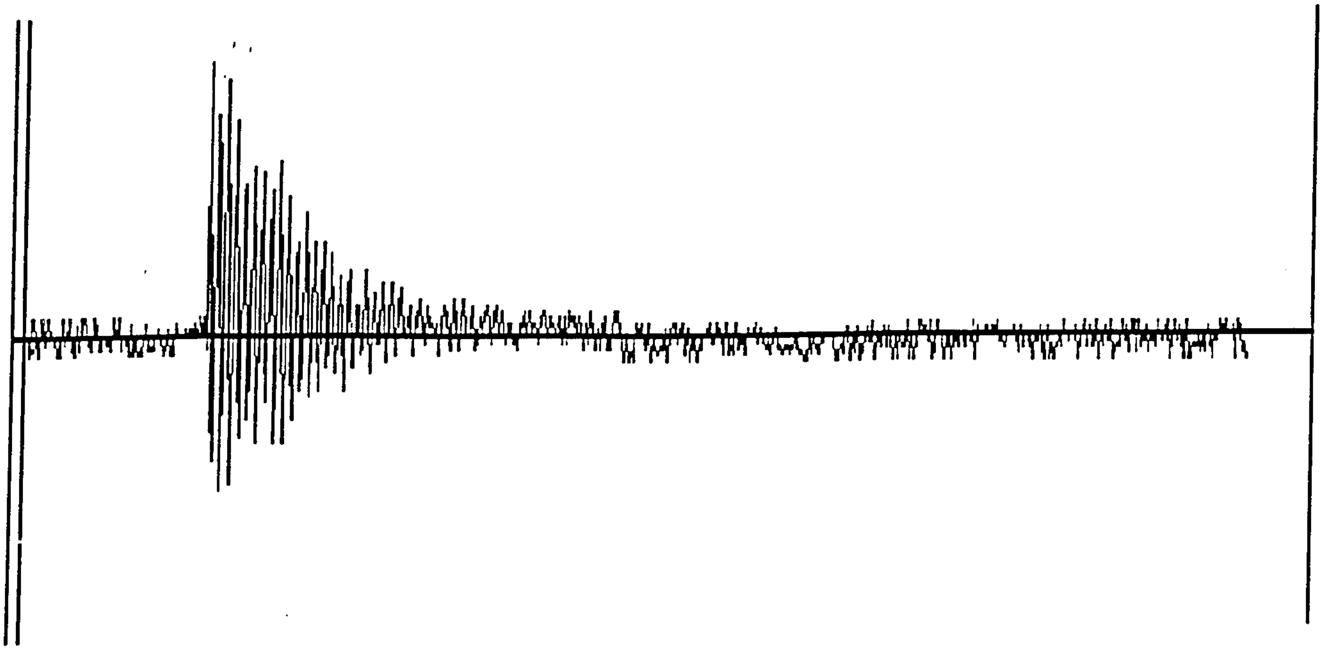


Figure 10. A typical response signal from the impact of the device with a single layered target. Drop height is six feet. Device comes to a halt within the target. Plot is based on using every other sample point.