

UNITED STATES AIR FORCE
SUMMER RESEARCH PROGRAM -- 1998
SUMMER FACULTY RESEARCH PROGRAM FINAL REPORTS

VOLUME 5B
WRIGHT LABORATORY

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

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Colonel Jan Cervený

Program Manager, RDL
Scott Licoscós

Program Administrator, RDL
Johnetta Thompson

Program Administrator, RDL
Rebecca Kelly-Clemmons

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base
Washington, D.C.

December 1998

20010319 055

AQM01-06-1209

PREFACE

Reports in this volume are numbered consecutively beginning with number 1. Each report is paginated with the report number followed by consecutive page numbers, e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3.

Due to its length, Volume 5 is bound in two parts, 5A and 5B. Volume 5A contains #1-17. Volume 5B contains reports #18-35. The Table of Contents for Volume 5 is included in all parts.

This document is one of a set of 15 volumes describing the 1998 AFOSR Summer Research Program. The following volumes comprise the set:

<u>VOLUME</u>	<u>TITLE</u>
1	Program Management Report
	<i>Summer Faculty Research Program (SFRP) Reports</i>
2	Armstrong Laboratory
3	Phillips Laboratory
4	Rome Laboratory
5A & 5B	Wright Laboratory
6	Arnold Engineering Development Center, United States Air Force Academy and Air Logistics Centers
	<i>Graduate Student Research Program (GSRP) Reports</i>
7	Armstrong Laboratory
8	Phillips Laboratory
9	Rome Laboratory
10	Wright Laboratory
11	Arnold Engineering Development Center and Wilford Hall Medical Center
	<i>High School Apprenticeship Program (HSAP) Reports</i>
12	Armstrong Laboratory
13	Phillips Laboratory
14	Rome Laboratory
15A, 15B & 15C	Wright Laboratory

REPORT DOCUMENTATION PAGE			Form Approved	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget.			AFRL-SR-BL-TR-00- and reviewing or Information	
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE December, 1998	3. RE 0781		
4. TITLE AND SUBTITLE 1998 Summer Research Program (SRP), Summer Faculty Research Program (SFRP), Final Reports, Volume 5B, Wright Laboratory		5. FUNDING NUMBERS F49620-93-C-0063		
6. AUTHOR(S) Gary Moore				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Research & Development Laboratories (RDL) 5800 Uplander Way Culver City, CA 90230-6608		8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research (AFOSR) 801 N. Randolph St. Arlington, VA 22203-1977		10. SPONSORING/MONITORING AGENCY REPORT NUMBER		
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release		12b. DISTRIBUTION CODE		
13. ABSTRACT (Maximum 200 words) The United States Air Force Summer Research Program (USAF-SRP) is designed to introduce university, college, and technical institute faculty members, graduate students, and high school students to Air Force research. This is accomplished by the faculty members (Summer Faculty Research Program, (SFRP)), graduate students (Graduate Student Research Program (GSRP)), and high school students (High School Apprenticeship Program (HSAP)) being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory (AFRL) Technical Directorates, Air Force Air Logistics Centers (ALC), and other AF Laboratories. This volume consists of a program overview, program management statistics, and the final technical reports from the SFRP participants at the Wright Laboratory.				
14. SUBJECT TERMS Air Force Research, Air Force, Engineering, Laboratories, Reports, Summer, Universities, Faculty, Graduate Student, High School Student			15. NUMBER OF PAGES	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to **stay within the lines** to meet **optical scanning requirements**.

Block 1. Agency Use Only (*Leave blank*).

Block 2. Report Date. Full publication date including day, month, and year, if available
(e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C - Contract	PR - Project
G - Grant	TA - Task
PE - Program Element	WU - Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es).
Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es).
Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. (*If known*)

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with....; Trans. of....; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."

DOE - See authorities.

NASA - See Handbook NHB 2200.2.

NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.

DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.

Leave blank.

NASA - Leave blank.

NTIS -

Block 13. Abstract. Include a brief (*Maximum 200 words*) factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code (*NTIS only*).

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

SFRP FINAL REPORT TABLE OF CONTENTS	i-x

1. INTRODUCTION	1
2. PARTICIPATION IN THE SUMMER RESEARCH PROGRAM	2
3. RECRUITING AND SELECTION	3
4. SITE VISITS	4
5. HBCU/MI PARTICIPATION	4
6. SRP FUNDING SOURCES	5
7. COMPENSATION FOR PARTICIPATIONS	5
8. CONTENTS OF THE 1996 REPORT	6

APPENDICIES:

A. PROGRAM STATISTICAL SUMMARY	A-1
B. SRP EVALUATION RESPONSES	B-1

SFRP FINAL REPORTS

SRP Final Report Table of Contents

Author	University/Institution Report Title	Armstrong Laboratory Directorate	Vol-Page
DR Harvey Babkoff	Ramat-Gan , Rechovoth Israel , The Impact Of Bright Light and a Moderate Caffeine Dose on Nocturnal Performance: A Preliminary Exp	AFRL/HEP _____	2- 1
DR Michael P Dooley	Iowa State University , Ames , IA Exposure of Female Rats to a 35GHz Electromagnetic Field on Day 1 of Gestation does not Alter Pregna	AFRL/HED _____	2- 2
DR Brent D Foy	Wright State University , Dayton , OH Kinetic Modeling of Slow Dissociation of Bromosulphophthalein from Albumin in Perfusedd Rat Liver:	AFRL/HES _____	2- 3
DR Kenneth A Graetz	University of Dayton , Dayton , OH Negotiation at a Distance: Why You Might Want to Use the Telephone	AFRL/HEN _____	2- 4
DR Verlin B Hinsz	North Dakota State University , Fargo , ND Conceptualizing Crew Performance in Dynamic Operational Environments A Hierachy of Embedded Acton-Co	AFRL/HEC _____	2- 5
DR Nandini Kannan	Univ of Texas at San Antonio , San Antonio , TX Statistical Models for Altitude Decompression Sickness	AFRL/HEP _____	2- 6
DR Ramaswamy Ramesh	Research Foundation of SUNY , Buffalo , NY Aircraft and DT: Modeling and Analysis of Training Effectiveness, Flight Tradeoffs, Costs and Resour	AFRL/HEA _____	2- 7
DR Nancy J Stone	Creighton University , Omaha , NE Reliability and Validity Testing of the Student Characteristics Scale	AFRL/HEJ _____	2- 8
DR Robin D Thomas	Miami University , Oxford , OH Preliminary Decisiion Analysis of the Data Exploitation,mission Planning and Communication (DEMPC) S	AFRL/HEJ _____	2- 9
DR Ram C Tripathy	Univ of Texas at San Antonio , San Antonio , TX The effect of Repeated Measurements of the Variance of The Estiamted od the ahalf Life of Dioxin in	AFRL/HED _____	2- 10
DR Edward W Wolfe	University of Florida , Gainesville , FL Detecting bidimensionality in Response Data: An Empirical Task Analysis Technique	AFRL/HEJ _____	2- 11

SRP Final Report Table of Contents

Author	University/Institution Report Title	Phillips Laboratory Directorate	Vol-Page
DR Graham R Allan	National Avenue , Las Vegas , NM Temporal Characterisation of a Synchronously-Pumped Periodically-Poled Lithium Niobate Optical Param	AFRL/DEL _____	3- 1
DR Mark J Balas	Univ of Colorado at Boulder , Boulder , CO Stable Controller Design for Deployable Precision Structures Using Perturbation Theory	AFRL/VSD _____	3- 2
DR Neb Duric	University of New Mexico , Albuquerque , NM Image Recovery Using Phase Diversity	AFRL/DEB _____	3- 3
DR Arthur H Edwards	University of N. C.- Charlotte , Charlotte , NC Theory of Hydrogen In Sio2	AFRL/VSS _____	3- 4
DR Claudio O Egalon	University of Puerto Rico , Mayaguez , PR Investigating The use of Optical Fiber as Optical Delay Line For Adaptive Optics Systems	AFRL/DEB _____	3- 5
DR Jeffrey F Friedman	University of Puerto Rico , San Juan , PR Low Light Level Adaptive Optics Applied to very High Resoluuiion Imaging	AFRL/DEB _____	3- 6
DR Vincent P Giannamore	Xavier University of Louisiana , New Orleans , LA Environmentally-Benign synthesis of 1,5-Hexadiyne and Related Studies	AFRL/DEB _____	3- 7
DR Gurnam S Gill	Naval Postgraduate School , Monterey , CA Partitioning of Power Aperature Product of Space Based Radar	AFRL/VSS _____	3- 8
DR Robert J Hinde	Univ of Tennessee , Knoxville , TN Computational Aspects of the Spectral Theory of Physical and Chemical Binding	AFRL/DEB _____	3- 9
DR Martin A Hunter	Holy Cross College , Worcester , MA Reaction of Electronically-Excited Nitrogen Atoms with Molecular Oxygen	AFRL/VSB _____	3- 10
DR Brian D Jeffs	Brigham Young University , Provo , UT Deterministic Methods for Blind Restoration of Adaptive Optics Images of Space Objects	AFRL/DES _____	3- 11

SRP Final Report Table of Contents

Author	University/Institution Report Title	Phillips Laboratory Directorate	Vol-Page
DR Donald J Leo	Virginia Tech , Blacksburg , VA self-Sensing Techniquir for Active Acoustic Attenuation	AFRL/VSD _____	3- 12
DR M. Arfin K Lodhi	Texas Tech University , Lubbock , TX Effect of Materials and Design Variations on Amtec Cell Losses	AFRL/VSD _____	3- 13
DR John P McHugh	University of New Hampshire , Durham , NH A Splitting Technique for the anelastic equations in atmospheric physics.	AFRL/VSF _____	3- 14
DR Stanly L Steinberg	University of New Mexico , Albuquerque , NM Lie-Algebraic Representations of Product Intgrals of Variable Matrices	AFRL/DEH _____	3- 15

SRP Final Report Table of Contents

Author	University/Institution Report Title	Rome Laboratory Directorate	Vol-Page
DR Ercument Arvas	Syracuse University, Syracuse, NY Design of a Microwave-To-Optical Link Amplifier For Radar Applications	AFRL/SDN	4- 1
DR Milica Barjaktarovic	Wilkes University, Wilkes Barre, PA Information Protection Tools and Methods	AFRL/IFG	4- 2
DR Stella N Batalama	SUNY Buffalo, Buffalo, NY Outlier Resistant DS-SS Signal Processing	AFRL/IFG	4- 3
DR Digendra K Das	SUNYIT, Utica, NY Modeling and Simulation of Mems Resonators	AFRL/IFT	4- 4
DR Venugopala R Dasigi	, Marietta, GA Toward an Architecture For A Global Information Base	AFRL/CA-I	4- 5
DR Kaliappan Gopalan	Purdue Research Foundation, West Lafayette, IN Amplitude and Frequency Modulation Characteristics of Stressed Speech	AFRL/IFE	4- 6
DR Donald L Hung	Washington State University, Richland, WA A Study on Accelerating the Ray/Triangular-Facet Intersection Computation in Xpatch	AFRL/IFSA	4- 7
DR Adam Lutoborski	Syracuse University, Syracuse, NY On a wavelet-based method of watermarking digital images	AFRL/IFE	4- 8
DR Brajendra N Panda	University of North Dakota, Grand Forks, ND A Model to Analyze Sensor Data For Detection of Multi-Source Attacks	AFRL/IFG	4- 9
DR Jerry L Potter	Kent State University, Kent, OH Architectures for Knowledge Bases	AFRL/IFT	4- 10
DR Salahuddin Qazi	NY Coll of Tech Utica/Rome, Utica, NY Modeling and Implementation of Low Data Rate Modem Using Matlab	AFRL/IFG	4- 11

SRP Final Report Table of Contents

Author	University/Institution Report Title	Rome Laboratory Directorate	Vol-Page
DR Richard R Schultz	University of North Dakota , Grand Forks , ND Image Registration Algorithm Based on the Projective Transformation Model	AFRL/IFE _____	4- 12
DR Kalpathi R Subramanian	University of N. C.- Charlotte , Charlotte , NC Enhancements to Cubeworld	AFRL/IFSA _____	4- 13
DR Shambhu J Upadhyaya	SUNY Buffalo , Buffalo , NY a Distributed Concurrent Intrusion Detection Scheme Based on Assertions	AFRL/IFG _____	4- 14
DR Robert E Yantorno	Temple University , Philadelphia , PA Co-Channel Speech and Speaker Identification Study	AFRL/IFE _____	4- 15

SRP Final Report Table of Contents

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Page
DR Farid Ahmed	Penn State Uni-Erie, Erie, PA Multiresolutional Information Feature for Dynamic Change Detecton in image Sequences	AFRL/SNA _____	5- 1
DR Kevin D Belfield	University of Central Florida, Orlando, FL Synthesis of 7-Benzothiazol-2YL-9,9-Didecylfluorene-2-Ylamine a versatile Intermediate for a New Ser	AFRL/ML _____	5- 2
DR Daniel D Bombick	Wright State University, Dayton, OH	AFRL/PRS _____	5- 3
DR Frank M Brown	University of Kansas, Lawrence, KS Recognizing Linearities In Manterials Databases	AFRL/ML _____	5- 4
DR Gregory A Buck	S Dakota School of Mines/Tech, Rapid City, SD Characterization of Acoustic Sources for Hypersonic Receptivity Research	AFRL/VAA _____	5- 5
DR Joe G Chow	Florida International Univ, Miami, FL Some Critical Issues of The Next Generation Transparency Program	AFRL/VAV _____	5- 6
DR Peter J Disimile	University of Cincinnati, Cincinnati, OH Documentation of the Airflow Patterns within and aircraft Engine Nacelle Simulator	AFRL/VAV _____	5- 7
DR Numan S Dogan	Tuskegee University, Tuskegee, AL Sensors for Focal Plane Array Passive Millimeter-Wave Imaging	AFRL/MN _____	5- 8
DR James M Fragomeni	Ohio University, Athens, OH Mechanical Strength Modeling of Particle strengthened Nickel-Aluminum Alloys Strengthened By Interne	AFRL/ML _____	5- 9
DR Zewdu Gebeyehu	Tuskegee University, Tuskegee, AL Synthesis & Characterization of Metal-Thioacid & Dihydrogen Phosphate Complexes Useful as Nonlinear	AFRL/MLP _____	5- 10
DR Patrick C Gilcrease	University of Wyoming, Laramie, WY Biocatalysis of Biphenyl and Diphenylacetylene to Synthesize Polymer Precursors	AFRL/ML _____	5- 11

SRP Final Report Table of Contents

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Page
DR David E Hudak	Ohio Northern University , Ada , OH Permanence Modeling and Scalability Analysis of the Navier-Stokes Solver FDL3DI Across Multiple Platfo	AFRL/VAA _____	5- 12
DR William P Johnson	University of Utah , Salt Lake City , UT Sorption of a Non-Ionic Suractant Versus a Dissolved Humic Substance to a Low Orgnaic Carbon Soil	AFRL/ML _____	5- 13
DR Jeffrey D Johnson	University of Toledo , Toledo , OH Using Neural Networks to Control a Tailless Fighter Aircraft	AFRL/VAC _____	5- 14
DR Jayanta S Kapat	University of Central Florida , Orlando , FL Fuel-Air Heat Exhcager For Cooled Cooling Air Systems with Fuel-Mist and Air-Jet Impingement	AFRL/PRT _____	5- 15
DR Vikram Kapila	Polytechnic Inst of New York , Brooklyn , NY Spacecraft Formation Flying: A Survey	AFRL/VAC _____	5- 16
DR Kenneth D Kihm	Texas Engineering Experiment Station , College Station , TX Micro-Scale Visualization of Thin Meniscus & Capillary Pore Flows of Capillary-Driven Heat Transfer	AFRL/VAV _____	5- 17
DR Lok C Lew Yan Voon	Worcester Polytechnic Inst , Worcester , MA Many-Body Theory of Quantum-Well Gain Spectra	AFRL/SND _____	5- 18
DR Rongxing Li	Ohio State University , Columbus , OH A Study fo Referencing Issues in Multiplatform and multisensor Based Object Location	AFRL/SNA _____	5- 19
DR Chun-Shin Lin	Univ of Missouri - Columbia , Columbia , MO Sensor Fusion w/Passive Millimeter Wave & Laser Radar for Target Detection	AFRL/MN _____	5- 20
DR Chaoqun Liu	Louisiana Tech University , Ruston , LA Boundary Conditions for Direct Numerical Simulation of Turbulent Flow	AFRL/VAA _____	5- 21
DR Carl E Mungan	University of Florida , Pensacola , FL Bidirectional Reflectance Distr. Functions Describing Firts-Surface Scattering	AFRL/MN _____	5- 22

SRP Final Report Table of Contents

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Page
DR Amod A Ogale	Clemson University , Clemson , SC Characterization of Microstructure Evolution in Pitch-Based Carbon Fibrers During Heat Treatment	AFRL/ML	5- 23
DR Carlos R Ortiz	Universidad Politecnica de Puerto Rico , Hato Rey , PR Simulation of the Antenna Pattern of Arbitrarily Oriented Very Large Phase/Time-Delay Scanned Antenn	AFRL/SNR	5- 24
DR Ramana M Pidaparti	Indiana U-Purdue at Indianap , Indianapolis , IN Flutter Prediction Methods for Aeroelastic Design Optimization	AFRL/VAS	5- 25
DR Stephen E Sadow	Mississippi State University , Mississippi State , MS Characterization of BN-Doped SiC Epitaxial Layers	AFRL/PRP	5- 26
DR Rathinam P Selvam	Univ of Arkansas , Fayetteville , AR Computer Modelling of Nonlinear Viscous Panel Flutter	AFRL/VAA	5- 27
DR Paavo Sepri	Florida Inst of Technology , Melbourne , FL A computational Study of Turbine Blade Interactions with Cylinder Wakes at Various Reynolds Numbers	AFRL/PRT	5- 28
DR Mo-How H Shen	Ohio State University , Columbus , OH Development of a Probabilistic Assessment Framework for High Cycle Fatigue Failures of gas Turbine E	AFRL/ML	5- 29
DR Hongchi Shi	Univ of Missouri - Columbia , Columbia , MO A Study of Models and Tools for Programming the VGI Parallel Computer	AFRL/MN	5- 30
DR Donald J Silversmith	Wayne State University , Detroit , MI Joule Heating Simulation of Poly-Silicon Thermal Micro-Actuators	AFRL/SNH	5- 31
DR Mehrdad Soumekh	SUNY Buffalo , Amherst , NY Alias-Free Processing of P-3 SAR Data	AFRL/SNR	5- 32
DR Joseph W Tedesco	Auburn University , Auburn , AL HIGH Velocity Penetration of Layered Grout Targets	AFRL/MN	5- 33

SRP Final Report Table of Contents

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Page
DR Mitch J Wolff	Wright State University , Dayton , OH Enhancements to A Direct Aeroclastic Stability Computational Model	AFRL/VAS _____	5- 34
DR Jeffrey L Young	University of Idaho , Moscow , ID A Detailed Study of the Numerical Properties of FDTD Algorithms for Dispersive Media	AFRL/VAA _____	5- 35

SRP Final Report Table of Contents

Author	University/Institution Report Title	Laboratory Directorate	Vol-Page
DR F. N. Albahadily	University of Central Oklahoma, Edmond, OH Effect of Environmental Variables on Aging Aircraft	OCALC _____	6 - 1
MS Shelia K Barnett	Mercer Univ, Macon, GA A Study of Scheduling and Tracking of Parts in the Plating Shop at Warner Robins Air Logistics Center	WRALC/TI _____	6 - 2
DR Ryan R Dupont	Utah State University, Logan, UT Natural Attenuation Evaluation Summary for a Chlorinated Solvent Plume, OUI, Hill AFB, Utah	OOALC/E _____	6 - 3
DR Carl L Enloe	James Madison Univ, Harrisonburg, VA A Device for Experimental Measurements of Elelctrostatic Shielding in a Spatially Non-Uniform Plasma	HQUSAF/D _____	6 - 4
DR Mark R Fisher	Southern Polytechnic State University, Marietta, GA Neural Network Control of Wind Tunnels for Cycle Time Reduction	AEDC _____	6 - 5
DR Sheng-Jen Hsieh	Pan American University, Edinbrg, TX Thermal Signature for Circuit Card Fault Identification	SAALC/TI _____	6 - 6
DR Suk B Kong	Incarnate Word College, San Antonio, TX Studies on The Amphetamine Derivatives and Analytical Standards	WHMC/59 _____	6 - 7
DR Kevin M Lyons	North Carolina State U-Raleigh, Raleigh, NC Filtered-Rayleigh Scattering in Reacting and Non-Reacting Flow	AEDC _____	6 - 8

1. INTRODUCTION

The Summer Research Program (SRP), sponsored by the Air Force Office of Scientific Research (AFOSR), offers paid opportunities for university faculty, graduate students, and high school students to conduct research in U.S. Air Force research laboratories nationwide during the summer.

Introduced by AFOSR in 1978, this innovative program is based on the concept of teaming academic researchers with Air Force scientists in the same disciplines using laboratory facilities and equipment not often available at associates' institutions.

The Summer Faculty Research Program (SFRP) is open annually to approximately 150 faculty members with at least two years of teaching and/or research experience in accredited U.S. colleges, universities, or technical institutions. SFRP associates must be either U.S. citizens or permanent residents.

The Graduate Student Research Program (GSRP) is open annually to approximately 100 graduate students holding a bachelor's or a master's degree; GSRP associates must be U.S. citizens enrolled full time at an accredited institution.

The High School Apprentice Program (HSAP) annually selects about 125 high school students located within a twenty mile commuting distance of participating Air Force laboratories.

AFOSR also offers its research associates an opportunity, under the Summer Research Extension Program (SREP), to continue their AFOSR-sponsored research at their home institutions through the award of research grants. In 1994 the maximum amount of each grant was increased from \$20,000 to \$25,000, and the number of AFOSR-sponsored grants decreased from 75 to 60. A separate annual report is compiled on the SREP.

The numbers of projected summer research participants in each of the three categories and SREP "grants" are usually increased through direct sponsorship by participating laboratories.

AFOSR's SRP has well served its objectives of building critical links between Air Force research laboratories and the academic community, opening avenues of communications and forging new research relationships between Air Force and academic technical experts in areas of national interest, and strengthening the nation's efforts to sustain careers in science and engineering. The success of the SRP can be gauged from its growth from inception (see Table 1) and from the favorable responses the 1997 participants expressed in end-of-tour SRP evaluations (Appendix B).

AFOSR contracts for administration of the SRP by civilian contractors. The contract was first awarded to Research & Development Laboratories (RDL) in September 1990. After completion of the 1990 contract, RDL (in 1993) won the recompetition for the basic year and four 1-year options.

2. PARTICIPATION IN THE SUMMER RESEARCH PROGRAM

The SRP began with faculty associates in 1979; graduate students were added in 1982 and high school students in 1986. The following table shows the number of associates in the program each year.

YEAR	SRP Participation, by Year			TOTAL
	SFRP	GSRP	HSAP	
1979	70			70
1980	87			87
1981	87			87
1982	91	17		108
1983	101	53		154
1984	152	84		236
1985	154	92		246
1986	158	100	42	300
1987	159	101	73	333
1988	153	107	101	361
1989	168	102	103	373
1990	165	121	132	418
1991	170	142	132	444
1992	185	121	159	464
1993	187	117	136	440
1994	192	117	133	442
1995	190	115	137	442
1996	188	109	138	435
1997	148	98	140	427
1998	85	40	88	213

Beginning in 1993, due to budget cuts, some of the laboratories weren't able to afford to fund as many associates as in previous years. Since then, the number of funded positions has remained fairly constant at a slightly lower level.

3. RECRUITING AND SELECTION

The SRP is conducted on a nationally advertised and competitive-selection basis. The advertising for faculty and graduate students consisted primarily of the mailing of 8,000 52-page SRP brochures to chairpersons of departments relevant to AFOSR research and to administrators of grants in accredited universities, colleges, and technical institutions. Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs) were included. Brochures also went to all participating USAF laboratories, the previous year's participants, and numerous individual requesters (over 1000 annually).

RDL placed advertisements in the following publications: *Black Issues in Higher Education*, *Winds of Change*, and *IEEE Spectrum*. Because no participants list either *Physics Today* or *Chemical & Engineering News* as being their source of learning about the program for the past several years, advertisements in these magazines were dropped, and the funds were used to cover increases in brochure printing costs.

High school applicants can participate only in laboratories located no more than 20 miles from their residence. Tailored brochures on the HSAP were sent to the head counselors of 180 high schools in the vicinity of participating laboratories, with instructions for publicizing the program in their schools.

High school students selected to serve at Wright Laboratory's Armament Directorate (Eglin Air Force Base, Florida) serve eleven weeks as opposed to the eight weeks normally worked by high school students at all other participating laboratories.

Each SFRP or GSRP applicant is given a first, second, and third choice of laboratory. High school students who have more than one laboratory or directorate near their homes are also given first, second, and third choices.

Laboratories make their selections and prioritize their nominees. AFOSR then determines the number to be funded at each laboratory and approves laboratories' selections.

Subsequently, laboratories use their own funds to sponsor additional candidates. Some selectees do not accept the appointment, so alternate candidates are chosen. This multi-step selection procedure results in some candidates being notified of their acceptance after scheduled deadlines. The total applicants and participants for 1998 are shown in this table.

1998 Applicants and Participants			
PARTICIPANT CATEGORY	TOTAL APPLICANTS	SELECTEES	DECLINING SELECTEES
SFRP	382	85	13
(HBCU/MI)	(0)	(0)	(0)
GSRP	130	40	7
(HBCU/MI)	(0)	(0)	(0)
HSAP	328	88	22
TOTAL	840	213	42

4. SITE VISITS

During June and July of 1998, representatives of both AFOSR/NI and RDL visited each participating laboratory to provide briefings, answer questions, and resolve problems for both laboratory personnel and participants. The objective was to ensure that the SRP would be as constructive as possible for all participants. Both SRP participants and RDL representatives found these visits beneficial. At many of the laboratories, this was the only opportunity for all participants to meet at one time to share their experiences and exchange ideas.

5. HISTORICALLY BLACK COLLEGES AND UNIVERSITIES AND MINORITY INSTITUTIONS (HBCU/MI)s

Before 1993, an RDL program representative visited from seven to ten different HBCU/MI's annually to promote interest in the SRP among the faculty and graduate students. These efforts were marginally effective, yielding a doubling of HBCU/MI applicants. In an effort to achieve AFOSR's goal of 10% of all applicants and selectees being HBCU/MI qualified, the RDL team decided to try other avenues of approach to increase the number of qualified applicants. Through the combined efforts of the AFOSR Program Office at Bolling AFB and RDL, two very active minority groups were found, HACU (Hispanic American Colleges and Universities) and AISES (American Indian Science and Engineering Society). RDL is in communication with representatives of each of these organizations on a monthly basis to keep up with their activities and special events. Both organizations have widely-distributed magazines/quarterlies in which RDL placed ads.

Since 1994 the number of both SFRP and GSRP HBCU/MI applicants and participants has increased ten-fold, from about two dozen SFRP applicants and a half dozen selectees to over 100 applicants and two dozen selectees, and a half-dozen GSRP applicants and two or three selectees to 18 applicants and 7 or 8 selectees. Since 1993, the SFRP had a two-fold applicant increase and a two-fold selectee increase. Since 1993, the GSRP had a three-fold applicant increase and a three to four-fold increase in selectees.

In addition to RDL's special recruiting efforts, AFOSR attempts each year to obtain additional funding or use leftover funding from cancellations the past year to fund HBCU/MI associates.

SRP HBCU/MI Participation, By Year				
YEAR	SFRP		GSRP	
	Applicants	Participants	Applicants	Participants
1985	76	23	15	11
1986	70	18	20	10
1987	82	32	32	10
1988	53	17	23	14
1989	39	15	13	4
1990	43	14	17	3
1991	42	13	8	5
1992	70	13	9	5
1993	60	13	6	2
1994	90	16	11	6
1995	90	21	20	8
1996	119	27	18	7

6. SRP FUNDING SOURCES

Funding sources for the 1998 SRP were the AFOSR-provided slots for the basic contract and laboratory funds. Funding sources by category for the 1998 SRP selected participants are shown here.

1998 SRP FUNDING CATEGORY	SFRP	GSRP	HSAP
AFOSR Basic Allocation Funds	67	38	75
USAF Laboratory Funds	17	2	13
Slots Added by AFOSR (Leftover Funds)	0	0	0
HBCU/MI By AFOSR (Using Procured Addn'l Funds)	0	0	N/A
TOTAL	84	40	88

7. COMPENSATION FOR PARTICIPANTS

Compensation for SRP participants, per five-day work week, is shown in this table.

1998 SRP Associate Compensation

PARTICIPANT CATEGORY	1991	1992	1993	1994	1995	1996	1997	1998
Faculty Members	\$690	\$718	\$740	\$740	\$740	\$770	\$770	\$793
Graduate Student (Master's Degree)	\$425	\$442	\$455	\$455	\$455	\$470	\$470	\$484
Graduate Student (Bachelor's Degree)	\$365	\$380	\$391	\$391	\$391	\$400	\$400	\$412
High School Student (First Year)	\$200	\$200	\$200	\$200	\$200	\$200	\$200	\$200
High School Student (Subsequent Years)	\$240	\$240	\$240	\$240	\$240	\$240	\$240	\$240

The program also offered associates whose homes were more than 50 miles from the laboratory an expense allowance (seven days per week) of \$52/day for faculty and \$41/day for graduate students. Transportation to the laboratory at the beginning of their tour and back to their home destinations at the end was also reimbursed for these participants. Of the combined SFRP and GSRP associates, 65 % claimed travel reimbursements at an average round-trip cost of \$730.

Faculty members were encouraged to visit their laboratories before their summer tour began. All costs of these orientation visits were reimbursed. Forty-three percent (85 out of 188) of faculty associates took orientation trips at an average cost of \$449. By contrast, in 1993, 58 % of SFRP associates elected to take an orientation visits at an average cost of \$685; that was the highest percentage of

associates opting to take an orientation trip since RDL has administered the SRP, and the highest average cost of an orientation trip.

Program participants submitted biweekly vouchers countersigned by their laboratory research focal point, and RDL issued paychecks so as to arrive in associates' hands two weeks later.

This is the third year of using direct deposit for the SFRP and GSRP associates. The process went much more smoothly with respect to obtaining required information from the associates, about 15% of the associates' information needed clarification in order for direct deposit to properly function as opposed to 7% from last year. The remaining associates received their stipend and expense payments via checks sent in the US mail.

HSAP program participants were considered actual RDL employees, and their respective state and federal income tax and Social Security were withheld from their paychecks. By the nature of their independent research, SFRP and GSRP program participants were considered to be consultants or independent contractors. As such, SFRP and GSRP associates were responsible for their own income taxes, Social Security, and insurance.

8. CONTENTS OF THE 1998 REPORT

The complete set of reports for the 1998 SRP includes this program management report (Volume 1) augmented by fifteen volumes of final research reports by the 1998 associates, as indicated below:

1998 SRP Final Report Volume Assignments

LABORATORY	SFRP	GSRP	HSAP
Armstrong	2	7	12
Phillips	3	8	13
Rome	4	9	14
Wright	5A, 5B	10	15
AEDC, ALCs, USAFA, WHMC	6	11	

APPENDIX A – PROGRAM STATISTICAL SUMMARY

A. Colleges/Universities Represented

Selected SFRP associates represented 169 different colleges, universities, and institutions, GSRP associates represented 95 different colleges, universities, and institutions.

B. States Represented

SFRP -Applicants came from 47 states plus Washington D.C. Selectees represent 44 states.

GSRP - Applicants came from 44 states. Selectees represent 32 states.

HSAP - Applicants came from thirteen states. Selectees represent nine states.

Total Number of Participants	
SFRP	85
GSRP	40
HSAP	88
TOTAL	213

Degrees Represented			
	SFRP	GSRP	TOTAL
Doctoral	83	0	83
Master's	1	3	4
Bachelor's	0	22	22
TOTAL	186	25	109

SFRP Academic Titles	
Assistant Professor	36
Associate Professor	34
Professor	15
Instructor	0
Chairman	0
Visiting Professor	0
Visiting Assoc. Prof.	0
Research Associate	0
TOTAL	85

Source of Learning About the SRP		
Category	Applicants	Selectees
Applied/participated in prior years	177	47
Colleague familiar with SRP	104	24
Brochure mailed to institution	101	21
Contact with Air Force laboratory	101	39
<i>IEEE Spectrum</i>	12	1
<i>BIIHE</i>	4	0
Other source	117	30
TOTAL	616	162

APPENDIX B – SRP EVALUATION RESPONSES

1. OVERVIEW

Evaluations were completed and returned to RDL by four groups at the completion of the SRP. The number of respondents in each group is shown below.

Table B-1. Total SRP Evaluations Received

Evaluation Group	Responses
SFRP & GSRPs	100
HSAPs	75
USAF Laboratory Focal Points	84
USAF Laboratory HSAP Mentors	6

All groups indicate unanimous enthusiasm for the SRP experience.

The summarized recommendations for program improvement from both associates and laboratory personnel are listed below:

- A. Better preparation on the labs' part prior to associates' arrival (i.e., office space, computer assets, clearly defined scope of work).
- B. Faculty Associates suggest higher stipends for SFRP associates.
- C. Both HSAP Air Force laboratory mentors and associates would like the summer tour extended from the current 8 weeks to either 10 or 11 weeks; the groups state it takes 4-6 weeks just to get high school students up-to-speed on what's going on at laboratory. (Note: this same argument was used to raise the faculty and graduate student participation time a few years ago.)

2. 1998 USAF LABORATORY FOCAL POINT (LFP) EVALUATION RESPONSES

The summarized results listed below are from the 84 LFP evaluations received.

1. LFP evaluations received and associate preferences:

Table B-2. Air Force LFP Evaluation Responses (By Type)

Lab	Evals Recv'd	How Many Associates Would You Prefer To Get ? (% Response)											
		SFRP				GSRP (w/Univ Professor)				GSRP (w/o Univ Professor)			
		0	1	2	3+	0	1	2	3+	0	1	2	3+
AEDC	0	-	-	-	-	-	-	-	-	-	-	-	-
WHMC	0	-	-	-	-	-	-	-	-	-	-	-	-
AL	7	28	28	28	14	54	14	28	0	86	0	14	0
USAF	1	0	100	0	0	100	0	0	0	0	100	0	0
PL	25	40	40	16	4	88	12	0	0	84	12	4	0
RL	5	60	40	0	0	80	10	0	0	100	0	0	0
WL	46	30	43	20	6	78	17	4	0	93	4	2	0
Total	84	32%	50%	13%	5%	80%	11%	6%	0%	73%	23%	4%	0%

LFP Evaluation Summary. The summarized responses, by laboratory, are listed on the following page. LFPs were asked to rate the following questions on a scale from 1 (below average) to 5 (above average).

2. LFPs involved in SRP associate application evaluation process:
 - a. Time available for evaluation of applications:
 - b. Adequacy of applications for selection process:
3. Value of orientation trips:
4. Length of research tour:
5.
 - a. Benefits of associate's work to laboratory:
 - b. Benefits of associate's work to Air Force:
6.
 - a. Enhancement of research qualifications for LFP and staff:
 - b. Enhancement of research qualifications for SFRP associate:
 - c. Enhancement of research qualifications for GSRP associate:
7.
 - a. Enhancement of knowledge for LFP and staff:
 - b. Enhancement of knowledge for SFRP associate:
 - c. Enhancement of knowledge for GSRP associate:
8. Value of Air Force and university links:
9. Potential for future collaboration:
10.
 - a. Your working relationship with SFRP:
 - b. Your working relationship with GSRP:
11. Expenditure of your time worthwhile:

(Continued on next page)

12. Quality of program literature for associate:
13. a. Quality of RDL's communications with you:
b. Quality of RDL's communications with associates:
14. Overall assessment of SRP:

Table B-3. Laboratory Focal Point Responses to above questions

	<i>AEDC</i>	<i>AL</i>	<i>USAFA</i>	<i>PL</i>	<i>RL</i>	<i>WHMC</i>	<i>WL</i>
<i># Evals Recv'd</i>	0	7	1	14	5	0	46
<i>Question #</i>							
2	-	86 %	0 %	88 %	80 %	-	85 %
2a	-	4.3	n/a	3.8	4.0	-	3.6
2b	-	4.0	n/a	3.9	4.5	-	4.1
3	-	4.5	n/a	4.3	4.3	-	3.7
4	-	4.1	4.0	4.1	4.2	-	3.9
5a	-	4.3	5.0	4.3	4.6	-	4.4
5b	-	4.5	n/a	4.2	4.6	-	4.3
6a	-	4.5	5.0	4.0	4.4	-	4.3
6b	-	4.3	n/a	4.1	5.0	-	4.4
6c	-	3.7	5.0	3.5	5.0	-	4.3
7a	-	4.7	5.0	4.0	4.4	-	4.3
7b	-	4.3	n/a	4.2	5.0	-	4.4
7c	-	4.0	5.0	3.9	5.0	-	4.3
8	-	4.6	4.0	4.5	4.6	-	4.3
9	-	4.9	5.0	4.4	4.8	-	4.2
10a	-	5.0	n/a	4.6	4.6	-	4.6
10b	-	4.7	5.0	3.9	5.0	-	4.4
11	-	4.6	5.0	4.4	4.8	-	4.4
12	-	4.0	4.0	4.0	4.2	-	3.8
13a	-	3.2	4.0	3.5	3.8	-	3.4
13b	-	3.4	4.0	3.6	4.5	-	3.6
14	-	4.4	5.0	4.4	4.8	-	4.4

3. 1998 SFRP & GSRP EVALUATION RESPONSES

The summarized results listed below are from the 120 SFRP/GSRP evaluations received.

Associates were asked to rate the following questions on a scale from 1 (below average) to 5 (above average) - by Air Force base results and over-all results of the 1998 evaluations are listed after the questions.

1. The match between the laboratories research and your field:
2. Your working relationship with your LFP:
3. Enhancement of your academic qualifications:
4. Enhancement of your research qualifications:
5. Lab readiness for you: LFP, task, plan:
6. Lab readiness for you: equipment, supplies, facilities:
7. Lab resources:
8. Lab research and administrative support:
9. Adequacy of brochure and associate handbook:
10. RDL communications with you:
11. Overall payment procedures:
12. Overall assessment of the SRP:
13.
 - a. Would you apply again?
 - b. Will you continue this or related research?
14. Was length of your tour satisfactory?
15. Percentage of associates who experienced difficulties in finding housing:
16. Where did you stay during your SRP tour?
 - a. At Home:
 - b. With Friend:
 - c. On Local Economy:
 - d. Base Quarters:
17. Value of orientation visit:
 - a. Essential:
 - b. Convenient:
 - c. Not Worth Cost:
 - d. Not Used:

SFRP and GSRP associate's responses are listed in tabular format on the following page.

Table B-4. 1997 SFRP & GSRP Associate Responses to SRP Evaluation

	Arnold	Brooks	Edwards	Eglin	Griffis	Hanscom	Kelly	Kirtland	Lackland	Robins	Tyndall	WPAFB	average
# res	6	48	6	14	31	19	3	32	1	2	10	85	257
1	4.8	4.4	4.6	4.7	4.4	4.9	4.6	4.6	5.0	5.0	4.0	4.7	4.6
2	5.0	4.6	4.1	4.9	4.7	4.7	5.0	4.7	5.0	5.0	4.6	4.8	4.7
3	4.5	4.4	4.0	4.6	4.3	4.2	4.3	4.4	5.0	5.0	4.5	4.3	4.4
4	4.3	4.5	3.8	4.6	4.4	4.4	4.3	4.6	5.0	4.0	4.4	4.5	4.5
5	4.5	4.3	3.3	4.8	4.4	4.5	4.3	4.2	5.0	5.0	3.9	4.4	4.4
6	4.3	4.3	3.7	4.7	4.4	4.5	4.0	3.8	5.0	5.0	3.8	4.2	4.2
7	4.5	4.4	4.2	4.8	4.5	4.3	4.3	4.1	5.0	5.0	4.3	4.3	4.4
8	4.5	4.6	3.0	4.9	4.4	4.3	4.3	4.5	5.0	5.0	4.7	4.5	4.5
9	4.7	4.5	4.7	4.5	4.3	4.5	4.7	4.3	5.0	5.0	4.1	4.5	4.5
10	4.2	4.4	4.7	4.4	4.1	4.1	4.0	4.2	5.0	4.5	3.6	4.4	4.3
11	3.8	4.1	4.5	4.0	3.9	4.1	4.0	4.0	3.0	4.0	3.7	4.0	4.0
12	5.7	4.7	4.3	4.9	4.5	4.9	4.7	4.6	5.0	4.5	4.6	4.5	4.6
Numbers below are percentages													
13a	83	90	83	93	87	75	100	81	100	100	100	86	87
13b	100	89	83	100	94	98	100	94	100	100	100	94	93
14	83	96	100	90	87	80	100	92	100	100	70	84	88
15	17	6	0	33	20	76	33	25	0	100	20	8	39
16a	-	26	17	9	38	23	33	4	-	-	-	30	
16b	100	33	-	40	-	8	-	-	-	-	36	2	
16c	-	41	83	40	62	69	67	96	100	100	64	68	
16d	-	-	-	-	-	-	-	-	-	-	-	0	
17a	-	33	100	17	50	14	67	39	-	50	40	31	35
17b	-	21	-	17	10	14	-	24	-	50	20	16	16
17c	-	-	-	-	10	7	-	-	-	-	-	2	3
17d	100	46	-	66	30	69	33	37	100	-	40	51	46

4. 1998 USAF LABORATORY HSAP MENTOR EVALUATION RESPONSES

Not enough evaluations received (5 total) from Mentors to do useful summary.

5. 1998 HSAP EVALUATION RESPONSES

The summarized results listed below are from the 23 HSAP evaluations received.

HSAP apprentices were asked to rate the following questions on a scale from
1 (below average) to 5 (above average)

1. Your influence on selection of topic/type of work.
2. Working relationship with mentor, other lab scientists.
3. Enhancement of your academic qualifications.
4. Technically challenging work.
5. Lab readiness for you: mentor, task, work plan, equipment.
6. Influence on your career.
7. Increased interest in math/science.
8. Lab research & administrative support.
9. Adequacy of RDL's Apprentice Handbook and administrative materials.
10. Responsiveness of RDL communications.
11. Overall payment procedures.
12. Overall assessment of SRP value to you.
13. Would you apply again next year? Yes (92 %)
14. Will you pursue future studies related to this research? Yes (68 %)
15. Was Tour length satisfactory? Yes (82 %)

	Arnold	Brooks	Edwards	Eglin	Griffiss	Hanscom	Kirtland	Tyndall	WPAFB	Totals
# resp	5	19	7	15	13	2	7	5	40	113
1	2.8	3.3	3.4	3.5	3.4	4.0	3.2	3.6	3.6	3.4
2	4.4	4.6	4.5	4.8	4.6	4.0	4.4	4.0	4.6	4.6
3	4.0	4.2	4.1	4.3	4.5	5.0	4.3	4.6	4.4	4.4
4	3.6	3.9	4.0	4.5	4.2	5.0	4.6	3.8	4.3	4.2
5	4.4	4.1	3.7	4.5	4.1	3.0	3.9	3.6	3.9	4.0
6	3.2	3.6	3.6	4.1	3.8	5.0	3.3	3.8	3.6	3.7
7	2.8	4.1	4.0	3.9	3.9	5.0	3.6	4.0	4.0	3.9
8	3.8	4.1	4.0	4.3	4.0	4.0	4.3	3.8	4.3	4.2
9	4.4	3.6	4.1	4.1	3.5	4.0	3.9	4.0	3.7	3.8
10	4.0	3.8	4.1	3.7	4.1	4.0	3.9	2.4	3.8	3.8
11	4.2	4.2	3.7	3.9	3.8	3.0	3.7	2.6	3.7	3.8
12	4.0	4.5	4.9	4.6	4.6	5.0	4.6	4.2	4.3	4.5
Numbers below are percentages										
13	60%	95%	100%	100%	85%	100%	100%	100%	90%	92%
14	20%	80%	71%	80%	54%	100%	71%	80%	65%	68%
15	100%	70%	71%	100%	100%	50%	86%	60%	80%	82%

MANY-BODY THEORY OF QUANTUM-WELL GAIN SPECTRA

Dr. Lok C. Lew Yan Voon

Assistant Professor

Department of Physics

Worcester Polytechnic Institute

100 Institute Road

Worcester, MA 01609

Final Report for:

Summer Faculty Research Program

Wright-Patterson Research Site

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, DC

and

Wright-Patterson Research Site

July 1998

MANY-BODY THEORY OF QUANTUM-WELL GAIN SPECTRA

Dr. Lok C. Lew Yan Voon
Assistant Professor
Department of Physics
Worcester Polytechnic Institute

Abstract

The aim of this summer project is to develop a theory and computer code for the group of Dr. John Loehr in the Sensors Directorate to assist them in analyzing fine details of the gain spectra of quantum-well laser diodes being measured. Such an in-house capability should help the Air Force Research Laboratory in developing new types of semiconductor lasers. We have critically studied the different theoretical models of material gain (particularly with respect to the many-body effects) currently available in the literature, developed a computer code for modelling the gain at a level comparable to the most advanced work reported so far, and uncovered deficiencies in current theories.

MANY-BODY THEORY OF QUANTUM-WELL GAIN SPECTRA

L. C. Lew Yan Voon

1 Introduction

The Sensors Directorate of the Air Force Research Laboratory is currently investigating the use of diode lasers operating at high temperatures (≈ 150 C) and high modulation frequencies (≈ 10 GHz) to replace the electrical back-plane of an aircraft with an optical back-plane. High-temperature operation is essential because cooling the laser requires energy and space, which are both very limited onboard an aircraft. Due to the importance of semiconductor lasers to the Air Force mission, we have aimed to bring our own contribution to the development of a theory of material gain. This will allow the Air Force researchers to have direct access to a sophisticated computer code in order to design and characterize new semiconductor quantum well (QW) lasers. Our first aim was to reproduce current models and test them at high temperatures. In the process, we have also uncovered extensions to those models which we will be investigating in future work.

2 Methodology

The fundamental property of a laser medium is the gain spectrum. The gain spectrum is solely determined by the external conditions (injection current, temperature, ...), the layer structure, and the interacting system of electrons, photons and phonons. This is a complex and, as yet, unsolved problem. The methodology of the current work partitions the problem into three considerations (which are, however, not independent): the dimensionality, the electronic band-structure problem, and the interactions (or many-body) problem. We call the final solution to the problem the many-body gain (MBG). In the following section, we will outline the microscopic derivation of the MBG expression based upon the semiconductor Bloch equations (SBE's). Our approach in the Results section is to then start from a simplified solution and to successively add contributions toward a full band-structure and many-body expression.

3 Many-Body Theory

3.1 Derivation of equations

3.1.1 Hamiltonian

The starting point is the second-quantized Hamiltonian for an interacting electron gas in a crystal with the inclusion of the electron-photon interaction. Since we are dealing with electrons in a crystal, we explicitly represent the field operators in terms of the orthonormalized Bloch functions and the corresponding annihilation operators:

$$\hat{\psi} = \sum_{n\nu\sigma\mathbf{k}} \phi_{n\nu\sigma\mathbf{k}}(\mathbf{r}) a_{n\nu\sigma}(\mathbf{k}). \quad (1)$$

We have chosen a basis set of wave functions which are labelled by a band index n , a subband index ν , a spin index σ , and a wave vector $\mathbf{k}$. The wave function will now be written explicit as a QW state:

$$\phi_{n\nu\sigma\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{A}} e^{i\mathbf{k}\cdot\boldsymbol{\rho}} \sum_i f_{n\nu\sigma\mathbf{k}}^i(z) u_{i0}(\mathbf{r}) \eta_{\sigma}, \quad (2)$$

where $\mathbf{k}$ is a 2D in-plane wave vector, $\boldsymbol{\rho}$ an in-plane position vector, A the in-plane normalization area, $f(z)$ the envelope function, $u_{i0}(\mathbf{r})$ the bulk zone-center cellular function, and η_{σ} a spinor. The Coulomb and electron-photon terms become

$$\begin{aligned} \hat{H}_C = & \frac{1}{2A} \sum_{\substack{n_1 \dots n_4 \\ \nu_1 \dots \nu_4}} \sum_{\sigma_1 \sigma_2} \sum_{\substack{\mathbf{k}' \mathbf{k}'' \\ \mathbf{q} \neq 0}} a_{n_1 \nu_1 \sigma_1}^{\dagger}(\mathbf{k}' + \mathbf{q}) a_{n_2 \nu_2 \sigma_2}^{\dagger}(\mathbf{k}'' - \mathbf{q}) a_{n_3 \nu_3 \sigma_2}(\mathbf{k}'') a_{n_4 \nu_4 \sigma_1}(\mathbf{k}') \\ & \times v(n_1 \nu_1 \sigma_1 \mathbf{k}' + \mathbf{q}, n_2 \nu_2 \sigma_2 \mathbf{k}'' - \mathbf{q}; n_3 \nu_3 \sigma_2 \mathbf{k}'', n_4 \nu_4 \sigma_1 \mathbf{k}'), \end{aligned} \quad (3)$$

with

$$\begin{aligned} & v(n_1 \nu_1 \sigma_1 \mathbf{k}' + \mathbf{q}, n_2 \nu_2 \sigma_2 \mathbf{k}'' - \mathbf{q}; n_3 \nu_3 \sigma_2 \mathbf{k}'', n_4 \nu_4 \sigma_1 \mathbf{k}') \\ & = \left(\frac{e^2}{2\epsilon_0 \epsilon_b q} \right) \sum_{i_1 \dots i_4} \iint dz dz' f_{n_1 \nu_1 \sigma_1 \mathbf{k}' + \mathbf{q}}^{i_1*}(z) f_{n_2 \nu_2 \sigma_2 \mathbf{k}'' - \mathbf{q}}^{i_2*}(z') f_{n_3 \nu_3 \sigma_2 \mathbf{k}''}^{i_3}(z') f_{n_4 \nu_4 \sigma_1 \mathbf{k}'}^{i_4}(z) e^{-q|z-z'|}, \\ & \equiv V_q^{(2D)} F_q(1, 2, 3, 4), \end{aligned} \quad (4)$$

and

$$\hat{H}_{e\gamma} = - \sum_{\substack{n_1 n_2 \\ \nu_1 \nu_2}} \sum_{\sigma_1 \sigma_2} \sum_{\mathbf{k} \mathbf{q}} a_{n_1 \nu_1 \sigma_1}^{\dagger}(\mathbf{k}) a_{n_2 \nu_2 \sigma_2}(\mathbf{k} - \mathbf{q}) \mathbf{d}_{n_1 n_2} \cdot \mathbf{E}_{n_1 \nu_1 \sigma_1, n_2 \nu_2 \sigma_2}(\mathbf{k}, \mathbf{q}, t), \quad (5)$$

with

$$\mathbf{E}_{n_1 \nu_1 \sigma_1, n_2 \nu_2 \sigma_2}(\mathbf{k}, \mathbf{q}, t) = \sum_{i_1 i_2} \int dz f_{n_1 \nu_1 \sigma_1 \mathbf{k}}^{i_1*}(z) f_{n_2 \nu_2 \sigma_2 \mathbf{k} - \mathbf{q}}^{i_2}(z) \mathbf{E}(\mathbf{q}, z, t). \quad (6)$$

3.1.2 Interband polarization

The macroscopic interband polarization induced by a coherent monochromatic classical light field is given by

$$\begin{aligned}
 \mathbf{P}(t) &= \int d\mathbf{r} \mathbf{P}(\mathbf{r}, t) = \int d\mathbf{r} \langle \psi^\dagger(\mathbf{r}, t) e \mathbf{r} \psi(\mathbf{r}, t) \rangle \\
 &= \sum_{\substack{n_1 \nu_1 \sigma_1 \\ n_2 \nu_2 \sigma_2 \mathbf{k}_2}} \sum_{ij} \langle a_{n_1 \nu_1 \sigma_1}^\dagger(\mathbf{k}) a_{n_2 \nu_2 \sigma_2}(\mathbf{k}) \rangle d_{n_1 n_2} \int dz f_{n_1 \nu_1 \sigma_1 \mathbf{k}}^{i*}(z) f_{n_2 \nu_2 \sigma_2 \mathbf{k}}^j(z) \\
 &= \sum_{\substack{n_1 \nu_1 \sigma_1 \\ n_2 \nu_2 \sigma_2 \mathbf{k}}} P_{\substack{n_1 \nu_1 \sigma_1 \\ n_2 \nu_2 \sigma_2}}(\mathbf{k}) d_{n_1 n_2} I(n_1 \nu_1 \sigma_1, n_2 \nu_2 \sigma_2, \mathbf{k}).
 \end{aligned} \tag{7}$$

In particular, for interband transitions,

$$\begin{aligned}
 \mathbf{P}(t) &= \sum_{\substack{\nu_1 \sigma_1 \\ \nu_2 \sigma_2 \mathbf{k}}} \left[P_{\substack{v \nu_1 \sigma_1 \\ c \nu_2 \sigma_2}}(\mathbf{k}) d_{cv}^* I(v \nu_1 \sigma_1, c \nu_2 \sigma_2, \mathbf{k}) + P_{\substack{c \nu_1 \sigma_1 \\ v \nu_2 \sigma_2}}(\mathbf{k}) d_{cv} I(c \nu_1 \sigma_1, v \nu_2 \sigma_2, \mathbf{k}) \right] \\
 &= \epsilon_0 \chi \mathbf{E}(\mathbf{r}, t) = \epsilon_0 \chi \left[\mathcal{E} e^{-i\omega t} + c.c. \right],
 \end{aligned} \tag{8}$$

giving

$$\chi = \frac{e^{i\omega t}}{\mathcal{E}} \sum_{\substack{\nu_1 \sigma_1 \\ \nu_2 \sigma_2 \mathbf{k}}} P_{\substack{v \nu_1 \sigma_1 \\ c \nu_2 \sigma_2}}(\mathbf{k}) d_{cv}^* I(v \nu_1 \sigma_1, c \nu_2 \sigma_2, \mathbf{k}). \tag{9}$$

3.1.3 Gain

The amplitude gain and the phase change are given by

$$g(E) - i \frac{d\phi}{dz} = \frac{i\omega}{2\eta_b c} \chi, \tag{10}$$

where η_b is the background refractive index and the phase change gives the carrier-induced refractive index change:

$$\frac{\delta n_b}{n_b} = -\frac{1}{q} \frac{d\phi}{dz}. \tag{11}$$

3.1.4 Equation of motion

Using the above Hamiltonian, we can now derive the equation of motion for the polarization

$$P_{\substack{v \nu \sigma \\ c \nu' \sigma'}}(\mathbf{k}) \equiv \langle a_{v \nu \sigma}^\dagger(\mathbf{k}) a_{c \nu' \sigma'}(\mathbf{k}) \rangle. \tag{12}$$

The complete general result is

$$\begin{aligned}
i\hbar \frac{\partial}{\partial t} P_{c\nu'\sigma'}(k) &= [E_{c\nu'\sigma'}(k) - E_{v\nu\sigma}(k)] P_{c\nu'\sigma'}(k) \\
&+ \sum_{n_1\nu_1\sigma_1} \sum_{\mathbf{q}} \mathbf{d}_{n_1\nu} \cdot \mathbf{E}_{n_1\nu_1\sigma_1}(\mathbf{k} + \mathbf{q}, \mathbf{q}, t) \langle a_{n_1\nu_1\sigma_1}^\dagger(\mathbf{k} + \mathbf{q}) a_{c\nu'\sigma'}(\mathbf{k}) \rangle \\
&- \sum_{n_1\nu_1\sigma_1} \sum_{\mathbf{q}} \mathbf{d}_{cn_1} \cdot \mathbf{E}_{c\nu'\sigma'}(\mathbf{k}, \mathbf{q}, t) \langle a_{v\nu\sigma}^\dagger(\mathbf{k}) a_{n_1\nu_1\sigma_1}(\mathbf{k} - \mathbf{q}) \rangle \\
&+ \frac{1}{2A} \sum_{\substack{n_1n_2n_3 \\ \nu_1\nu_2\nu_3, \sigma_1}} \sum_{\mathbf{k}', \mathbf{q} \neq 0} \{ v(n_1\nu_1\sigma_1\mathbf{k}' + \mathbf{q}, n_2\nu_2\sigma_1\mathbf{k} - \mathbf{q}; v\nu\sigma\mathbf{k}, n_3\nu_3\sigma_1\mathbf{k}') \\
&\times \langle a_{n_1\nu_1\sigma_1}^\dagger(\mathbf{k}' + \mathbf{q}) a_{n_2\nu_2\sigma_1}^\dagger(\mathbf{k} - \mathbf{q}) a_{n_3\nu_3\sigma_1}(\mathbf{k}') a_{c\nu'\sigma'}(\mathbf{k}) \rangle \\
&- v(n_1\nu_1\sigma_1\mathbf{k} + \mathbf{q}, n_2\nu_2\sigma_1\mathbf{k}' - \mathbf{q}; n_3\nu_3\sigma_1\mathbf{k}', v\nu\sigma\mathbf{k}) \\
&\times \langle a_{n_1\nu_1\sigma_1}^\dagger(\mathbf{k} + \mathbf{q}) a_{n_2\nu_2\sigma_1}^\dagger(\mathbf{k}' - \mathbf{q}) a_{n_3\nu_3\sigma_1}(\mathbf{k}') a_{c\nu'\sigma'}(\mathbf{k}) \rangle \\
&+ v(c\nu'\sigma'\mathbf{k}, n_1\nu_1\sigma_1\mathbf{k}' - \mathbf{q}; n_2\nu_2\sigma_1\mathbf{k}', n_3\nu_3\sigma_1\mathbf{k} - \mathbf{q}) \\
&\times \langle a_{v\nu\sigma}^\dagger(\mathbf{k}) a_{n_1\nu_1\sigma_1}^\dagger(\mathbf{k}' - \mathbf{q}) a_{n_2\nu_2\sigma_1}(\mathbf{k}') a_{n_3\nu_3\sigma_1}(\mathbf{k} - \mathbf{q}) \rangle \\
&- v(n_1\nu_1\sigma_1\mathbf{k}' + \mathbf{q}, c\nu'\sigma'\mathbf{k}; n_2\nu_2\sigma_1\mathbf{k} + \mathbf{q}, n_3\nu_3\sigma_1\mathbf{k}') \\
&\times \langle a_{v\nu\sigma}^\dagger(\mathbf{k}) a_{n_1\nu_1\sigma_1}^\dagger(\mathbf{k}' + \mathbf{q}) a_{n_2\nu_2\sigma_1}(\mathbf{k} + \mathbf{q}) a_{n_3\nu_3\sigma_1}(\mathbf{k}') \rangle \}.
\end{aligned} \tag{13}$$

We now make the Hartree-Fock (HF) approximation. A key result of the HF approximation is the lack of carrier-carrier scattering. This would have been present in the polarization equation as an imaginary polarization term. This can also be seen in the renormalization of the energies by a real static self-energy. Scattering terms will therefore have to be obtained separately. The equation of motion is then

$$\begin{aligned}
i\hbar \frac{\partial}{\partial t} P_{c\nu'\sigma'}(k) &= [E_{c\nu'\sigma'}(k) - E_{v\nu\sigma}(k)] P_{c\nu'\sigma'}(k) \\
&+ \sum_{n_1\nu_1\sigma_1} \sum_{\mathbf{q}} \mathbf{d}_{n_1\nu} \cdot \mathbf{E}_{n_1\nu_1\sigma_1}(\mathbf{k} + \mathbf{q}, \mathbf{q}, t) \langle a_{n_1\nu_1\sigma_1}^\dagger(\mathbf{k} + \mathbf{q}) a_{c\nu'\sigma'}(\mathbf{k}) \rangle \\
&- \sum_{n_1\nu_1\sigma_1} \sum_{\mathbf{q}} \mathbf{d}_{cn_1} \cdot \mathbf{E}_{c\nu'\sigma'}(\mathbf{k}, \mathbf{q}, t) \langle a_{v\nu\sigma}^\dagger(\mathbf{k}) a_{n_1\nu_1\sigma_1}(\mathbf{k} - \mathbf{q}) \rangle \\
&+ \frac{1}{A} \sum_{\substack{n_1n_2n_3 \\ \nu_1\nu_2\nu_3, \sigma_1}} \sum_{\mathbf{q} \neq 0} \{ v(n_1\nu_1\sigma_1\mathbf{k}, n_2\nu_2\sigma_1\mathbf{k} - \mathbf{q}; v\nu\sigma\mathbf{k}, n_3\nu_3\sigma_1\mathbf{k} - \mathbf{q}) \\
&\times \langle a_{n_2\nu_2\sigma_1}^\dagger(\mathbf{k} - \mathbf{q}) a_{n_3\nu_3\sigma_1}(\mathbf{k} - \mathbf{q}) \rangle \langle a_{n_1\nu_1\sigma_1}^\dagger(\mathbf{k}) a_{c\nu'\sigma'}(\mathbf{k}) \rangle \\
&- v(c\nu'\sigma'\mathbf{k}, n_1\nu_1\sigma_1\mathbf{k} - \mathbf{q}; n_2\nu_2\sigma_1\mathbf{k}, n_3\nu_3\sigma_1\mathbf{k} - \mathbf{q}) \\
&\times \langle a_{v\nu\sigma}^\dagger(\mathbf{k}) a_{n_2\nu_2\sigma_1}(\mathbf{k}) \rangle \langle a_{n_1\nu_1\sigma_1}^\dagger(\mathbf{k} - \mathbf{q}) a_{n_3\nu_3\sigma_1}(\mathbf{k} - \mathbf{q}) \rangle \}.
\end{aligned} \tag{14}$$

Equation (14) shows that, in the absence of the Coulomb potential, the polarization $P_{nm}(\mathbf{k})$ satisfies a simple first-order differential equation whose solution is the well-known free-carrier polarization. However, when the Coulomb potential is retained, $P_{nm}(\mathbf{k})$ is coupled to $P_{nm}(\mathbf{q})$ and to other $P_{n'm'}(\mathbf{k}')$, leading to a coupled set of integro-differential equations. This problem is extremely complicated and has not even been referred to in the literature. We start by solving the simplified problem considered by others and which goes under the name of the semiconductor Bloch equations (SBE).

3.1.5 Semiconductor Bloch equations

The SBE can be obtained as approximations to our Eq. (14). The traditional derivation (see, e.g., Ref. [1]) starts from a simplified Hamiltonian, whose Coulomb term consists only of the same two conduction-band indices and of the same two valence-band indices. It is then straightforwardly transformed into the electron-hole representation. Using Eq. (14), one first uses the electric dipole and two-band approximations giving

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} P_{vc}(\mathbf{k}) = & \left[\left\{ E_c(\mathbf{k}) - \frac{1}{A} \sum_{\substack{n_2 n_3 \\ \mathbf{q} \neq 0}} v(c\mathbf{k}, n_2\mathbf{k} - \mathbf{q}; c\mathbf{k}, n_3\mathbf{k} - \mathbf{q}) \langle a_{n_2}^\dagger(\mathbf{k} - \mathbf{q}) a_{n_3}(\mathbf{k} - \mathbf{q}) \rangle \right\} \right. \\ & - \left. \left\{ E_v(\mathbf{k}) - \frac{1}{A} \sum_{\substack{n_2 n_3 \\ \mathbf{q} \neq 0}} v(v\mathbf{k}, n_2\mathbf{k} - \mathbf{q}; v\mathbf{k}, n_3\mathbf{k} - \mathbf{q}) \langle a_{n_2}^\dagger(\mathbf{k} - \mathbf{q}) a_{n_3}(\mathbf{k} - \mathbf{q}) \rangle \right\} \right] P_{vc}(\mathbf{k}) \\ & + (n_{c\mathbf{k}} - n_{v\mathbf{k}}) \left[\mathbf{d}_{cv} \cdot \mathbf{E}_y(\mathbf{k}, 0, t) + \frac{1}{A} \sum_{\substack{n_2 n_3 \\ \mathbf{q} \neq 0}} v(c\mathbf{k}, n_2\mathbf{k} - \mathbf{q}; v\mathbf{k}, n_3\mathbf{k} - \mathbf{q}) \langle a_{n_2}^\dagger(\mathbf{k} - \mathbf{q}) a_{n_3}(\mathbf{k} - \mathbf{q}) \rangle \right]. \end{aligned}$$

The single-particle energies are then replaced by renormalized ones, only terms linear in the particle densities are kept, and collision terms (which are absent at the HF level) are added.

$$\boxed{\begin{aligned} \left[i\hbar \frac{\partial}{\partial t} - (\epsilon_c(\mathbf{k}) - \epsilon_v(\mathbf{k})) \right] P_{vc}(\mathbf{k}) = & (n_{c\mathbf{k}} - n_{v\mathbf{k}}) \left[\mathbf{d}_{cv} \cdot \mathbf{E}_y(\mathbf{k}, 0, t) \right. \\ & + \left. \frac{1}{A} \sum_{\mathbf{q} \neq 0} v(c\mathbf{k}, v\mathbf{k} - \mathbf{q}; v\mathbf{k}, c\mathbf{k} - \mathbf{q}) P_{vc}(\mathbf{q}) \right] - i\hbar \left[\gamma_{\mathbf{k}} P_{vc}(\mathbf{k}) - \sum_{\mathbf{q}} \gamma_{\mathbf{k}, \mathbf{q} - \mathbf{k}} P_{vc}(\mathbf{q}) \right]. \end{aligned}} \quad (15)$$

There are equivalent equations for the carrier distributions but we will not be interested in them due to the time scales involved.

3.1.6 Many-body gain

We now solve Eq. (15) to obtain the gain formula. The polarization function can be formally written as

$$P_{vc}(\mathbf{k}, t) = \int_{-\infty}^t dt' e^{(i\omega_{\mathbf{k}} + \gamma_{\mathbf{k}})(t'-t)} \left\{ -i\Omega_{\mathbf{k}}(z, t') \Delta_{\mathbf{k}}(t') - \sum_{\mathbf{q}} \gamma_{\mathbf{k}, \mathbf{q}-\mathbf{k}} P_{vc}(\mathbf{q}, t') \right\}, \quad (16)$$

where

$$\omega_{\mathbf{k}} = \frac{1}{\hbar} [\epsilon_c(\mathbf{k}) - \epsilon_v(\mathbf{k})], \quad \Delta_{\mathbf{k}}(t) = n_{c\mathbf{k}}(t) - n_{v\mathbf{k}}(t) \quad (17)$$

$$\Omega_{\mathbf{k}}(z, t) = \frac{1}{\hbar} \left[\mathbf{d}_{cv} \cdot \mathbf{E}_c(\mathbf{k}, 0, t) + \frac{1}{A} \sum_{\mathbf{q} \neq 0} v(c\mathbf{k}, v\mathbf{k} - \mathbf{q}; v\mathbf{k}, c\mathbf{k} - \mathbf{q}) P_{vc}(\mathbf{q}, t) \right]. \quad (18)$$

We now make the quasiequilibrium approximation for which the carrier densities vary little with time and the rotating-wave approximation which is valid at optical frequencies and consists in dropping the integrated part of the polarization which is not resonant. Then

$$P_{vc}(\mathbf{k}, t) = -\frac{i}{\hbar} \Delta_{\mathbf{k}}(t) \mathbf{d}_{cv} \cdot \mathbf{E}_c(\mathbf{k}, 0, t) \frac{1}{i(\omega_{\mathbf{k}} - \omega) + \gamma_{\mathbf{k}}} + \sum_{\mathbf{q} \neq 0} \left[-\frac{i}{\hbar} \frac{\Delta_{\mathbf{k}}(t)}{A} v(c\mathbf{k}, v\mathbf{k} - \mathbf{q}; v\mathbf{k}, c\mathbf{k} - \mathbf{q}) - \gamma_{\mathbf{k}, \mathbf{q}-\mathbf{k}} \right] \int_{-\infty}^t dt' e^{(i\omega_{\mathbf{k}} + \gamma_{\mathbf{k}})(t'-t)} P_{vc}(\mathbf{q}, t'). \quad (19)$$

Finally, one makes use of the Pade approximation. One assumes that the polarization can be iterated in terms of the Coulomb potential (the Mott regime) and the consequent series resummed using the lowest-order Pade approximation. The zeroth-order (free-carrier) polarization is

$$P_{vc}^{(0)}(\mathbf{k}, t) = -\frac{i}{\hbar} \Delta_{\mathbf{k}}(t) \mathbf{d}_{cv} \cdot \mathbf{E}_c(\mathbf{k}, 0, t) \frac{1}{i(\omega_{\mathbf{k}} - \omega) + \gamma_{\mathbf{k}}} \equiv \epsilon_0 \chi_{\mathbf{k}}^{(0)}(t) \mathcal{E} e^{-i\omega t}, \quad (20)$$

where

$$\chi_{\mathbf{k}}^{(0)}(t) = -\frac{i d_{cv}(\mathbf{k}) [n_{c\mathbf{k}}(t) - n_{v\mathbf{k}}(t)]}{\epsilon_0 \hbar [i(\omega_{\mathbf{k}} - \omega) + \gamma_{\mathbf{k}}]}, \quad (21)$$

and $d_{cv} \equiv \epsilon \cdot \mathbf{d}_{cv}$. The next term is

$$P_{vc}^{(1)}(\mathbf{k}, t) = \epsilon_0^2 \chi_{\mathbf{k}}^{(0)}(t) \mathcal{E} e^{-i\omega t} \frac{1}{d_{cv}(\mathbf{k}) A} \sum_{\mathbf{q} \neq 0} \left[V_{|\mathbf{k}-\mathbf{q}|}^{cvc} + \frac{i\hbar \gamma_{\mathbf{k}, \mathbf{q}-\mathbf{k}}}{\Delta_{\mathbf{k}}(t)} \right] \chi_{\mathbf{q}}^{(0)}(t) \quad (22)$$

$$= \epsilon_0^2 \chi_{\mathbf{k}}^{(0)}(t) \mathcal{E} e^{-i\omega t} Q(\mathbf{k}, t), \quad (23)$$

where the Coulomb potential

$$V_{|\mathbf{k}-\mathbf{q}|}^{cvc} = v(c\mathbf{k}, v\mathbf{k} - \mathbf{q}; v\mathbf{k}, c\mathbf{k} - \mathbf{q}),$$

and the Coulomb enhancement

$$Q(\mathbf{k}, t) = \frac{1}{d_{cv}(\mathbf{k})A} \sum_{\mathbf{q} \neq 0} \left[V_{|\mathbf{k}-\mathbf{q}|} + \frac{i\hbar\gamma_{\mathbf{k},\mathbf{q}-\mathbf{k}}}{\Delta_{\mathbf{k}}(t)} \right] \chi_{\mathbf{q}}^{(0)}(t).$$

One finally has

$$g(E) - i\frac{d\phi}{dz} = \frac{q}{2\varepsilon_0\varepsilon_b\hbar V} \sum_{\mathbf{k}} \frac{|d_{cv}(\mathbf{k})|^2(f_{c\mathbf{k}} - f_{v\mathbf{k}})}{i(\omega_{\mathbf{k}} - \omega) + \gamma_{\mathbf{k}}} \frac{1}{1 - \varepsilon_0 Q(\mathbf{k}, t)}, \quad (24)$$

where the Pade approximation has been used.

3.2 Analysis of band-structure and many-body effects

Having derived one form of the gain equation, we would like to identify a number of features present which derive from the dimensionality, the electronic dispersion relation, and the many-body theory. For the dimensionality, there are two considerations: ideal 2D and quasi-2D (q2D). For the electronic dispersion, there are a whole host of issues: infinite or finite barriers, parabolic bands, nonparabolicity or full band structure, constant or k -dependent optical matrix elements. The many-body theory gives rise to plasma screening, Coulomb potential, band-gap renormalization, Coulomb enhancement, carrier-carrier scattering, and carrier-phonon scattering. Recall that these considerations are not all independent. Indeed, we will be mainly interested in how the first two affect the last one.

3.2.1 Finite barriers

The simplest model of electron confinement is that due to an infinite barrier, for which there are exact solutions for the confinement energies of a single QW:

$$E_{\nu} = \frac{\hbar^2 \pi^2 \nu^2}{2L_w^2 m_{\nu}^*}. \quad (25)$$

A much better fit to transition energies can be obtained by root finding the solutions to the dispersion relation for a finite-barrier QW:

$$\frac{k_w}{m_w^*} \tan\left(\frac{k_w L_w}{2}\right) = \frac{\kappa_b}{m_b^*}, \quad (26)$$

where

$$k_w = \sqrt{\frac{2m_w^* E}{\hbar^2}}, \quad \kappa_b = \sqrt{\frac{2m_b^* (V_b - E)}{\hbar^2}},$$

and m_w^* (m_b^*) is the well (barrier) effective mass and V_b is the potential barrier height.

3.2.2 Nonparabolicity

We have introduced nonparabolicity into the conduction band by using a Kane-type modification of the effective mass into an energy-dependent one:

$$\frac{m^*(E)}{m^*} = \frac{E_g}{E_g + E}, \quad E = \frac{\hbar^2 k^2}{2m^*(E)}, \quad (27)$$

giving the electron energy as

$$E = \left[-1 + \sqrt{1 + \frac{4}{E_g} \frac{\hbar^2 k^2}{2m^*}} \right] \frac{E_g}{2}. \quad (28)$$

3.2.3 Optical matrix elements

As a first approximation, one can use bulk zone-center matrix elements multiplied by a polarization factor for the selection rules involving heavy-hole and light-hole states, and an overlap factor between the envelope functions. Within the infinite-barrier approximation, the overlap factor leads to the selection rule $\Delta\nu = 0$. The dipole matrix element is then given by

$$d_{cv}^2(\nu, \mathbf{k}) = \begin{cases} d_{\text{bulk}}^2(\mathbf{k})(1 - |\hat{\mathbf{e}} \cdot \hat{\mathbf{k}}|^2) & c - hh \\ d_{\text{bulk}}^2(\mathbf{k})(\frac{1}{3} + |\hat{\mathbf{e}} \cdot \hat{\mathbf{k}}|^2) & c - lh \end{cases}, \quad (29)$$

where $\hat{\mathbf{e}}$ is the unit polarization vector and $\hat{\mathbf{k}}$ the growth direction.

One can also assume constant bulk matrix elements or introduce a Kane-type k dependence:

$$d_{mn}(\mathbf{k}) = d_{mn}(0) \frac{E_{mn}(0)}{E_{mn}(\mathbf{k})}. \quad (30)$$

3.2.4 2D vs. quasi-2D

One of the places where current theories is most lacking is in a proper treatment of the quasi-2D nature of the theory. Even the latest and most sophisticated published work in the field do not appear to consider the quasi-2D nature fully and consistently [1, 2, 3, 4, 5]. The basic difference between a pure 2D and a q2D theory of the electrons lies in the neglect of the envelope function and in the subband structure of the band structure. While some workers have taken the subband nature of the electrons into consideration in computing the single-particle energies, they nonetheless still used a 2D form for the Coulomb potential.

We here give the exact 2D form factor for the infinite-barrier case:

$$\begin{aligned}
F_q(lmnp) &= qL_w \left\{ \frac{\delta_{l-p+m-n,0} + \delta_{l-p-m+n,0} - \delta_{l+p+m-n,0} - \delta_{l+p-m+n,0}}{[(m-n)^2\pi^2 + q^2L_w^2]} \right. \\
&\quad \left. - \frac{\delta_{l-p+m+n,0} + \delta_{l-p-m-n,0} - \delta_{l+p+m+n,0} - \delta_{l+p-m-n,0}}{[(m+n)^2\pi^2 + q^2L_w^2]} \right\} \\
&\quad - 16\pi^4 q^2 L_w^2(lmnp) \\
&\quad \times \frac{1 + (-1)^{m+n+l+p} - [(-1)^{l+p} + (-1)^{m+n}]e^{-qL_w}}{[(m+n)^2\pi^2 + q^2L_w^2][(m-n)^2\pi^2 + q^2L_w^2][(l+p)^2\pi^2 + q^2L_w^2][(l-p)^2\pi^2 + q^2L_w^2]}. \\
&\equiv I_q(lmnp) + H_q(lmnp).
\end{aligned} \tag{31}$$

$$\tag{32}$$

This result had previously been given by Ell and Haug [6].

3.2.5 Quasi-Fermi levels

We assume rapid thermalization within the conduction and the valence bands such that the carrier distributions can be described by quasi-Fermi-Dirac distributions and one quasi-Fermi level (or chemical potential) for each band.

The simplest model is for a perfect 2D system with parabolic bands. There is an exact expression for the chemical potential if there is only one occupied subband:

$$\beta\mu = \ln \left[e^{\frac{\pi\hbar^2\beta}{m^*}n_i} - 1 \right], \tag{33}$$

where n_i is the areal carrier density. If there are N_ν occupied subbands, the chemical potential has to be found from the root of the following equation:

$$n = \sum_{\nu} \frac{m_{\nu}^*}{\pi\hbar^2\beta} \ln \left[1 + e^{\beta(\mu - \Delta E_{\nu}(0))} \right], \tag{34}$$

where n is the net areal carrier density in the given band, and $\Delta E_{\nu}(0)$ is the energy of the bottom of each subband ν with respect to the bottom of the lowest subband (the reference energy for the definition of μ). In the electron-hole picture, the Fermi-Dirac distribution is

$$f_{\nu}(\mathbf{k}) = \frac{1}{e^{\beta(E_{\nu}(\mathbf{k}) + \Delta E_{\nu}(0) - \mu)} + 1}. \tag{35}$$

3.2.6 Plasma screening

In order to incorporate correlation effects, one can screen the Coulomb potential using the Lindhard dielectric function. In practical terms, one often make use of the static plasmon-pole approximation:

$$\frac{1}{\epsilon(\mathbf{q})} = 1 - \frac{\omega_p^2}{\omega_q^2}, \tag{36}$$

where

$$\omega_p^2 = L_w \sum_{\nu} \frac{n_{\nu}}{m_{\nu}} q^2 V_q(\nu), \quad (37)$$

$$\omega_q^2 = \omega_p^2 \left(1 + \frac{q}{\kappa}\right) + \frac{C}{N_{\nu}^2} \left(\sum_{\nu} \frac{\hbar q^2}{2m_{\nu}}\right)^2, \quad (38)$$

$$\kappa = q \sum_{\nu} V_q(\nu) \frac{\partial n_{\nu}}{\partial \mu_{\nu}} = q \sum_{\nu} V_q(\nu) \frac{m_{\nu}}{\pi \hbar^2} \left(1 - e^{-\frac{\pi \hbar^2}{m_{\nu}} \beta n_{\nu}}\right), \quad (39)$$

where L_w is the well width, n_{ν} the areal carrier density in subband ν , N_{ν} the total number of subbands in the problem, and C a constant usually taken between 1 and 4. These screening parameters have been derived for the ideal 2D and parabolic band cases. This should be a suitable approximation.

3.2.7 Hartree-Fock self energy

We obtain the following exchange energy within the HF approximation:

$$\Delta E_{n\nu\sigma}^{\text{exc}}(\mathbf{k}) = -\frac{1}{A} \sum_{2, \mathbf{q} \neq 0} v(v\nu\sigma\mathbf{k}, n_2\nu_2\sigma\mathbf{k} - \mathbf{q}; v\nu\sigma\mathbf{k}, n_2\nu_2\sigma\mathbf{k} - \mathbf{q}) f_{n_2\nu_2\sigma}(\mathbf{k} - \mathbf{q}). \quad (40)$$

Note that all the spin indices are the same, confirming that it is the exchange contribution. This is a generalization of the exchange self-energy expression (due to the summation over the 2 quantum numbers) found in the many-body gain literature, although we have found an identical result in the band-gap renormalization work of Ryan and Reinecke [7]. In practice, one also screens the Coulomb potential.

3.2.8 Coulomb hole self-energy

At the HF level, the Coulomb hole (CH) self energy is introduced *ad hoc* to account for electron correlation (screening) due to the Coulomb interaction. It is defined to be

$$\Delta E^{\text{CH}} \equiv \frac{1}{L^D} \sum_{\mathbf{q} \neq 0} [V_s(q) - V(q)] = V_s(\mathbf{r} = 0) - V(\mathbf{r} = 0). \quad (41)$$

In Fourier space, it is seen to represent the difference between a screened Coulomb energy and the unscreened one; it is thus a correction to the HF theory which does not contain screening effects. In real space, one sees that it is the difference in the self-energy of a charge interacting with itself due to the screened and unscreened Coulomb potential; of course, each self-energy is divergent but the difference is well-defined.

3.2.9 Scattering terms

Carrier-carrier scattering

$$\gamma_{\mathbf{k}} = \frac{1}{\hbar} (\Gamma_{\mathbf{k}}^{\nu_e} + \Gamma_{\mathbf{k}}^{\nu_h}) \quad (42)$$

$$\begin{aligned} \Gamma_{\mathbf{k}}^{\nu} &= \frac{1}{A^2} \sum_{\nu', \nu'', \nu'''} \sum_{\mathbf{q}, \mathbf{k}'} 2\pi V_{s,q}^2 (\nu', \nu'''; \nu'', \nu) \delta(\epsilon_{\nu}(\mathbf{k}) + \epsilon_{\nu''}(\mathbf{k}') - \epsilon_{\nu'}(\mathbf{k} + \mathbf{q}) - \epsilon_{\nu'''}(\mathbf{k}' - \mathbf{q})) \\ &\times [f_{\nu'\mathbf{k}+\mathbf{q}}(1 - f_{\nu''\mathbf{k}'})f_{\nu'''\mathbf{k}'-\mathbf{q}} + (1 - f_{\nu'\mathbf{k}+\mathbf{q}})f_{\nu''\mathbf{k}'}(1 - f_{\nu'''\mathbf{k}'-\mathbf{q}})], \end{aligned} \quad (43)$$

$$\gamma_{\mathbf{k},\mathbf{q}} = \frac{1}{\hbar} (\Gamma_{\mathbf{k},\mathbf{q}}^{\nu_e} + \Gamma_{\mathbf{k},\mathbf{q}}^{\nu_h}) \quad (44)$$

$$\begin{aligned} \Gamma_{\mathbf{k},\mathbf{q}}^{\nu} &= \frac{1}{A} \sum_{\nu', \nu'', \nu'''} \sum_{\mathbf{k}'} 2\pi V_{s,q}^2 (\nu', \nu'''; \nu'', \nu) \delta(\epsilon_{\nu}(\mathbf{k}) + \epsilon_{\nu''}(\mathbf{k}') - \epsilon_{\nu'}(\mathbf{k} + \mathbf{q}) - \epsilon_{\nu'''}(\mathbf{k}' - \mathbf{q})) \\ &\times [(1 - f_{\nu\mathbf{k}})(1 - f_{\nu''\mathbf{k}'})f_{\nu'''\mathbf{k}'-\mathbf{q}} + f_{\nu\mathbf{k}}f_{\nu''\mathbf{k}'}(1 - f_{\nu'''\mathbf{k}'-\mathbf{q}})]. \end{aligned} \quad (45)$$

These results are taken from Ref. [8].

Carrier-LO phonon scattering

We consider the Fröhlich mechanism between the carriers and the 3D phonons:

$$\begin{aligned} \Gamma_{LO,\mathbf{k}}^{\nu} &= \frac{\pi e^2 \hbar \omega_q L_w}{2\epsilon_0^2 V} \left(\frac{1}{\epsilon_{\infty}} - \frac{1}{\epsilon_s} \right) \sum_{\mathbf{q}_{||}} \frac{1}{q_{||}} [(N_{ph} + 1) \delta(\epsilon_{\nu}(\mathbf{k}) - \epsilon_{\nu}(\mathbf{k} - \mathbf{q}) - \hbar \omega_q) F^{\nu}(\mathbf{k}, -\mathbf{q}) \\ &+ N_{ph} \delta(\epsilon_{\nu}(\mathbf{k}) - \epsilon_{\nu}(\mathbf{k} + \mathbf{q}) + \hbar \omega_q) F^{\nu}(\mathbf{k}, \mathbf{q})], \end{aligned} \quad (46)$$

where

$$F^{\nu}(\mathbf{k}, \mathbf{q}) = \sum_{i_1 \dots i_4} \int \int dz dz' f_{\nu\mathbf{k}}^{i_1*}(z) f_{\nu\mathbf{k}}^{i_2*}(z') f_{\nu\mathbf{k}-\mathbf{q}}^{i_3}(z') f_{\nu\mathbf{k}-\mathbf{q}}^{i_4}(z) e^{-q_{||}|z-z'|}. \quad (47)$$

3.3 Results

We now list some of the key results. Our canonical system is a 100 Å GaAs infinite-barrier quantum well, using a two-band parabolic approximation with a constant linewidth broadening factor. External conditions are a temperature of 300K and a carrier density of $3 \times 10^{18} \text{ cm}^{-3}$.

- Band-gap renormalization redshifts the gain spectrum.

The SX contribution is k -dependent and smaller than the CH energy [Fig. 1, 2]. The BGR is reduced if form factors are used for the Coulomb matrix elements. With the q2D form factor within the infinite-barrier approximation, the BGR is of the order of 40-50 meV.

- The many-body gain is enhanced over the free-carrier value [Fig. 3].

This is predominantly due to the Coulomb enhancement factor.

- A nonparabolic conduction band increases the high-energy slope of the gain spectrum.
- Use of a constant optical matrix element (instead of a k dependent one) leads to a slight lower gain, the effect being more pronounced for the many-body gain [Fig. 3].
- The intrasubband carrier-carrier scattering is dominated by the h - h scattering over e - e and both are reduced by the form factors. They add up to ≈ 15 meV [Fig. 4].
- The intrasubband e -LO phonon Fröhlich interaction contributes ≈ 5 meV to the linewidth [Fig. 5].

4 Summary

Using an infinite-barrier solution to the electron problem, we have evaluated the many-body gain of quantum-well structures. This has allowed a detailed understanding of the influence of various factors on the many-body effects and on the gain.

A number of issues were identified for further study:

- Incorporate a realistic band structure.

We are currently adapting a four-band band-structure code and concurrently upgrading our gain program. Once these tasks are completed, we will combine the two.

- Continuum states.

High-temperature operation of the laser will require consideration of the continuum states. This would constitute novel work.

- Fully-quantized Hamiltonian.

This would constitute a generalization of the formalism presented in the current report and would allow one to study other effects at the same level (e.g., spontaneous emission and phonons).

- Electron-phonon interaction.

We have here only considered optical phonons with flat dispersion relations. It is known that the electron-phonon scattering rate is modified if realistic phonon dispersions are used.

Acknowledgments

I would like to thank Dr. John Loehr and Dr. Rocky Sherriff for their collaboration on this project.

References

- [1] W. W. Chow, S. W. Koch, and M. Sargent III. *Semiconductor-laser physics*. Springer-Verlag Berlin (1994).
- [2] D. Campi, P. J. Bradley, R. Calvani, and R. Caponi. Modeling of nonlinear absorption and refraction in quantum-well structures for all-optical switching. *IEEE J. Quantum Elec.* **29**, 1144 (1993).
- [3] C.-F. Hsu, P. S. Zory, C.-H. Wu, and M. A. Emanuel. Coulomb enhancement in InGaAs-GaAs quantum-well lasers. *IEEE J. Selected Topics in Quantum Electronics* **3**, 158 (1997).
- [4] W. W. Chow, A. Knorr, S. Hughes, A. Girndt, and S. W. Koch. Carrier correlation effects in a quantum-well semiconductor laser medium. *IEEE J. Selected Topics in Quantum Electronics* **3**, 136 (1997).
- [5] B. P. C. Tsou and D. L. Pulfrey. The influence of Coulomb enhancement on the modulation properties of quantum-well lasers. *IEEE J. Quantum Elec.* **34**, 318 (1998).
- [6] C. Ell and H. Haug. Absorption and optical gain spectra and band gap renormalization of highly excited quantum well systems. *Phys. Stat. Sol. (b)* **159**, 117 (1990).
- [7] J. C. Ryan and T. L. Reinecke. Band-gap renormalization of optically excited semiconductor quantum wells. *Phys. Rev. B* **47**, 9615 (1993).
- [8] W. W. Chow, P. M. Smowton, P. Blood, A. Girndt, F. Jahnke, and S. W. Koch. Comparison of experimental and theoretical GaInP quantum well gain spectra. *Appl. Phys. Lett.* **71**, 157 (1997).

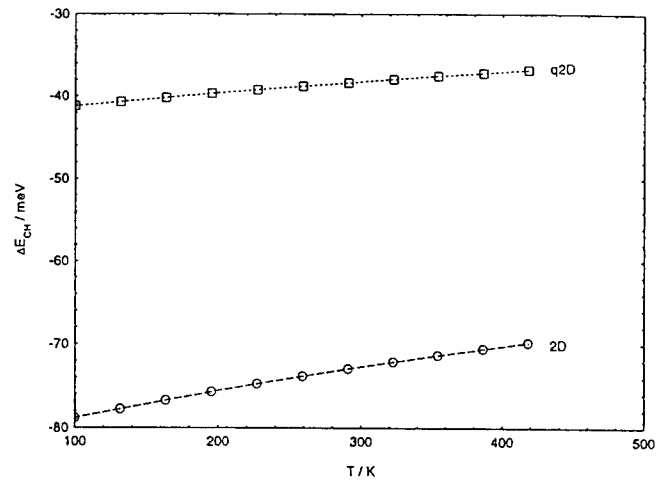


Figure 1: Influence of form factor and temperature on the Coulomb hole gap.

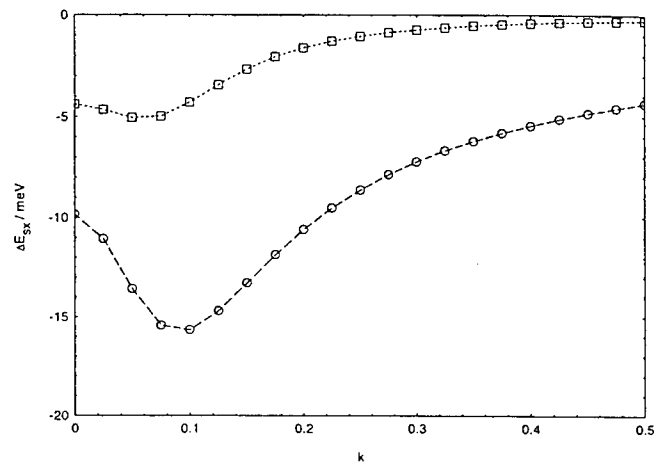


Figure 2: Influence of form factor and temperature on the screened exchange gap.

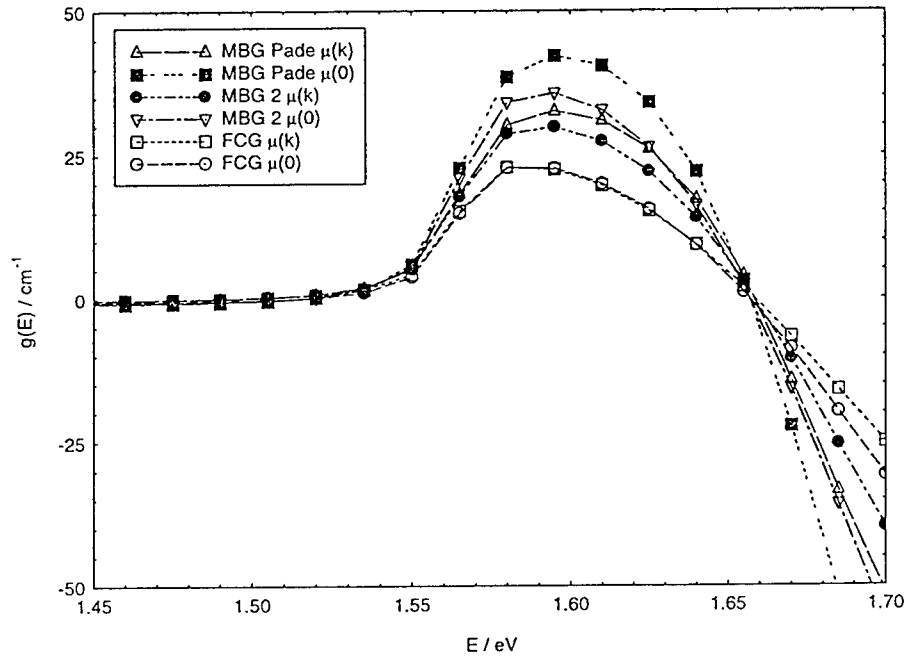


Figure 3: Free-carrier gain and many-body gain (using a two-term iteration and Pade resummation) with constant and k -dependent optical matrix elements.

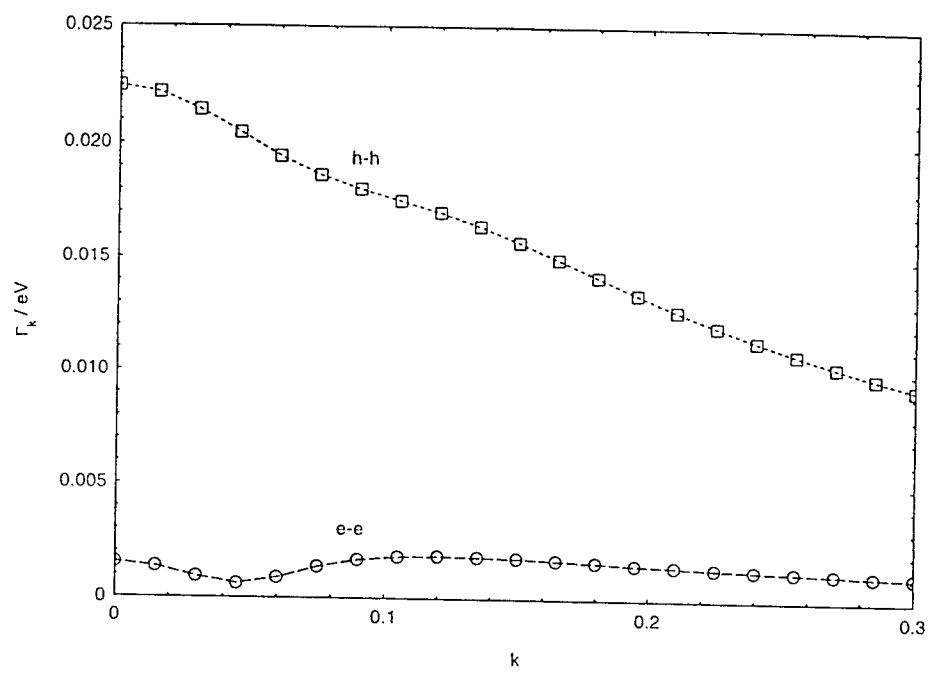


Figure 4: k dependence of electron-electron and hole-hole scatterings.

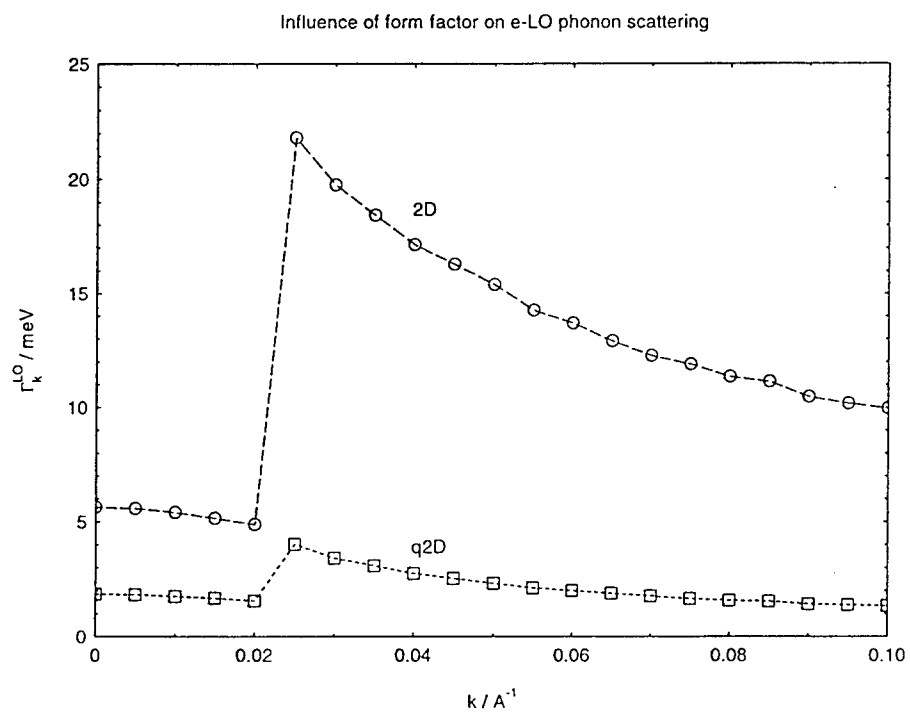


Figure 5: Influence of form factor and wave vector on electron-LO phonon scattering.

**A STUDY OF REFERENCING ISSUES IN MULTIPLATFORM AND
MULTISENSOR BASED OBJECT LOCATION**

Rongxing (Ron) Li, Ph.D.
Associate Professor
Department of Civil and Environmental Engineering and Geodetic Science
The Ohio State University
470 Hitchcock Hall, 2070 Neil Venue
Columbus, Ohio 43210-1275

Final Report
for
AFOSR Summer Faculty Research Program
Air Force Research Laboratory, WPAFB

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

September 1998

A STUDY OF REFERENCING ISSUES IN MULTIPLATFORM AND MULTISENSOR BASED OBJECT LOCATION

Rongxing (Ron) Li, Ph.D.

Associate Professor

Department of Civil and Environmental Engineering and Geodetic Science

The Ohio State University

470 Hitchcock Hall, 2070 Neil Venue

Columbus, Ohio 43210-1275

Abstract

This report summarizes the result of a two-month research effort in investigation of georeferencing in multiplatform and multisensor based object location. The environment used in this report can be characterized by a set of dynamic platforms such as satellites, aircraft, and terrestrial vehicles with various sensors, for example, GPS, INS, SAR, EO, and FLIR. The sensors may be installed on the platforms separately or combined. It is assumed that communication between the platforms is available, so that real-time applications demanding coordination among the platforms can be carried out. The object to be located is simplified as a point on the ground. Data acquired by this kind of integrated sensor system can be characterized by its vast volume, comprehensive spatial relationships, and associated temporal dynamics.

After a brief introduction, chapter 2 reviews mathematical models of least squares adjustment and gives a comparison between two methods used in modeling sensor uncertainties. A conceptual model of dynamic object location by a multiplatform and multisensor system is introduced in chapter 3. In chapter 4, design of an experiment using optical images, a SAR image, and a terrain model is given to test the concept in chapter 3. Finally, conclusions are drawn in chapter 5.

A STUDY ON REFERENCING ISSUES IN MULTIPLATFORM AND MULTISENSOR BASED OBJECT LOCATION

Rongxing (Ron) Li

1. Introduction

This study investigates referencing issues related to object location using data such as imagery and navigation data acquired by a multiplatform and multisensor system. The environment used in this report can be characterized by a set of dynamic platforms such as satellites, aircraft, and terrestrial vehicles with various sensors, for example, GPS, INS, SAR, EO, and FLIR. The sensors may be installed on the platforms separately or combined. It is assumed that communication between the platforms is available, so that real-time applications demanding coordination among the platforms can be carried out. The object to be located is simplified as a point on the ground. Data acquired by this kind of integrated sensor system can be characterized by its vast volume, comprehensive spatial relationships, and associated temporal dynamics.

In principle, one sensor or a combination of sensors allow determination of the location of the object relative to the sensor(s) (Moffitt, and Mikhail 1980, Curlander and McDonough 1991, Layne and Blasch 1997). The navigation sensors add absolute information into the system. Thus, the relative location of the object is referenced to a ground coordinate system (Li 1996). Since various sensors have their strengths and weakness, a well-designed strategy is necessary to minimize the uncertainty of the object location using the given platforms and sensors. Particularly, issues of individual sensor and platform uncertainty, correlation between sensors and platforms, and final object location uncertainty are concerned.

This report discusses Least Squares models in Chapter 2, concept of an improved dynamic model in Chapter 3, design of an experiment in Chapter 4, and conclusions in Chapter 5.

2. Least Squares Model for Object Location

Suppose that l is a vector representing observations such as ranges and range rates from SAR and image coordinates measured from electro-optic or FLIR images. Vector X (state) contains all unknown parameters such as platform locations and coordinates of the object. In principle, the relationship between the observations and the state can be described by

$$l + V = F(X) \quad (1)$$

where V is a vector of white noise (corrections) associated with the observation l . Linearization of (1) using a Taylor series leads to

$$l + V = F(X_0) + A \Delta X + O(\Delta X). \quad (2)$$

Here $X = X_0 + \Delta X$, with X_0 being the approximate value of X and ΔX being increment based on X_0 . Matrix A is a Jacobian matrix defined as

$$A = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_m} \\ \vdots & \dots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \dots & \frac{\partial f_n}{\partial x_m} \end{pmatrix}. \quad (3)$$

$O(\Delta X)$ is higher order terms. Assume that ΔX is small and we neglect $O(\Delta X)$. The observation equation can be written as

$$V = A \Delta X + F(X_0) - l, \quad (4)$$

or

$$V = A \Delta X + L. \quad (5)$$

Each observation will make one row in matrix equation (5). The uncertainty of observation l_i is represented by its variance σ_i^2 . Similarly the covariance between l_i and l_j is σ_{ij} . The covariance matrix of the observation vector l is

$$R = \Sigma_{ll} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{pmatrix}. \quad (6)$$

The weight matrix is defined as $W = \sigma_0^{-2} R^{-1}$. Here σ_0^2 is the unit weight variance.

$$W = \sigma_0^{-2} R^{-1} = \begin{pmatrix} w_1 & w_{12} & \dots & w_{1n} \\ w_{21} & w_2 & \dots & w_{2n} \\ \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_n \end{pmatrix}. \quad (7)$$

If the observations are uncorrelated, $\sigma_{ij} = 0$. Thus, Σ_{ll} is a diagonal matrix, so is W , with $w_{ij} = 0$ ($i \neq j$) and $w_i = \sigma_0^2 / \sigma_i^2$.

The objective is to estimate the unknown vector ΔX from the observations L (actually l) by Equation (5) that is usually overdetermined (more equations/observations than unknowns). A Least Squares Adjustment (LSA) is going to determine the ΔX so that $V^T W V = \text{Sum}(w_i v_i^2) = \min$. A normal equation is built as

$$\begin{aligned} & (A^T W A) \Delta X + A^T W L = 0 \\ \text{or} \quad & N \Delta X + A^T W L = 0. \end{aligned} \quad (8)$$

The LSA estimate is

$$\Delta X = -N^{-1} A^T W L. \quad (9)$$

The covariance matrix of ΔX is

$$\Sigma_X = \sigma_o^2 N^{-1} = \sigma_o^2 (A^T W A)^{-1}. \quad (10)$$

There are different forms of the LSA model including some extended models, for example, those discussed in Moffitt and Mikhail (1980), Leick (1995), Mikhail and Gracie. Extension of LSA using Kalman filter for handling GPS data and other dynamically collected data can be found in Leack (1995).

The object to be located has coordinates (X, Y, Z). Its covariance submatrix, the upper-left 3X3 submatrix, can be extracted from the covariance matrix of the state

$$\Sigma_X = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} & & & \\ \sigma_{yx} & \sigma_y^2 & \sigma_{yz} & & & \\ \sigma_{zx} & \sigma_{zy} & \sigma_z^2 & & & \\ & & & \sigma_{x_{s1}}^2 & & \\ & & & & \ddots & \\ & & & & & \sigma_{k_{s1}}^2 \end{pmatrix}.$$

We can describe the location uncertainty by an ellipsoid that is derived from the covariance submatrix (Mikhail and Gracie).

2.1 General LS model

A general LS model has an observation equation

$$A X + B V = f, \quad (11)$$

B is a coefficient matrix for V that is unit matrix -I in Equation (5). Assume that $V_e = BV$ and $I_e = BI$. $R_e = BRB^T$ Equation (11) becomes

$$AX + V_e = f.$$

Comparing to Equation (5), the LS solution is

$$X = [A^T (BRB^T)^{-1} A]^{-1} [A^T (BRB^T)^{-1} f] \text{ and} \\ R_x = [A^T (BRB^T)^{-1} A]^{-1}.$$

2.2 Application in object location

Suppose that we have observations from n sensors (or one sensor measuring n times) to locate an object. The state vector is $X = (X_t, Y_t, Z_t, X_{S1}, Y_{S1}, Z_{S1}, \dots, X_{Sn}, Y_{Sn}, Z_{Sn})^T$. Here we compare two ways of using LS to solve the problem. Method I splits the state vector to two vectors, object state $X_1 = (X_t, Y_t, Z_t)^T$ and sensor state $X_2 = (X_{S1}, Y_{S1}, Z_{S1}, \dots, X_{Sn}, Y_{Sn}, Z_{Sn})^T$. X_2 is then combined with noise vector V . The observation equation becomes

$$(A_1 \ A_2) \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + V = A_1 X_1 + (A_2 \ I) \begin{pmatrix} X_2 \\ V \end{pmatrix} = A_1 X_1 + B \begin{pmatrix} X_2 \\ V \end{pmatrix} = f. \quad (12)$$

The general LS model is applied to treat X_1 as unknowns and $(X_2 \ V)^T$ as noise. The LS solution is

$$X_1 = \{A_1^T [(A_2 \ I) R (A_2 \ I)^T]^{-1} A_1\}^{-1} \{A_1^T [(A_2 \ I) R (A_2 \ I)^T]^{-1} f\} \\ = [A_1^T (B R B^T)^{-1} A_1]^{-1} [A_1^T (B R B^T)^{-1} f]. \\ R_{X1} = \{A_1^T [(A_2 \ I) R (A_2 \ I)^T]^{-1} A_1\}^{-1} \\ = [A_1^T (B R B^T)^{-1} A_1]^{-1}. \quad (13)$$

Comments are: a) the direct measurements are ranges, range rates, and others, not the sensor locations X_2 ; b) the sensor state should be treated as signals (unknowns to be estimated), instead of a part of noise, the squared sum of which is minimized by LSA; c) depending on prior variances given to X_2 , the estimate of X_1 may be adjusted either too much or less. We should note that measurements are made between X_1 and X_2 . There is usually correlation between them. Also, there should be correlation between the observation l , noise V , X_1 and X_2 . Any assumption of independence between them needs to be proven.

Method II proposes to treat both object location and sensor locations as unknowns to be estimated. The observation equation is

$$(A_1 \ A_2) \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + V = f. \quad (14)$$

The LS solution is

$$X = (X_1 \ X_2)^T = [(A_1 \ A_2)^T R^{-1} (A_1 \ A_2)]^{-1} [(A_1 \ A_2)^T R^{-1} f]$$

$$R_X = \begin{pmatrix} R_{x_1} & R_{x_{12}} \\ R_{x_{21}} & R_{x_2} \end{pmatrix} = [(A_1 \ A_2)^T R^{-1} (A_1 \ A_2)]^{-1}. \quad (15)$$

The covariance matrix of the object location R_{X1} is a 3x3 upper-left submatrix in Equation (15). Unless someone can prove that Equations (13) and (15) are identical, the estimates from Method I and II will be different. The advantages of Method II are that correlations between object and sensor locations are considered; sensor locations are estimated as the results of LSA; and the uncertainty estimated should be more realistic.

2.3 An example

Laser ranges are used to locate an object as described in Figure 1. Measured are ranges r_m from sensor (X_s, Y_s, Z_s) to the object at (X, Y, Z) , with noise of V_{r_m} .

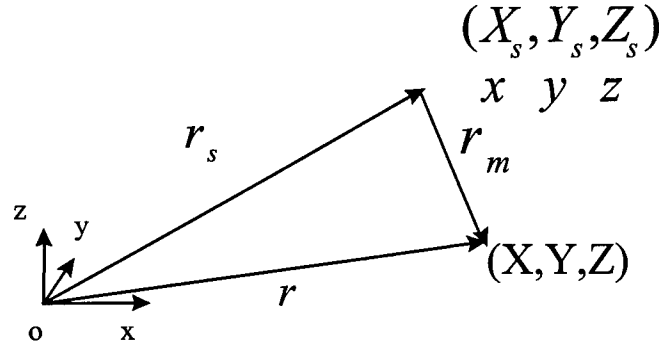


Figure 1. Laser ranges for object location (X, Y, Z)

The observation equation is

$$\begin{aligned} r_m + V_{r_m} &= |r - r_s| = \left[(X - X_s)^2 + (Y - Y_s)^2 + (Z - Z_s)^2 \right]^{1/2} \\ &= ULOS_x dX + ULOS_y dY + ULOS_z dZ - ULOS_x dX_s - ULOS_y dY_s \\ &\quad - ULOS_z dZ_s + r_o + O(\Delta) \end{aligned}$$

Here $ULOS_i$ are direction cosines of the unit vector of the Line of Sight (LOS):

$$\begin{aligned} ULOS_x &= \frac{(X - X_s)}{\sqrt{(X - X_s)^2 + (Y - Y_s)^2 + (Z - Z_s)^2}} \\ ULOS_y &= \frac{(Y - Y_s)}{\sqrt{(X - X_s)^2 + (Y - Y_s)^2 + (Z - Z_s)^2}} \end{aligned}$$

$$ULOS_z = \frac{(Z - Z_s)}{\sqrt{(X - X_s)^2 + (Y - Y_s)^2 + (Z - Z_s)^2}} \text{ and}$$

$$r_o = \sqrt{(X - X_{s_o})^2 + (Y - Y_{s_o})^2 + (Z - Z_{s_o})^2}.$$

The linearized observation equation is

$$r_m - r_o = (ULOS_x \ ULOS_y \ ULOS_z \ -ULOS_x \ -ULOS_y \ -ULOS_z) \begin{pmatrix} dX \\ dY \\ dZ \\ dX_s \\ dY_s \\ dZ_s \end{pmatrix} + (-V_{r_m}).$$

Note that $r_m - r_o$ is small, but usually not zero. dX , dY , dZ , dX_s , dY_s , and dZ_s are unknowns to be estimated. They are signals, not noises in V_{r_m} .

If Method I would be used, the observation equation should be:

$$r_m - r_o = (ULOS_x \ ULOS_y \ ULOS_z) \begin{pmatrix} dX \\ dY \\ dZ \end{pmatrix} + (-ULOS_x - ULOS_y - ULOS_z - 1) \begin{pmatrix} dX_s \\ dY_s \\ dZ_s \end{pmatrix} + V_{r_m}.$$

$$= AX + BV.$$

In comparison to Equation (13) for Method I,

$$f = r_m - r_o, \quad X = \begin{pmatrix} dX \\ dY \\ dZ \end{pmatrix}, \quad V = \begin{pmatrix} dX_s \\ dY_s \\ dZ_s \\ V_{r_m} \end{pmatrix}.$$

The solution is then

$$X = [A^T (BRB^T)^{-1} A]^{-1} [A^T (BRB^T)^{-1} f]$$

$$R_x = [A^T (BRB^T)^{-1} A]^{-1}. \quad (16)$$

To compare with Method I, we reformat the observation equation as

$$(A \ B) (X_1 \ X_2)^T + V = (A \ B) X + V = f$$

Method II gives a LS estimate and covariance matrix of X

$$\begin{aligned} X &= [(A \ B)^T R^{-1} (A \ B)]^{-1} [(A \ B)^T R^{-1} f] \\ R_x &= [(A \ B)^T R^{-1} (A \ B)]^{-1} \end{aligned} \quad (17)$$

Note that matrix B in Equations (15) and (17) are different. The upper-left 3x3 *submatrix* is the covariance matrix R_x of the object location X_1 . This *submatrix* is usually not the same of R_x in Equation (16).

2.4 Discussion

In observation equation (1) or (4), l contains measurements (available); V is white noise associated with l (can be determined after LSA). Once X_0 is given, $L=F(X_0) - l$ can be calculated; A is a "design" matrix (to be calculated before LSA). Finally, ΔX is unknown (to be determined). There are three phases of LSA.

Phase 0 (Planning): Before actual measurements, l is not available and ΔX cannot be calculated. However, given expected quality of observations l , the covariance matrix, which is often diagonal under the assumption that observations are uncorrelated, can be derived. The weight matrix W is then available, given unit weight variance σ_0^2 . The design matrix A depends on the configuration of the overall system such as arrangement of sensors, sensor types, and other conditions. X_0 is needed to calculate A , but not ΔX and l . Consequently, the covariance matrix of ΔX , Σ_x , can be computed using Equation (10). Furthermore, for any vector variable Z that is a function of X , $Z=BX$, its covariance matrix can be derived by an error propagation:

$$\Sigma_Z = B \Sigma_x B^T. \quad (18)$$

In general, the task in the planning phase is to investigate the expected quality (covariance matrix) of the unknowns and other derivatives without actual observations. If coordinates (X_p, Y_p, Z_p) of an object P to be determined are contained in $X=(\dots X_p, Y_p, Z_p)$, their variances and covariances are contained in Σ_x . The error ellipsoid that represents the uncertainty of the expected location is characterized by its three axes.

Efforts should be made to analyze the quality and combinations of the individual error sources σ_i^2 of observation l_i represented in the weight matrix of observations and configuration of sensors reflected in the design matrix A . An estimation of the object location uncertainty can then be obtained. Another way is to fix the uncertainty of the object location so that the covariance matrix of the unknowns is fixed. Inversely the weight matrix or design matrix can be calculated.

Phase 1 (Data Processing): It deals with evaluation of the vector ΔX from observations l . Since the number of equations is usually larger than the number unknowns, a least squares solution is given by Equation (9). The estimate of the unknowns is then $X=X_0+\Delta X$. Note that the high order terms in Equation (2) were neglected. This requires

ΔX to be determined iteratively with the updated X being new approximate value of X_0 . This procedure will continue until ΔX is significantly small.

Phase 2 (Uncertainty Analysis): In Equation (10) an estimate of the covariance matrix of X is given. It describes the variances of the estimated unknowns and covariances between unknown parameters. Uncertainty obtained in Phase 0 is the predicted uncertainty. However, uncertainty estimated in Phase 2 is the one associated with the LS solution of X in Phase 1. For example, error ellipsoids of estimated object locations can be visualized and analyzed. In addition, uncertainties of parameters that are functions of the unknowns can also be calculated. Statistic tests can also be used to test the data processing result from Phase 1 to conclude if the result is acceptable or not (Mikhail and Gracie).

3. Concept of an Improved Dynamic Model

Object location is critical to a wide range of applications, such as military targeting, rescue operations, emergency management, and environmental monitoring. Real-time operations are often required. Sensors involved in such a real-time system (Figure 2) are from multiple platforms and of various classes such as SAR, laser, infrared, electron-optic array, hyperspectral, GPS, INS and others. To meet the goal of accurate object location, a good understanding of the characteristics of individual sensors and their collective impact on the object location is needed. To solve the problem of multiplatform and multisensor based target location, we should address a series of issues on analysis of object location uncertainty, simulation of sensor-object correlation, uncertainty minimization, and optimization of platform and sensor configuration.

This requires an innovative mathematical model for characterizing uncertainties of individual sensors as well as those associated with sensor and target locations to be estimated. A dynamic model for target location and uncertainty propagation will accurately update the target location and uncertainty in real-time based on multiplatform and multisensor observations collected up to the current time. If the location uncertainty is too high, the system will suggest additional observations using the available sensors and platforms within the configuration. An extended least squares estimator will determine the final target location using all observations collected to give the highest accuracy possible.

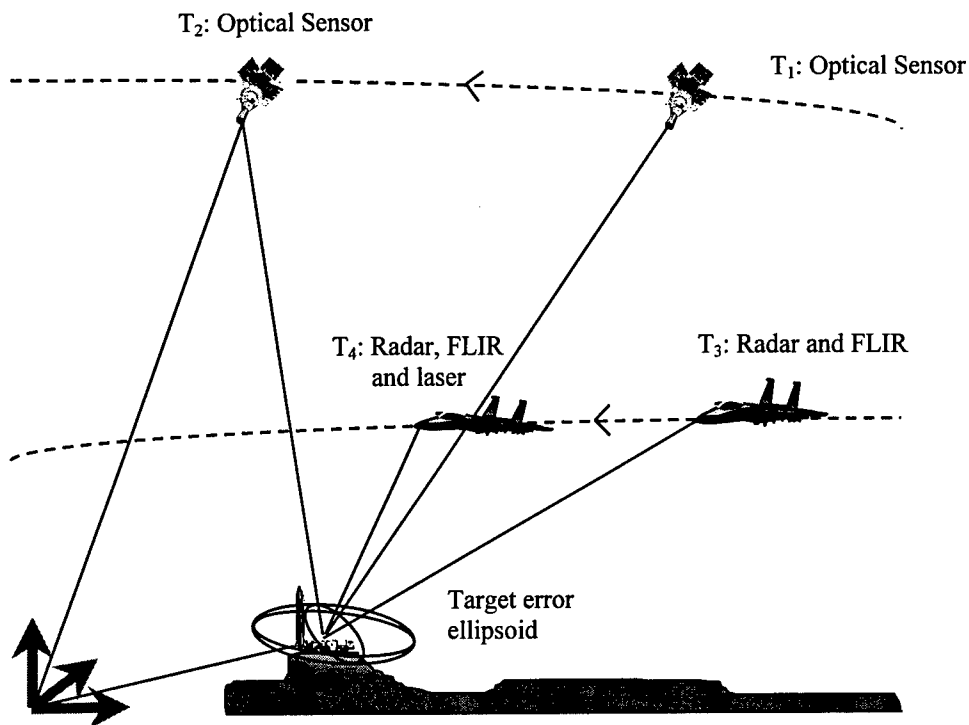


Figure 2. Object location in an integrated sensor system

The strategy to handle the comprehensive multiplatform and multisensor observations is to separate direct measurements (e.g. time) of sensors from their resulting geometric output (e.g. range), so that existing sophisticated sensor processing software will deal with sensor internal errors to compute geometric output and estimate associated errors. Algorithms and software systems for processing signals within a sensor are usually provided by manufactures or other sources. Geometric observations from sensors are, for instance, range and range rate from SAR, range from laser, image coordinates from airborne or satellite optical/infrared images, locations, angles and velocity from GPS and INS. A covariance matrix of the observations is usually simplified to a diagonal matrix, indicating non-correlation between the observations. The task is then to examine correlations within a sensor (for example range and range rate), within a platform, and within a sensor system. Depending on the research results the observation covariance matrix may or may not be diagonal. This will certainly affect the target location and uncertainty.

3.1 Target uncertainty in an integrated sensor system

The uncertainty of an object location, represented by a 3-D ellipsoid, is usually large, based on a reconnaissance resource (e.g. satellite in Figure 2). But this approximate location provides an initial location for platforms with higher accurate sensors to reduce the location uncertainty. For example, given GPS and INS navigation accuracies, aircraft takes radar range measurements of the object at t_k and t_{k+1} (may not be “continuous”, Figure 3 (a)). The resulting ellipsoid of the object is relatively flat (almost parallel to XY-plane), namely good vertical accuracy and relatively poor horizontal accuracy. If a FLIR image either from the same platform of the radar or another platform is available and used in determination of the target location, the corresponding ellipsoid will become smaller in horizontal plane and have relatively homogeneous errors in all three directions (Figure 3(b)). It should be noted that the size and orientation of the object location error ellipsoid are functions of a number of factors in the sensor system such as types of measurements, from where they are taken with respect to the object, and combined impact of all sensors. Such a comprehensive error behavior demands the development of a mathematical model to allow flexible analysis of all the factors involved for the optimal design of sensor configuration, mission planning, and accurate object location.

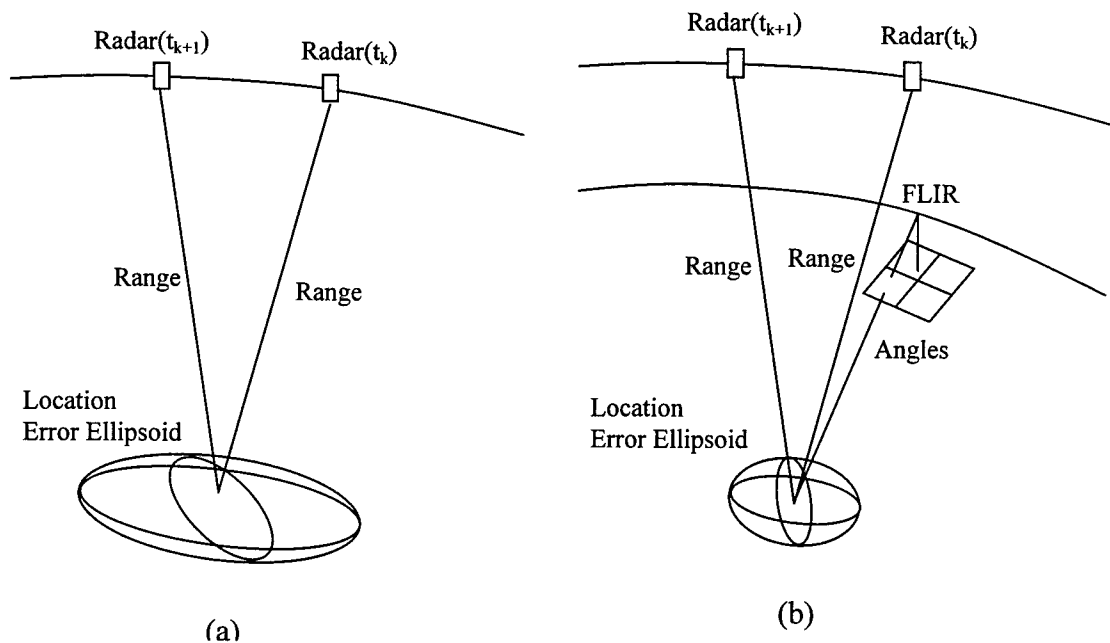


Figure 3. An example of uncertainty reduction, (a) ellipsoid by a two-range triangulation, (b) adding a FLIR measurement to reduce the uncertainty.

3.2 Dynamic model

Suppose that we start with a rough position of the object determined by a reconnaissance resource. The estimated object location can be improved when new observations become available. Each platform is defined by a body position and orientation which are

determined by an INS unit. Offsets, including translation and rotation, between the body and other sensors are usually estimated by a calibration procedure. Sensor orientation is derived from the platform orientation and calibration parameters. The accumulative nature of the pointing errors will be investigated.

In a dynamic model, observations are those from sensors. The state vector includes three coordinates of the object, three coordinates and three orientation angles of each platform (instead of each sensor) so that the size of the state vector is effectively reduced. The model estimates the state vector and covariance matrix at t_{k+1} based on their values at t_k . The covariance matrix of observations is given according to a priori knowledge of the quality of each sensor class. The dynamic behavior of the state (including object location) and uncertainty during transition will be studied. A dynamic estimator, such as Kalman filter, will be used to minimize the object/sensor location uncertainty and to estimate the optimal object location.

In contrast, a static model will treat all observations as static observations, no matter from which platform, which sensor, and when taken. The observations along with a covariance matrix will be input into a Least Squares (LS) adjustment system with a very large observation vector and a very large state vector. The optimal state vector is estimated by minimizing the sum of squared noises of all observations. An efficient algorithm will be developed to simplify the LS model to solve the 3x1 target state vector and 3x3 target location covariance submatrix. The static model supplies the optimal target location and associated covariance matrix once all observations are taken.

Based on above dynamic and static models, we are able to simulate a multiplatform and multisensor mission for an optimal platform and sensor configuration given a desirable object uncertainty. The models can also be used for determining the accurate object location if actual measurements are taken.

3.3 Equations

In this dynamic model, the state vector at any time t is defined as

$$\begin{aligned} X(t) = (X, Y, Z, \dots, X_{p1}, Y_{p1}, Z_{p1}, \dot{X}_{p1}, \dot{Y}_{p1}, \dot{Z}_{p1}, \omega_{p1}, \varphi_{p1}, \kappa_{p1}, \dot{\omega}_{p1}, \dot{\varphi}_{p1}, \dot{\kappa}_{p1}, \dots, \\ X_{pn}, Y_{pn}, Z_{pn}, \dot{X}_{pn}, \dot{Y}_{pn}, \dot{Z}_{pn}, \omega_{pn}, \varphi_{pn}, \kappa_{pn}, \dot{\omega}_{pn}, \dot{\varphi}_{pn}, \dot{\kappa}_{pn})^T. \end{aligned} \quad (19)$$

Here (X, Y, Z) are coordinates of the object to be located. The rest of the vector are position, velocity, orientation and angular velocity of n platforms at time t . Once calibrated, position and orientation parameters of sensors on each platform can be derived based on the parameters of the platform, calibration parameters (positional and angular offsets), and gimbal readings if any. The dimension of $X(t)$ is $3+12n$. Observations at time t are

$$l(t) = (x_{\text{push}}, 0, x_{\text{FLIR}}, y_{\text{FLIR}}, r_{\text{SAR}}, \dot{r}_{\text{SAR}}, r_{\text{LASER}}, \dot{r}_{\text{LASER}}, \dots, x_{pi}, y_{pi}, z_{pi}, \dot{x}_{pi}, \dot{y}_{pi}, \dot{z}_{pi}, \omega_{pi}, \varphi_{pi}, \kappa_{pi}, \dot{\omega}_{pi}, \dot{\varphi}_{pi}, \dot{\kappa}_{pi}, \dots)^T. \quad (20)$$

Here the observations may be coordinates of the object in a push broom satellite image and/or a FLIR image, a range and range rate from a SAR image, a laser range, GPS and INS measurements of position, velocity, orientation angles, and angular velocity of the i th platform. In the dynamic model R_k is the covariance matrix of l at k th iteration. L_o is the approximate value of L . $Z = L - L_o$ is usually very small, but not zero. The dynamic states of the platforms can be estimated by a Kalman filter with the observation equation at time t_k :

$$Z_k = H_k X_k + V_k. \quad (21)$$

The covariance of V_k is $E(V_k V_k^T)$. H is a Jacobian matrix containing partial derivatives of Z with respect to X . V_k is white noise of Z_k . The dynamic update of X is performed by

$$X_{k+1} = \Phi_k X_k + W_k, \quad (22)$$

Where Φ_k is state transition matrix to be determined, W_k is white noise of X_k with known covariance matrix of P_k . $Q_k = E(W_k W_k^T)$ will be developed.

The iterative solution of the problem starts with given prior covariance matrix P_k^- and the approximate/prior estimate X_k^- of X_k . The Kalman gain is defined as

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1}. \quad (23)$$

The state vector and covariance matrix are updated by

$$\begin{aligned} X_k &= X_k^- + K_k (Z_k - H_k X_k^-), \text{ and} \\ P_k &= (I - K_k H_k) P_k^-. \end{aligned} \quad (24)$$

To repeat the process, the “project ahead” parameters are

$$\begin{aligned} X_{k+1}^- &= \phi_k X_k, \text{ and} \\ P_{k+1}^- &= \phi_k P_k \phi_k^T + Q_k. \end{aligned} \quad (25)$$

H_k is usually a function of X and should be updated along with Z_k each time when a new value of X is obtained. From Equation (24) it is obvious that diagonal elements of $K_k H_k$ must be smaller than 1 to make a variance great than 0. Therefore, in design stage alternative configuration of platforms and sensors can be made to produce different matrix H_k and Kalman gain so that the object location error or the trace of P_k will be minimized.

3.4 Object recognition with georeference information

Object recognition is a process to find an object from a set of sensed data, such as an optical image or SAR image. Tremendous efforts have been made by scientists to solve this problem. The progress in this research area is yet modest in comparison to other fast growing areas because of the complexity of the problem. Target recognition can be improved by using georeference information in the following ways:

- The known sensor geometry and object/sensor uncertainty will greatly reduce target searching dimensions and ranges. For example, a two-dimensional search in image domain can be reduced to an one-dimensional search along an epipolar line using epipolar geometry derived from known camera orientation information;
- If the position of the target is found in one image, its position in other images (even from different sensors) can be predicted by using a DTED or a second image. This provides a very strong geometric constraint that will be valuable for candidate validation or matching between images; and
- Since sensor orientation parameters are known, a great deal of information in scene domain can be used to assist target recognition in image domain. For instance, if a 3-D design model of the object is available, through a transformation from scene to image domain, a comparison between the design model and the target candidate in images can be performed. Once the object is recognized in all images, its accurate location in the scene domain can then be determined using the above static LS model.

4. Design of an Experiment

To verify some of the above concepts, an experiment with actual data will be carried out. The data set includes one scene of aerial photograph taken by a mapping camera and a SAR image taken by a spotlight radar system. Both images cover an area with objects that have known coordinates. The photograph was scanned to produce a digital image of 4,000 rows and 9,000 columns. Camera focal length, principal point location, and lens distortion parameters are given. The SAR image has a dimension of 1800x1800 pixels. It covers a part of the aerial photograph. Pixel resolution, scene center, and other parameters are available. This experiment is a) to calculate horizontal coordinates of known objects based on one SAR data under the assumption that the terrain is flat in the area, b) to use both the aerial photograph and SAR image for object location, and c) to find either another aerial photographs or a DTED of the area, plus the aerial photograph and SAR image, for precision object location.

4.1 Spotlight SAR for horizontal coordinates calculation

From pixel to image coordinates

An objects is measured by its coordinates of (row,column)=(i,j) in pixel. The coordinates are then converted to image coordinates (x,y):

$$x = -(i-900) dx \text{ and } y = -(j-900) dy. \quad (26)$$

(i,j)=(900,900) is the location of the scene center in the image. dx and dy are pixel size in row and column direction respectively.

Correction from the display plane to horizontal plane

There is an angle β_o between the display plane and the horizontal plane. Image coordinates (x,y) on the display plane is transformed to those (x',y') on the horizontal plane by

$$\begin{aligned} x' &= x \\ y' &= y \cos\beta_o. \end{aligned} \quad (27)$$

Transformation to the Processed Scene System

If the ground squint angle is θ , the angle that rotates the image plane to the PCS is

$$\theta_{ac} = 180^\circ - \theta.$$

The coordinates in PCS are calculated as

$$\begin{aligned} X_p &= x' \cos\theta_{ac} + y' \sin\theta_{ac} \\ Y_p &= -x' \sin\theta_{ac} + y' \cos\theta_{ac}. \end{aligned} \quad (28)$$

Suppose the terrain in the area is flat, then (X_p, Y_p) are coordinates of the object in PCS.

If a terrain model is available, corrections on coordinates can be computed according to the elevation difference between the point and the ground plane of the image – elevation of the scene center.

Transformation from PCS to ECEF

Transformation PCS to Earth Center Earth Fixed (ECEF) coordinate system consists of a rotation and a translation. The rotation is defined by a rotation matrix R_{PSS}^{ECEF} that is a function of longitude, latitude, and heading of the planed flight. The amount of translation is determined by coordinates of the scene center $(X_{PSC}, Y_{PSC}, Z_{PSC})$ in ECEF:

$$\begin{pmatrix} X_{ECEF} \\ Y_{ECEF} \\ Z_{ECEF} \end{pmatrix} = R_{PCS}^{ECEF} \begin{pmatrix} X_p \\ Y_p \\ Z_p \end{pmatrix} + \begin{pmatrix} X_{PSC} \\ Y_{PSC} \\ Z_{PSC} \end{pmatrix}. \quad (29)$$

4.2 Aerial photograph and SAR for object location

The objective is to compute the vector r at point $P(X,Y,Z)$ by measuring its image point $p(x,y)$, assuming that the camera exposure center location and orientation are known.

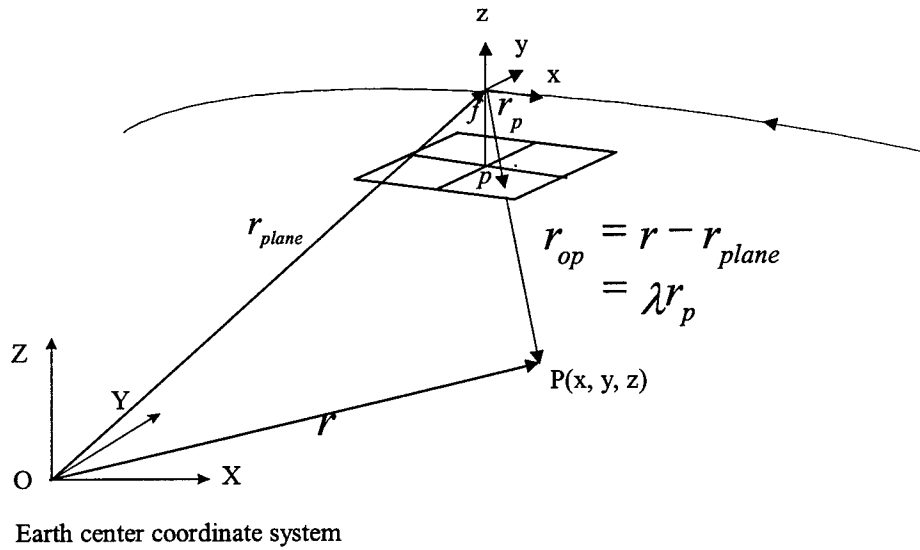


Figure 4. Aerial photograph for object location

The observation equation is

$$r_{op} = r - r_{plane} = \lambda R_{sensor}^{earth} \begin{pmatrix} x \\ y \\ -f \end{pmatrix} \text{ or}$$

$$\begin{pmatrix} x \\ y \\ -f \end{pmatrix} = \frac{1}{\lambda} R_{earth}^{sensor} \begin{pmatrix} X - X_{plane} \\ Y - Y_{plane} \\ Z - Z_{plane} \end{pmatrix}. \quad (30)$$

Here (x,y) are measured image coordinates of the object. R_{earth}^{sensor} is a rotation matrix from the earth coordinate system to the sensor coordinate system and λ is a scaling parameter. Dividing the first two equations by the third one in Equation (30) results in collinearity equations:

$$\begin{aligned}
x &= -f \frac{R_{11}(X-X_{plane})+R_{12}(Y-Y_{plane})+R_{13}(Z-Z_{plane})}{R_{31}(X-X_{plane})+R_{32}(Y-Y_{plane})+R_{33}(Z-Z_{plane})} \\
y &= -f \frac{R_{121}(X-X_{plane})+R_{22}(Y-Y_{plane})+R_{23}(Z-Z_{plane})}{R_{31}(X-X_{plane})+R_{32}(Y-Y_{plane})+R_{33}(Z-Z_{plane})}
\end{aligned} \tag{31}$$

Additional parameters such as lens distortion can be added to the above equation. For each object location we will have two equations for each aerial photograph and two equations for the SAR. Thus, we can solve for three unknown coordinates of the object (X,Y,Z).

4.3 Precision object location

If more observations such as image coordinates of the object in another aerial photograph, a DTED, or another SAR image are available, the object location can be determined with more redundant observations so that the reliability and accuracy of the location can be enhanced. Statistic tests will be used to check the object location estimate. Its final error ellipsoid is derived from the covariance matrix of the unknown vector of LS adjustment.

5. Conclusions

Based on the above research results, the following conclusion can be drawn:

- High accurate and reliable object location in a multiplatform and multisensor system requires precise modeling of individual sensor behavior, geometric configuration, and combined impact of the integrated system;
- Considerations should be given to the entire procedure of simulation and prior uncertainty estimation, optimal object location estimation, and post analysis;
- A dynamic model for the multiplatform and multisensor system should be developed and implemented to meet demand from a number of real-time applications; and
- Georeferencing can enhance capabilities of object recognition and other AI applications.

References

- Bar-Shalom, Y. and X.-R. Li 1993. Estimation and Tracking: Principles, Techniques, and Software. Artech, Boston and London.
- Brown, R.G. and P.Y.C. Hwang 1992. Introduction to Random Signals and Applied Kalman Filter. John Wiley & Sons, Inc.
- Curlander, J.C. and R.N. McDonough 1991. Synthetic Aperture Rada Systems and Signal Processing. John Wiley & Sons, Inc.
- Layne, J.R. and E.P. Blasch 1997. Integrated Synthetic Aperture Rada and Navigation Systems for Targeting Applications. Final Report WL-TR-97-1185, Air Force Research Laboratory, 47 pages.
- Leick, A. 1995. GPS Satellite Surveying. John Wiley & Sons, Inc.
- Li, R. 1997. Mobile Mapping: An Emerging Technology for Spatial Data Acquisition. Journal of Photogrammetric Engineering and Remote Sensing, Vol. 63, No. 9, pp.1085-1092.
- Mikhail, E.M. and G. Gracie. Analysis and Adjustment of Survey Measurements, Van Nostrand Reinhold Company.
- Moffitt, F.H. and E.M. Mikhail 1980. Photogrammetry. Harper&Row, Publishers, Inc., New York.

**SENSOR FUSION WITH PASSIVE MILLIMETER WAVE (PMMW)
AND LASER RADAR (LADAR) FOR TARGET DETECTION**

**Chun-Shin Lin
Associate Professor
Department of Electrical Engineering**

**University of Missouri – Columbia
Columbia, Missouri 65211**

**Final Report for:
Summer Faculty Research Program
Air Force Research Lab.
Munitions Directorate
Advanced Guidance Division (AFRL/MNG)
Eglin AFB, FL**

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

and

**Air Force Research Lab.
Munitions Directorate
Advanced Guidance Division (AFRL/MNG)
Eglin AFB, FL**

July 17, 1998

SENSOR FUSION WITH PASSIVE MILLIMETER WAVE (PMMW) AND LASER RADAR (LADAR) FOR TARGET DETECTION

Chun-Shin Lin
Associate Professor
Department of Electrical Engineering
University of Missouri – Columbia

Abstract

Advanced sensors and guidance techniques are required in killing mobile offensive and defensive systems. Many different sensors such as radar, video camera, laser radar (LADAR), millimeter wave (MMW) systems, infrared images, acoustic sensors, etc. are available for such usage. However, no single sensor seems to provide completely satisfactory capabilities. While some sensors have complimentary capabilities, integration of multiple sensors for kill can relax the task difficulty and provide more reliable results. The use of multiple sensors can also reduce the possibility of being defeated by countermeasures. In this project, we studied the framework and investigated potential techniques for integration and fusion of information from passive millimeter wave (PMMW) and laser radar (LADAR) systems. The focus has been on target detection. The PMMW is used to detect metal objects and the LADAR checks those regions of interest for other evidence of existence of a target. Advantages obtained by integrating these two sensors include reduction of task complexity and improvement of reliability, both due to efficient localization of regions of interests from the PMMW. Since PMMW possess weather penetration capabilities through fog, cloud, smoke, etc., the combined system has a near all weather capability. A LADAR provides three-dimensional (3-D) information for checking target details. It should be used as the primary sensor for target selection upon target detection. The framework of the fusion is based on the Dempster-Shafer decision method. The fusion may be done in the algorithm level and sensor level. With the Dempster-Shafer method as the framework, new sensors or new decision components can be integrated into the existent system easily.

SENSOR FUSION WITH PASSIVE MILLIMETER WAVE (PMMW) AND LASER RADAR (LADAR) FOR TARGET DETECTION

Chun-Shin Lin

I. Introduction

Advanced sensors and guidance techniques are required in killing mobile offensive and defensive systems. Possible sensors include radar, video camera, laser radar (LADAR), millimeter wave (MMW) systems, infrared images, acoustic sensors, etc. However, no single sensor seems to provide completely satisfactory capabilities. For instance, the laser range sensor and video camera are susceptible to adverse weather (such as the situations with fog, rain or cloud) as well as smoke and dust screen. On the other hand, radar and millimeter wave signals can penetrate fog, smoke, etc. but cannot provide as good spatial resolution as that from optical sensors. While some sensors have complementary capabilities, integration of multiple sensors for kill can relax the task difficulty and provide more reliable results. The use of multiple sensors also reduces the possibility of being defeated by countermeasures.

The objective of this project is to investigate frameworks and potential techniques for integration and fusion of information from passive MMW (PMMW) [1][2] and LADAR systems [3][4]. Although the study focuses only on the two sensors, further fusion with other sensors can be implemented.

In this report, the PMMW and LADAR are first briefly introduced in Section 2. The suggested framework is given in Section 3. More details on target detection through fusion are presented in Section 4. Section 5 gives experimental results. Data fusion and the use of the Dempster-Shafer method are discussed in Section 6. Discussion and conclusions are in Section 7.

II. PMMW and LADAR

2.1 Passive Millimeter Wave (PMMW)

Millimeter wave frequency bands span from 30 to 300 GHz. Due to atmospheric attenuation, the primary windows with minimal loss occur at 35, 94, 140 and 220GHz. Millimeter wave sensors at these frequencies provide a *near* all-weather capability. They represent a compromise between microwave

systems and the electro-optical systems. The MMW systems do not have the all-weather capabilities of microwave systems, but provide a better resolution. Compared to eletro-optical systems, MMW systems cannot give high resolution but are superior in situations with fog, cloud, smoke, dust screen, etc.

A passive millimeter wave system receives the signal reflected and emitted by observed objects and does not require an MMW source. This reduces the cost and system complexity, and also reduces the possibility of being detected in a missile attack mission.

Antenna temperature is one means for describing the amplitude of the sensed signal. The sources that determine the antenna temperature include the apparent brightness temperature of the sky, natural terrain materials and metal objects. The apparent brightness temperature of a source contains both the emitted and reflected energy from the source surface. Different materials have different reflection and emitting properties. One useful property for target detection is that metal hardly emits but does well reflect. Figure 1 shows PMMW images and visible light photos of a scene under good and poor weather conditions. It is seen that the fog does not affect the PMMW. For more a detailed introduction and discussion on PMMW, please refer to reference 1.

2.2 Laser Radar (LADAR)

Laser radar uses laser light for measuring distance. It can provide both a range image and a reflectance image and both can be used in missile guidance. The range image provides 3-D information and has high angle resolution. The sensor is especially good for target classification/recognition but is susceptible to adverse weather and cannot image through thick cloud, fog, etc. Due to its narrower field-of-view angle, it is also less adequate for searching targets in a broad area. The contrast in a range image is usually not large but the 3-D information is there.

Calibration Field Tests With Passive Metal Reflectors and Active Beacon

BRUNSWICK, MAINE NAS—JULY 21, 1990
MEDIUM FOG, VISIBILITY ~100 METER
SIZE OF A METAL PANEL 4' x 4'

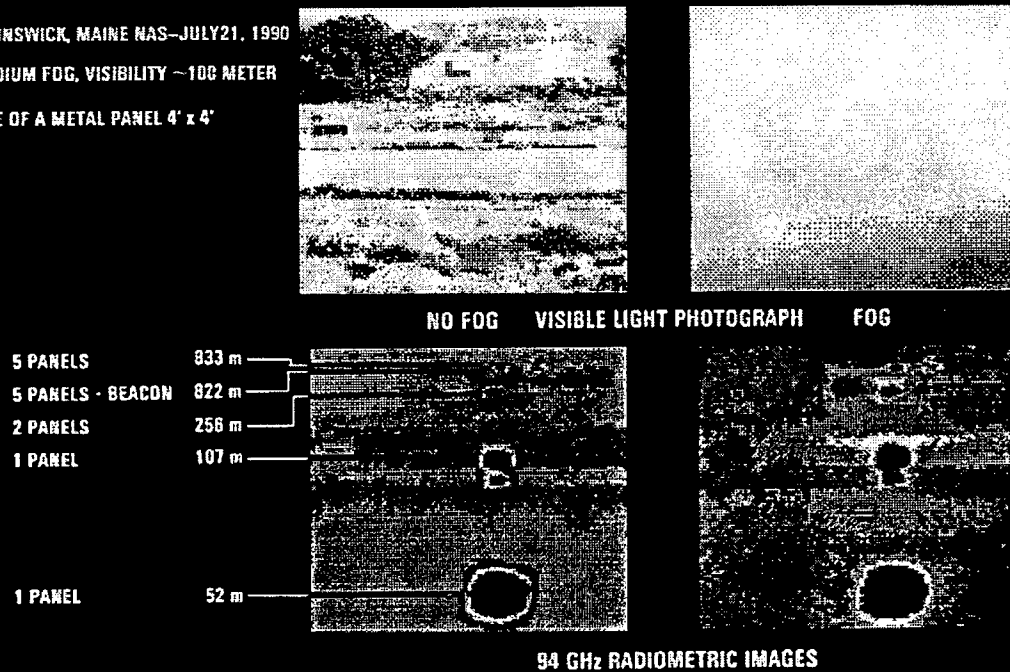


Figure 1. Visible light photos and PMMW images (from TRW)

2.3 Advantages from Integration of PMMW and LADAR

PMMW and LADAR have complementary capabilities and fusion with these two sensors will provide the following advantages:

- Make the detection task easier and the results more reliable; reduce the search time.

The PMMW has good capability to distinguish metal from other materials and the LADAR can easily detect objects above the ground level from 3-D information. Combining information from these two sensors, the target detection task becomes easier. In the situation with clutter in background, with only a video image or a range image, target detection is extremely difficult and requires a lot of complicated work [5][6]. However, with the use of both PMMW and LADAR, the task should become much easier. In addition, the result should be much more reliable.

- LADAR provides 3-D information, which is the most complete one for checking target details. LADAR can be used as the major sensor during the further step in target selection.
- The combination will create the near-all-weather capabilities.
- PMMW could provide some size information of targets in a shorter distance. Such information may become useful in case LADAR fails due to weather or other problems.

III. Framework for Integration/Fusion with PMMW and LADAR

There are two types of sensor integration – the competitive sensor integration and the complementary sensor integration. The former uses sensors returning essentially the same information. One example is the use of identical sensors to measure a pressure. The readings can be slightly different due to either noise or sensor inaccuracy. One can obtain a more reliable value through fusion among them. Complementary sensor integration uses sensors returning partial information on the physical entity sensed. The use of PMMW and LADAR belongs to this type.

In this development, the following capabilities/features are desired:

- Integration should best utilize the information from each sensor, extract the most useful features/evidence for decision making (either target detection or selection).
- The system should degrade only slightly when one sensor system fails due to adverse weather, countermeasures, or other causes.
- New sensors or new classifier components should be able to be integrated into the existent system easily. This means that the design should be in modules and the fusion should give soft decisions (i.e., possibility for each proposition instead of only “yes” or “no”).

Figure 2 shows a general diagram of sensor fusion with capability of adding sensors or algorithm components (for evidence detection). The diagram shows two levels of fusion – the algorithm/evidence level and the sensor level. The evidence/algorithm level permits ease of software modification. For example, one can combine the results from two types of classifiers and add additional classifier components without requiring major change to previously designed software. The sensor level makes it possible to fuse information with additional sensors, either internal or external.

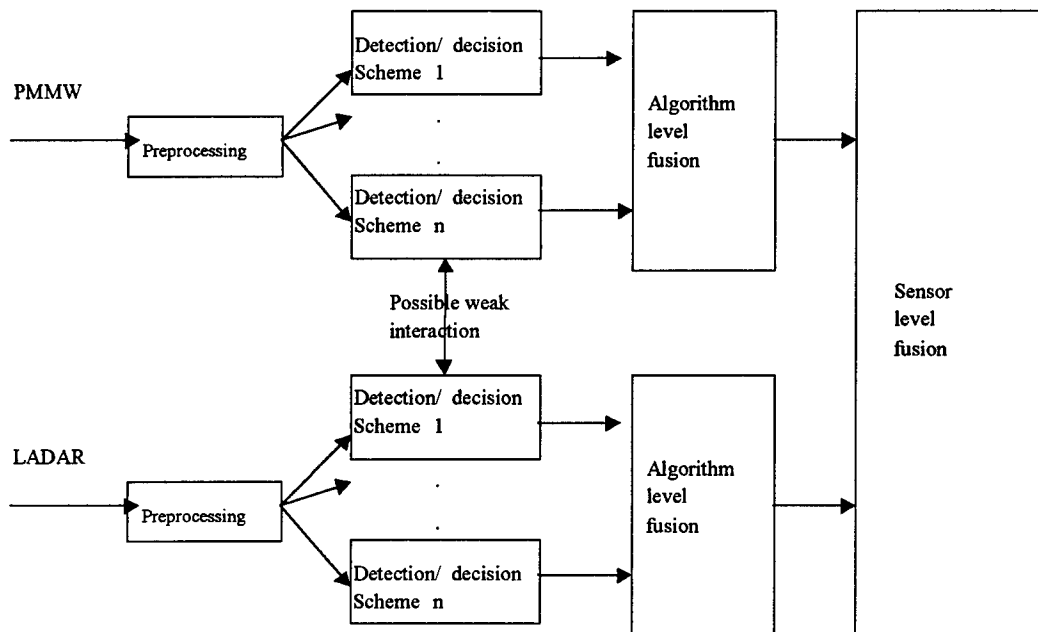


Figure 2. A general structure for sensor fusion

IV. Evidences for Target Detection

4.1 Evidences from PMMW and LADAR

In the millimeter-wave spectral region, metal objects have extremely low emissivity values and, hence are very reflective of millimeter wave energy. Metal objects mainly reflect the very small amount of energy coming from the sky, which can be thought of as radiometrically very cold. In contrast, the soil, trees and most ground objects emit MMW and appear relatively hot. Consequently, the cold temperature detected in a PMMW image is an evidence of a possible metal target.

LADAR range images on the other hand provide 3-D information. Objects above the ground, structured edges/surfaces, surfaces vertical to the ground, specific dimensions, etc., indicate the existence of possible targets.

4.2 Detection of Evidences

Three types of evidences were used in the illustration of the experimental results described below. These were cold temperature regions in PMMW images, and jump edges and vertical surfaces in LADAR

images. The cold areas in a PMMW image provide the regions of interest (ROI) for the LADAR system. The LADAR system checks each of these cold regions for the jump edge and vertical surface evidences. The detection of cold temperature employs the histogram information extracted from the PMMW image. A cutoff is selected near the location at which the histogram curve starts to increase quickly.

Jump edges are detected along each column of pixels. The range difference $R(\text{row}, \text{column}) - R(\text{row}+1, \text{column})$ is used to see if there is a jump edge (R denotes the range value). For a tall object, the range jump will be large. If the heights of the targets to be detected are known, the range differences for possible jump edges can be estimated. This can be used to eliminate objects that are either too low or too high from target detection consideration. By placing limits on the range difference, jump edges of a building can be removed.

Another evidence used in the experiments is the existence of vertical surfaces. The absolute value of the range difference, $|R(\text{row}, \text{column}) - R(\text{row}+1, \text{column})|$, is used. A small range difference is a criterion used in detecting vertical surfaces.

4.3 Diagram for Target Detection

For target detection, the PMMW is used to detect metal objects and the LADAR looks into the details in those areas for verification. The diagram with three evidences mentioned above is shown in Figure 3 to illustrate the detection task. Each evidence is detected/evaluated using a program module. The results are integrated through the fusion process.

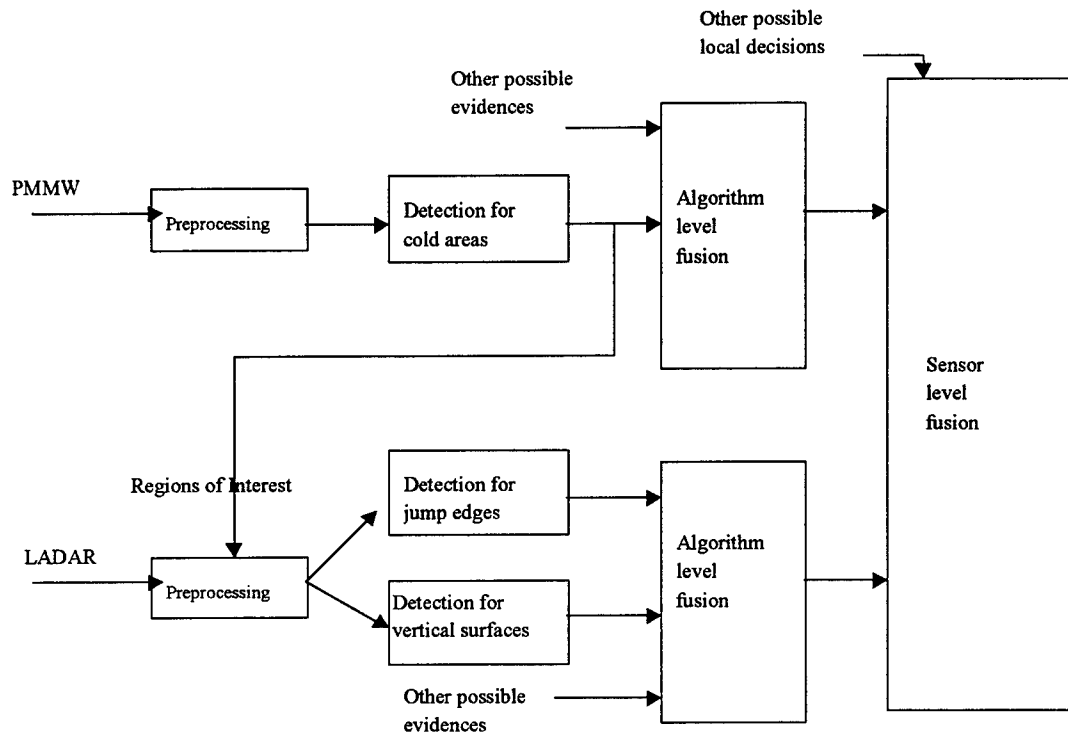
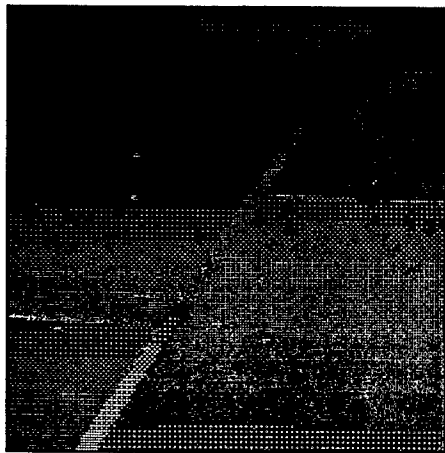


Figure 3. Diagram for target detection

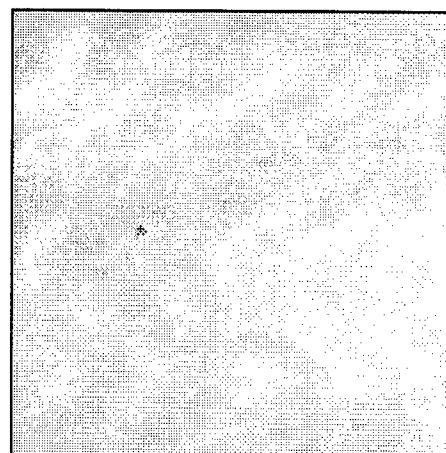
V. Experimental Results

For experiments, five PMMW and five laser range images were generated employing the IRMA multi-spectral target and background scene simulation. All images were for the same scene at different distances. The depression angle was 20 degrees and the distances were 1000m, 800m, 600m, 400m and 200m. These images are shown in Figures 4 and 5.

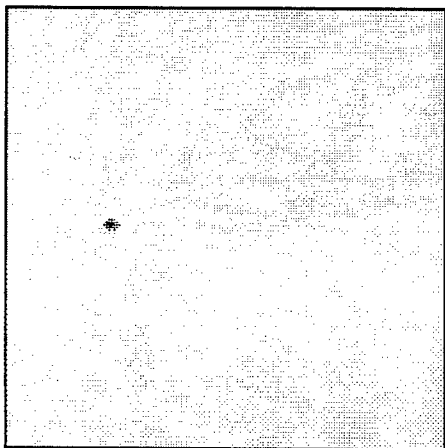
In a PMMW image, low energy areas can be obtained by simply selecting an appropriate cutoff value according to the image histogram. This value should be located near the point where the histogram curve starts to sharply incline. One typical histogram curve is shown in Figure 6. When the distance is longer, the image becomes less clear and the cutoff point should be selected at the higher end. Figure 7(a1), (b1), ... (e1) show the results. It is seen that in Figure 7(e1) some blobs do not belong to targets and some targets are missing. This is due to the stronger effects of blurring because of the relatively smaller target image at the longer distance. Validation should begin with the right-size blobs if computational time is critical.



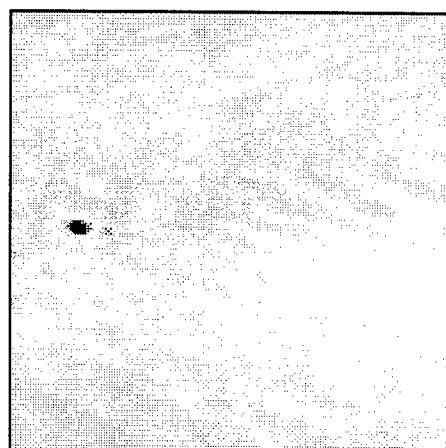
(a) The scene with 12 targets



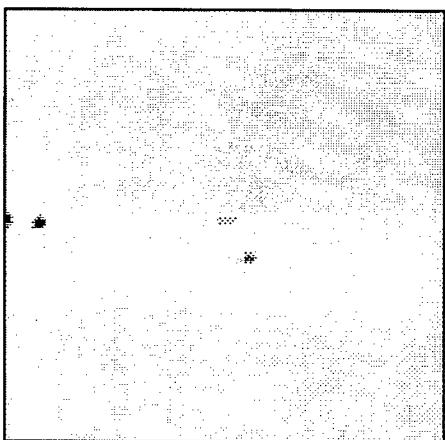
(b) PMMW image at 1000m



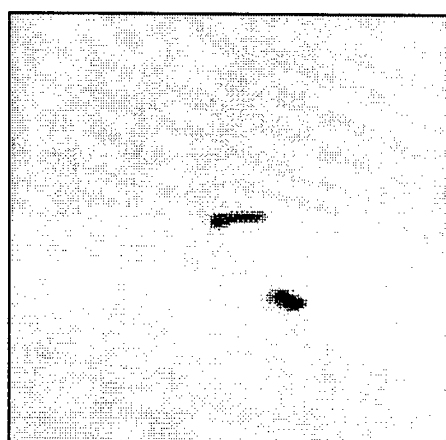
(c) PMMW image at 800m



(d) PMMW image at 600m

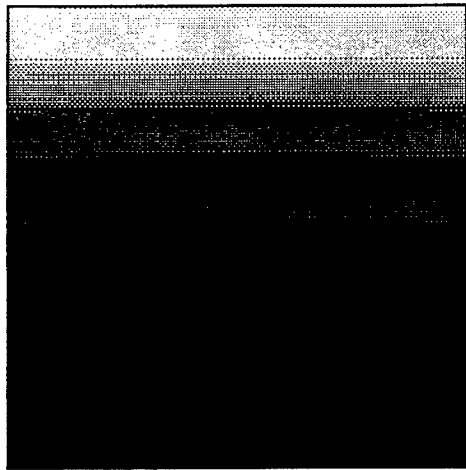


(e) PMMW image at 400m

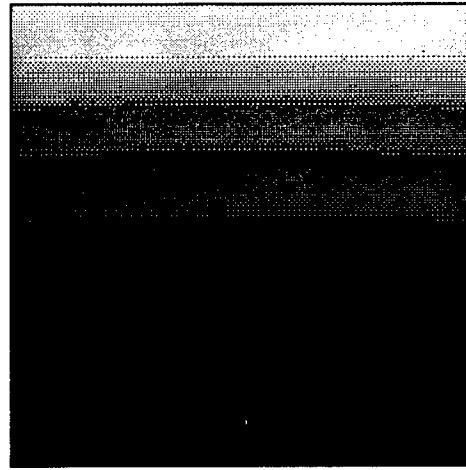


(f) PMMW image at 200m

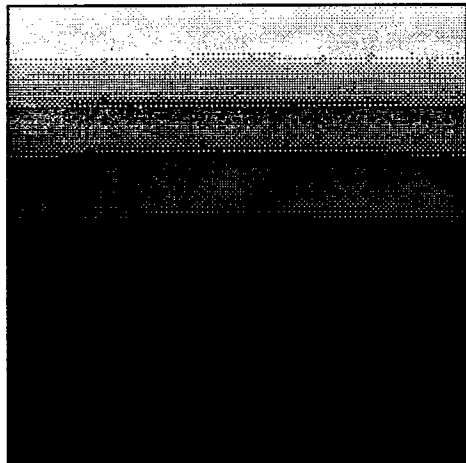
Figure 4. The scene and the PMMW images



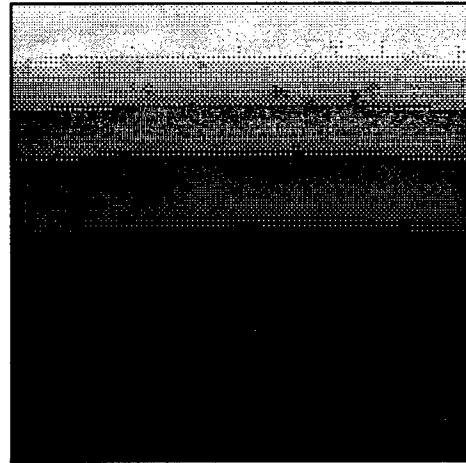
(a) LADAR image (1000m)



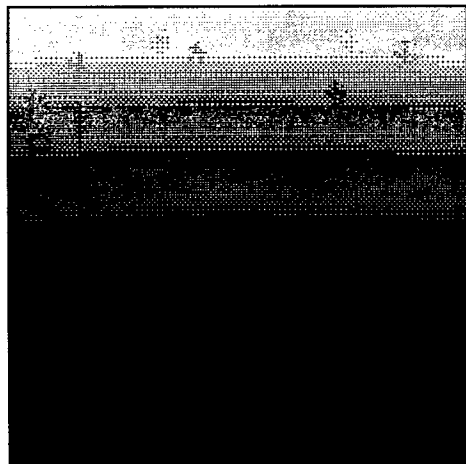
(b) LADAR image (800m)



(c) LADAR image (600m)



(d) LADAR image (400m)



(e) LADAR image (200m)

Figure 5. LADAR images at
 (a) 1000m
 (b) 800m
 (c) 600m
 (d) 400m and
 (e) 200m

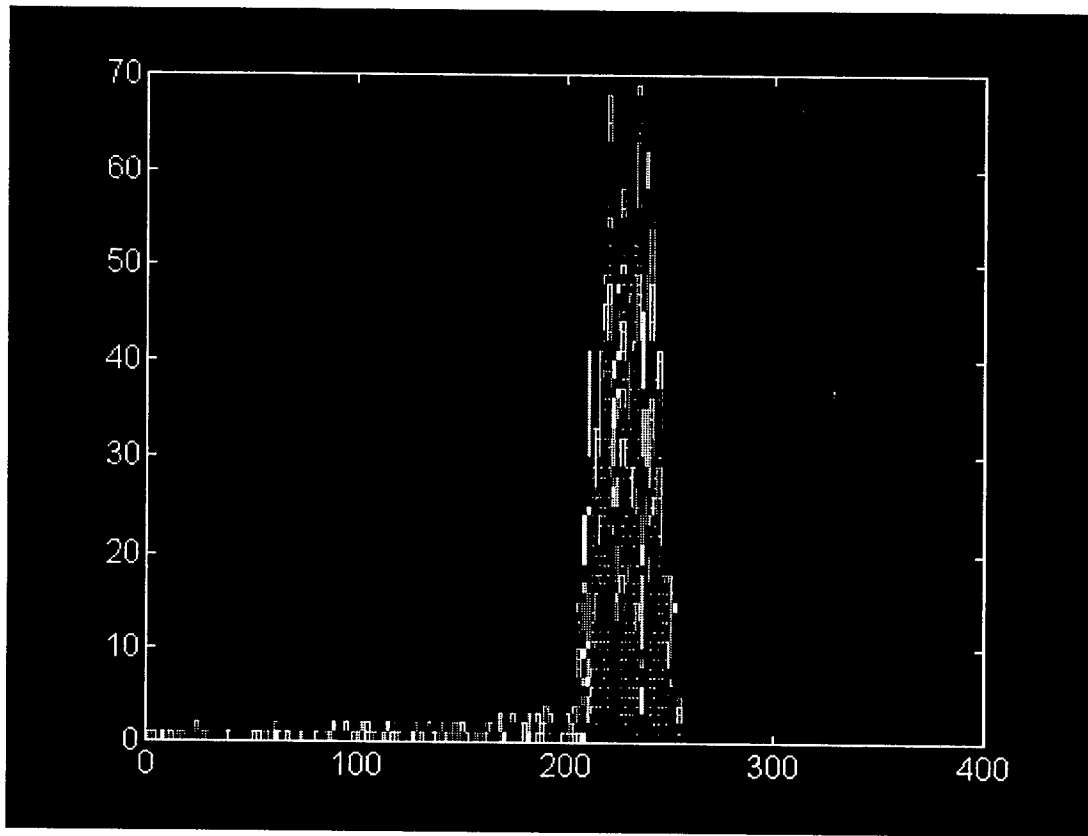
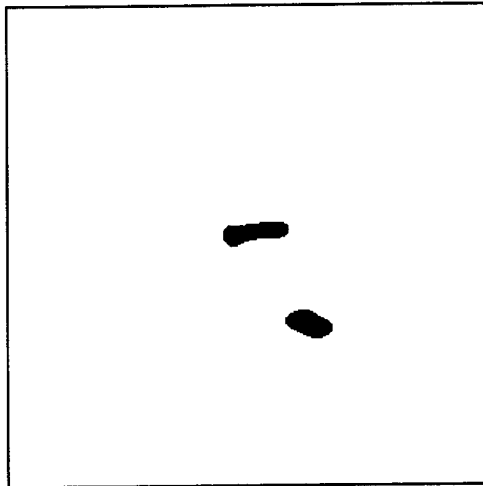
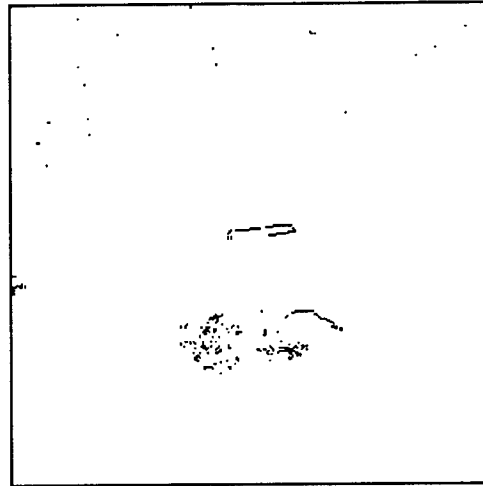


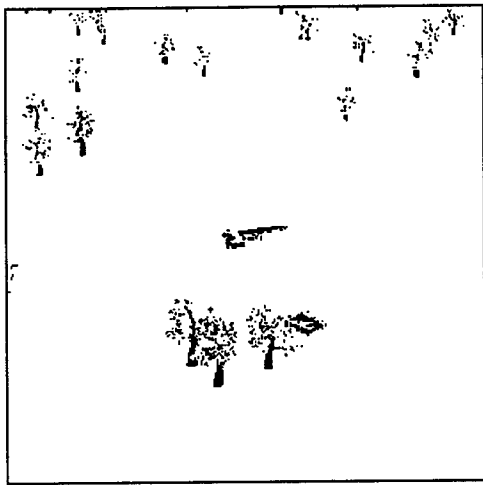
Figure 6. Typical histogram curve for the observed scene



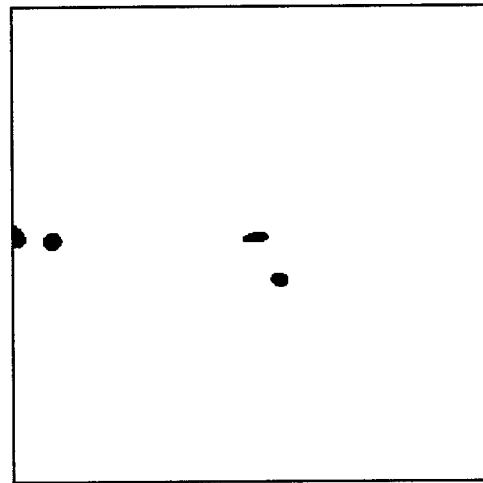
(a1) PMMW after thresholding (200m)



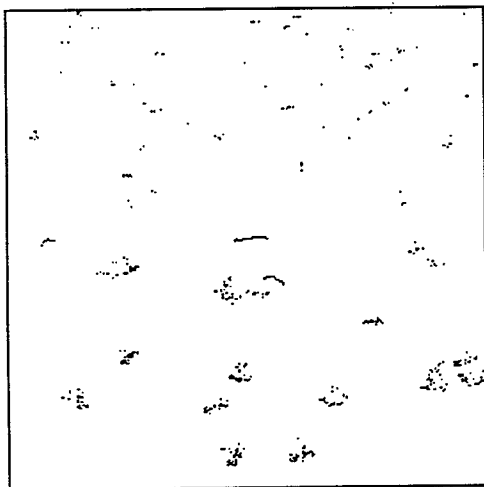
(a2) Jump edges from LADAR (200m)



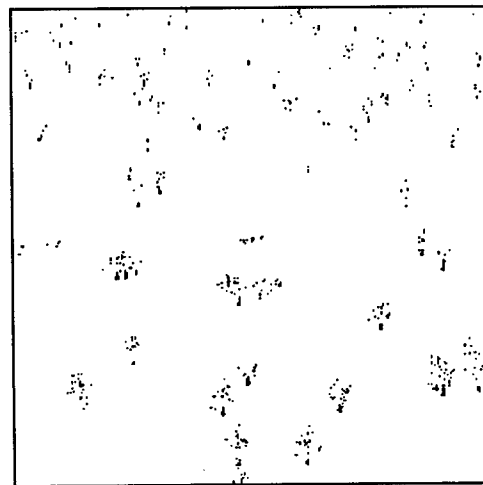
(a3) Vertical surfaces from LADAR (200m)



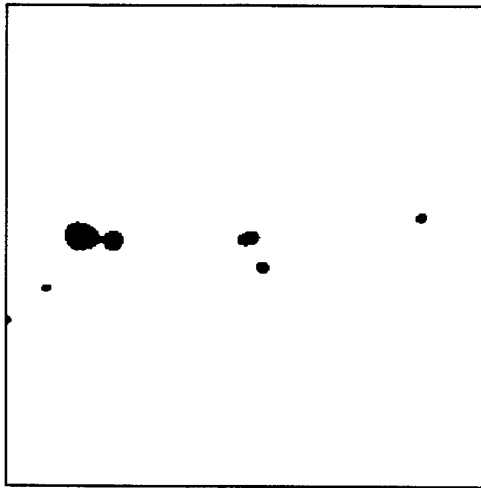
(b1) PMMW after thresholding (400m)



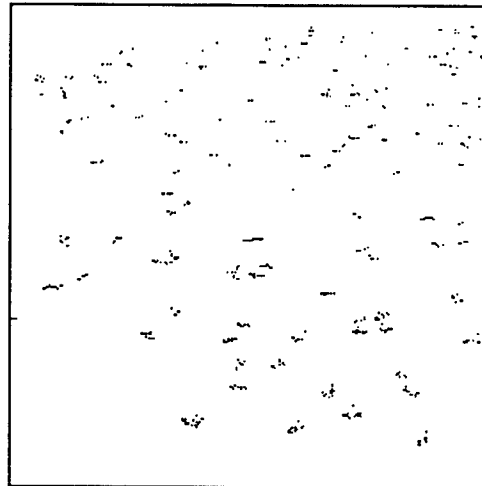
(b2) Jump edges from LADAR (400m)



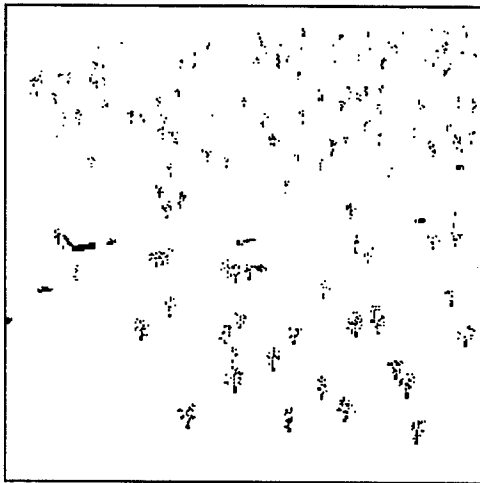
(b3) Vertical surfaces from LADAR (400m)



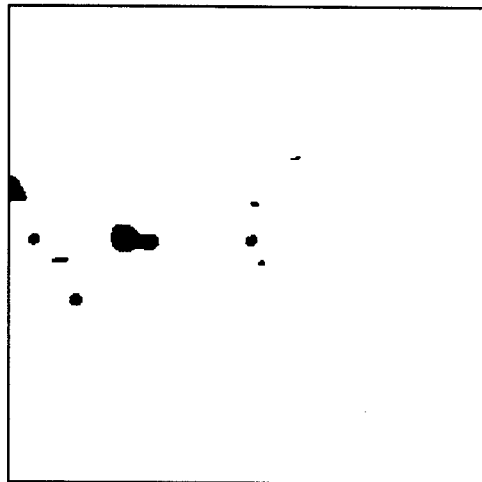
(c1) PMMW after thresholding (600m)



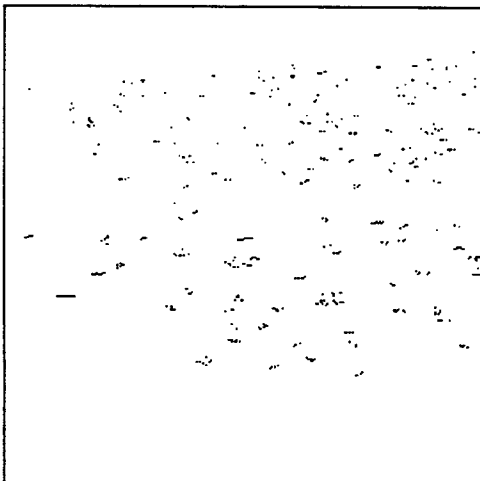
(c2) Jump edges from LADAR (600m)



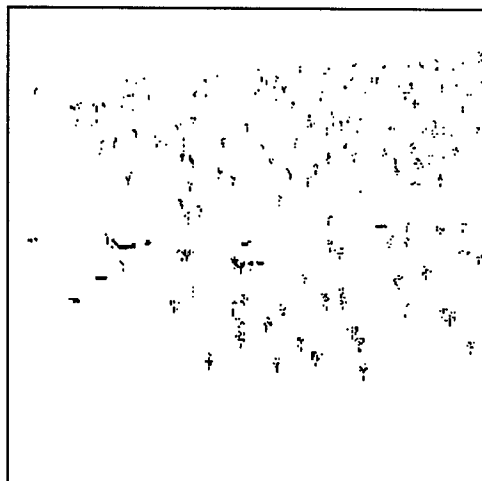
(c3) Vertical surfaces from LADAR (600m)



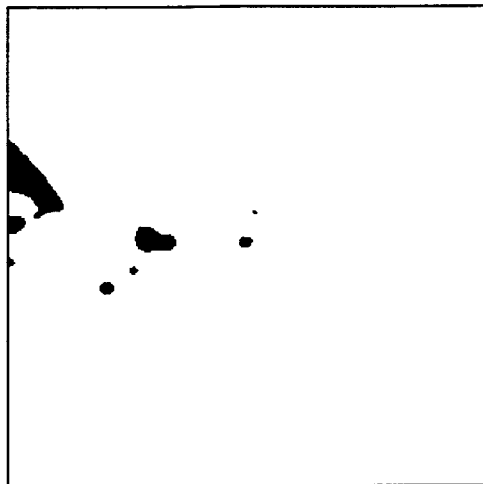
(d1) PMMW after thresholding (800m)



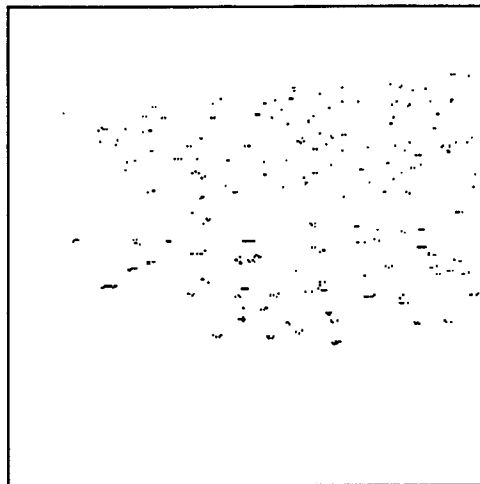
(d2) Jump edges from LADAR (800m)



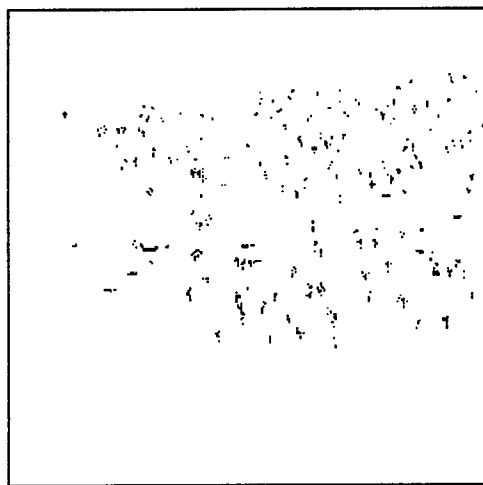
(d3) Vertical surfaces from LADAR (800m)



(e1) PMMW after thresholding (1000m)



(e2) Jump edges from LADAR (1000m)



(e3) Vertical surfaces from LADAR (1000m)

Figure 7. Evidences of targets after image processing (at 200,400,600,800 and 1000m)

(a1)(b1)(c1)(d1)(e1)(e2)	Cold regions from PMMW images
(a2)(b2)(c2)(d2)(e2)	Jump edges from LADAR images
(a3)(b3)(c3)(d3)(e3)	Vertical surfaces from LADAR images

VI. Data Fusion

Fusion is necessary when separate decision modules make decisions. Majority vote, the AND-OR technique, Bayesian inference, neural network approach and the Dempster-Shafer method are some typical ways for data fusion [7]. Among these, the first two give hard decisions and the others provide soft decisions. A soft decision is one that assigns a possibility to each proposition.

In the following, the Dempster-Shafer method [8][9] is introduced. Assume that there are two decision modules i and j , and A , X , and Y are propositions. The possibility mass $m_i(X)$ is assigned for X by i and $m_j(Y)$ is assigned for Y by j . The Dempster-Shafer method will combine the information from i and j and calculate the possibility mass $m_{ij}(A)$ for A as

$$m_{ij}(A) = \frac{\sum_{X \cap Y = A} m_i(X)m_j(Y)}{1 - \sum K_p} = \frac{\sum_{X \cap Y = A} m_i(X)m_j(Y)}{1 - \sum_{X \cap Y = \phi} m_i(X)m_j(Y)} \quad (1)$$

where $\sum K_p$ is the sum for the conflicting ones (i.e., $X \cap Y = \phi$). Some examples for propositions are {tank}, {tank $\cup$ truck}, {building $\cup$ truck}, etc. Intersection of two sets usually gives a smaller set or even an empty set. For instance, {tank} $\cap$ {tank $\cup$ truck} = {tank} and {tank} $\cap$ {building $\cup$ truck} = ϕ .

The method can be used to combine the decisions from individual sensors to determine the overall possibility of an interesting spot being or not being a particular target. For example, since the metal-roof building blob in our experimental PMMW images is large, the PMMW system would assign a small value, say 0.1, to the possibility of it being the desired target, i.e., $m_1(\text{target}) = 0.1$. However, since the area is a metal, the system may not want to exclude the possibility of it being the target. Thus it may assign $m_1(\text{nontarget}) = 0.2$. Consequently, the unassigned value 0.7 (=1-0.1-0.2) will go to $m_1(\Theta)$, where Θ denotes either target or nontarget (i.e., uncertain). On the other hand, the LADAR may find that the object (the building) is too high and has good confidence to assign a big value for the possibility of it being a nontarget. The possibility values may be $m_2(\text{target}) = 0.1$, $m_2(\text{nontarget}) = 0.8$ and $m_2(\Theta) = 0.1$.

The Dempster-Shafer method can then combine the information from these two sensor systems. The results are shown in Table 1.

Table 1. Calculation based on Dempster-Shafer combination

	$m_2(\text{target})=0.1$	$m_2(\text{nontarget})=0.8$	$m_2(\Theta)=0.1$
$m_1(\text{target})=0.1$	$m_{12}(\text{target})=0.01/0.9$	$K1=0.08$	$m_{12}(\text{target})=0.01/0.9$
$m_1(\text{nontarget})=0.2$	$K2=0.02$	$m_{12}(\text{nontarget})=0.16/0.9$	$m_{12}(\text{nontarget})=0.02/0.9$
$m_1(\Theta)=0.7$	$m_{12}(\text{target})=0.07/0.9$	$m_{12}(\text{nontarget})=0.56/0.9$	$m_{12}(\Theta)=0.07/0.9$

All entries in the table are calculated using equation (1). $K1$ and $K2$ are for empty sets (conflicting propositions) and are not valid possibility mass. In this example, the value of $(1-\sum K_p)$ is 0.9. The total possibility mass for each of $m_{12}(\text{target})$, $m_{12}(\text{nontarget})$ and $m_{12}(\Theta)$ is the sum of those corresponding ones in the above table, i.e.,

$$m_{12}(\text{target}) = 0.01/0.9 + 0.07/0.9 + 0.01/0.9 = 0.1$$

$$m_{12}(\text{nontarget}) = 0.16/0.9 + 0.56/0.9 + 0.02/0.9 = 0.74/0.9$$

$$m_{12}(\Theta) = 0.07/0.9$$

Note that the sum of the above three, $m_{12}(\text{target}) + m_{12}(\text{nontarget}) + m_{12}(\Theta)$, remains equal to 1 due to the normalization through the division of $(1-\sum K_p)$. The method to assign m_i 's must be designed. One possible approach is to use fuzzy rules.

The Dempster's rules of combination are both communicative and associative [10]. The order of combination is not important to the final results.

VII. Discussion and Conclusions

7.1 Possible Extension on the Study

Further extension can be in two aspects -- a thorough study on detail target detection, and the study on target selection.

Target Detection

In the short ten-week period, an algorithmic framework was developed and employed to illustrate the conceptual advantages of integrating PMMW and LADAR for target detection. Further study should be directed at the following:

- Possibility value assignment

The Dempster-Shafer method was suggested for data fusion. This method requires the assignment of a possibility value for each proposition. One possible approach for generating the possibility mass at the individual sensor level is to use fuzzy rules. The design of this portion will be critical to the overall system performance.

- Refinement based on current accomplishment

The following are example problems that can be investigated further.

- Extending the detection distance

The detection beyond 1000m becomes harder by using the simple thresholding technique. A more sophisticated technique may be needed to increase the detection distance.

- Optimizing parameters used in detection

One example is the selection of cutoff for detecting “cold” areas in PMMW images.

- Using additional evidences

Only three evidences were used in the study. Questions to be answered include whether more evidences will help and what is the “best” set of evidences, and to what degree does selection of a particular set of evidences affect performance and computational loads.

- Exploring more complicated environments

Testing schemes in more complicated environments should be performed. Through such study, it is possible to better understand the performance of a design and also discover potential problems that must be solved.

Target Selection

This 10-week study focused on target detection. After targets are detected, one of them will be selected for attack. The missile system will need to know more details about targets if it wants to do

selection. For instance, the missile may prefer to attack a TEL or tank than some other type of vehicle. Since LADAR has high resolution, it is more adequate for this type of decision making. Additionally, the LADAR can provide a range image as well as a reflectance image. Fusion using these two types of images is a research problem deserving further investigation. The fusion could provide some advantages yet unknown. At shorter distance, the PMMW may be able to acquire some degree of size information. Such information could become useful in case LADAR fails due to weather conditions or other causes.

7.2 Conclusions

In this study, fusion of PMMW and LADAR data was performed. The focus was on target detection. PMMW is first used to detect metal objects and then LADAR checks those regions of interest for other evidences of existence of a target. Advantages achieved by integrating these two sensors include the reduction of task complexity and the improvement of reliability, both due primarily to the efficient localization of regions of interests employing the PMMW. Since PMMW can penetrate fog, cloud, smoke, etc., the combined system has a near all-weather capability. The framework of the fusion is based on the Dempster-Shafer decision method. The fusion may be done at the algorithm level and at the sensor level. Determination of possibility mass for each proposition for each sensor is a required work, and represents a major topic for suggested study extension. Other issues on further extension were addressed in the previous subsection.

Acknowledgment

I would like to express my great appreciation to organizations and people who sponsored and helped this study. The AFOSR Summer Faculty Research Program sponsored this research. The Advanced Guidance Division of the Air Force Research Lab Munitions Directorate provided local support in the Eglin Air Force Base. Mr. Sengvieng Amphay sponsored my visit, arranged research resources, and provided research documentation and valuable suggestions. Mr. Bryce Sundstrom also provided research documentation, helps, and valuable discussion. Dr. Mike Richards of Nichols Research

helped generate simulation images. Dr. Dennis Goldstein managed the program in the Eglin Air Force Base, helped process necessary paperwork and arranged housing.

References

1. R. M. Smith, K. D. Trott and B. M. Sundstrom, "The Passive MM-Wave Scenario," *Microwave Journal*, March 1996.
2. R. Smith, S. W. Worrell and R. M. Knox, "Passive Millimeter-wave Imaging (PMMWI)," March, 1994, *Proceedings from Infrared Information Symposium (IRIS)*, July 1995.
3. J. E. Bevington, *Laser Radar ATR Algorithms: Phase III final report*, Tech. Rep., Alliant Techsystems, Inc., May, 1992.
4. M. R. Stevens and J. R. Beveridge, "Precise Matching of 3-D Target Models to Multisensor Data," *IEEE Transactions on Image Processing*, Vol. 6, No. 1, January 1997, pp126-142.
5. D. Casasent and A. Ye, "Detection Filters and Algorithm Fusion for ATR," *IEEE Transactions on Image Processing*, Vol. 6, No. 1, January 1997, pp114-125
6. B. Javidi, G. Zhang, F. Parchekani, and P. Refregier, "Performance of Minimum-mean-square-error Filter for Spatially Nonoverlapping Target and Input-scene Noise," *Applied Optics*, Vol. 33, No. 35, December 10, 1994, pp. 8197-8239
7. D. Hall, *Mathematical Techniques in Multisensor Data Fusion*, 1992. Artech House, Inc, Norwood, MA..
8. T. D. Garvey, J. D. Lowrance and M. A. Fischler, "An Inference Technique for Integrating Knowledge from Disparate Sources, Proc. 7th Int. Joint Conf. Artificial Intell., pp. 319-325, Vancouver. 1981.
9. M. A. Abidi and R. C. Gonzalez (editors), *Data Fusion in Robotics and Machine Intelligence*, 1992. Academic Press, Inc., San Diego, CA .
10. J. D. Lowrance and T. D. Garvey, "Evidence Reasoning: A developing concept," *Proceedings of the IEEE International Conference on Cybernetics and Society*, Oct. 1982, pp. 6-9.

BOUNDARY CONDITIONS FOR DIRECT NUMERICAL SIMULATION OF TURBULENT FLOW

Chaoqun Liu

Professor

College of Engineering and Science

Louisiana Tech University

P.O. Box 10348, Ruston, LA 71272

Final Report for:
Summer Faculty Research Program
Wright-Patterson Research Site

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC
and
Wright-Patterson Research Site

August 1998

Boundary Conditions for Direct Numerical Simulation of Turbulent Flow

Chaoqun Liu
Professor
College of Engineering and Science
Louisiana Tech University

Abstract

The computational results of FDL3D flow solver using the high-order compact scheme and high-order filter for some fundamental flows are reported for code verification. The boundary conditions in a curvilinear coordinates for inflow, outflow, far-field, and solid-wall surface is derived and discussed with some computational results.

1 Introduction

Most important flows, such as transition, turbulence, vortex dynamics, etc. which Air Force is interested in, are involved in a time-dependent process that requires a time-accurate numerical prediction, such as direct numerical simulation (DNS) or large eddy simulation (LES).

As well-understood now, DNS and LES are required to have minimum errors in wave magnitude and phase. However, low order up-winding, hybrid, or other schemes can introduce serious dissipation and dispersion even for a single wave and are therefore not encouraged to use for DNS or LES. Although second-order LES calculation is still popular in literature nowadays, but is fundamentally incorrect.

As pointed by Hirsch (1997), a historic mistake was made (Fromm, 1964) and was quoted by many text-books (Roache, 1972) that stated central differencing is not stable for convection-diffusion equations unless the Peclet number is less than 2. This wrong statement prevented CFD community from using high-order central differencing for decades. In fact, the Peclet number has nothing to do with the stability and many central differencing schemes are stable when the Peclet number is greater than 2, but may get some non-physical oscillating which is caused by some unresolved high-frequency waves. However, these high-frequency waves can be restricted or filtered by using proper filters. The difference between artificial damping and filter is that the former adds some artificial viscosity to the Navier-Stokes equations or changes the governing Navier-Stokes equations, but the later only change the solution by filtering some unresolved waves which is only related to the solution and has no changes to the Navier-Stokes equations. High-order central differencing scheme with filter is much more reasonable approach than up-winding or central differencing tied with artificial damping approaches.

The high-order compact central differencing approach has appeared for years (Strikwerda, 1989), but Lele (1992) first systematically used it with filter for CFD. Recently Visbal and Gaitonde (1997, 1998) gave general approach to develop both high-order compact scheme and high-order filter (up to 10th order). If this approach is verified and validated, it could be a new direction to conduct CFD calculation for time-dependent flow and may replace many old approaches such as up-winding, TVD, central with damping, hybrid, etc. The verification and validation is one of the purposes of my work for the summer.

Actually, my group used the high-order scheme with special multi-dimensional filter (we call them fine-coarse-fine grid mapping) to successfully simulate the whole process of flow transition on flat plate 5 years ago (Liu et al, 1993, 1995) and then for flow transition around airfoils with all speeds recently (Liu et al, 1997, 1998). After 5 years, it is still hard to find anyone else who can repeat our work.

DNS/LES for complex geometry was criticized, but now an exponentially increasing number of researchers are working on this direction. The fundamental problem with DNS/LES

seems to be cleared and our experience shows DNS can be used for flow transition and LES can be used for fully developed turbulent flow for a spectrum of engineering applications.

The other purpose of this work is to develop appropriate boundary conditions for time-dependent Navier-Stokes equations which is compatible and can maintain high-order accuracy for the solution. The characteristic boundary conditions for Euler system can be found in a number of literatures (Whitfield, 1983; Thompson, 1987, 1990). For Navier-Stokes system, Poinot and Lele (1992) have tested some approaches similar to characteristic boundary conditions, but for Cartesian coordinates only. For time-dependent flow with perturbation, many authors suggested to use sponge boundary condition (Fasel, 1990; Kloker et al, 1993; Collis et al, 1996). We should develop characteristic-like boundary conditions for curvilinear coordinates with sponge boundary.

2 Boundary conditions with compatibility

The characteristic boundary conditions for Euler system cannot be directly used for Navier-Stokes equations. Actually, there is no characteristic relation in Navier-Stokes system. However, in some certain circumstances, such as free stream, far field, etc. where the viscosity is less important, the characteristic relation can still be used. Poinot and Lele (1992) gave some ideas which use the original N-S equation but replace some convection terms by characteristic relation. However, they only gave the formula for Cartesian coordinates. Here, we derive these relations in curvilinear coordinates.

2.1 Governing Equations

The three-dimensional Navier-Stokes equations in generalized curvilinear coordinates (ξ, η, ζ) can be written as

$$\frac{1}{J} \frac{\partial Q}{\partial t} + \frac{\partial(E - E_v)}{\partial \xi} + \frac{\partial(F - F_v)}{\partial \eta} + \frac{\partial(G - G_v)}{\partial \zeta} = 0 \quad (1)$$

The flux vectors for compressible flow are

$$Q = \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ \rho w \\ e \end{pmatrix}$$

$$E = \frac{1}{J} \begin{pmatrix} \rho U \\ \rho U u + p \xi_x \\ \rho U v + p \xi_y \\ \rho U w + p \xi_z \\ U(e + p) \end{pmatrix}$$

$$\begin{aligned}
F &= \frac{1}{J} \begin{pmatrix} \rho V \\ \rho V u + p \eta_x \\ \rho V v + p \eta_y \\ \rho V w + p \eta_z \\ V(e + p) \end{pmatrix} \\
G &= \frac{1}{J} \begin{pmatrix} \rho W \\ \rho W u + p \zeta_x \\ \rho W v + p \zeta_y \\ \rho W w + p \zeta_z \\ W(e + p) \end{pmatrix} \\
E_v &= \frac{1}{J} \begin{pmatrix} 0 \\ \tau_{xx} \xi_x + \tau_{yx} \xi_y + \tau_{zx} \xi_z \\ \tau_{xy} \xi_x + \tau_{yy} \xi_y + \tau_{zy} \xi_z \\ \tau_{xz} \xi_x + \tau_{yz} \xi_y + \tau_{zz} \xi_z \\ q_x \xi_x + q_y \xi_y + q_z \xi_z \end{pmatrix} \\
F_v &= \frac{1}{J} \begin{pmatrix} 0 \\ \tau_{xx} \eta_x + \tau_{yx} \eta_y + \tau_{zx} \eta_z \\ \tau_{xy} \eta_x + \tau_{yy} \eta_y + \tau_{zy} \eta_z \\ \tau_{xz} \eta_x + \tau_{yz} \eta_y + \tau_{zz} \eta_z \\ q_x \eta_x + q_y \eta_y + q_z \eta_z \end{pmatrix} \\
G_v &= \frac{1}{J} \begin{pmatrix} 0 \\ \tau_{xx} \zeta_x + \tau_{yx} \zeta_y + \tau_{zx} \zeta_z \\ \tau_{xy} \zeta_x + \tau_{yy} \zeta_y + \tau_{zy} \zeta_z \\ \tau_{xz} \zeta_x + \tau_{yz} \zeta_y + \tau_{zz} \zeta_z \\ q_x \zeta_x + q_y \zeta_y + q_z \zeta_z \end{pmatrix}
\end{aligned}$$

where

$$\begin{aligned}
U &= u \xi_x + v \xi_y + w \xi_z \\
V &= u \eta_x + v \eta_y + w \eta_z \\
W &= u \zeta_x + v \zeta_y + w \zeta_z
\end{aligned} \tag{2}$$

J is the coordinate transformation Jacobian matrix:

$$J = \frac{\partial(\xi, \eta, \zeta)}{\partial(x, y, z)} \tag{3}$$

$$e = \frac{p}{\gamma - 1} + \frac{1}{2} \rho (u^2 + v^2 + w^2) \tag{4}$$

$$\begin{aligned}
\tau_{xx} &= \frac{2\mu}{3R_e} [2(\xi_x \frac{\partial u}{\partial \xi} + \eta_x \frac{\partial u}{\partial \eta} + \zeta_x \frac{\partial u}{\partial \zeta}) - (\xi_y \frac{\partial v}{\partial \xi} + \eta_y \frac{\partial v}{\partial \eta} + \zeta_y \frac{\partial v}{\partial \zeta}) - (\xi_z \frac{\partial w}{\partial \xi} + \eta_z \frac{\partial w}{\partial \eta} + \zeta_z \frac{\partial w}{\partial \zeta})] \\
\tau_{yy} &= \frac{2\mu}{3R_e} [2(\xi_y \frac{\partial v}{\partial \xi} + \eta_y \frac{\partial v}{\partial \eta} + \zeta_y \frac{\partial v}{\partial \zeta}) - (\xi_x \frac{\partial u}{\partial \xi} + \eta_x \frac{\partial u}{\partial \eta} + \zeta_x \frac{\partial u}{\partial \zeta}) - (\xi_z \frac{\partial w}{\partial \xi} + \eta_z \frac{\partial w}{\partial \eta} + \zeta_z \frac{\partial w}{\partial \zeta})] \\
\tau_{zz} &= \frac{2\mu}{3R_e} [2(\xi_z \frac{\partial w}{\partial \xi} + \eta_z \frac{\partial w}{\partial \eta} + \zeta_z \frac{\partial w}{\partial \zeta}) - (\xi_y \frac{\partial v}{\partial \xi} + \eta_y \frac{\partial v}{\partial \eta} + \zeta_y \frac{\partial v}{\partial \zeta}) - (\xi_x \frac{\partial u}{\partial \xi} + \eta_x \frac{\partial u}{\partial \eta} + \zeta_x \frac{\partial u}{\partial \zeta})]
\end{aligned}$$

$$\begin{aligned}
\tau_{xy} &= \frac{\mu}{Re} [(\xi_y \frac{\partial u}{\partial \xi} + \eta_y \frac{\partial u}{\partial \eta} + \zeta_y \frac{\partial u}{\partial \zeta}) + (\xi_x \frac{\partial v}{\partial \xi} + \eta_x \frac{\partial v}{\partial \eta} + \zeta_x \frac{\partial v}{\partial \zeta})] \\
\tau_{xz} &= \frac{\mu}{Re} [(\xi_z \frac{\partial u}{\partial \xi} + \eta_z \frac{\partial u}{\partial \eta} + \zeta_z \frac{\partial u}{\partial \zeta}) + (\xi_x \frac{\partial w}{\partial \xi} + \eta_x \frac{\partial w}{\partial \eta} + \zeta_x \frac{\partial w}{\partial \zeta})] \\
\tau_{yz} &= \frac{\mu}{Re} [(\xi_y \frac{\partial w}{\partial \xi} + \eta_y \frac{\partial w}{\partial \eta} + \zeta_y \frac{\partial w}{\partial \zeta}) + (\xi_z \frac{\partial v}{\partial \xi} + \eta_z \frac{\partial v}{\partial \eta} + \zeta_z \frac{\partial v}{\partial \zeta})]
\end{aligned} \tag{5}$$

$$\begin{aligned}
q_x &= u\tau_{xx} + v\tau_{xy} + w\tau_{xz} + \frac{\mu}{(\gamma - 1)M_r^2 Re Pr} (\xi_x \frac{\partial T}{\partial \xi} + \eta_x \frac{\partial T}{\partial \eta} + \zeta_x \frac{\partial T}{\partial \zeta}) \\
q_y &= u\tau_{xy} + v\tau_{yy} + w\tau_{yz} + \frac{\mu}{(\gamma - 1)M_r^2 Re Pr} (\xi_y \frac{\partial T}{\partial \xi} + \eta_y \frac{\partial T}{\partial \eta} + \zeta_y \frac{\partial T}{\partial \zeta}) \\
q_z &= u\tau_{xz} + v\tau_{yz} + w\tau_{zz} + \frac{\mu}{(\gamma - 1)M_r^2 Re Pr} (\xi_z \frac{\partial T}{\partial \xi} + \eta_z \frac{\partial T}{\partial \eta} + \zeta_z \frac{\partial T}{\partial \zeta})
\end{aligned} \tag{6}$$

All the equations are cast in non-dimensional forms, and the four dimensionless parameters resulted from non-dimensionalization are defined as:

$$M_r = \frac{U_r^*}{\sqrt{\gamma R T_r^*}} \tag{7}$$

$$Re = \frac{\rho_r^* U_r^* L^*}{\mu_r^*} \tag{8}$$

$$Pr = \frac{C_p \mu_r^*}{\kappa_r^*} \tag{9}$$

$$\gamma = \frac{C_p}{C_v} \tag{10}$$

The sup-star '*' represents reference values. R , C_p and C_v are the ideal gas constant, the specific heats at constant pressure and constant volume. Sutherland's law in dimensionless form is,

$$\mu = T^{\frac{3}{2}} \frac{1 + S}{T + S}, \quad S = \frac{110.3}{T_r^*} \tag{11}$$

2.2 Compatible boundary conditions

Using the characteristic analysis [15], in which hyperbolic terms in ξ direction of Eq. (1) are modified, we can recast this system as:

$$\begin{aligned}
\frac{\partial \rho}{\partial t} + d_1 + V \frac{\partial \rho}{\partial \eta} + \rho(\eta_x \frac{\partial u}{\partial \eta} + \eta_x \frac{\partial u}{\partial \eta} + \eta_x \frac{\partial u}{\partial \eta}) + W \frac{\partial \rho}{\partial \zeta} + \rho(\zeta_x \frac{\partial u}{\partial \zeta} + \zeta_x \frac{\partial u}{\partial \zeta} + \zeta_x \frac{\partial u}{\partial \zeta}) + vis_1 &= 0 \\
\frac{\partial u}{\partial t} + d_2 + V \frac{\partial u}{\partial \eta} + \frac{1}{\rho} \eta_x \frac{\partial p}{\partial \eta} + W \frac{\partial u}{\partial \zeta} + \frac{1}{\rho} \zeta_x \frac{\partial p}{\partial \zeta} + vis_2 &= 0 \\
\frac{\partial v}{\partial t} + d_3 + V \frac{\partial v}{\partial \eta} + \frac{1}{\rho} \eta_y \frac{\partial p}{\partial \eta} + W \frac{\partial v}{\partial \zeta} + \frac{1}{\rho} \zeta_y \frac{\partial p}{\partial \zeta} + vis_3 &= 0 \\
\frac{\partial w}{\partial t} + d_4 + V \frac{\partial w}{\partial \eta} + \frac{1}{\rho} \eta_z \frac{\partial p}{\partial \eta} + W \frac{\partial w}{\partial \zeta} + \frac{1}{\rho} \zeta_z \frac{\partial p}{\partial \zeta} + vis_4 &= 0 \\
\frac{\partial p}{\partial t} + d_5 + V \frac{\partial p}{\partial \eta} + \gamma p(\eta_x \frac{\partial u}{\partial \eta} + \eta_x \frac{\partial u}{\partial \eta} + \eta_x \frac{\partial u}{\partial \eta}) + W \frac{\partial p}{\partial \zeta} + \gamma p(\zeta_x \frac{\partial u}{\partial \zeta} + \zeta_x \frac{\partial u}{\partial \zeta} + \zeta_x \frac{\partial u}{\partial \zeta}) + vis_5 &= 0
\end{aligned} \tag{12}$$

where vector d is obtained from characteristic analysis, and can be expressed as

$$\begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \end{pmatrix} = \begin{pmatrix} \frac{1}{c^2} [\frac{1}{2}(\mathcal{L}_1 + \mathcal{L}_5) + \mathcal{L}_2] \\ \frac{\xi_x}{2\beta\rho c}(\mathcal{L}_5 - \mathcal{L}_1) - \frac{1}{\beta^2}(\xi_y\mathcal{L}_3 + \xi_z\mathcal{L}_4) \\ \frac{\xi_y}{2\beta\rho c}(\mathcal{L}_5 - \mathcal{L}_1) + \frac{1}{\beta^2\xi_x}[(\xi_x^2 + \xi_z^2)\mathcal{L}_3 - \xi_z\xi_y\mathcal{L}_4] \\ \frac{\xi_z}{2\beta\rho c}(\mathcal{L}_5 - \mathcal{L}_1) - \frac{1}{\beta^2\xi_x}[\xi_y\xi_z\mathcal{L}_3 - (\xi_x^2 + \xi_y^2)\mathcal{L}_4] \\ \frac{1}{2}(\mathcal{L}_1 + \mathcal{L}_5) \end{pmatrix} \quad (13)$$

where c is sound speed, and

$$\beta = \sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2} \quad (14)$$

$\mathcal{L}'_i$ s are the amplitudes of characteristic waves corresponding to each characteristic velocity λ_i , which are

$$\lambda_1 = U - C_\xi \quad (15)$$

$$\lambda_2 = \lambda_3 = \lambda_4 = U \quad (16)$$

$$\lambda_5 = U + C_\xi \quad (17)$$

C_ξ is defined as Eq. (??) and $\mathcal{L}'_i$ s can be expressed as

$$\begin{aligned} \mathcal{L}_1 &= (U - C_\xi) \left[-\frac{\rho c}{\beta} \left(\xi_x \frac{\partial u}{\partial \xi} + \xi_y \frac{\partial v}{\partial \xi} + \xi_z \frac{\partial w}{\partial \xi} \right) + \frac{\partial p}{\partial \xi} \right] \\ \mathcal{L}_2 &= U \left(c^2 \frac{\partial \rho}{\partial \xi} - \frac{\partial p}{\partial \xi} \right) \\ \mathcal{L}_3 &= U \left(-\xi_y \frac{\partial u}{\partial \xi} + \xi_x \frac{\partial v}{\partial \xi} \right) \\ \mathcal{L}_4 &= U \left(-\xi_z \frac{\partial u}{\partial \xi} + \xi_x \frac{\partial w}{\partial \xi} \right) \\ \mathcal{L}_5 &= (U + C_\xi) \left[\frac{\rho c}{\beta} \left(\xi_x \frac{\partial u}{\partial \xi} + \xi_y \frac{\partial v}{\partial \xi} + \xi_z \frac{\partial w}{\partial \xi} \right) + \frac{\partial p}{\partial \xi} \right] \end{aligned} \quad (18)$$

The terms vis'_i s in Eq. (12) represent viscous terms, they are given by transformation of initial viscous terms in Eq.(1)

$$\begin{pmatrix} vis_1 \\ vis_2 \\ vis_3 \\ vis_4 \\ vis_5 \end{pmatrix} = -J\hat{P} \left(\frac{\partial E_v}{\partial \xi} + \frac{\partial F_v}{\partial \eta} + \frac{\partial G_v}{\partial \zeta} \right) \quad (19)$$

where $\hat{P}$ is matrix as

$$\hat{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -\frac{u}{\rho} & \frac{1}{\rho} & 0 & 0 & 0 \\ -\frac{v}{\rho} & 0 & \frac{1}{\rho} & 0 & 0 \\ -\frac{w}{\rho} & 0 & 0 & \frac{1}{\rho} & 0 \\ \frac{\gamma-1}{2}(u^2 + v^2 + w^2) & (1-\gamma)u & (1-\gamma)v & (1-\gamma)w & \gamma-1 \end{pmatrix} \quad (20)$$

In other two directions, analysis should be the same. In ζ direction, the modified N-S equation should be

$$\begin{aligned}
\frac{\partial \rho}{\partial t} + d_1 + V \frac{\partial \rho}{\partial \eta} + \rho(\eta_x \frac{\partial u}{\partial \eta} + \eta_x \frac{\partial u}{\partial \eta} + \eta_x \frac{\partial u}{\partial \eta}) + U \frac{\partial \rho}{\partial \xi} + \rho(\xi_x \frac{\partial u}{\partial \xi} + \xi_x \frac{\partial u}{\partial \xi} + \xi_x \frac{\partial u}{\partial \xi}) + vis_1 &= 0 \\
\frac{\partial u}{\partial t} + d_2 + V \frac{\partial u}{\partial \eta} + \frac{1}{\rho} \eta_x \frac{\partial p}{\partial \eta} + U \frac{\partial u}{\partial \xi} + \frac{1}{\rho} \xi_x \frac{\partial p}{\partial \xi} + vis_2 &= 0 \\
\frac{\partial v}{\partial t} + d_3 + V \frac{\partial v}{\partial \eta} + \frac{1}{\rho} \eta_y \frac{\partial p}{\partial \eta} + U \frac{\partial v}{\partial \xi} + \frac{1}{\rho} \xi_y \frac{\partial p}{\partial \xi} + vis_3 &= 0 \quad (21) \\
\frac{\partial w}{\partial t} + d_4 + V \frac{\partial w}{\partial \eta} + \frac{1}{\rho} \eta_z \frac{\partial p}{\partial \eta} + U \frac{\partial w}{\partial \xi} + \frac{1}{\rho} \xi_z \frac{\partial p}{\partial \xi} + vis_4 &= 0 \\
\frac{\partial p}{\partial t} + d_5 + V \frac{\partial p}{\partial \eta} + \gamma p(\eta_x \frac{\partial u}{\partial \eta} + \eta_x \frac{\partial u}{\partial \eta} + \eta_x \frac{\partial u}{\partial \eta}) + U \frac{\partial p}{\partial \xi} + \gamma p(\xi_x \frac{\partial u}{\partial \xi} + \xi_x \frac{\partial u}{\partial \xi} + \xi_x \frac{\partial u}{\partial \xi}) + vis_5 &= 0
\end{aligned}$$

$$\begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \end{pmatrix} = \begin{pmatrix} \frac{1}{c^2} [\frac{1}{2}(\mathcal{L}_1 + \mathcal{L}_5) + \mathcal{L}_2] \\ \frac{\zeta_x}{2\beta\rho c}(\mathcal{L}_5 - \mathcal{L}_1) + \frac{1}{\beta^2 \zeta_x} [(\zeta_y^2 + \zeta_z^2)\mathcal{L}_3 - \zeta_x \zeta_y \mathcal{L}_4] \\ \frac{\zeta_y}{2\beta\rho c}(\mathcal{L}_5 - \mathcal{L}_1) - \frac{1}{\beta^2 \zeta_y} [(\zeta_x \zeta_y \mathcal{L}_3 - (\zeta_x^2 + \zeta_z^2)\mathcal{L}_4] \\ \frac{\zeta_z}{2\beta\rho c}(\mathcal{L}_5 - \mathcal{L}_1) - \frac{1}{\beta^2 \zeta_z} [\zeta_x \mathcal{L}_3 + \zeta_y \mathcal{L}_4] \\ \frac{1}{2}(\mathcal{L}_1 + \mathcal{L}_5) \end{pmatrix} \quad (22)$$

$$\begin{aligned}
\mathcal{L}_1 &= (W - C_\zeta) \left[-\frac{\rho c}{\beta} (\zeta_x \frac{\partial u}{\partial \zeta} + \zeta_y \frac{\partial v}{\partial \zeta} + \zeta_z \frac{\partial w}{\partial \zeta}) + \frac{\partial p}{\partial \zeta} \right] \\
\mathcal{L}_2 &= W(c^2 \frac{\partial \rho}{\partial \zeta} - \frac{\partial p}{\partial \zeta}) \\
\mathcal{L}_3 &= W(-\zeta_x \frac{\partial w}{\partial \zeta} + \zeta_z \frac{\partial u}{\partial \zeta}) \\
\mathcal{L}_4 &= W(-\zeta_y \frac{\partial w}{\partial \zeta} + \zeta_z \frac{\partial v}{\partial \zeta}) \\
\mathcal{L}_5 &= (W + C_\zeta) \left[\frac{\rho c}{\beta} (\zeta_x \frac{\partial u}{\partial \zeta} + \zeta_y \frac{\partial v}{\partial \zeta} + \zeta_z \frac{\partial w}{\partial \zeta}) + \frac{\partial p}{\partial \zeta} \right]
\end{aligned} \quad (23)$$

$$\beta = \sqrt{\zeta_x^2 + \zeta_y^2 + \zeta_z^2} \quad (24)$$

$$C_\zeta = c\beta \quad (25)$$

In η direction, modified N-S equation should be

$$\begin{aligned}
\frac{\partial \rho}{\partial t} + d_1 + W \frac{\partial \rho}{\partial \zeta} + \rho(\zeta_x \frac{\partial u}{\partial \zeta} + \zeta_x \frac{\partial u}{\partial \zeta} + \zeta_x \frac{\partial u}{\partial \zeta}) + U \frac{\partial \rho}{\partial \xi} + \rho(\xi_x \frac{\partial u}{\partial \xi} + \xi_x \frac{\partial u}{\partial \xi} + \xi_x \frac{\partial u}{\partial \xi}) + vis_1 &= 0 \\
\frac{\partial u}{\partial t} + d_2 + W \frac{\partial u}{\partial \zeta} + \frac{1}{\rho} \zeta_x \frac{\partial p}{\partial \zeta} + U \frac{\partial u}{\partial \xi} + \frac{1}{\rho} \xi_x \frac{\partial p}{\partial \xi} + vis_2 &= 0 \\
\frac{\partial v}{\partial t} + d_3 + W \frac{\partial v}{\partial \zeta} + \frac{1}{\rho} \zeta_y \frac{\partial p}{\partial \zeta} + U \frac{\partial v}{\partial \xi} + \frac{1}{\rho} \xi_y \frac{\partial p}{\partial \xi} + vis_3 &= 0 \quad (26)
\end{aligned}$$

$$\begin{aligned}\frac{\partial w}{\partial t} + d_4 + W \frac{\partial w}{\partial \zeta} + \frac{1}{\rho} \zeta_z \frac{\partial p}{\partial \zeta} + U \frac{\partial w}{\partial \xi} + \frac{1}{\rho} \xi_z \frac{\partial p}{\partial \xi} + vis_4 &= 0 \\ \frac{\partial p}{\partial t} + d_5 + W \frac{\partial p}{\partial \zeta} + \gamma p (\zeta_x \frac{\partial u}{\partial \zeta} + \zeta_x \frac{\partial u}{\partial \zeta} + \zeta_x \frac{\partial u}{\partial \zeta}) + U \frac{\partial p}{\partial \xi} + \gamma p (\xi_x \frac{\partial u}{\partial \xi} + \xi_x \frac{\partial u}{\partial \xi} + \xi_x \frac{\partial u}{\partial \xi}) + vis_5 &= 0\end{aligned}$$

$$\begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \end{pmatrix} = \begin{pmatrix} \frac{1}{c^2} [\frac{1}{2}(\mathcal{L}_1 + \mathcal{L}_5) + \mathcal{L}_2] \\ \frac{\eta_x}{2\beta\rho c}(\mathcal{L}_5 - \mathcal{L}_1) + \frac{1}{\beta^2\eta_y}[(\eta_y^2 + \eta_z^2)\mathcal{L}_3 - \eta_x\eta_z\mathcal{L}_4] \\ \frac{\eta_y}{2\beta\rho c}(\mathcal{L}_5 - \mathcal{L}_1) - \frac{1}{\beta^2}[\eta_x\mathcal{L}_3 + \eta_z\mathcal{L}_4] \\ \frac{\eta_z}{2\beta\rho c}(\mathcal{L}_5 - \mathcal{L}_1) - \frac{1}{\beta^2\eta_y}[\eta_x\eta_z\mathcal{L}_3 - (\eta_x^2 + \eta_y^2)\mathcal{L}_4] \\ \frac{1}{2}(\mathcal{L}_1 + \mathcal{L}_5) \end{pmatrix} \quad (27)$$

$$\begin{aligned}\mathcal{L}_1 &= (V - C_\eta) \left[-\frac{\rho c}{\beta} \left(\eta_x \frac{\partial u}{\partial \eta} + \eta_y \frac{\partial v}{\partial \eta} + \eta_z \frac{\partial w}{\partial \eta} \right) + \frac{\partial p}{\partial \eta} \right] \\ \mathcal{L}_2 &= V \left(c^2 \frac{\partial \rho}{\partial \eta} - \frac{\partial p}{\partial \eta} \right) \\ \mathcal{L}_3 &= V \left(-\eta_x \frac{\partial v}{\partial \eta} + \eta_y \frac{\partial u}{\partial \eta} \right) \\ \mathcal{L}_4 &= V \left(-\eta_z \frac{\partial v}{\partial \eta} + \eta_y \frac{\partial w}{\partial \eta} \right) \\ \mathcal{L}_5 &= (V + C_\eta) \left[\frac{\rho c}{\beta} \left(\eta_x \frac{\partial u}{\partial \eta} + \eta_y \frac{\partial v}{\partial \eta} + \eta_z \frac{\partial w}{\partial \eta} \right) + \frac{\partial p}{\partial \eta} \right]\end{aligned} \quad (28)$$

$$\beta = \sqrt{\eta_x^2 + \eta_y^2 + \eta_z^2} \quad (29)$$

$$C_\eta = c\beta \quad (30)$$

2.2.1 Subsonic Inflow Boundary Condition

For subsonic inflow at $\xi = 1$, four quantities should be imposed. We choose u, v, w, T as these quantities, but density ρ is obtained by Eq. (12). At this boundary four characteristic waves $\mathcal{L}_2, \mathcal{L}_3, \mathcal{L}_4, \mathcal{L}_5$, are entering the computational domain, while $\mathcal{L}_1$ is going outward. Then $\mathcal{L}_1$ can be calculated from interior points using Eq. (18) by a third order finite difference scheme. $\mathcal{L}_2, \mathcal{L}_3, \mathcal{L}_4, \mathcal{L}_5$ are given by [15] [12]

$$\mathcal{L}_3 = \xi_y \frac{\partial u}{\partial t} - \xi_x \frac{\partial v}{\partial t} \quad (31)$$

$$\mathcal{L}_4 = \xi_z \frac{\partial u}{\partial t} - \xi_x \frac{\partial w}{\partial t} \quad (32)$$

$$\mathcal{L}_5 = \mathcal{L}_1 + \frac{2\beta\rho c}{\xi_x} \left[\frac{1}{\beta^2} (\xi_y \mathcal{L}_3 + \xi_z \mathcal{L}_4) - \frac{\partial u}{\partial t} \right] \quad (33)$$

$$\mathcal{L}_2 = \frac{\rho}{M_r^2} \frac{\partial T}{\partial t} + \frac{1}{2}(\mathcal{L}_1 + \mathcal{L}_5) \quad (34)$$

2.2.2 Subsonic Outflow Boundary Condition

For subsonic outflow at $\xi = N_x$, static pressure p_∞ of far field is given. At this boundary, four characteristic waves $\mathcal{L}_2, \mathcal{L}_3, \mathcal{L}_4, \mathcal{L}_5$, are leaving the computational domain, while the left one $\mathcal{L}_1$ is entering the field. So that $\mathcal{L}_2, \mathcal{L}_3, \mathcal{L}_4, \mathcal{L}_5$ can be calculated from interior points using Eq. (18) by a third order finite difference scheme, while the other one is given by

$$\mathcal{L}_1 = K(p - p_\infty) \quad (35)$$

2.2.3 Far Field Boundary Condition

For far field boundary at $\zeta = N_z$, the directions of characteristic waves are determined automatically by local values, then the $\mathcal{L}_i$ of outgoing waves are calculated from interior points using Eq. (23), and those going inward waves are set to zero. i.e.

$$\mathcal{L}_i = \begin{cases} \mathcal{L}_i & \text{for } \lambda_i > 0 \\ 0 & \text{for } \lambda_i < 0 \end{cases} \quad (36)$$

2.2.4 No-Slip Wall Boundary Condition

For no-slip wall at $\zeta = 1$, velocities u, v, w are set to zero, $\mathcal{L}_3, \mathcal{L}_4$ are given by

$$\mathcal{L}_3 = \frac{1}{\rho} [(\eta_z \zeta_x - \eta_x \zeta_z) \frac{\partial p}{\partial \eta} + (\xi_z \zeta_x - \xi_x \zeta_z) \frac{\partial p}{\partial \xi}] \quad (37)$$

$$\mathcal{L}_4 = \frac{1}{\rho} [(\eta_z \zeta_y - \eta_y \zeta_z) \frac{\partial p}{\partial \eta} + (\xi_z \zeta_y - \xi_y \zeta_z) \frac{\partial p}{\partial \xi}] \quad (38)$$

The characteristic wave $\mathcal{L}_1$ is going outward, so it can be calculated from interior points, while the in-going wave $\mathcal{L}_5$ is given by

$$\mathcal{L}_5 = \mathcal{L}_1 + \frac{2\rho\beta c}{\zeta_z} \left[\frac{1}{\beta^2} (\zeta_x \mathcal{L}_3 + \zeta_y \mathcal{L}_4) - \frac{1}{\rho} (\eta_z \frac{\partial p}{\partial \eta} + \xi_z \frac{\partial p}{\partial \xi}) \right] \quad (39)$$

for adiabatic wall, $\mathcal{L}_2$ should be

$$\mathcal{L}_2 = 0 \quad (40)$$

for isothermal wall, $\mathcal{L}_2$ should be

$$\mathcal{L}_2 = \frac{1}{2}(\gamma - 1)(\mathcal{L}_1 + \mathcal{L}_5) \quad (41)$$

When all $\mathcal{L}_i$ at each boundary are obtained, d_i can be easily calculated from Eq. (13), Eq. (22), Eq. (27). In Eq. (12), Eq. (21), Eq. (26), apart from d_i , all other terms are computed just as interior points. Boundary Eq. (12) Eq. (21), Eq. (26) are linearized and solved implicitly with interior points.

2.3 Sponge boundary conditions

Using sponge layer actually leads to an exponential decay of reflection. The idea to use sponge boundary condition to eliminate wave reflection at inflow, outflow and far-field was initiated by Fasel (1990). The basic idea can be described as follows:

If the governing equation can be abbreviated as

$$\frac{\partial U}{\partial t} + LU = 0 \quad (42)$$

$$(43)$$

In the outflow sponge layer we may change the equation to:

$$\frac{\partial U}{\partial t} + LU - f_d(x)U' = 0 \quad (44)$$

$$(45)$$

where U' is a perturbation of U ($U' = U - U_0$). This will no doubt add very large damping to the perturbation in the sponge layer. Here we can use a formula given by Collis & Lele (1996):

$$f_d(x) = \begin{cases} A_s \left(\frac{x-x_s}{x_0-x_s} \right)^{N_s} & x \in (x_s, x_0] \\ 0 & \text{otherwise} \end{cases} \quad (46)$$

$$(47)$$

Here x_s denotes the start of the sponge and x_0 denotes the location of the outflow boundary. The constants, A_s and N_s , control the amplitude and strength of the sponge function. We expect that the sponge function only help eliminate the non-physical wave reflection from the boundary but has little effect to physical solution outside the sponge. Usually, we take one T-S wavelength as the sponge layer.

The sponge layers can be embedded in the inflow and outflow sections to eliminate the reflection in the streamwise direction. For subsonic flow, the inflow sponge is necessary, while for supersonic flow, the inflow sponge can be removed.

3 Code verification with different filters and different boundary conditions

In order to make sure the FDL3D code can give converged results for steady flow without perturbation, a number of simple flows was selected to verify the code. The flow and boundary conditions and convergence are described as follows:

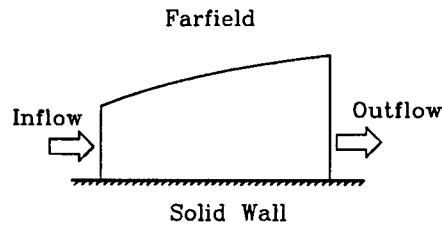


Figure 1: Boundary conditions

Flow test conditions: $XM1=0.1$, $RE=1000$, $TW=1.0$, where $XM1$, RE , and TW refer to Mach number, Reynolds number, and temperature of the wall respectively.

1. Periodic Boundary in x-, y-, and z- directions

Case I: uniform flow with all periodic boundaries

grids=32x13x13, uniform grids in y-direction

$dx=0.0967$, $dy=0.0833$, $dz=0.0833$, $dt=0.01$

$x=0-3$, $y=0-1$, $z=0-1$

max. error=0.44935E-05 after 1000 iterations

Conclusion: The compact scheme code is converged for stretched grids

Case II: uniform flow with stretched grids

grids=32x13x13, stretched grids in y-direction

periodic in x, y, and z

$dx=0.17$, $dy=0.126$ ($j=1$), $dz=0.315$, $dt=0.02$

$x=335.9 - 341.1$, $y=0-22$, $z=0-3.785$

max. error=0.111e-5 after 1000 iterations

Conclusion: The compact scheme code is converged

Case III: uniform flow with large dx

grids=32x13x13, stretched grids in y-direction

periodic in x, y, and z

$dx=1.7$, $dy=0.126$ ($j=1$), $dz=0.315$, $dt=0.02$

$x=3359 - 3411$, $y=0-22$, $z=0-3.785$

max. error=0.1977e-1 after 1000 iterations

Conclusion: The convergence is seriously degenerated when dx is larger

Case IV: uniform flow with small dx
 grids=32x13x13, stretched grids in y-direction
 periodic in x, y, and z
 dx=0.017, dy=0.126 (j=1), dz=0.315, dt=0.02
 x=3359 - 3411, y=0-22, z=0-3.785
 max. error=0.1347e-4 after 1000 iterations
 Conclusion: The convergence is worse when dx is smaller

Grid test conclusion:

The code is convergent for periodic boundaries, but is sensitive to aspect ratio of grids

2. Wall Bounded on top and bottom, periodic in x and z directions (Channel flow):

Case I: Eighth-order filter with low order filter at boundary points
 grids=32x13x13, uniform grids in y-direction
 periodic in x and z but wall bounded in y
 dx=0.0967, dy=0.0833 (j=1), dz=0.0833, dt=0.01
 x=0-3, y=0-1, z=0-1
 max. error=0.67235E-02 after 1000 iterations
 It is not converged.

Case II: Same as Case I, but the filter is changed to second order in y-direction
 max. error=0.24073E-03 after 1000 iterations
 max. error=0.30272E-05 after 2000 iterations
 It is converged well.

Case III: Same as Case I, but change the filter to forth order in J-direction but no filter on j=2 and j=jdm-1
 max. error=0.11050E-02 after 1000 iterations
 max. error=0.63993E-03 after 2000 iterations and starts oscillating after around 1500 iterations.

Case IV: Same as Case I, but change the filter to forth order in J-direction but second-order filter on j=2 and j=jdm-1
 max. error=0.68211E-2 after 1000 iterations
 max. error=0.46214E-02 after 2000 iterations
 Converges, but very slow with some oscillating

Case V: Same as Case I, but change the filter to eighth order in J-direction but no filter on j=2, 4 and j=jdm-1, jdm-2, jdm-3

max. error=0.81015E-03 after 1000 iterations
max. error=0.41929E-03 after 2000 iterations
max. error=0.28442E-03 after 3770iterations
Converges with some oscillating

Conclusion: The current filter treatment, high-order for interior points but low-order for boundary points does not work, but it works if we use a filter with same order (e.g. 2). It gives a clue that we have to develop new approach for filter at boundary points. It is better than the current approach if we do not use filter on near boundary points.

Suggestions: Not use filter on near boundary points or develop a new approach for filter on boundary points which is compatible and will not generate noises.

3. Wall Bounded on bottom, symmetric on top, and periodic in x and z directions (half channel flow):

Case I : grids=32x13x13, uniform grids in y-direction

periodic in x and z but wall bounded on bottom and symmetric on top

dx=0.0967, dy=0.0833 (j=1), dz=0.0833, dt=0.01

x=0-3, y=0-1, z=0-1

Second-order filter

max. error= 0.16020E-02 after 1000 iterations

max. error= 0.47519E-03 after 2000 iterations

Converges well with some oscillating and the solution sounds reasonable (half channel solution).

4. Wall Bounded on bottom, symmetric on top, periodic in z- direction, characteristic at inflow, and extrapolation at outlet (half channel flow).

Case I :

grids=32x13x13, uniform grids in y-direction

dx=0.0967, dy=0.0833 (j=1), dz=0.0833, dt=0.01

x=0-3, y=0-1, z=0-1

Characterisctic at inflow (Whitfield's formula)

Second-order filter

max. error= 0.57943E-01 after 1000 iterations

max. error= 0.24919E-01 after 2000 iterations

It does not converge.

Case II :

grids=32x13x13, uniform grids in y-direction

$dx=0.0967$, $dy=0.0833$ ($j=1$), $dz=0.0833$, $dt=0.01$

$x=0-3$, $y=0-1$, $z=0-1$

Characteristic at inflow (Lele's formula)

Second-order filter

max. error= 0.16017E-02 after 1000 iterations

max. error= 0.45652E-03 after 2000 iterations

Converges slow with some oscillating. The solution has no negative v near inflow.

Conclusion: Whitfield's formula cannot be used for viscous flow near the wall. Lele's formula for N-S equations can be used for viscous flow, but converges very slow. We may need to use implicit inflow boundary conditions.

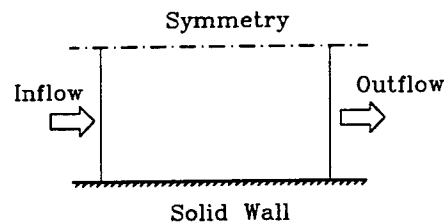


Figure 2: Channel Flow

4 Conclusion and suggestions

We conclude the report as follows :

1. The FDL3D code with high-order compact scheme and high-order filter can work for both steady and time-dependent Navier-Stokes equations
2. The code (FDL3D) needs to be improved for high aspect ratio grids
3. The code needs new filter for boundary points
4. We can use symmetric boundary condition for far-field (similar to half channel)

5. We can use extrapolation for outflow boundary
6. The current wall boundary condition is fine
7. Lele's compatible boundary condition with characteristic relation can be used for Navier-Stokes equations (viscous flow), but may need to be implicit.
8. Use of sponge can improve converges which is more important for time-dependent flow to eliminate non-physical reflection

We need to continue this effort to develop appropriate filters and compatible boundary conditions for the FDL3D code.

5 Acknowledgment

The author are grateful to AFOSR for sponsoring this summer faculty research program and for the financial support. The author are also grateful to Drs. M. Visbal, D. Gaitonde, and J. Shang for their technical assistance and helpful discussions and suggestions.

References

- [1] Collis, S.S. and Lele, S.K., A computational approach to swept leading-edge receptivity. *AIAA* 96-0180, 1996.
- [2] Fasel, H., Numerical investigation of the three-dimensional development in boundary-layer transition, *AIAA Journal*, **28**, pp 29-37, 1990.
- [3] Fromm, J., The time dependent flow of an incompressible viscous fluid, *Methods in Computational Physics*, **3**, Academic Press, New York, pp 345-382, 1964.
- [4] Gaitonde, D. V., and Shang, J.S., Optimized compact-difference-based finite volume scheme for linear wave phenomena, *J. Comput. Phys.* **138**, pp. 617-643, 1997
- [5] Hirsch,C., Numerical computation of internal and external flows, Vol. 1 :*Fundamentals of numerical discretization*, pp 403-404, 1997.
- [6] Kloker, M., Konzelmann, U., and Fasel, H., Outflow boundary conditions for spatial Navier-Stokes simulations of transition boundary layers, *AIAA Journal*, **31**, pp 620-628, 1993.
- [7] LeLe, S. K. Compact finite difference schemes with spectral-like resolution. *J. Comput. Phys.* **103**, pp.16-42, 1992.

- [8] Liu, C., Liu, Z., and McCormick, S., Multigrid methods for flow transition in three-dimensional boundary layers with surface roughness, NASA Contractor Report 4540, September, 1993.
- [9] Liu, C., and Liu, Z., Multigrid mapping and box relaxation for simulation of the whole process of flow transition in 3-D boundary layers, *J. Comput. Phys.*, **119**, pp.325-341, 1995.
- [10] Liu, C. and Liu, Z., Direct Numerical Simulation for Flow Transition Around Airfoils, Proceedings of First AFOSR International Conference on DNS/LES, Louisiana Tech University, Ruston, Louisiana, August 4-8, 1997, pp13-28, Edited by C. Liu, Z. Liu and L. Sakell.
- [11] Liu, Z., Xiong, G., and Liu, C., A contravariant velocity based implicit multilevel method for simulating the whole process of incompressible flow transition around Joukowski airfoils, *J. of Applied Mechanics and Engineering*, pp111-161, No. 1, Vol. 3, 1998, with Z. Liu and G. Xiong.
- [12] Poinso, T. J., LeLe, S. K. Boundary conditions for direct simulations of compressible viscous flows. *J. Comput. Phys.* **101**, pp 104-129, 1992.
- [13] Roache, P.J., Computational Fluid Dynamics, Hermosa, Albuquerque, New Mexico, 1972.
- [14] Strikwerda, J., Finite Difference Schemes and Partial Differential Equations, **Chapman & Hall**, pp 67-69, 1989.
- [15] Thompson, K. W., Time dependent boundary conditions for hyperbolic systems, I, *J. Comput. Phys.* **68**, pp 1-24, 1987.
- [16] Thompson, K. W., Time dependent boundary conditions for hyperbolic systems, I, *J. Comput. Phys.* **89**, pp 439-461, 1990.
- [17] Visbal, M. R. and Gaitonde, D. V., High-order accurate methods for unsteady vortical flows on curvilinear meshes, AIAA Paper 98-0131, January 12-15, 1998.
- [18] Whitfield, D., Three-dimensional unsteady Euler equation solutions using flux vector splitting, *Mississippi State University Report*, 1983.

BIDIRECTIONAL REFLECTANCE DISTRIBUTION FUNCTIONS
DESCRIBING FIRST-SURFACE SCATTERING

Carl E. Mungan
Assistant Professor
Department of Physics

The University of West Florida
11000 University Parkway
Pensacola, FL 32514-5751

Final Report for:
Summer Faculty Research Program
Air Force Research Laboratory / Munitions Directorate

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

and

Air Force Research Laboratory / Munitions Directorate
Eglin Air Force Base, FL

August 1998

BIDIRECTIONAL REFLECTANCE DISTRIBUTION FUNCTIONS DESCRIBING FIRST-SURFACE SCATTERING

Carl E. Mungan
Assistant Professor
Department of Physics
The University of West Florida

Abstract

The bidirectional reflectance distribution function (BRDF) is a measure of the amount of light scattered by some medium from one direction into another. Integrating it over specified incident and reflected solid angles defines the reflectance, which can be easily related to the absorptance (or emissivity) of a sample. The BRDF can thus be taken as a fundamental quantity for the optical characterization of an object and it correspondingly is important in a large variety of applications. When a beam of electromagnetic radiation (visible, infrared, radar, etc.) strikes a body, it can scatter off the top or first surface, as well as from the volume or subsequent interfaces. However, the total amount of light reflected from the first surface depends primarily on the complex index of refraction of the illuminated medium (relative to that of the incident medium) and is often sufficiently large that this dominates the scattering from the material. On the other hand, the topography of this interface determines the angular distribution of the scattered radiation—smooth surfaces reflect almost entirely into the specular direction, while with increasing roughness the light tends to diffract into all possible directions. Ultimately an object will appear equally bright throughout the outgoing hemisphere if its surface is perfectly diffuse (i.e., Lambertian). Measuring and modeling the BRDF can thus give valuable information about the nature of a target sample.

The paper is organized as follows. The introduction lists some domains of study in which reflectance plays an important role, as well as the relationship between it and other quantities of importance in optics. Next, the nomenclature needed to define and characterize the BRDF is presented, along with some related issues. Following this, a long section reviews the principal analytical and numerical models used to describe first-surface scattering; this comprises the heart of the present paper and considerable effort has been expended to unify the often disparate notations and points of view in the literature. Finally, this report ends with a summary of select experimental measurements, most of which are quite recent and varied in style and purpose; this gives a flavor for the ongoing efforts in this field.

BIDIRECTIONAL REFLECTANCE DISTRIBUTION FUNCTIONS DESCRIBING FIRST-SURFACE SCATTERING

Carl E. Mungan

Introduction

A variety of models and measurements of reflectance have been described in the literature but their range of validity is generally restricted by the domain of interest of the authors. Roughly speaking, these domains can be described as follows:

- Remote sensing—aircraft or satellite measurements of terrestrial vegetation and geography, usually under ambient atmospheric lighting conditions;
- Astronomy—telescopic measurements of planets and moons in the solar system illuminated directly or indirectly by the sun;
- Computer graphics—development of techniques intended to photorealistically simulate the surface appearance of a wide variety of materials;
- Military/commercial—field measurements under carefully controlled conditions of man-made objects, typically painted or bare-metal air or ground based targets;
- Optics—goniometric laboratory measurements of small, well-characterized samples.

Select models and experiments applicable to each of these five domains are discussed in this paper.

In general, theoretical models begin by considering light to be incident on a spot on the surface of an object from a range of solid angles Ω_i (which can vary from a delta function for a fully directional beam to 2π for the hemisphere) centered on polar (zenith) angle θ_i (defined with respect to the macroscopic surface normal) and azimuthal angle ϕ_i (conveniently defined with respect to some sample feature in the case of an anisotropic surface or arbitrarily and irrelevantly for an isotropic one). This light must be either transmitted, absorbed, or reflected by the object; the fraction of the incident flux P (radiant power in W) which is subsumed by each of these mechanisms is specified by the dimensionless ratios called the transmittance τ , the absorptivity (or absorptance) α (not to be confused with the absorption coefficient given by $-\ln[1-\alpha]/L$ where L is the sample length), and the reflectance ρ , respectively. I will always assume that the illuminated object is opaque, so that $\tau = 0$; accordingly, $\alpha + \rho = 1$. The reflected light (which can be more generally described as scattered light) is collected by a detector spanning a solid angle Ω_r centered on angles θ_r and ϕ_r . The emissivity ϵ of the sample is defined to be the dimensionless ratio of the total radiant flux emitted by the sample to that of a blackbody having the same geometry and temperature. Kirchhoff's law says that $\epsilon = \alpha$, as follows by imagining the sample to be enclosed inside an opaque cavity whose temperature is equal to that of the sample: this equality means that there can be no net gain from or loss to the blackbody environment by the object. Furthermore, by interposing narrowband filters between the sample and cavity, it is evident that the spectral emissivity $\epsilon(\lambda)$, defined in terms of the ratio of emitted fluxes in a unit wavelength interval centered around λ , must equal the spectral absorptivity $\alpha(\lambda)$, whose definition is similarly a ratio of spectral fluxes. [Note that I have eschewed the notation ϵ_λ , to avoid confusion with the spectral emission coefficient defined in Sec. 13.1.1 of Hecht (1998).] An object is said to be a graybody if $\epsilon(\lambda)$ is a constant less than 1. Nicodemus (1965, 1970) has argued that equality also holds for the directional quantities, $\epsilon(\theta, \phi) = \alpha(\theta, \phi)$, for emission into some direction and absorption from the same direction. [One could then call ϵ the hemispherical emissivity, equal to the average of the directional emissivity over all projected solid angles, analogous to Eq. (2) below.] Thus the directional and spectral dependence of the emitted light can be related to the reflectance and temperature (via the Stefan-Boltzmann law) of the sample. Furthermore the polarization of the emission is seen to be the complement of that of the reflected radiation at the same wavelength, a result which can also be understood by visualizing an emitting center as lying slightly below the surface of the sample, so that the light escapes in

accordance with the Fresnel relations [cf. Eq. (6) below]; in fact, Sandus (1965) has argued that this holds even for emitting centers lying on the surface. Consequently, I will mainly focus attention on the reflectance for the remainder of this paper, even in the infrared region where the sample emission is significant at the ambient temperatures of interest.

Reflectance Nomenclature

The incident flux P per unit illuminated area A of an object is known as the irradiance (or illuminance) $H \equiv dP/dA$ in units of W/m^2 ; the corresponding emitted or scattered quantity is called the emittance or exitance. [Note that the element dA of area is taken to be small on the macroscopic scales of interest, but large compared to the sub-resolution irregularities in the sample surface, and is taken to be perpendicular to the macroscopic surface normal (Nicodemus, 1965).] For simplicity, both the illuminance and the exitance will be referred to as the irradiance of the surface, to be distinguished from the intensity of the irradiating beam, $I \equiv dP/(dA \cos \theta)$, where the $\cos \theta$ term projects the element of surface area dA into the direction of propagation of the beam, which is inclined at polar angle θ to the (macroscopic) surface normal. The irradiance per unit projected solid angle is called the radiance (or luminance, which corresponds to the photometric concept of brightness), $L \equiv dH/(\cos \theta d\Omega) = dI/d\Omega$ in units of $W m^{-2} sr^{-1}$. Here the element of solid angle is $d\Omega \equiv \sin \theta d\theta d\phi$; sometimes (Nicodemus *et al.*, 1977) a separate symbol is also introduced for the element of projected solid angle $\cos \theta d\Omega$, though I will not do so, to keep things simpler. Let the subscript i refer to incident quantities and r to reflected (scattered) terms. The bidirectional reflectance distribution function (BRDF) can now be defined as $f(\theta_i, \phi_i; \theta_r, \phi_r) \equiv dL_r/dH_i = dP_r/(dP_i \cos \theta_i d\Omega_r)$ with units of sr^{-1} ; here dP_i is the incident power illuminating dA from direction (θ_i, ϕ_i) and dP_r is the radiant power reflected from dA into the outgoing solid angle $d\Omega_r$, centered about the direction (θ_r, ϕ_r) . Usually the surface is taken to be (macroscopically) planar and the incident beam to have a spatially uniform intensity profile, so that one can integrate over a macroscopic area A of the sample to get $f = dP_r/(P_i \cos \theta_i d\Omega_r)$ where now P_i is the incident power illuminating A and dP_r is the flux reflected from A into $d\Omega_r$. For an isotropic surface, f is a function only of $\phi_r - \phi_i$ and not of the two azimuthal angles individually. Unlike reflectances, which cannot exceed unity, the BRDF can be very large, becoming unbounded for a purely specular reflector—cf. Eq. (5) below. When necessary, a double subscript can be added to f to denote the polarizations of first the source and second the detector relative to the planes of incidence and reflection, respectively. In that case, these four values of f can be thought of as the elements of a 2×2 matrix $\mathbf{f}$ and the component intensities of the incident and reflected radiation as two-vectors such that $d\mathbf{L}_r = \mathbf{f} d\mathbf{H}_i$. However, this description does not suffice to determine the scattering for other states of incident and reflected polarization, such as circular, because the phase information of the fields has not been retained. One approach to circumventing this limitation is to work instead with the (complex) components of the incident and reflected electric fields, $\mathbf{E}_i$ and $\mathbf{E}_r$, respectively, which are connected by a 2×2 scattering matrix $\mathbf{S}$ (i.e., $\mathbf{E}_r = \mathbf{S} \mathbf{E}_i$), which is known as the Jones calculus (Ruck *et al.*, 1970). Alternatively, the Mueller matrix and Stokes vector representation can be used, wherein $\mathbf{f}$ is expressed as a 4×4 matrix and the beams as intensity 4-vectors (Flynn and Alexander, 1995). This latter approach is needed for describing partially polarized radiation, while the Jones notation is used for coherent illumination and reflection. If the radiation is partially coherent, then a treatment in terms of the coherency (or polarization) matrix is necessary. For a nice discussion of the relation between the Jones, Mueller, and coherency matrices, with particular reference to the ensemble averaging necessary for handling scattering from statistical media, see Kim *et al.* (1987).

By integrating the BRDF over all scattered angles, one gets the dimensionless directional-hemispherical reflectance

$$\rho(\theta_i, \phi_i; 2\pi) = \int_{2\pi} f(\theta_i, \phi_i; \theta_r, \phi_r) \cos \theta_r d\Omega_r, \quad (1)$$

which gives the fractional amount of flux reflected into the entire hemisphere out of that incident from a particular direction. (The integral over 2π is of course a shorthand meaning integrate θ

from 0 to $\pi/2$ and ϕ from 0 to 2π , for a total solid angle of 2π .) If instead one averages over all incident angles, the hemispherical-directional reflectance is obtained as

$$d\rho(2\pi; \theta_r, \phi_r) = \frac{\cos \theta_r d\Omega_r}{\pi} \int_{2\pi} f(\theta_i, \phi_i; \theta_r, \phi_r) \cos \theta_i d\Omega_i. \quad (2)$$

Note that the denominator is π and not 2π because of the $\cos \theta_i$ projection factor. It should be clear from inspection of these two equations how to write other quantities of interest, such as the conical-hemispherical reflectance, the bihemispherical reflectance (or albedo), and so on—Nicodemus *et al.* (1977) tabulate a variety of such entities. If the prefactor before the integral in Eq. (2) is left off, the resulting quantity is instead called the hemispherical-directional reflectance factor, which is defined as the ratio of the flux reflected by the sample to that which would be reflected by an ideal Lambertian reflector for the same geometry, because $f = 1/\pi$ for the latter according to the discussion following Eq. (9).

From the definition of the BRDF, the scattered radiance can be related to the incident radiance by

$$L_r(\theta_r, \phi_r) = \int_{\Omega_i} L_i(\theta_i, \phi_i) f(\theta_i, \phi_i; \theta_r, \phi_r) \cos \theta_i d\Omega_i, \quad (3)$$

which can be rather complicated to evaluate analytically. The emissivity is given by Kirchhoff's law as

$$\varepsilon(\theta_i, \phi_i) = 1 - \int_{2\pi} f(\theta_i, \phi_i; \theta_r, \phi_r) \cos \theta_r d\Omega_r, \quad (4)$$

and is thus simply related to the (directional-hemispherical) reflectance, as discussed in the introduction.

According to the Helmholtz reciprocity theorem, $f(\theta_i, \phi_i; \theta_r, \phi_r) = f(\theta_r, \phi_r; \theta_i, \phi_i)$, so that either direction may be that of the incident beam with the other the reflected beam. Note from Eqs. (1) and (2) that this would imply that the directional-hemispherical reflectance equals the hemispherical-directional reflectance factor; interestingly enough, however, equality of these two reflectance terms remains true even when Helmholtz reciprocity does not (Nicodemus, 1970). There has been considerable discussion about the conditions under which reciprocity holds (Clarke and Parry, 1985; Kriebel, 1996; Shirley *et al.*, 1997; Snyder, 1998)—for example, it clearly does not apply to a Faraday isolator. It is generally accepted however that the theorem is obeyed by most materials (including compound objects) under ordinary conditions and hence models which are not reciprocal are generally considered unphysical and experimental failures are often attributed to measurement errors or limitations.

BRDF Models of Surface Scattering

It is often possible to distinguish surface scattering from volume scattering. Some samples (e.g., bare metals) have a penetration depth (equal to the reciprocal of the absorption coefficient) which is so small that all of the reflection can be assumed to occur at its illuminated or first surface. Other samples (e.g., a painted object) exhibit scattering both from the first surface (and possibly subsequent surfaces, such as the paint-substrate interface) as well as from the bulk interior (e.g., due to defects or paint pigments). Still other samples (e.g., a forest imaged from an airplane) do not have a well-defined first surface at all and could perhaps be best described as pure volume scatterers (Snyder and Wan, 1998). Volumetric reflection almost invariably involves multiple scattering (although a sufficiently cratered surface can as well). This greatly adds to the complexity of the problem and is not fully understood. For this reason, I will restrict attention to first-surface scattering.

Begin by considering the simplest case. A purely specular reflector (i.e., a perfectly smooth and planar mirror) has a BRDF equal to zero unless $\theta_r = \theta_i$ and $\phi_r = \phi_i + \pi$, in which case the reflectance is described by the Fresnel equations, so that (Ellis, 1994)

$$f = R(\theta_i) \frac{\delta(\theta_r - \theta_i) \delta(\phi_r - \phi_i - \pi)}{\cos \theta_r \sin \theta_r}, \quad (5)$$

assuming that the illuminated area of the mirror is large enough (compared to the wavelength of the light) that diffraction from it into off-specular directions is negligible. For nonmagnetic materials (Hecht, 1998),

$$R_s = |r_s|^2 \quad \text{with} \quad r_s = \frac{\cos \theta_i - \sqrt{n^2 - \sin^2 \theta_i}}{\cos \theta_i + \sqrt{n^2 - \sin^2 \theta_i}} \quad (6a)$$

and

$$R_p = |r_p|^2 \quad \text{with} \quad r_p = \frac{n^2 \cos \theta_i - \sqrt{n^2 - \sin^2 \theta_i}}{n^2 \cos \theta_i + \sqrt{n^2 - \sin^2 \theta_i}} \quad (6b)$$

for polarization perpendicular ("senkrecht," also called horizontal or TE and denoted by an h , $-$, or $\perp$) and parallel (sometimes referred to as vertical or TM and symbolized by v , $+$, or $\parallel$) to the plane of incidence, respectively. Here $n \equiv n_t/n_i$ where n_i and n_t are the (complex) refractive indices of the incident medium and the transmitted medium (i.e., the top layer of the sample), respectively; $n_i = 1$ and $n_t \equiv n$ correspond to the usual case of light in space scattering off a medium of index n . Any incident beam can of course be decomposed into s and p components, provided that proper accounting is made of their relative phase. Recall that these Fresnel reflectances are strongly dependent upon the angle of incidence—for example, R_p falls to zero at the Brewster angle for a dielectric (or has a minimum at the pseudo-Brewster angle for a conducting medium) but rises to unity for grazing incidence. Substitution of Eq. (5) into Eq. (1) identifies $R(\theta_i)$ as the directional-hemispherical reflectance; if the mirror is ideal (i.e., has infinite conductivity) then this reflectance is unity for any polarization and wavelength (Born and Wolf, 1965). Note that a flat specular reflector does not depolarize incident s or p radiation, although it will reverse the handedness of circular light and rotate the plane of polarization of light linearly polarized in directions other than s and p , as discussed in more detail later.

All real surfaces have some roughness however. If the characteristic size scales (to be defined more precisely below) of this roughness are large compared to the wavelength of the light, then the approximation of geometrical optics holds. In that case, the simplest model for the surface scattering is that of Torrance and Sparrow (1967), who pictured the surface as being comprised of small, randomly oriented, specular facets. Neglecting shadowing and masking of the facets by each other, the BRDF for single scattering from this isotropic surface is easily shown to be

$$f = R(\beta) \frac{\Xi(\theta_n)}{4 \cos \theta_i \cos \theta_r}, \quad (7)$$

where 2β is the angle between the incident and reflected directions, known as the bistatic angle ($2\beta = 0$ corresponds to monostatic or back scattering, while $2\beta = \pi$ refers to forward scattering pertinent to the optical extinction theorem for example), and $\Xi(\theta_n)$ is the density function (in units of sr^{-1}) of facet normals pointing in the direction θ_n relative to the macroscopic surface normal, i.e., $\Xi(\theta_n) d\Omega_n dA$ is the total (not the projected) surface area of the facets spanned by dA which have normals lying within solid angle $d\Omega_n$ (Maxwell and Weiner, 1974). Clearly this density function must be normalized such that

$$2\pi \int_0^{\pi/2} \Xi(\theta_n) \cos \theta_n \sin \theta_n d\theta_n = 1. \quad (8)$$

Torrance and Sparrow took Ξ to be a Gaussian distribution function with zero mean. It can be shown that $\cos 2\beta = \cos \theta_i \cos \theta_r + \sin \theta_i \sin \theta_r \cos(\phi_r - \phi_i)$, $\cos \theta_n = (\cos \theta_i + \cos \theta_r)/2 \cos \beta$, and $d\Omega_n = d\Omega/4 \cos \beta$ using spherical geometry. Observe that Eq. (7) explicitly satisfies Helmholtz reciprocity. It can be rewritten in terms of experimentally measurable quantities by noting that $\Xi(\theta_n)$ is proportional to the monostatic BRDF $f(\theta_n, 0; \theta_n, 0)$. The polarization dependence is

determined solely by the Fresnel coefficients; hence, in the plane of incidence there are only 4 independent Mueller matrix elements—e.g., the 11, 12, 33, and 34 terms, which can be related to the real and imaginary parts of the 2 diagonal elements, $r_s(\beta)$ and $r_p(\beta)$, of the Jones scattering matrix. Videen *et al.* (1992) have shown that this simple polarization prediction is in quite good agreement with experimental measurements on scratched or sandblasted copper and aluminum. Sung and Eberhardt (1978) have improved on the model by allowing the facets to be slightly curved and calculating a perturbative correction using the Rayleigh-Rice method described later. The above BRDF needs to be multiplied by a geometric attenuation factor $G(\hat{\mathbf{i}}, \hat{\mathbf{n}}, \hat{\mathbf{r}})$ to correct for shadowing and masking, the former referring to the partial illumination of a facet shadowed by an adjacent one and the latter to the partial visibility of a facet occluded by another. Here $\hat{\mathbf{i}}$, $\hat{\mathbf{n}}$, and $\hat{\mathbf{r}}$ refer to unit vectors in the source, facet normal, and detector directions respectively, as seen from the illuminated spot; note that $2\beta = \cos^{-1}(\hat{\mathbf{i}} \cdot \hat{\mathbf{r}})$. For simplicity, Torrance and Sparrow supposed that each facet forms the side of a symmetric V-groove cavity whose axes are parallel to the macroscopic surface (but have random azimuthal orientations) and whose upper edges all lie in the same plane. In that case, the fraction G of a given facet surface which contributes to the reflected flux (i.e., which is both illuminated and visible) can be found geometrically in terms of a range of unobscured angles of the projections of $\hat{\mathbf{i}}$ and $\hat{\mathbf{r}}$ into the plane defined by $\hat{\mathbf{n}}$ and the macroscopic surface normal; the grooves are assumed to be long enough that end effects can be neglected. Although this model for the geometric attenuation is very simple, it successfully predicts that for large angles of incidence the peak in the BRDF in the plane of incidence occurs at an angle of reflection larger than specular. Such off-specular peaks are seen experimentally, both for coarsely roughened metals and dielectrics.

Finally, Torrance and Sparrow assumed that the multiply reflected light (and any volume scattering) was purely diffused. A purely diffuse (Lambertian) scatterer has two properties. First, its reflected radiance L_r is independent of the viewing angle and hence is equal to the hemispherical exitance H_r divided by π (again not 2π). The reflected flux per unit solid angle is proportional to $\cos\theta_r$, which is called Lambert's law; since the projected area of a surface element also varies as $\cos\theta_r$, this is consistent with the fact that L_r is independent of the viewing angle. Equation (3) then implies that f is independent of θ_r and ϕ_r , and so Eq. (1) becomes $f = \rho(\theta_i, \phi_i; 2\pi)/\pi$. Second, its directional-hemispherical reflectance is independent of the angle of incidence and thus is equal to the bihemispherical reflectance; this property is a statement of reciprocity. The BRDF is then related to the bihemispherical reflectance according to

$$f = \frac{\rho(2\pi; 2\pi)}{\pi} \quad (9)$$

and is independent of both the incident and reflected angles. If the surface is ideal (i.e., has unit albedo) then the BRDF further simplifies to $1/\pi$. Incidentally, note that a blackbody has zero albedo according to Eq. (4) and hence obeys Lambert's law in emission; this explains why the sun appears as a uniformly bright disk even though it actually is a sphere. It should be emphasized that Eq. (9) does *not* imply that L_r is independent of the incident angle: on the contrary, Eq. (3) says that the reflected radiance due to a well-collimated incident beam of fixed radiance is proportional to $\cos\theta_i$ because the flux incident on a unit area of the surface varies in this way (Torrance and Sparrow, 1967); this means that a spherical Lambertian scatterer illuminated from a fixed direction will appear shaded around the edges (Oren and Nayar, 1995), not flat as some authors have incorrectly claimed (Ellis, 1994), in striking contrast to a spherical Lambertian emitter. It should also be borne in mind that many rough objects, such as an ordinary piece of paper, are reasonably Lambertian at near-normal viewing angles, but become increasingly specular at grazing angles due to the increase in the first-surface Fresnel reflectance as well as to the compression of the surface roughness in the viewing direction—this latter effect gives rise, for example, to the horizontally streaked reflections commonly seen off waxed tile floors (Shirley *et al.*, 1997). Finally, note that by substituting Eq. (5) or (9) into (1) or (2), the directional-hemispherical reflectances of an ideal mirror and of an ideal diffuser are found to be equal, and likewise for the hemispherical-directional reflectances (Judd, 1967).

Multiple scattering from the surface or volume of an object tends to depolarize the incident light; even two reflections from a valley can cross s or p polarization, as nicely illustrated in Fig. 14 of O'Donnell and Mendez (1987). For this reason, the light scattered by a Lambertian surface is typically assumed to have random polarization. However, this is not necessarily true, because the two defining properties which led to Eq. (9) do not explicitly require multiple scattering. For example, one could imagine choosing the density function in Eq. (7) to give a BRDF that is at least approximately independent of incident and reflected angles, thus characterizing a Lambertian diffuser, although only singly scattered, specular reflections from isotropic facets are involved which consequently cannot depolarize s or p light. Even volume scattering from a paint does not fully randomize the polarization of the radiation, as Ellis (1996b) has demonstrated experimentally. Therefore, one should be careful about assuming that diffusers always reduce the degree of polarization of a beam to nearly zero, in the absence of detailed knowledge of the scattering processes.

Oren and Nayar (1995) have developed an alternative form of Torrance and Sparrow's model in which each facet is assumed to reflect purely diffusely rather than specularly. In effect, this presumes that the surface has two distinct roughness scales and hence is said to be compositely rough (Barrick, 1970): a coarse range specified by macroscopically flat facets whose dimensions are large compared to the wavelength of the light, and a fine range responsible for Lambertian scattering from each facet. [A very different two-roughness-scale model has been developed by Leader (1979).] Their model was developed to describe the reflectances of materials such as plaster, sandpaper, and cloth for computer rendering applications, and they performed experimental measurements which gave fairly good agreement with the calculations. Although one might suppose that the overall effect of the two roughness scales in the model is merely to further roughen the surface and hence leave the overall scattering approximately Lambertian, this is incorrect: the reflectance is markedly non-Lambertian and in fact has a strong backscattering peak for large angles of incidence which gets cut off for reflection angles beyond the source direction (i.e., $\theta_r < -\theta_i$). Notice that this is completely opposite to the strong forward scattering peak predicted by the Torrance and Sparrow model.

The cause of the backscattering peak is reminiscent of Hapke's (1963) shadow-hiding explanation for the opposition effect of the moon, wherein its brightness peaks at full moon when the sun is directly behind the earth. In the lunar case, the relevant scattering is volumetric, because the dust on the moon is very porous and is taken to be an open network of particles. Particles closer to the surface cast shadows down on the lower particles, which thus reduces the reflected radiance in every direction except the incident one where the shadows are hidden by the particles that created them. To put it another way, the light will be attenuated both as it penetrates the medium and as it leaves it after being reflected, unless it backscatters into the incident direction in which case it can retrace its path out and escape without being blocked. It is assumed that the albedo ρ of the individual particles is small enough that multiple scattering is negligible and that the particles are large enough that diffraction around their edges is minimal, since either of these would tend to wash out the effect. After making a number of simplifying approximations, the BRDF of the medium can then be expressed as the product of three factors, each of which satisfies Helmholtz reciprocity,

$$f = \frac{1}{\cos \theta_i + \cos \theta_r} \cdot R_B(\beta, g) \cdot S(\beta). \quad (10)$$

The first factor is the well-known Lommel-Seeliger scattering term arising from the total distance traveled through the attenuating medium in the course of reflecting from a typical particle. The second term is called the retroreflection function and decreases monotonically in value from 2 for $2\beta = 0$ (backscattering) to 1 for $2\beta \geq \pi/2$. It depends on a constant, g , of the order of unity, which is called the compaction parameter because it is related to the fractional volume of the medium occupied by the particles. The last factor is the scatter function, defined as the ratio of the scattered power per unit solid angle to the power incident on a particle by a plane wave, averaged over all orientations and shapes of the particles. Taking the incident plane wave

to be of unit intensity, this implies

$$S(\beta) \equiv \left\langle \frac{dP_r}{A_\perp d\Omega_r} \right\rangle \quad (11)$$

where $A_\perp$ is the cross-sectional area of the particle (i.e., the projected area of the particle onto an incident wave front). Notice that $S(\beta)$ is approximately equal to the ensemble average of the cosine-corrected BRDF (Stover, 1990) of the particles, $dP_r/(dP_i d\Omega_r)$; the equality becomes exact for pancake-shaped (i.e., planar) particles. The scatter function is written by Hapke as $S(\beta) = \rho \Sigma(\beta)$, where $\Sigma(\beta)$ has been normalized such that

$$4\pi \int_0^{\pi/2} \Sigma(\beta) \sin(2\beta) d\beta = 1. \quad (12)$$

Both $S(\beta)$ and $\Sigma(\beta)$, just like f , have units of sr^{-1} . For example, $S(\beta) = 1/4\pi$ describes isotropic scattering from smooth, infinitely conducting (i.e., ideally specular) spheroids—if the conductivity is not infinite, then Eq. (6) implies that there will be a broad forward scattering peak (Van de Hulst, 1957) since the Fresnel reflectances only become equal to unity at grazing angles—while $\Sigma(\beta) = 2[\sin(2\beta) + (\pi - 2\beta)\cos(2\beta)]/3\pi^2$ describes backscattering from Lambertian spheres, as first calculated by Schönberg. As can be seen from Hapke Fig. 6, substituting this latter scattering function into Eq. (10) results in a strong backscattering peak. This is also the explanation for the glory seen when looking down on clouds from an airplane with the sun at one's back and for the hot spots which appear on vegetation in remote sensing.

Experimental retroreflectance measurements of MgCO_3 , BaSO_4 , sulfur, and white, red, blue, and black Nextel paints were later performed by Egan and Hilgeman (1976) using a cube beamsplitter, which was cleverly followed (rather than preceded) by a chopper so as to reject light scattered by the room or the prism. Both a bandpass-filtered tungsten iodide lamp and a 632.8-nm HeNe laser were used as sources. In all cases an opposition effect was observed that could reasonably be described by Eq. (10), except for the blue and black paints under laser illumination which exhibited an anomalously strong backscattering peak attributed to interference effects. Presciently, the opposition effect of the moon is now known to result not from shadow hiding but instead from coherent backscatter (Hapke *et al.*, 1993). This phenomenon, related to weak localization, arises as follows (Wolf and Maret, 1985). Suppose that a wave of incident propagation vector $\mathbf{k}_0$ experiences m elastic scattering events, where $m \geq 2$. Let $\mathbf{k}_i$ denote the propagation vector after the i^{th} event, so that $\mathbf{k}_m$ points in the direction of observation. In the case of backscattering, $\mathbf{k}_m = -\mathbf{k}_0$, and hence some of the incident field can also follow the time-reversed path $-\mathbf{k}_m \rightarrow -\mathbf{k}_{m-1} \rightarrow \dots \rightarrow -\mathbf{k}_0$. The phase difference between these two paths is obviously zero and thus they will interfere constructively, giving double the signal calculated for the otherwise incoherent addition of intensities. Convincing evidence that this enhancement effect is responsible for lunar backscattering was provided by examining the circular polarization ratio of scattered radiation from soil samples with incident circular laser light. Shadow hiding involves primarily single scattering so that the helicity should be reversed, while coherent backscatter involves multiple scatterings, many of which are into the forward direction, and hence the original polarization should be partially preserved, and this is what is in fact observed. The angular width of this coherent backscatter peak is supposed to approximately equal λ/l , where l is the transport mean free path for photons in the medium; for strongly absorbing particles as in the case of the moon, l is roughly equal to the average spacing between scatterers. Analysis of the data implies $l = 1 \mu\text{m}$, in contrast to the $40 \mu\text{m}$ mean particle size in the lunar dust; hence, the scatterers must be small asperities on the grain surfaces, rather than the particles themselves. This is consistent with the observation of coherent backscatter from wavelength-sized roughness on well-characterized surfaces (O'Donnell and Mendez, 1987).

Returning to Oren and Nayar's model, suppose that a V-cavity is illuminated from the right. Then the left facet will be brighter than the right one because it receives more light. Viewed from the left, an observer sees principally the darker right facet and comparatively little of the foreshortened left facet. But as he moves toward the source direction, the fraction of the brighter

area increases while that of the darker decreases and hence the BRDF increases in the backscattering direction and is inherently non-Lambertian. Specifically, the BRDF for a single Lambertian facet whose normal is inclined at polar angle θ_n and azimuthal angle ϕ_n relative to the macroscopic surface normal is

$$f = \frac{\rho}{\pi} \cos \theta_n [1 + \tan \theta_i \tan \theta_n \cos(\phi_i - \phi_n)] [1 + \tan \theta_r \tan \theta_n \cos(\phi_r - \phi_n)], \quad (13)$$

where ρ is the albedo of the facet. This expression is then multiplied by a geometric attenuation factor $G(\hat{\mathbf{i}}, \hat{\mathbf{n}}, \hat{\mathbf{r}})$ to account for shadowing and masking, and the result is averaged over ϕ_n assuming a uniform distribution of azimuthal orientations of the V-grooves. Finally, that is multiplied by the density function of facet normals $\Xi(\theta_n)$ and integrated with respect to $\sin \theta_n d\theta_n$ over the hemisphere. Oren and Nayar chose $\Xi(\theta_n) \cos \theta_n$ to be a Gaussian with zero mean, which differs slightly from Torrance and Sparrow (1967) or Maxwell and Weiner (1974) who chose $\Xi(\theta_n)$ itself to be such. Finally, two-bounce interreflections were taken into account by integrating over all positions on the two faces of a V-groove which connect directions $\hat{\mathbf{i}}$ and $\hat{\mathbf{r}}$ geometrically, where shadowing and masking determine the limits of integration, and finally again integrating over ϕ_n and θ_n . The singly scattered and doubly scattered BRDFs were then added to give the total BRDF, which exhibits reciprocity. Some of the required integrals are fairly complicated and were evaluated numerically or approximated functionally. The resulting BRDF is nearly Lambertian for small angles of incidence, wherein both facets of every V-cavity have similar irradiance, as well as for azimuthal angles 90° out of the plane of incidence since the relative irradiance of both facets is then approximately constant with respect to the polar angle of reflection. But there is a strong backscattering peak when the source and viewing directions coincide, as well as an interreflection enhancement in the forward direction, and the rendered image of a curved surface illuminated from the viewer direction can be made very flat, mimicking the appearance of the moon or of a photographed clay vase for example. Of course the model reduces to pure Lambertian scattering if the roughness (as parametrized by the Gaussian standard deviation in the facet density function) is set equal to zero.

The facets in Torrance and Sparrow's or Oren and Nayar's model are assumed to be large compared to the wavelength λ of the incident light. More specifically, any non-composite, isotropic surface (or any individual component of a composite surface) can be characterized by two roughness scales—an out-of-plane length quantified by the rms surface height variation σ and an in-plane distance parametrized by the correlation length l . In the facet models, both of these parameters (for the coarse component in Oren and Nayar's case) must be much larger than λ . Mathematically (Bennett and Porteus, 1961), the surface height profile is taken to be $z = \zeta(x, y)$ with the zero level set equal to the mean,

$$\langle z \rangle \equiv \lim_{X, Y \rightarrow \infty} \frac{1}{XY} \int_{-Y/2}^{Y/2} \int_{-X/2}^{X/2} \zeta(x, y) dx dy = 0, \quad (14)$$

where the sample has been taken to be infinite in area, macroscopically spanning the xy -plane, for simplicity. With these definitions, the mean square surface height is given by the variance,

$$\sigma^2 \equiv \langle z^2 \rangle \equiv \lim_{X, Y \rightarrow \infty} \frac{1}{XY} \int_{-Y/2}^{Y/2} \int_{-X/2}^{X/2} \zeta^2(x, y) dx dy, \quad (15)$$

and the autocorrelation (or autocovariance) function is (Hecht, 1998)

$$A(s, t) \equiv \lim_{X, Y \rightarrow \infty} \frac{1}{XY} \int_{-Y/2}^{Y/2} \int_{-X/2}^{X/2} \zeta(x, y) \zeta(x - s, y - t) dx dy, \quad (16)$$

so that $A(0, 0) = \sigma^2$. For an isotropic surface, $A(s, t) = A(t, s)$ so that the correlation function is only a function of a single variable $\tau \equiv (s^2 + t^2)^{1/2}$ known as the lag; the correlation length l is defined as that lag for which A diminishes to $1/e$ of its peak value (i.e., to σ^2/e). There are two commonly used statistical distributions for rough surfaces (Barrick, 1970). The first assumes Gaussian correlations,

$$A(s,t) = \sigma^2 e^{-(s^2+t^2)/l^2}, \quad (17)$$

for which the rms slope of the surface can be shown to be

$$m \equiv \sqrt{\left\langle \left(\frac{\partial \zeta}{\partial x} \right)^2 + \left(\frac{\partial \zeta}{\partial y} \right)^2 \right\rangle} = \frac{2\sigma}{l} \quad (18)$$

(which equals the rms value of $\tan \theta_n$ in the facet models and hence is determined in turn by the density function). The other supposes the surface height autocorrelation function to be exponential,

$$A(s,t) = \sigma^2 e^{-\sqrt{s^2+t^2}/l}, \quad (19)$$

in which case m turns out to be undefined, because such surfaces are jagged with many vertical facets, describing, as an example in remote sensing, an urban area including buildings.

The requirement that $\sigma \gg \lambda$, as well as the naive model of V-grooves having coplanar top edges, can be relaxed by invoking wave rather than geometrical optics. Nevertheless, many simplifying assumptions remain in order to make the physics tractable; the general problem of optical scattering by an arbitrarily specified surface has not been solved, even without considering volumetric scattering of the portion of the beam transmitted below the first surface. Briefly, the goal is to find the reflected electromagnetic field for some known incident field and characterized surface by satisfying the required boundary conditions. Typically shadowing, masking, multiple reflections, and surface waves are neglected. One approach has been developed in great detail by the radar scattering community (Beckmann and Spizzichino, 1987); note that the (differential) radar cross section σ (having units of area) is related to the BRDF by $f = \sigma/4\pi A \cos \theta_i \cos \theta_r$, where A is the (macroscopic) illuminated surface area of the sample (Ruck *et al.*, 1970). O'Donnell and Mendez (1987) have verified that this theory is in good agreement with optical measurements on appropriately fabricated samples.

In detail, Maxwell's equations imply that the electric field satisfies a wave equation, and if the time dependence is separated out (Fourier analyzing the constituent frequencies if necessary), the Helmholtz equation is obtained for any scalar component of the field (typically the incident and reflected radiation are decomposed into s and p polarizations). Green's theorem can be used to recast this differential equation as the Kirchhoff integral, which expresses the field anywhere in space in terms of the electric field and its gradient at every point on any closed surface enclosing the spatial point of interest. In the present application, the surface is taken to be that of the object together with an enclosing hemisphere at infinity, with the latter giving zero contribution to the result. In accord with Huygens' principle, each point on the material surface is taken to be the source of a spherical wave, and the Fraunhofer far-field limit is considered, in which the incident and scattered beams are taken to be plane waves. Finally, the Kirchhoff approximation is invoked to obtain the electric field and its normal derivative on the reflecting surface, by assuming that the scattering from any point on the surface is described by the Fresnel amplitude reflection coefficients, r in Eq. (6), from a smooth plane tangent to that point. This is the most serious restriction of the validity of this approach, as it obviously requires that the radii of curvature of any surface irregularities be large compared to the wavelength, or equivalently, that $l \gg \lambda$, implying a gently rolling surface. A formal integral solution to the problem of scattering by a rough surface of finite conductivity is then obtained. Unfortunately it is too complicated to evaluate for real situations of interest unless the relevant amplitude reflection coefficient, which depends on the angle of incidence relative to the local surface normal, is constant across the surface. In practice, this occurs either because the conductivity can be taken to be infinite or because the surface roughness is sufficiently small that we can replace the reflection coefficient by its average value. For the simpler, perfectly conducting case, integration by parts gives the scattered field in terms of a Fourier transform over the area A of the isotropic surface,

$$\frac{E_r}{E_{r,smooth}} = F(\theta_i, \theta_r, \phi_r - \phi_i) \frac{1}{A} \iint_A e^{i(\mathbf{k}_i - \mathbf{k}_r) \cdot \mathbf{r}_s} dx dy \quad (20)$$

neglecting edge effects, where $\mathbf{k}_i$ and $\mathbf{k}_r$ are the incident and reflected propagation vectors, respectively, with magnitude $k \equiv 2\pi/\lambda$, $E_{r,smooth}$ is the field which would be diffracted into the specular direction if the finite-sized, infinitely-conducting surface were smooth and the incident wave were *s*-polarized, $F = [1 + \cos\theta_i \cos\theta_r + \sin\theta_i \sin\theta_r \cos(\phi_r - \phi_i)] / [\cos\theta_i (\cos\theta_i + \cos\theta_r)]$, and $\mathbf{r}_s = (x, y, \zeta)$ is a point on the surface. Accounting for the Huygens-Fresnel Principle (cf. Sec. 10.3.1 of Hecht, 1998), the projected area of the sample ($A \cos\theta_i$), and the solid angle subtended at the detector ($d\Omega_r$), I find that the BRDF is given by

$$f = \frac{A \cos\theta_i}{\lambda^2 \cos\theta_r} \left| \frac{E_r}{E_{r,smooth}} \right|^2, \quad (21)$$

which satisfies Helmholtz reciprocity. For a statistically rough surface, Eq. (21) needs to be averaged by multiplying it by the normalized height distribution $p(\zeta)$ and integrating over all ζ . The usual assumption is that this distribution is Gaussian, i.e.,

$$p(\zeta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\zeta^2/2\sigma^2}. \quad (22)$$

Notice that Eq. (21) will involve a quadruple integral (over say $dA dA'$) which can be recast in terms of the autocorrelation function. Assuming that this is also Gaussian and given by Eq. (17), a rather complicated final expression for the scattering results.

Two special cases are of most interest, however. For a slightly rough surface ($\sigma \ll \lambda$), the scattering is predominantly specular and the directional-conical reflectance is approximately

$$\rho_{spec}(\theta_i, \phi_i; \Delta\Omega_{spec}) = e^{-(2k\sigma \cos\theta_i)^2}, \quad (23)$$

where $\Delta\Omega_{spec}$ spans the specular lobe centered about $\theta_r = \theta_i$ and $\phi_r = \phi_i + \pi$ and is roughly equal to λ^2/A due to diffraction from the finite-sized surface A (which is determined, for example, by the usual product of sinc-squared functions for a rectangular surface). Equation (23) is a well-known result nicely derived from the Fraunhofer diffraction formula by Davies (1954). Bennett and Porteus (1961) have experimentally verified that it gives the specular reflectance, at sufficiently long wavelengths, of a slightly rough surface relative to a smooth surface of the same material. Good agreement was obtained with measurements in the mid-infrared for glass disks roughened with $\sim 10\text{-}\mu\text{m}$ grit, overcoated with aluminum, and characterized by stylus profilometry. Normal incidence was employed, thus minimizing effects such as shadowing and any polarization dependence. These measurements can also be described in terms of the total integrated scatter (TIS), defined as the ratio of the diffusely to the specularly scattered powers. Writing the specularly, diffusely, and total scattered powers as P_{spec} , P_{diff} , and $P_{tot} \equiv P_{spec} + P_{diff}$, respectively, one sees for a slightly rough surface that

$$\text{TIS} = \frac{P_{tot} - P_{spec}}{P_{spec}} = \frac{1 - \rho_{spec}}{\rho_{spec}} \approx (2k\sigma \cos\theta_i)^2, \quad (24)$$

where the second equality presumes that slight roughening merely redistributes the total power scattered by a smooth sample without diminishing it, as follows from the assumption of infinite conductivity. Note that to the level of the approximations used above, the TIS is also equal to $1 - (P_{spec}/P_{tot}) = 1 - (P_{spec}/P_{spec,smooth})$, where, recalling Eq. (5), $P_{spec,smooth} = R(\theta_i)P_i$ is the specularly reflected power by a smooth surface of the same material, thus explaining other definitions of the TIS used in the literature (e.g., Elson and Bennett, 1979b). Church *et al.* (1977) have shown, from the Rayleigh-Rice perturbation theory discussed below, that Eq. (24) holds regardless of the specific form chosen for the height distribution function $p(\zeta)$ in Eq. (22) and hence is not dependent on the assumption of Gaussian statistics—see, for example, the discussion in connection with Eq. (34) below. A particularly nice feature of TIS measurements is that no reference sample is necessary: the diffusely reflected power can be measured using an integrating sphere and one detector, while a small hole (coincident with the input hole for normal

incidence) and a second detector is used to measure the specularly reflected power (Stover, 1990). Also note that Eq. (24) provides a quantitative basis for the Rayleigh criterion for the roughness of a sample, wherein a surface is respectively considered rough or smooth if $4\pi\sigma\cos\theta_i/\lambda$, which equals the phase difference between two specular rays reflected from points separated by a height σ on the surface, is large or small compared to $\pi/2$ or approximately 1 (Beckmann and Spizzichino, 1987).

Equation (23), giving the specular reflectance of a slightly rough surface, is in fact proportional to $|\langle E_r \rangle|^2$. However, the total scattered intensity is more exactly given by $\langle |E_r|^2 \rangle = |\langle E_r \rangle|^2 + \langle |E_r - \langle E_r \rangle|^2 \rangle$. The angular distribution of the diffusely scattered radiation is determined by the second term on the right-hand side of this equality, which defines the variance of the complex scalar amplitude E_r , given in turn by the sum of the variances of its real and imaginary parts. In general, this depends upon the joint probability distribution of finding the height to be ζ_2 at a lag τ away from a point where the height is ζ_1 . However, for a slightly rough surface the integral of interest can be expanded to first order in ζ/λ and σ/λ , with the result that the variance depends only on the autocorrelation function of Eq. (17). The diffuse BRDF is then approximately found to be (Beckmann and Spizzichino, 1987)

$$f_{diff} = \frac{k^4 l^2 \sigma^2}{4\pi} \frac{[1 + \cos\theta_i \cos\theta_r + \sin\theta_i \sin\theta_r \cos(\phi_r - \phi_i)]^2}{\cos\theta_i \cos\theta_r} \times \exp[-k^2 l^2 \{\sin^2\theta_i + \sin^2\theta_r + 2\sin\theta_i \sin\theta_r \cos(\phi_r - \phi_i)\}/4]. \quad (25)$$

Similar results have been obtained by Davies (1954) and Leader (1979) except for the angular prefactor—as discussed by Stover (1990), as well as in a footnote on page 75 and in Appendix A of Beckmann and Spizzichino (1987), the form of this angular term depends on the obliquity factor used, of which considerable variation exists in the literature.

The other important case of Eq. (21) averaged using Eq. (22) is very rough scattering ($\sigma \gg \lambda$), for which the reflection is predominantly diffuse and given by the BRDF

$$f = \frac{1}{\pi m^2} \frac{[1 + \cos\theta_i \cos\theta_r + \sin\theta_i \sin\theta_r \cos(\phi_r - \phi_i)]^2}{\cos\theta_i \cos\theta_r (\cos\theta_i + \cos\theta_r)^4} \times \exp\left[-\frac{\sin^2\theta_i + \sin^2\theta_r + 2\sin\theta_i \sin\theta_r \cos(\phi_r - \phi_i)}{m^2 (\cos\theta_i + \cos\theta_r)^2}\right] \quad (26)$$

where, as defined in Eq. (18), m is the rms surface slope. In Davies (1954), the argument of the exponential is larger by a factor of 2, apparently stemming from a different assumed form for the surface-height joint probability distribution.

It is emphasized again that the tangent-plane approximation inherent in the above physical-optics method implies that the solution will only be correct in the high-optical-frequency limit ($\lambda \ll l$) and in this respect is not more accurate than geometrical optics, although it can be applied to a wider variety of target geometries. A perturbation approach can be used to approximate the scattering at lower frequencies ($\lambda \lesssim l$; obviously at very low frequencies a rough surface will ultimately appear smooth and scatter specularly) and has the added advantage that it is explicitly vector based. However, the relevant perturbation is of the surface height and hence the method only applies to slightly rough samples ($\sigma \ll \lambda$). The basic idea was first proposed by Rayleigh in 1895 and later extended by Rice (1951), so that it is today known as the Rayleigh-Rice method. The height profile $\zeta(x,y)$ is Fourier analyzed into its sinusoidal components $Z(k_x, k_y)$. Any individual component acts like a periodic grating, scattering an incident plane wave into a set of directions given by the 2D grating equations. Assuming the surface is only slightly rough, the diffuse scattering will be dominated by the first-order peaks, so that

$$\begin{aligned} (\mathbf{k}_r - \mathbf{k}_i) \cdot \hat{\mathbf{x}} &= k(\sin\theta_r \cos\phi_r - \sin\theta_i) = k_x \\ (\mathbf{k}_r - \mathbf{k}_i) \cdot \hat{\mathbf{y}} &= k \sin\theta_r \sin\phi_r = k_y \end{aligned} \quad (27)$$

where k_x and k_y can be either positive or negative so as to include all four first-order peaks, and where I have followed the usual convention (e.g., Beckmann and Spizzichino, 1987; Church and Zavada, 1975) of orienting the x -axis so that $\phi_i = -\pi$. Equation (27) can be viewed as an expression of conservation of linear momentum. The total electric field above the surface (which for simplicity is taken to be infinitely conducting) in the case of incident s -polarization (which avoids coupling to surface waves) is then

$$\mathbf{E} = \mathbf{E}_{0i} e^{i\mathbf{k}_i \cdot \mathbf{r}} - \mathbf{E}_{0r} e^{i\mathbf{k}_r(0,0) \cdot \mathbf{r}} - \sum_{k_x, k_y} \mathbf{E}_{0r}(k_x, k_y) e^{i\mathbf{k}_r(k_x, k_y) \cdot \mathbf{r}}, \quad (28)$$

where the 0 subscripts on the electric fields denote the complex amplitudes, suppressing the unimportant time dependences. The first term on the right-hand side of Eq. (28) is the incident field; the second is the specularly reflected field, whose amplitude is equal to that of the incident field to lowest order in the roughness and whose sign arises from the 180° phase shift for external TE reflection from Eq. (6a); and the last term is the diffusely reflected field, given by the sum of the first-order diffraction peaks. (Note that Rice writes $k_x = mK$ and $k_y = nK$, where K is some arbitrarily small fundamental spatial frequency, which in practice can be chosen by requiring $2\pi/K$ to be of the order of the largest experimentally observable spatial wavelength, namely the sample or beam diameter. Furthermore, there are upper limits on the values of k_x and k_y , beyond which the diffracted orders disappear into the surface; for example, one must have $(k_x^2 + k_y^2)^{1/2} \leq k$ at normal incidence, i.e., $\lambda \leq l$. Thus, the above summation can actually be considered to be over a finite set of integers m and n .) Maxwell's equation $\nabla \cdot \mathbf{E} = 0$ requires that each component wave in the scattered field be transverse, $\mathbf{k}_r \cdot \mathbf{E}_{0r} = 0$. Furthermore, electromagnetic boundary conditions must be imposed at the surface. For example, the two tangential components of the total electric field must vanish on the surface, $\hat{\mathbf{n}} \times \mathbf{E}|_{z=\zeta} = 0$, where $\hat{\mathbf{n}}$ is the local surface normal given by

$$\hat{\mathbf{n}} = \frac{(-\partial\zeta/\partial x, -\partial\zeta/\partial y, 1)}{\sqrt{1 + (\partial\zeta/\partial x)^2 + (\partial\zeta/\partial y)^2}}, \quad (29)$$

or approximately $(-\partial\zeta/\partial x, -\partial\zeta/\partial y, 1)$ using the perturbation assumption (which implies the surface slopes are gentle, again consistent with $\lambda \lesssim l$). These first derivatives of the surface profile are given by the Fourier transforms of $k_x Z(k_x, k_y)$ and $k_y Z(k_x, k_y)$. A matrix equation is thus obtained for the Cartesian components of the unknown amplitudes $\mathbf{E}_{0r}$, which (as for $\hat{\mathbf{n}}$) are expanded to lowest nonzero order in ζ . The statistical character of the surface profile is now introduced by assuming that $Z(k_x, k_y)$ is distributed normally about zero and that the variables k_x and k_y are independent. In particular, the ensemble average of the squared modulus of $Z(k_x, k_y)$ defines the power spectral density (PSD),

$$W(k_x, k_y) = \frac{1}{A} \left\langle \left| \iint_A \zeta(x, y) e^{i(k_x x + k_y y)} dx dy \right|^2 \right\rangle \quad (30)$$

where A denotes the area of the sample. (In the present context, the $1/A$ prefactor normalizes the scattered power to the incident power.) According to the Wiener-Khinchin relation, Eq. (30) can also be expressed as the Fourier transform of the autocorrelation function. In polar coordinates, the azimuthal integration yields a Bessel function of order zero for an isotropic surface (Elson and Bennett, 1979b), so that W is only a function of $(k_x^2 + k_y^2)^{1/2}$. Specifically, assuming a Gaussian autocorrelation function as given by Eq. (17), the radial integration can also be performed, giving

$$W(k_x, k_y) = \pi l^2 \sigma^2 e^{-(k_x^2 + k_y^2) l^2 / 4}, \quad (31)$$

while for the exponential surface height correlations of Eq. (19), one obtains

$$W(k_x, k_y) = \frac{2\pi l^2 \sigma^2}{[1 + (k_x^2 + k_y^2) l^2]^{3/2}} \quad (32)$$

which becomes a Lorentzian in the case of 1D roughness, $\zeta(x)$. Note that, strictly speaking,

Eq. (32) cannot be used in the present context because it implies (infinitely) steep surface slopes, as discussed in connection with Eq. (19); however, many researchers have ignored this point and used it anyway, implicitly assuming that the correlation function eventually becomes non-exponential for lags below the range of measurement (which is on the order of λ in optical experiments and of the radius of the stylus when using mechanical profilometry). Specifically, the autocorrelation function must have zero slope at zero lag to be physically meaningful (Elson and Bennett, 1979a).

Elson (1975; also see Elson and Ritchie, 1974) has performed an analysis which is essentially equivalent to the preceding Rayleigh-Rice vector theory, except that he has used the full electromagnetic boundary conditions rather than assuming infinite conductivity. The wave equations were expanded to first order in $\zeta(x,y)$ and solved by Green function methods. His oft-quoted result for the diffuse BRDF is

$$f_{diff} = \frac{k^4}{\pi^2} \cdot \cos \theta_i \cos \theta_r \cdot Q \cdot W(k_x, k_y) \quad (33)$$

assuming that $m \ll 1$. If W is given by Eq. (31), this result reduces to Eq. (25) provided that the form of Q applicable to an infinitely conducting surface is adopted and that the polarization factor of Eq. (36) is accounted for. Note that some authors (e.g., Church *et al.*, 1977) employ the symmetric form of the Fourier transform in Eq. (30), and hence the first term on the right becomes $4k^4$ instead. Here, k^4 is the usual "Rayleigh blue-sky" factor; the angles comprise an obliquity factor and have a significant effect near grazing; Q , which is proportional to the scattering matrix elements, is called the optical factor and depends on the relevant material and polarization properties—it has been tabulated for a variety of cases of interest by Barrick (1970) and Church *et al.* (1977), reducing to the Fresnel reflectances R of Eq. (6) for small-angle (i.e., near specular) scattering, and should be summed (averaged) for a polarization-insensitive detector (unpolarized source); finally, in the present context $W(k_x, k_y)$ from Eq. (30) is called the surface factor (Elson and Bennett, 1979a). Since each Fourier component of the surface diffracts the incident light into a unique set of first-order directions, it makes intuitive sense that there is a one-to-one mapping between the angular distribution of the scattered light (as given by f) and the power spectral density of the roughness profile (as given by W). Specifically, consider a 1D sinusoidal grating given by $z(x) = 2^{1/2} \sigma \sin(Kx)$. Then the PSD is proportional to a product of two Dirac delta functions,

$$W(k_x, k_y) = 2\pi^2 \sigma^2 \delta(k_x \pm K) \delta(k_y), \quad (34)$$

where $k_x = \pm K$ gives the two first-order peaks. Using this to evaluate f in Eq. (33), assuming small-angle scattering [which by Eq. (27) implies $K/k \ll \sin \theta_i$, i.e., a low-spatial-frequency grating and an angle of incidence not too far off-normal], integrating it with respect to $\cos \theta_r d\Omega_r$ over the outgoing hemisphere (implicitly excluding the specular beam), and dividing that by $R(\theta_i)$ gives the TIS, which agrees with Eq. (24). This result can also be written, to lowest order in σ , as $TIS = 2p$, where p is called the diffraction efficiency and is defined as the ratio of the power diffracted into the +1 and 0 orders (Stover, 1975; Church *et al.*, 1977), with the factor of 2 arising from the fact that the same amount of power is diffracted into the +1 and -1 orders in the small-angle-scattering limit.

Equation (33) has been compared to experimental measurements (Germer *et al.*, 1997) on a fabricated silicon microroughness standard having a pseudorandom distribution of circular pits (with $\sim 1\text{-}\mu\text{m}$ diameters and 1-nm depths) using a cw doubled-Nd:YAG laser. Excellent agreement resulted for an incident angle of $\theta_i = 45^\circ$, reflected polar angles of $\theta_r = 30^\circ, 45^\circ$, and 60° and azimuthal angles of $\phi_r = 10^\circ\text{--}170^\circ$, and the four polarization combinations *ss*, *sp*, *ps*, and *pp*. An interesting feature of the data is that the *pp* scattering vanishes for certain out-of-plane angles (e.g., $\phi_r \approx 60^\circ$ for $\theta_r = 45^\circ$) which are the bidirectional analogs of the Brewster angle. Furthermore, ellipsometry indicates that, for incident linearly polarized light, the out-of-plane (lateral) scattered light remains strongly linearly polarized, albeit with a rotated plane of vibration in general. Both of these effects can be used to probe other scattering mechanisms,

such as those due to surface particles or bulk defects (Germer, 1997), whose signals might otherwise be masked by the microroughness surface scattering.

The optical factor Q describes the polarization of the reflected beam relative to that of the incident beam and can be considered a generalization of the Fresnel reflectances of Eq. (6). In particular, the cross-polarization terms Q_{sp} and Q_{ps} are zero for scattering into the plane of incidence. For the common case of in-plane (longitudinal) ss scattering, it is convenient to note (Stover, 1990) that the optical factor is simply the geometric mean of $R_s(\theta_i)$ and $R_s(\theta_r)$. If the incident light is polarized but neither s nor p (i.e., linearly polarized at an oblique angle, circularly polarized, or elliptically polarized), then for both the specular and diffuse components even the singly scattered light is depolarized. To put this on a more precise footing, the polarization factor p of a wave is defined as E_{0p}/E_{0s} , where the complex amplitude E_0 has been resolved into its vertical and horizontal components (Beckmann and Spizzichino, 1987). I choose to follow the sign convention wherein any scalar component of the field is written as

$$E = |E_0| e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t + \epsilon)} = E_0 e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}, \quad (35)$$

where $\omega = ck$ in vacuum, while some authors choose the opposite signs for $\mathbf{k} \cdot \mathbf{r}$ and ωt . Accordingly, $\text{Im}\{p\} = 0$ implies linear polarization (e.g., $p = 0$ is horizontal and $p = \infty$ is vertical), while $\text{Im}\{p\}$ greater or less than 0 refers to left-handed or right-handed polarization, respectively (e.g., $p = i$ is left circular and $p = -i$ is right circular). If the polarization of the incident wave is described by p_i and that of the reflected wave by p_r , then the (complex) depolarization factor q is defined by the relation $p_r = qp_i$; specifically, $q = 1$ implies no depolarization, while $q = 0$ or ∞ corresponds to a filter, scattering only one component of the incident light, as occurs at the Brewster angle for example. [Note carefully that depolarization here refers to a change in the *state* of polarization of a beam, not in its *degree* of polarization. Both the incident and scattered waves are assumed to be fully polarized. However, many workers (e.g., Renau *et al.*, 1967) use the term to refer instead to an increase in the randomly polarized fraction of a beam.] For an ideal specular reflector, $q = -1$, so that an infinite, flat, perfect conductor will depolarize all incident polarizations except s and p , changing both the orientation of the polarization ellipse and its handedness; this same result holds for longitudinal scattering by a rough surface (of infinite conductivity) if the tangent-plane approximation holds (i.e., $l \gg \lambda$). If a smooth plane has finite conductivity instead, then $q = r_p(\theta_i)/r_s(\theta_i)$ and one sees that an obliquely linearly polarized beam (i.e., neither s nor p) incident at an oblique angle (i.e., neither normal to nor grazing the surface) gives rise to a (longitudinal) specularly reflected beam which in the case of a dielectric remains linearly polarized but with a rotated plane of vibration, and which in the case of a metal is elliptically polarized (Jenkins and White, 1957); in contrast, $q = -1$ at normal incidence and $+1$ at grazing incidence.

In the case of lateral scattering, even incident s or p light is depolarized. Consider, for example, a vertically polarized wave at near grazing incidence scattering off a perfectly conducting surface element tilted upwards about the x axis by 45° . Resolving the incident field into tangential and normal components and applying the boundary conditions, the reflected wave is found to be horizontally polarized. This effect is responsible for the depolarization of television waves by tin rooves in cities. The general result for specular reflection off facets of finite conductivity is shown from spherical geometry (Beckmann and Spizzichino, 1987) to be

$$p_r = \frac{p_i(r_s \tan \beta_1 \tan \beta_2 + r_p) + (r_p - r_s) \tan \beta_1}{r_s + r_p \tan \beta_1 \tan \beta_2 - p_i(r_s \tan \beta_2 - r_p \tan \beta_1)}, \quad (36a)$$

where

$$\beta_1 \equiv \sin^{-1} \left[\frac{\sin \theta_i \sin \phi_r}{\sqrt{1 - (\cos \theta_i \cos \theta_r - \sin \theta_i \sin \theta_r \cos \phi_r)^2}} \right] \quad (36b)$$

and

$$\beta_2 \equiv \cos^{-1} [\cos \beta_1 \cos \phi_r - \sin \beta_1 \cos \theta_i \sin \phi_r], \quad (36c)$$

with $\phi_i = -\pi$, as in Eq. (27). It is easy to check that for longitudinal scattering ($\phi_r = 0$), one gets $\beta_1 = \beta_2 = 0$, so that $q = r_p/r_s$ as above. These expressions are already rather complicated, without even considering the depolarization of waves diffracted into non-specular directions, as occurs in the Rayleigh-Rice model for example. Note that knowledge of the optical factors is not sufficient to describe scattering from and to states of arbitrary polarization, because the four Q values involve intensity rather than amplitude ratios and hence the necessary phase information is not available. For example, a beam linearly polarized at 45° to the vertical and a circularly polarized wave both have equal-intensity s and p components and hence cannot be distinguished on that basis alone. Unfortunately, no general analytical solution to the problem of depolarization by a rough surface exists.

To circumvent this situation and the limitations inherent in the Kirchhoff method (i.e., the tangent-plane approximation $l \gg \lambda$ and the assumption of infinite conductivity) or in the Rayleigh-Rice theory (namely, the assumption of slight roughness $\sigma \ll \lambda$ and gentle slopes $\lambda \leq l$), direct numerical solutions of the electromagnetic equations have been undertaken. In essence, one starts from a vector form of the Kirchhoff integral (Jackson, 1975)—also known as the Chu-Stratton integrals (Ruck *et al.*, 1970)—for the electric and magnetic fields, $\mathbf{E}$ and $\mathbf{H}$, which express the total fields (or more simply, the scattered fields if the incident fields are subtracted off) at any point in space in terms of their values on the sample surface and a free-space Green function. In turn, the surface fields can be related to the induced surface current density $\mathbf{J}$. Finally, $\mathbf{J}$ can be obtained from a similar integral by relating the scattered fields on the surface to the incident fields via the boundary conditions, taking care to handle the singularity when the argument of the Green function is zero (i.e., when the source point is identified with the field point). These integral equations can also be obtained more directly from the optical extinction theorem (Soto-Crespo and Nieto-Vesperinas, 1989). To solve them numerically, the finite surface is chopped into a discrete number of sampling points, say N in all. The elements of surface area, $\sec\theta_n dx dy$, and the surface-normal unit vectors $\hat{\mathbf{n}}$ (where $\hat{\mathbf{n}} \cdot \hat{\mathbf{z}} \equiv \cos\theta_n$) are determined by $\zeta(x,y)$ from Eq. (29). The integrals expressing the boundary conditions can then be written as two $N \times N$ matrix equations for the N values of $\mathbf{E}_i$ and $\mathbf{H}_i$ on the surface in terms of the N values of the two tangential components of $\mathbf{J}$. These equations are inverted to give $\mathbf{J}$, which is inserted in turn into the finite approximations for the integrals for $\mathbf{E}_r$ and $\mathbf{H}_r$, completing the solution. This is known as the point-matching method. The N sampled values of the surface roughness profile are generated by a Monte Carlo procedure constructed to give zero mean with the desired root variance σ and autocorrelation function $A(\tau)$. The scattered power is proportional to $\langle |\mathbf{E}_r|^2 \rangle$, where the average is over an ensemble of such sequences of sampled values.

For example, Soto-Crespo and Nieto-Vesperinas (1989) have computed the scattering from 1D infinitely-conducting random (Gaussian-correlated) or periodic surfaces. For this geometry, the Green function is a zeroth-order Hankel function of the first kind. Interestingly, there is an intermediate range of roughness (between the regimes of mostly specular scattering and enhanced backscattering at small and large values of σ/λ , respectively) for which the reflection is very nearly Lambertian, at least for near-normal angles of incidence. Furthermore, these researchers compared their numerical results to the analytical predictions of the Kirchhoff approximation and were thus able to prepare a useful graph indicating the range of values of σ/l versus l/λ for given angles of incidence for which the two integrated scattered powers agree with each other to within a specified percentage. Saillard and Maystre (1990) have considered 1D random surfaces of finite but large conductivity, applicable for instance to gold at infrared (IR) wavelengths. In this case, the kernels of the integral equations are sharply peaked and hence difficult to accurately evaluate by the above numerical procedures. So instead, an "impedance boundary condition" is imposed, whereby a local linear relationship between the total field in the metal and its normal derivative on the surface is assumed. Knotts *et al.* (1993) have numerically evaluated the complete Mueller matrix for these same conditions and compared the results to experimental measurements on well-defined samples, prepared by overcoating a photoresist plate exposed to a suitable 1D laser pattern. Qualitative agreement resulted when Gaussian statistics

were employed, but detailed agreement required that the sequences of surface-height values be taken directly from digitized profilometer data rather than be generated by the Monte Carlo technique, indicating that rather subtle statistical properties of a surface can play a significant role in determining the scattering. This emphasizes the most important limitation of such numerical techniques: they must be repeated from scratch for each specific set of sample properties and illumination characteristics.

Experimental Measurements of Surface BRDFs

In this section, an assortment of experimental papers are briefly discussed, but no attempt at completeness is made—for this purpose, reviews such as the one by Asmail (1991) should be consulted. It is logical to begin with the reflectometers used to measure BRDFs in the laboratory. Typical examples are described by Roche and Pelletier (1984), Zaworski *et al.* (1993), and Sandmeier *et al.* (1997). At their hearts are a goniometer which permits one to choose values of θ_i , ϕ_i , θ_r , and/or ϕ_r in certain ranges, to within specified angular resolutions and source and detector solid angles. Two common design geometries are: type (i) where the source and detector are each mounted on a pair of arc rails so that they can be scanned throughout the hemisphere above a stationary sample; and type (ii) for which the source is fixed in place, the sample rotates about the x and y axes defining its surface, and the detector orbits around the sample in a single plane. Various choices for the collimation or focusing of the incident light are discussed in Nicodemus *et al.* (1977). Sometimes the sample is rapidly spun about its normal direction to average out inhomogeneities such as laser speckle, and the incident beam chopped so that lock-in detection can be used to eliminate scattering by ambient light. Oppenheim *et al.* (1994) have made measurements on various recommended standards of diffuse (nearly Lambertian) reflectance in the IR used in calibrating such instruments.

Moving on to applications in the domains of interest outlined in the introduction, Snyder and Wan (1996) have used a Fourier-transform IR spectrometer mounted on a type (i) goniometer to investigate soil samples. It proved necessary to make measurements both with the source on and off, in order to correct for sample heating by it. The ultimate goal is to determine the BRDF for analysis of satellite-based thermometric imaging of earth's surface. Stavridi *et al.* (1997) have examined brick, tile, and concrete with a CCD camera in a type (ii) geometry for architectural applications. As is typical of many experimental papers, these researchers have fitted their data by combining several models discussed in the previous section. Specifically, they chose a linear combination of the Torrance-Sparrow and Oren-Nayar facet models. There is no theoretical justification for such *ad hoc* combinations, and so they are probably best viewed as empirical fitting functions which work because they contain a sufficient number of free parameters. Newell and Keski-Kuha (1996) have looked at extreme-ultraviolet (UV) scatterers and baffles of interest to astronomers. Perhaps not surprisingly, materials which perform well in the IR and visible spectral ranges often make poor choices for the UV. Similarly, Watkins *et al.* (1993) made s - and p -polarized measurements of high-reflectance dielectric coatings using laser sources and an InGaAs or photomultiplier detector in a type (ii) configuration. Superpolishing of the substrates can reduce the BRDF, but generally only at the design wavelength—scatter characteristics at different wavelengths tend to be uncorrelated. Bickel *et al.* (1987; also see Iafelice and Bickel, 1987) have measured the complete Mueller matrices for smooth and distorted metal surfaces by periodically modulating the incident polarization state from a laser and measuring the fundamental and second harmonic of the scattered light in a type (ii) setup; this demonstrates the sensitivity of polarimetric optical scattering to surface perturbations. An alternative technique for finding the Mueller matrix uses a fixed linear polarizer and a variable retarder in both the source and receiver arms (Sornsin and Chipman, 1996); an integrating sphere scrambles the analyzed state to compensate for any polarization sensitivity of the detector. Burnell *et al.* (1994) have studied oxidized nickel collected from the interior of furnace tubes for the purposes of distinguishing emitted and reflected radiation in pyrometric measurements. After finding that the data could not be fit with physically realistic values of the parameters in the Torrance-Sparrow

model, they also modified it in an *ad hoc* fashion to give a function which worked better but could not be interpreted physically. This is reminiscent of the approach used to devise the coupled model (Shirley *et al.*, 1997) or the (similar) modified Beard-Maxwell model (Ellis, 1996a) for glossy surfaces.

References

- C. Asmail (1991), "Bidirectional scattering distribution function (BSDF): A systematized bibliography," J. Res. Natl. Inst. Stand. Technol. **96**, 215.
- D.E. Barrick (1970), "Rough Surfaces" in *Radar Cross Section Handbook*, edited by G.T. Ruck (Plenum Press, NY), Vol. 2.
- P. Beckmann and A. Spizzichino (1987), *The Scattering of Electromagnetic Waves from Rough Surfaces* (Artech House, Norwood MA).
- H.E. Bennett and J.O. Porteus (1961), "Relation between surface roughness and specular reflectance at normal incidence," J. Opt. Soc. Am. **51**, 123.
- W.S. Bickel, R.R. Zito, and V.J. Iafelice (1987), "Polarized light scattering from metal surfaces," J. Appl. Phys. **61**, 5392.
- M. Born and E. Wolf (1965), *Principles of Optics*, 3rd ed. (Pergamon, London).
- J.G. Burnell, J.V. Nicholas, and D.R. White (1994), "Scattering model for rough oxidized metal surfaces applicable to radiation thermometry of reformer furnaces," Opt. Eng. **34**, 1749.
- E.L. Church and J.M. Zavada (1975), "Residual surface roughness of diamond-turned optics," Appl. Opt. **14**, 1788.
- E.L. Church, H.A. Jenkinson, and J.M. Zavada (1977), "Measurement of the finish of diamond-turned metal surfaces by differential light scattering," Opt. Eng. **16**, 360.
- F.J.J. Clarke and D.J. Parry (1985), "Helmholtz reciprocity: Its validity and application to reflectometry," Light. Res. & Tech. **17**, 1.
- H. Davies (1954), "The reflection of electromagnetic waves from a rough surface," Proc. I.E.E. **101**, 209.
- W.G. Egan and T. Hilgeman (1976), "Retroreflectance measurements of photometric standards and coatings," Appl. Opt. **15**, 1845.
- K.K. Ellis (1994), "Reflectance phenomenology and modeling tutorial," <http://www.erim.org/on-line-docs/GUIDE/guide.frm.html>.
- K.K. Ellis (1996a), "Beard-Maxwell BRDF model improvements for polarized signature calculations," Proceedings of the Workshop on Infrared and Millimeter Wave Polarimetry.
- K.K. Ellis (1996b), "Polarimetric bidirectional reflectance distribution function of glossy coatings," J. Opt. Soc. Am. A **13**, 1758.
- J.M. Elson and R.H. Ritchie (1974), "Diffuse scattering and surface-plasmon generation by photons at a rough dielectric surface," Phys. Stat. Sol. (b) **62**, 461.
- J.M. Elson (1975), "Light scattering from semi-infinite media for non-normal incidence," Phys. Rev. B **12**, 2541.
- J.M. Elson and J.M. Bennett (1979a), "Relation between the angular dependence of scattering and the statistical properties of optical surfaces," J. Opt. Soc. Am. **69**, 31.
- J.M. Elson and J.M. Bennett (1979b), "Vector scattering theory," Opt. Eng. **18**, 116.
- D.S. Flynn and C. Alexander (1995), "Polarized surface scattering expressed in terms of a bidirectional reflectance distribution function matrix," Opt. Eng. **34**, 1646.
- T.A. Germer (1997), "Angular dependence and polarization of out-of-plane optical scattering from particulate contamination, subsurface defects, and surface microroughness," Appl. Opt. **36**, 8798.
- T.A. Germer, C.C. Asmail, and B.W. Scheer (1997), "Polarization of out-of-plane scattering from microrough silicon," Opt. Lett. **22**, 1284.
- B.W. Hapke (1963), "A theoretical photometric function for the lunar surface," J. Geophys. Res. **68**, 4571.
- B.W. Hapke, R.M. Nelson, and W.D. Smythe (1993), "The opposition effect of the moon: The contribution of coherent backscatter," Science **260**, 509.
- E. Hecht (1998), *Optics*, 3rd ed. (Addison Wesley, Reading MA).
- V.J. Iafelice and W.S. Bickel (1987), "Polarized light-scattering matrix elements for select perfect and perturbed optical surfaces," Appl. Opt. **26**, 2410.
- J.D. Jackson (1975), *Classical Electrodynamics*, 2nd ed. (John Wiley, NY).
- F.A. Jenkins and H.E. White (1957), *Fundamentals of Optics*, 3rd ed. (McGraw-Hill, NY).
- D.B. Judd (1967), "Terms, definitions, and symbols in reflectometry," J. Opt. Soc. Am. **57**, 445.
- K. Kim, L. Mandel, and E. Wolf (1987), "Relationship between Jones and Mueller matrices for random media," J. Opt. Soc. Am. A **4**, 433.
- M.E. Knotts, T.R. Michel, and K.A. O'Donnell (1993), "Comparisons of theory and experiment in light scattering from a randomly rough surface," J. Opt. Soc. Am. A **10**, 928.
- K.-T. Kriebel (1996), "On the limited validity of reciprocity in measured BRDFs," Remote Sens. Environ. **58**, 52.
- J.C. Leader (1979), "Analysis and prediction of laser scattering from rough-surface materials," J. Opt. Soc. Am. **69**, 610.

- J.R. Maxwell and S.F. Weiner (1974), "Polarized emittance. Volume I: Polarized bidirectional reflectance with Lambertian or non-Lambertian diffuse components," NTIS Report No. AD-782178.
- M.P. Newell and R.A.M. Keski-Kuha (1996), "Bidirectional reflectance distribution function of diffuse extreme ultraviolet scatterers and extreme ultraviolet baffle materials," *Appl. Opt.* **36**, 5471.
- F.E. Nicodemus (1965), "Directional reflectance and emissivity of an opaque surface," *Appl. Opt.* **4**, 767.
- F.E. Nicodemus (1970), "Reflectance nomenclature and directional reflectance and emissivity," *Appl. Opt.* **9**, 1474.
- F.E. Nicodemus, J.C. Richmond, J.J. Hsia, I.W. Ginsberg, and T. Limperis (1977), "Geometrical considerations and nomenclature for reflectance," NBS Monograph 160.
- K.A. O'Donnell and E.R. Mendez (1987), "Experimental study of scattering from characterized random surfaces," *J. Opt. Soc. Am. A* **4**, 1194.
- U.P. Oppenheim, M.G. Turner, and W.L. Wolfe (1994), "BRDF reference standards for the infrared," *Infrared Phys. Technol.* **35**, 873.
- M. Oren and S.K. Nayar (1995), "Generalization of the Lambertian model and implications for machine vision," *Int. J. Comput. Vision* **14**, 227.
- J. Renau, P.K. Cheo, and H.G. Cooper (1967), "Depolarization of linearly polarized EM waves backscattered from rough metals and inhomogeneous dielectrics," *J. Opt. Soc. Am.* **57**, 459.
- S.O. Rice (1951), "Reflection of electromagnetic waves from slightly rough surfaces," *Commun. Pure Appl. Math.* **4**, 351.
- P. Roche and E. Pelletier (1984), "Characterizations of optical surfaces by measurement of scattering distribution," *Appl. Opt.* **23**, 3561.
- G.T. Ruck, D.E. Barrick, and W.D. Stuart (1970), "Theory" in *Radar Cross Section Handbook*, edited by G.T. Ruck (Plenum Press, NY), Vol. 1.
- M. Saillard and D. Maystre (1990), "Scattering from metallic and dielectric rough surfaces," *J. Opt. Soc. Am. A* **7**, 982.
- S. Sandmeier, C. Müller, B. Hosgood, and G. Andreoli (1997), "Sensitivity studies of bidirectional reflectance data using the EGO/JRC goniometer facility," International ISPRS Symposium on Physical Measurements and Signatures in Remote Sensing.
- O. Sandus (1965), "A review of emission polarization," *Appl. Opt.* **4**, 1634.
- P. Shirley, H. Hu, B. Smits, E. Lafortune (1997), "A practitioners' assessment of light reflection models," Proceedings of the Pacific Graphics Conference.
- W.C. Snyder and Z. Wan (1996), "Surface temperature correction for active infrared reflectance measurements of natural materials," *Appl. Opt.* **35**, 2216.
- W.C. Snyder (1998), "Reciprocity of the bidirectional reflectance distribution function (BRDF) in measurements and models of structured surfaces," *IEEE Trans. Geosci. Remote Sensing* **36**, 685.
- W.C. Snyder and Z. Wan (1998), "BRDF models to predict spectral reflectance and emissivity in the thermal infrared," *IEEE Trans. Geosci. Remote Sensing* **36**, 214.
- E.A. Sornsin and R.A. Chipman (1996), "Polarization BRDF of satellite materials," Proceedings of the Workshop on Infrared and Millimeter Wave Polarimetry.
- J.M. Soto-Crespo and M. Nieto-Vesperinas (1989), "Electromagnetic scattering from very rough random surfaces and deep reflection gratings," *J. Opt. Soc. Am. A* **6**, 367.
- M. Stavridi, B. van Ginneken, J.J. Koenderink (1997), "Surface bidirectional reflection distribution function and the texture of bricks and tiles," *Appl. Opt.* **36**, 3717.
- J.C. Stover (1975), "Roughness characterization of smooth machined surfaces by light scattering," *Appl. Opt.* **14**, 1796.
- J.C. Stover (1990), *Optical Scattering: Measurement and Analysis* (McGraw-Hill, NY).
- C.C. Sung and W.D. Eberhardt (1978), "Explanation of the experimental results of light backscattered from a very rough surface," *J. Opt. Soc. Am.* **68**, 323.
- K.E. Torrance and E.M. Sparrow (1967), "Theory for off-specular reflection from roughened surfaces," *J. Opt. Soc. Am.* **57**, 1105.
- H.C. Van de Hulst (1957), *Light Scattering by Small Particles* (John Wiley, NY).
- G. Videen, J.-Y. Hsu, W.S. Bickel, and W.L. Wolfe (1992), "Polarized light scattered from rough surfaces," *J. Opt. Soc. Am. A* **9**, 1111.
- S.E. Watkins, J.P. Black, and B.J. Pond (1993), "Optical scatter characteristics of high-reflectance dielectric coatings and fused-silica substrates," *Appl. Opt.* **32**, 5511.
- P.-E. Wolf and G. Maret (1985), "Weak localization and coherent backscattering of photons in disordered media," *Phys. Rev. Lett.* **55**, 2696.
- J.R. Zaworski, J.R. Welty, and M.K. Drost (1993), "Measurement techniques for bidirectional reflectance of engineering materials" in *Radiative Heat Transfer: Theory and Applications*, ASME Vol. 244.

CHARACTERIZATION OF MICROSTRUCTURE EVOLUTION IN PITCH-BASED CARBON FIBERS DURING HEAT TREATMENT

Amod A. Ogale
Professor
Department of Chemical Engineering

Clemson University
209 Earle Hall
Clemson, SC 29634-0909

Final Report for:
Summer Faculty Research Program
Wright Patterson Air Force Research Labs

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base

and

Wright Patterson Air Force Research Labs

August 1998

CHARACTERIZATION OF MICROSTRUCTURE EVOLUTION IN PITCH-BASED CARBON FIBERS DURING HEAT TREATMENT

Amod A. Ogale
Professor
Department of Chemical Engineering
Clemson University

Abstract

The evolution of structure and mechanical properties with various heat treatment conditions was investigated in this study for pitch-based carbon fibers. AR-mesophase and SCF-mesophase were obtained in the form of pitch fibers. The fibers were first stabilized to render them infusible and then subjected to heat treatment temperatures (HTT) that ranged from 300°C to 3000°C.

The unique aspect of the present study was the characterization of dimensional changes of individual fibers in terms of their length and diameter. The dimensional measurements for AR fibers indicate very clearly that starting from an oxidized state, the length of the fibers shrinks about 8% at a HTT of 900°C. Above 900°C, the length does not reduce any further, instead it increases slightly. The final shrinkage of the length (relative to the oxidized state) is about 6%. The slight increase can be explained by the alignment of the graphene layer planes along the fiber axis that result in the shrinkage of fibers in the transverse direction (diameter) but an expansion along the longitudinal direction. Considerably larger changes are observed in the diameter of the fibers. Starting from the oxidized state, a significant drop of 15% is observed till 900°C HTT. From 900°C till 1500°C, there appears to be a slight increase in the diameter. Beyond 1500°C, there is again a significant reduction of the diameter with the ultimate shrinkage being about 20% at HTT approaching 3000°C.

The fiber microstructure was characterized by scanning electron microscopy (SEM). Consistent with the dimensional measurements, the SEM examinations reveal that the first major change is observed at 900°C, where a radial texture is observed. At higher HTT, the development of radial texture is more pronounced and by 2400°C the graphene-layer planes are seen very clearly. Wide-angle x-ray scattering (WAXS) measurements indicate that at 2875°C HTT, the fibers have a d_{002} spacing of 0.3354 nm indicating that the material is almost fully graphitic. Single-filament testing indicates that the tensile modulus and strength increase with increasing HTT till 2100°C, where a slight drop is observed followed by further increase till the maximum HTT of 2875°C. The maximum modulus was measured to be about 385 GPa whereas the maximum strength was found to be about 1.8 GPa.

The present study establishes that in AR mesophase pitches, the first significant change in dimensions and microstructure occurs at about 900°C. Whereas microstructure and properties continue to evolve over higher heat treatment temperatures, the next significant change is observed above 2400°C. These results should help in the establishment of optimum processing conditions for carbon fiber/carbon matrix composites.

CHARACTERIZATION OF MICROSTRUCTURE EVOLUTION IN PITCH-BASED CARBON FIBERS DURING HEAT TREATMENT

Amod A. Ogale

Introduction

Carbon fibers/carbon composites have unique thermal and mechanical properties [1-3]. However, their high processing costs have limited their use to such applications where performance is the critical controlling factor [4]. The high costs have prevented their use in a number of military and civilian applications.

One approach to reducing processing costs is the simultaneous carbonization of the "green" fibers and the "green" matrix. This approach eliminates one expensive carbonization step, because this approach combines two carbonization steps into one. The science of simultaneous carbonization of fibers and matrix, however, is not well understood.

The processing and characterization of composites has been a subject of great interest. Several techniques have been proposed to monitor the structure and the stresses generated during processing due to a mismatch of thermal expansion coefficients of the various phases as well as the chemical changes taking place in the fiber/matrix. In carbon-carbon composites, the stresses are primarily caused by the shrinkage of the precursor material when noncarbon constituents (oxygen, nitrogen, CO, CO₂) are evolved during carbonization and further heat treatment steps. This shrinkage results in matrix/fiber cracks and leads to deterioration of physical and mechanical properties.

The microstructure and properties of pitch-based fibers have been investigated by a number of researchers [5-10]. However, only a limited number of studies have dealt with dimensional changes [11,12].

Objectives

To optimize processing conditions, it is important to develop a clear understanding of the dimensional and microstructural changes taking place in the mesophase/precursor during the various heat treatment steps. Therefore, this study was directed at the characterization of the evolution

of microstructure in pitch-based fibers for various heat treatment conditions. Specific objectives were:

- (i) to quantitatively measure the dimensional changes in the fiber;
- (ii) to characterize the fiber; and
- (iii) measure the tensile properties of the fiber to obtain a relationship between dimensional changes, microstructure evolution, and mechanical properties.

Experimental

Materials

All of the studies were conducted on pitch-based carbon fibers. Two grades of pitches were investigated: AR-mesophase, provided by Mitsubishi Gas Chemical Company, Japan, and mesophase fraction separated by supercritical fluid (SCF) extraction. The SCF precursor was provided courtesy of Prof. Mark Thies, and the pitch fibers were spun courtesy of Prof. Dan Edie, both at Clemson University.

Processing

Stabilization

The as-spun pitch fibers were first stabilized by an air-oxidation process. In this process, the fibers are heated to a temperature below their softening point in the presence of air. Whereas thin (10 micrometer diameter) AR pitch fibers can be stabilized in a temperature range of 180°C to 280°C within a few hours, the fibers in the present study, about 50-100 μm , had to be oxidized for about 3 days at 200°C. The thicker fibers were needed in the present study to facilitate the accurate measurement of length and diameter for the dimensional study (described in the next subsection). The longer stabilization time is consistent with the results of Singer and Mitchell [13] who showed that the stabilization process is controlled by the diffusion of oxygen in the thick fibers, and that the stabilization time required for 100 μm diameter fibers is more than an order of magnitude larger than that required for 10 μm fibers. The SCF fibers were stabilized in two steps: 12 hrs at 180°C followed by a 10°C/min heating to 280°C, a holding time of approximately 17 hours, and a cooldown at 10°C/min. At the end of the stabilization cycle, the infusible state of the fibers was confirmed by contacting the fibers with a soldering iron heated to 400°C.

Carbonization/Graphitization

The stabilized fibers were heated to a final temperature ($T_{\max}$) ranging from 300°C to 3000°C in increments of 300°C (10 sets) at a nominal heating rate of 3°C/min. Heat treatment temperatures up to 900°C were attained in a Mellen furnace (Webster, New Hampshire), whereas those above 900°C were attained in a graphitizing furnace (Materials Research Furnace Inc., NH).

Characterization

Dimensional

The first approach that was explored was the *in situ* measurement of length changes taking place in a single fiber during heat treatment. A preliminary design was conceived for a fixture that could be mounted in a microbalance. One end of the fiber would be attached to the sapphire rod of the instrument, whereas the other end would be held rigidly in the carbon/carbon composite fixture. However, the major problem with the fixture was its inability to grip the fiber ends without crushing them. Modifications to the fixture would have required considerable machining effort. Since a machine shop was not available at the AFRL Labs, these modifications will be attempted in a follow-up project at Clemson University. However, another approach was adopted that was *ex situ*. The dimensions were measured by a Nikon Microphot-FXL optical microscope, both before and after the heat treatment cycle. The fibers were placed individually on glass slides and the length and diameter were measured with the aid of a Microcode indexer with a resolution of 1 micrometer. For heat treatment, the fibers were placed on a graphite plate (approx. 2 ft diameter) and the position of each fiber was carefully noted. After heat treatment in the graphitization furnace, the fibers were carefully removed from the graphite plate and placed on the glass slides.

Microstructural

The microstructure was characterized by a scanning electron microscope (Personal SEM, RJ Lee Instruments, Trafford, PA). High magnification images were obtained for a limited number of samples using a Hitachi S900 high resolution microscope. The crystalline content and structure (d_{002} spacing, L_c , and L_a) of fibers processed at 2875°C were measured by wide angle X-ray scattering (WAXS) using a Huber Instrument.

Tensile Properties

The tensile modulus, strength, and strain-to-failure of fibers were measured by conducting single-fiber tests in a Sintech tensile testing machine. The gage length was 25.4 mm and the extension rate was 0.0508 mm/min. To obtain statistically significant values, the number of samples required for single-fiber testing is fairly large, often exceeding 20 replicates for a given experimental condition. Since time constraints did not allow for the testing of such large number of samples during this summer project, about 5 samples were tested for each condition. The larger set will be tested in a follow-up project at Clemson University.

Results and Discussion

Dimensional Changes

Since carbonization involves the removal of noncarbon species from the organic precursors, significant dimensional changes (typically shrinkage) take place during processing. One of the primary objectives of the present study was the careful measurement of such changes accompanying various heat treatment conditions.

Figure 1 displays a calibration plot of the readings from the Microcode indexer plotted against the markings from a Bosch & Lomb calibration slide. The Microcode measurement were accurate to within ± 1 μm . Next, Table 1 displays the length measurements for four different stabilized AR fibers. The fiber lengths used in the present study ranged from about 10 mm up to about 40 mm. Different lengths were purposely used to enable unambiguous identification of the fibers as they were placed on and removed from the graphite plate. For a given conformation of the fiber, the length could be measured consistently with an error of no more than ± 10 μm . The diameter measurements could be performed consistently to within ± 1 μm . Repeated measurements of fiber lengths (after removing and replacing the fibers on the glass slide) have an error of no more than 0.2%, with much of the error arising from the waviness of fibers.

After heat treatment, about half of the AR fibers remained fairly straight, whereas the other half curled slightly. In most cases the curvature was small and the length could be accurately determined by adding the length of 5 to 6 individual segments. For SCF fibers, the extent of curling was significant, the reason being their diameter was only about 30-50 μm as compared with 50-100 μm for AR fibers. The ratio of the fiber length before and after heat treatment is reported as "L/Lo".

Figure 2 displays the length ratio as a function of the heat treatment temperatures for AR and SCF fibers. As stated above, the SCF fiber tended to curl up during the heat treatment and as a result the length measurement had considerable variations; the changes in the length ratio from one temperature to the next are statistically not significant at 95% confidence level. Therefore, the results should be treated as semiquantitative trends. However, for AR fibers the measurements were quite accurate. Starting from the stabilized state, the length of the fibers shrinks about 5% at a HTT of 600°C and about 8% at a HTT of 900°C. From 900°C to 1200°C, there appears to be a slight increase in the length, but this difference is not statistically significant at the 95% confidence level. A summary of the statistics is presented in Table 2. Beyond 1200°C, the length does not shrink, instead it increases slightly. The increase from 1200°C to 1500°C is statistically significant. The changes from 1500°C to 1800°C and 1800°C to 2400°C are statistically not significant. However, at the highest processing temperatures, the increase in the length ratios is statistically significant. Thus, at the highest heat treatment temperature of 2875°C, the fibers display a shrinkage of about 6% as compared with 8% at the intermediate temperature of 900°C.

Considerably larger changes are observed in the diameter of the fibers, as displayed in Figure 3. The diameter measurements for the SCF fibers were more consistent than their length counterparts. However, because the SCF fibers were thinner than the AR fibers, the SCF measurements tend to have more error than do their AR counterparts. To verify if the large shrinkage in the diameter relative to that along the length was due to surface oxidation (from fugitive oxygen) during heat treatment, the diameter ratio was plotted as a function of the reciprocal diameter. For increasing diameter, the extent of surface oxidation decreases. Thus, in the limiting case (as D becomes large and the reciprocal of D approaches zero), the intercept on the ordinate could provide a measure of the shrinkage in the absence of oxidative loss. Representative plots for 600, 1800, 2400, and 2700°C are presented in Figure 4. However, the diameter ratio was not found to have any dependence on the reciprocal diameter. Therefore, the role of surface oxidation was not significant and the values reported in Figure 3 for the diameter ratio were the average values (from various do) for any given HTT.

Starting from the oxidized state, a significant drop of 15% is observed in the diameter ratio till a HTT of 900°C. From 900°C to 1200°C, there is an increase in the diameter ratio, but the increase from 1200°C to 1500°C is statistically not significant for either of AR or SCF fibers. The large

drop in the diameter ratio at 1800°C is statistically significant but cannot be explained by a physical explanation. Beyond 1800°C, there is a slight increase of the diameter for both AR and SCF fibers with the ultimate shrinkage being about 20% for the AR fibers and 30% for the SCF fibers at the highest HTT of 2875°C.

Microstructural Characterization

Scanning electron microscopy (SEM) was used to characterize the microstructure of the fibers at the various HTTs. The micrographs depict typical lateral surfaces of the fibers as well as the cross-sections of fibers failed in tension. Because of space constraints, it is not possible to include all of the micrographs (about 30) in this report; only those at 3 conditions will be presented (as-spun, 900°C, and 2875°C).

Figure 5 displays a micrograph of an as-spun AR pitch fiber. The diameter of the fiber is about 80 μm and the cross-section as well as the lateral surface of the fiber are devoid of any texture. The fractured surface of the fiber is typical of a brittle failure. The presence of bubbles and voids generated by the off-gassing of the AR mesophase is evident in this micrograph as well as others. For a stabilized AR fiber, the cross-section is again typical of a brittle failure. For smaller stabilized fiber (about 40 μm diameter), some radial texture can be observed. The origin of the texture is very likely the radial orientation of the discotic AR mesophase molecules generated during fiber spinning. Also longitudinal cracks result from weaker properties of the material in the hoop direction.

After heat treatment at 300°C, there is virtually no additional texture developed. In fibers treated to 600°C, the presence of a slight radial texture can be detected, but the lateral fiber surface is still fairly smooth.

The first clear evidence of a radial texture is observed at a HTT of 900°C, as illustrated in Fig. 6. The splitting of the fibers in a "Pac-man" pattern is typical of the anisotropic pitch-based fibers. The alignment of the discotic mesophase molecules in the radial direction results in the material being fairly weak in the hoop direction, and the hoop stresses generated during the heat treatment are relieved by the "Pac-man" split. Surface striations are also observed for the first time at 900. At 1200°C, the existence of a lamellar structure has started to emerge, and at 1500°C, the texture is slightly more developed as compared to that at 1200°C. At higher HTTs of 1800 and 2100°C (figures 17, 18, and 19), there is a gradual sharpening of the domains or the texture.

At 2400°C, the presence of graphene-layer planes is very clear. The microdomains (consisting of graphene layers) can be easily seen as can the folding of the graphene planes. At 2700°C, there is further refinement of the structure. However, the presence of some unusual fibers was also detected. The fibers appear to have a hollow core, although the length of the core could not be determined. In these fiber too, the graphene planes are well-developed and aligned normal to the inside surface of the hollow core. At the highest HTT that could be attained in the present study, 2875°C, the microdomains consisting of sheet-like graphene planes are very clearly seen in Fig. 7. Some unusual features were also identified such as the presence of a peeled skin and a hollow core.

SEM investigation was also conducted on SCF fibers. The observation are similar to those for the AR fibers. As-spun SCF fibers are smooth and have no texture. Stabilized fibers too are fairly devoid of texture, although in some fibers a very slight radial texture may be observed. At a HTT of 300°C, a slight radial texture may be observed in some fibers; also splitting is seen. At a HTT of 600°C, some radial texture is observed in the skin, but not in the core, but the lateral surface is still smooth.

The emergence of lamellar graphene planes is apparent at 900°C as is the presence of surface striations. At a HTT of 1200°C, the existence of a radial structure is clear, and by 1500°C the radial structure is observed throughout the cross-section of small fibers, but not in a thicker. At 1800°C, the layered structure of graphene planes is apparent. At 2400°C, 2700°C, and 2875°C, the presence of microdomains with graphene planes is very clear. The unusual hollow core observed for AR is also observed for some SCF fibers.

High resolution SEM was also used for selected samples to obtain details of the microstructure. For fibers heat treated to 2875°C, the well developed graphene planes are evident. Images could be obtained at magnifications as high as 50,000X and 100,000X. The high resolution images were also obtained at fibers carbonized at 900°C. These micrographs illustrate that even at the low processing temperature of 900°C, graphene-like planes exist. These planes appear to be surrounded by a continuous matrix like material, in contrast to the fibers treated at 2875°C, where the graphene layer appear to be present discretely.

The microstructure of the fibers heat treated to 2875°C was also characterized by wide angle x-ray scattering (WAXS). Only one sample could be investigated because the control and data-acquisition unit of the XRD instrument was being repaired. The data from the 2-theta and

azimuthal scans conducted in the symmetrical transmission mode indicate that the d_{002} spacing for the fibers heat treated at 2875°C was 0.3354 nm. This indicates that the material is almost fully graphitic since the limiting value of d_{002} spacing is reported as 0.3354 nm. The stacking height, L_c , was determined to be about 36 nm, whereas the L_a was found to be about 48 nm. The presence of a (103) peak confirmed the existence of 3-dimensional crystallinity (as opposed to turbostratic structure).

Tensile Properties

The tensile properties were measured by single-filament testing. To obtain statistically significant conclusions, the number of replicates required in single-filament tests often exceed 20. However, due to time constraints, only 5 samples were tested for each HTT condition. Also, the tensile fixture was slightly damaged and as a result many samples broke while cutting the paper tab. In addition, the SCF fibers were significantly wavy after heat treatment above 200°C and broke in the paper tab while being straightened. The limited tensile strength and modulus data for SCF fibers are displayed in Figure 8. The differences in values at various HTTs are statistically not significant. Therefore, only a qualitative trend can be inferred from the graph that the strength and modulus increase from the lowest HTT of 600°C to the highest HTT of 2875°C.

The tensile strength and modulus data for AR fibers are displayed in Figure 9. Starting from the lowest HTT of 600°C, the modulus increases for higher HTTs of 900 and 1200°C; the increases are statistically significant at the 95% confidence level. The moduli increase from a HTT of 1200 to 1500°C is statistically not significant, but those from 1500 to 1800°C and from 1800 to 2100°C are. From 2100°C to 2400°C, the modulus decreases and the difference is statistically significant. The change in modulus from 2400 to 2700°C is not significant, but the modulus increase from 2700 to 2875°C is statistically significant. The maximum modulus was measured to be 55 Msi (380 GPa).

The tensile strength data for AR fibers follow a similar trend as that displayed by the modulus data, although the peak appears slightly earlier at 1800°C. From 1800°C to 2400°C, there is a slight decrease in tensile strength, but beyond 2400°C, the strength increases, and a maximum strength of 253 ksi (1.8 GPa) is observed at the highest HTT of 2875°C. The strain-to-failure data for AR and SCF fibers are presented in Figure 10. Because of the experimental error associated with these measurements, most of the differences are statistically not significant.

The strength and modulus values are somewhat lower than the values of 2-3 GPa and 300-800 GPa reported in the studies of Mochida et al [7] and Pencock, et al. [8]. The primary reason for the lower values is that the fibers chosen for this study were considerably thicker (ranging from 30 micrometers to 100 micrometers) than those used in most other studies. Also, as was illustrated in the SEM micrographs, the thick fibers tend to have more defects in the form of gas-bubbles, cracks, and the characteristics "Pac-man" splitting. However, it should be recalled that the thicker fibers facilitated accurate measurement of the dimensional changes and were purposely chosen for the present study where the primary emphasis was characterization of the dimensional changes accompanying heat treatment.

Conclusions

The following conclusions are drawn from this study:

1. Changes in fiber length can be reliably measured for fibers that have diameters in the range 50 to 100 μm .
2. For the AR mesophase pitch fibers, the maximum length shrinkage of 8% occurs at a heat treatment temperature of about 900°C. For the highest HTT of 2875°C, the total shrinkage in length was only 6%. The slight increase from 900 to 2875°C may result from the alignment of the graphene planes from the initial out-of-plane orientation to the very highly aligned state at 2875°C. The shrinkage in diameter is about 20% at the highest HTT of 2875°C, corresponding to an area reduction of almost 36% in the transverse plane. These values clearly establish the anisotropic nature of the dimensional changes during carbonization.
3. High resolution scanning electron microscopy images indicate that significant development of graphene-like layers begins at about 900°C and continues till about 2400°C. XRD results indicate that the AR material heat treated to 2875°C has a d_{002} spacing of 0.3354 nm and is almost fully graphitic.

Future Work

The most significant aspect of the present study was the characterization of dimensional changes together with supporting evidence from SEM, XRD, and tensile measurements. Whereas the dimensional measurements could be successfully conducted *ex situ*, the measurement of dimensional changes need to be measured *in situ* to obtain an estimate of the changes at the actual elevated processing temperatures. A modified carbon/carbon fixture suitable for mounting in a TGA apparatus needs to be developed as a part of the future work. Also, XRD data and tensile data need to be generated for the AR fibers. Finally, the methodology developed in this project should be verified on SCF and other fibers.

Acknowledgments

Funding from AFOSR through the Summer Faculty Research Program Contact Number F49620-93-C-0063 is gratefully acknowledged. I would also like to acknowledge the technical and organizational support provided by the Carbon Group (Focus area 1) of the Nonmetallic branch of the Materials Directorate, Wright Patterson Air Force Research Labs, Dayton. Special thanks to Kris Kearns, Scott Thiebert (Lab Focal Point), Dave Anderson, Bill Price, Ken Davidson, Allan Crasto, Ken Lindsay, Katie Thorpe, and Bill Ragland, Amanda Colleary, and Darren Wells for the help and interaction while conducting the research and making the summer visit not only very pleasant but also a very productive one.

References

1. D. Schmidt, K. E. Davidson, L. S. Thiebert, "Evolution of Carbon-Carbon Composites," SAMPE Journal, 32 (4), 44-50, 1996.
2. E. Fitzer and L. M. Manocha, "Carbon Reinforcements and Carbon/Carbon Composites," Springer-Verlag Publishers, 1998.
3. J. D. Buckley and D. D. Edie, "Carbon-Carbon Materials and Composites," NASA Reference Publication 1254, 1992.
4. H. G. Mahs, W. L. Vauhn, "Four Advances in C-C Composites," NASA CP 3249 361 (1994).
5. I. Mochida, et al., "Microstructure of Mesophase Pitch-Based Carbon Fiber and Its Control," Carbon, 34 (8), 941-956, 1996.

6. S. P. Jones, C. C. Fain, and D. D. Edie, "Structural Development in Mesophase Pitch Based Carbons Fibers Produced From Naphthalene," Carbon, 35(11), 1553-43, 1997.
7. I. Mochida, K. Shimizu, Y. Korai, and Y. Sakai, "Mesophase Pitch Catalytically Prepared From Anthracene," Carbon, 30(1), 55-61, 1992..
8. G. M. Pencock, G. H. Taylor, and J. D. Fitzgerald, "Microstructure in a Series of Mesophase Pitch-Based Fibers from du Pont," Carbon, 31 (4), 591-609, 1993.
9. X. Bourrat, E. J. Roche, and J. G. Lavin, "Structure of Mesophase Pitch Fibers," Carbon, 28 (2/3), 435-446, 1990.
10. D. J. Johnson, I. Tomizuka, and O. Watanabe, "The Fine Structure of Pitch-Based Carbon Fibres," Carbon, 13, 529-534, 1975.
11. L. M. Manocha, H. Bhatt, and S. M. Manocha, "Development of Carbon-Carbon Composites by Co-Carbonization of Phenolic Resin and Oxidized PAN Fibers," 34 (7), 841-849, 1996.
12. V. Liedtke and K. J. Huttinger, "Mesophase Pitches as Matrix Precursor of Carbon Fiber Reinforced Carbon: III. Mechanical Properties of Composites," Carbon, 34 (9), 1081-86, 1996.
13. L. S. Singer and S. Mitchell, "Diffusion of Oxygen in Pitch," Carbon, 35 (5), 599-604, 1997.

Table 1. Replicate length values, mm, of stabilized AR fibers.

Sample #	L1 AR Stabil	Average L1	Std Dev L1	Repl Avg L2	Std Dev L2	L1/L2
1	28.443					
	28.462	28.452	0.0096	28.462	0.0031	0.999
	28.451					
2	31.024					
	31.041	31.041	0.0174	31.08	0.0041	0.998
	31.059					
3	17.337					
	17.327	17.334	0.0061	17.357	0.0093	0.998
	17.338					
4	17.708					
	17.74	17.728	0.017	17.76	0.0059	0.998
	17.736					

Table 2. Comparison of length values for as-spun and stabilized AR fibers.

Sample	L AR As-Spun	Std Dev L AsSp	Lstab/Lassp
1	28.464	0.0108	0.9999
2	31.06	0.0081	1.0006
3	17.3227	0.0106	1.002
4	17.711	0.0148	1.0027

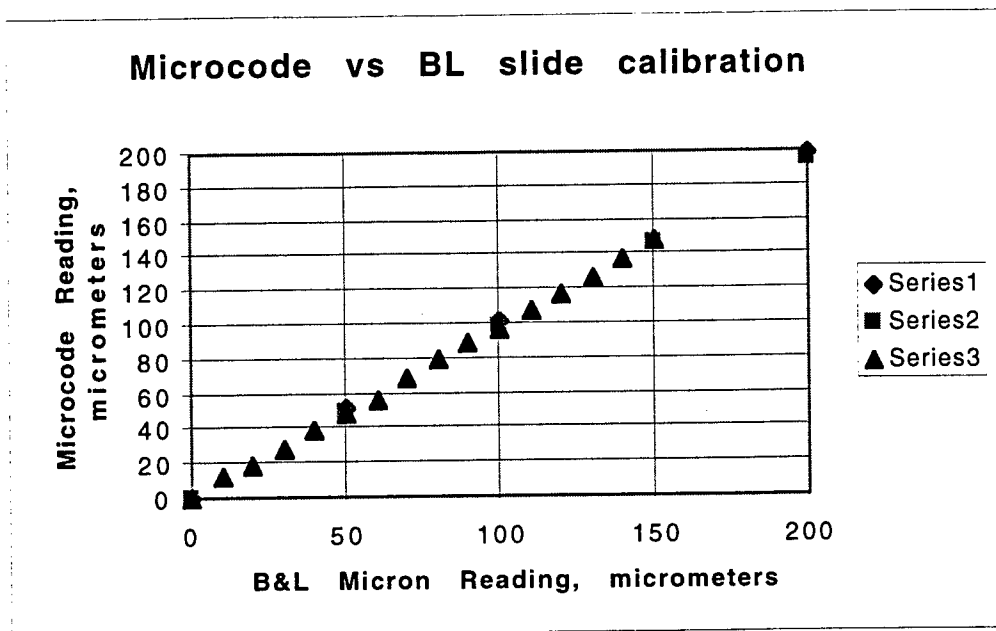


Figure 1. A calibration curve for the Microcode Indexer Vs Readings from the Bosch & Laumb calibration microslide.

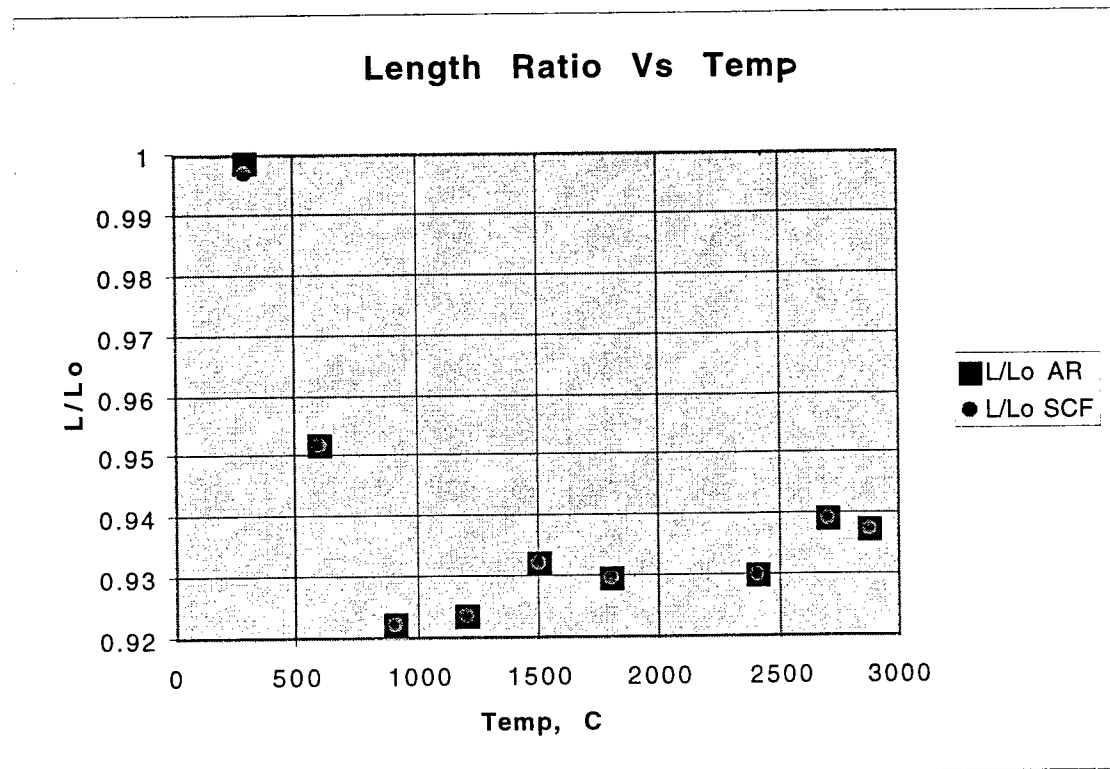


Figure 2. Length ratio after heat treatment as a function of the heat treatment temperature.

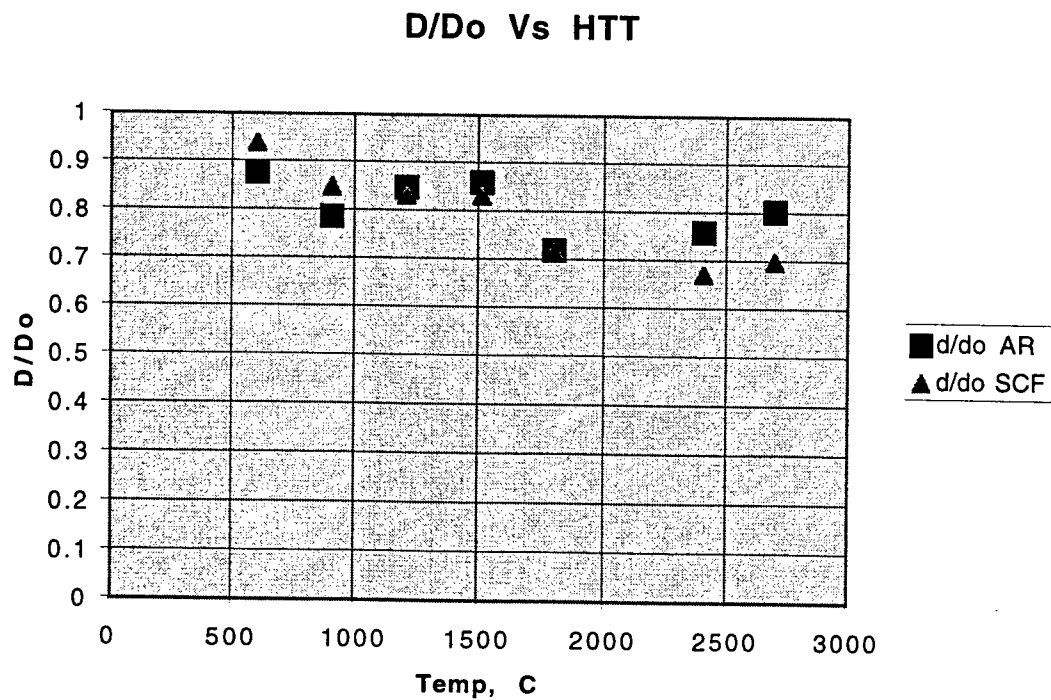


Figure 3. Fiber diameter ratio as a function of heat treatment temperatures.

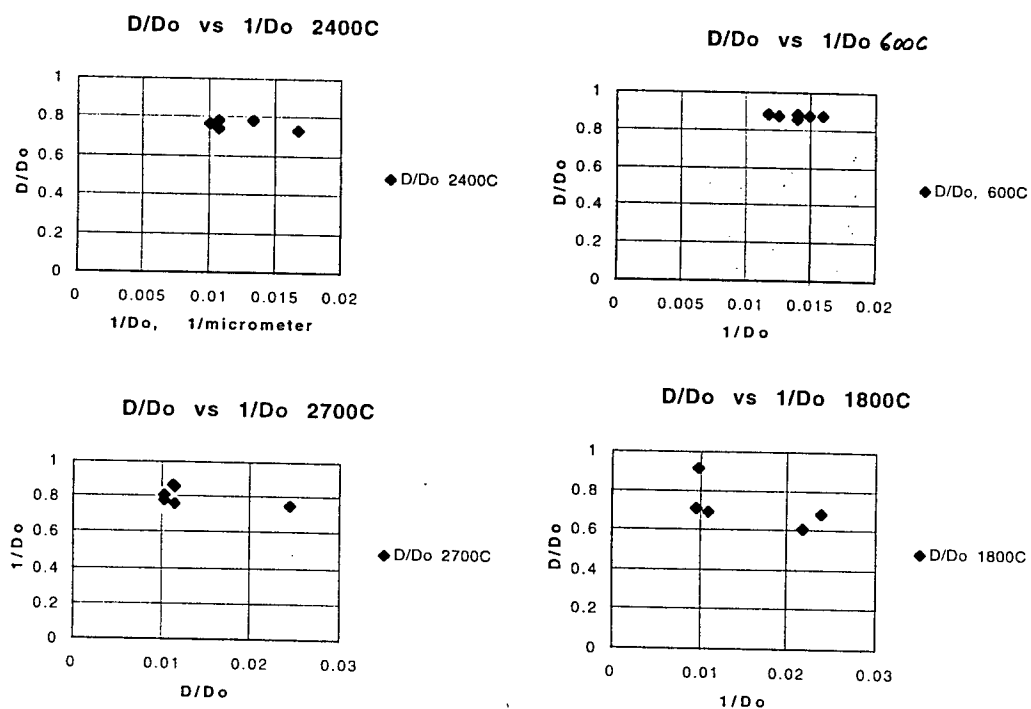


Figure 4. D/D_o ratio plotted as a function of reciprocal diameters.

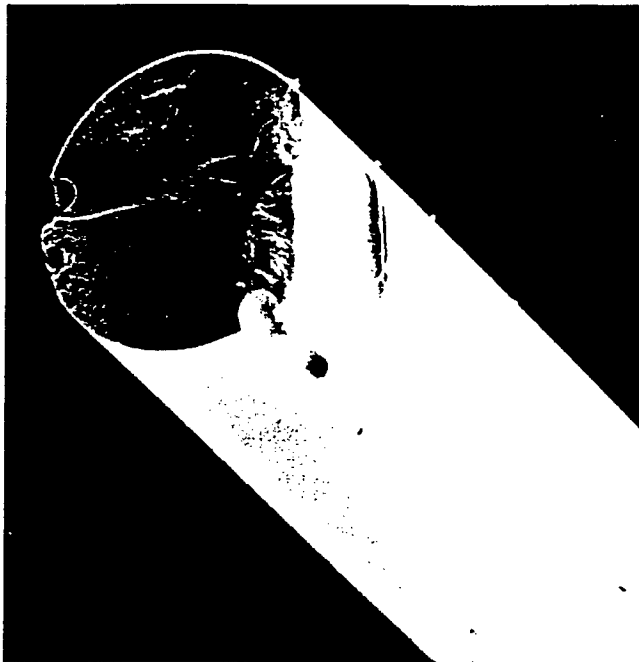


Figure 5. SEM micrograph of as-spun AR pitch fiber at a magnification of 500X.

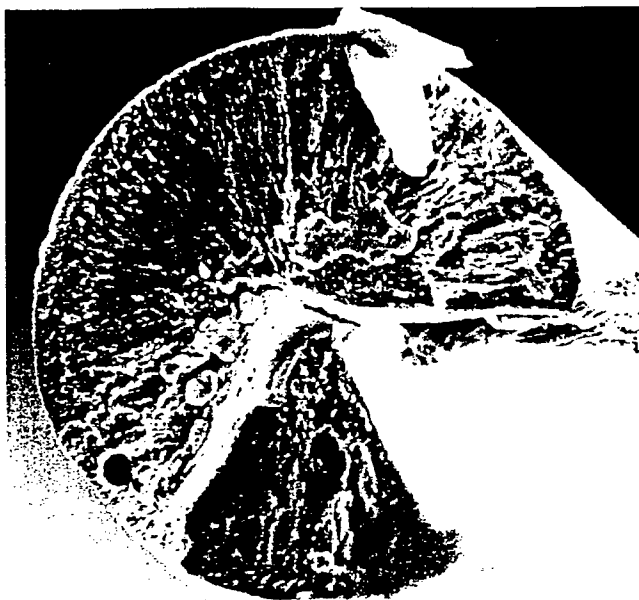


Figure 6. SEM micrograph of a AR fiber heat treated at 900°C at a magnification of 1000X.



Figure 7. SEM micrograph of AR fiber heat treated to 2875°C at a magnification of 1000X.

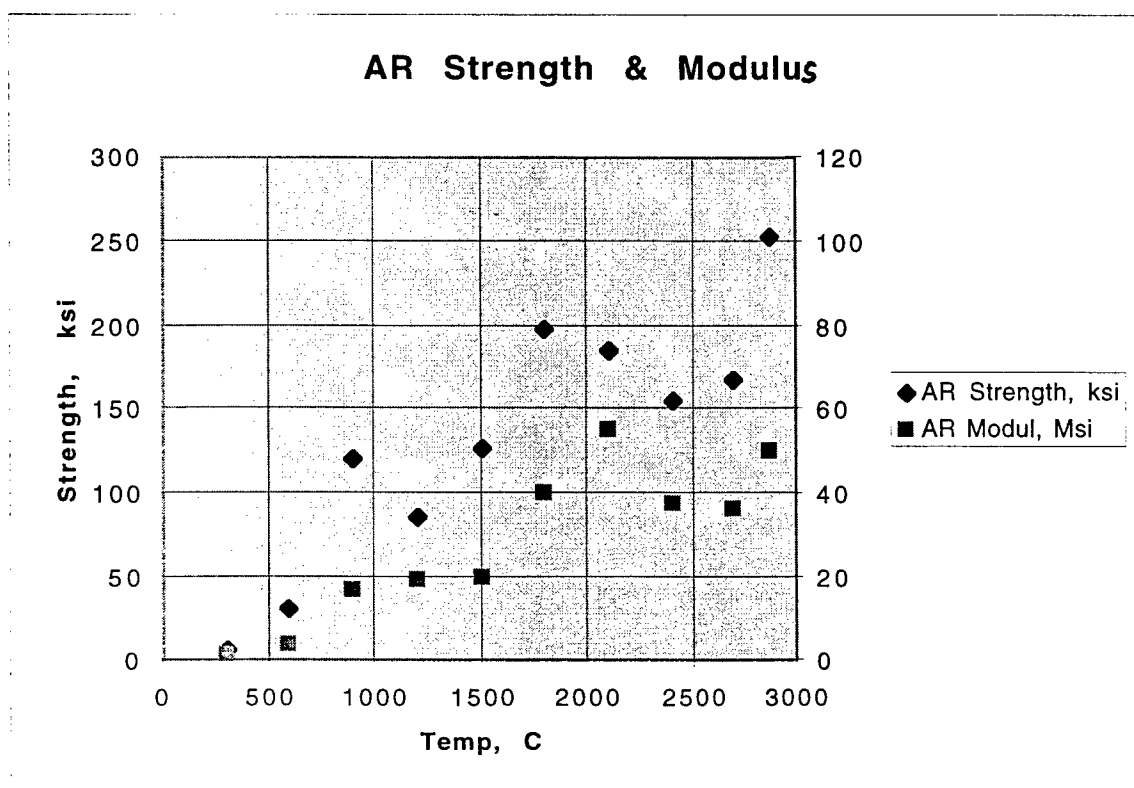


Figure 8. Strength and Modulus of heat treated AR fibers.

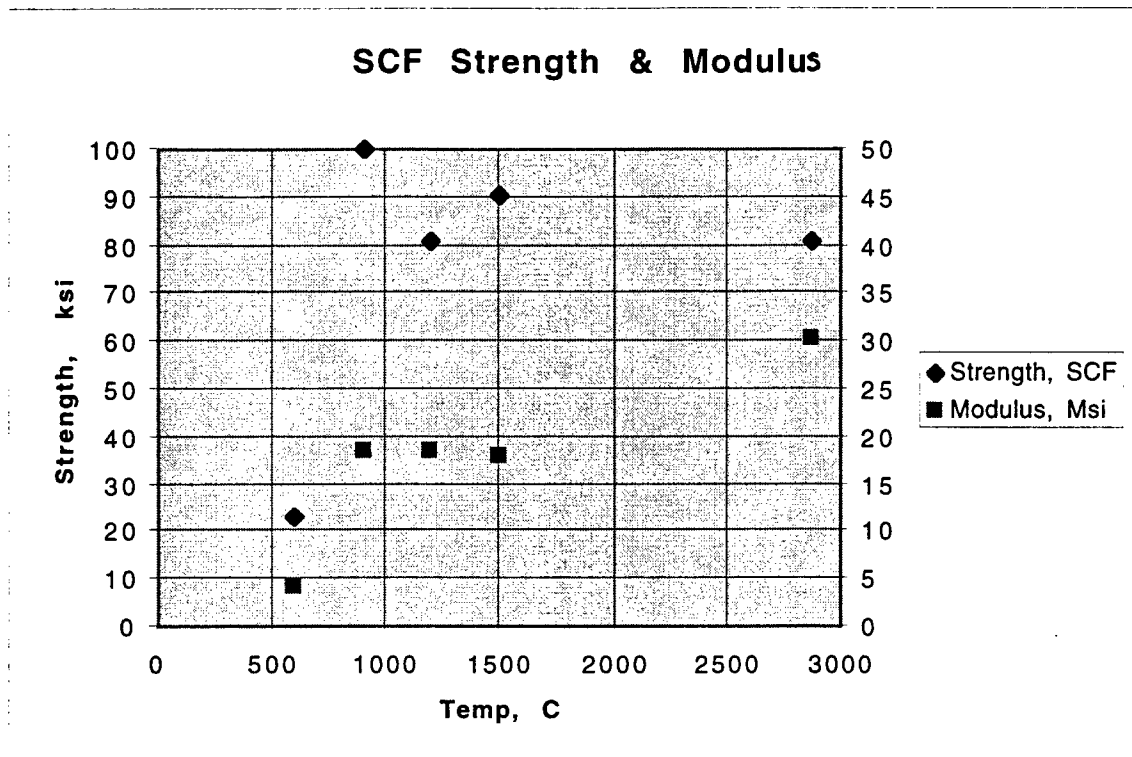


Figure 9. Strength and Modulus of heat treated SCF carbon fibers.

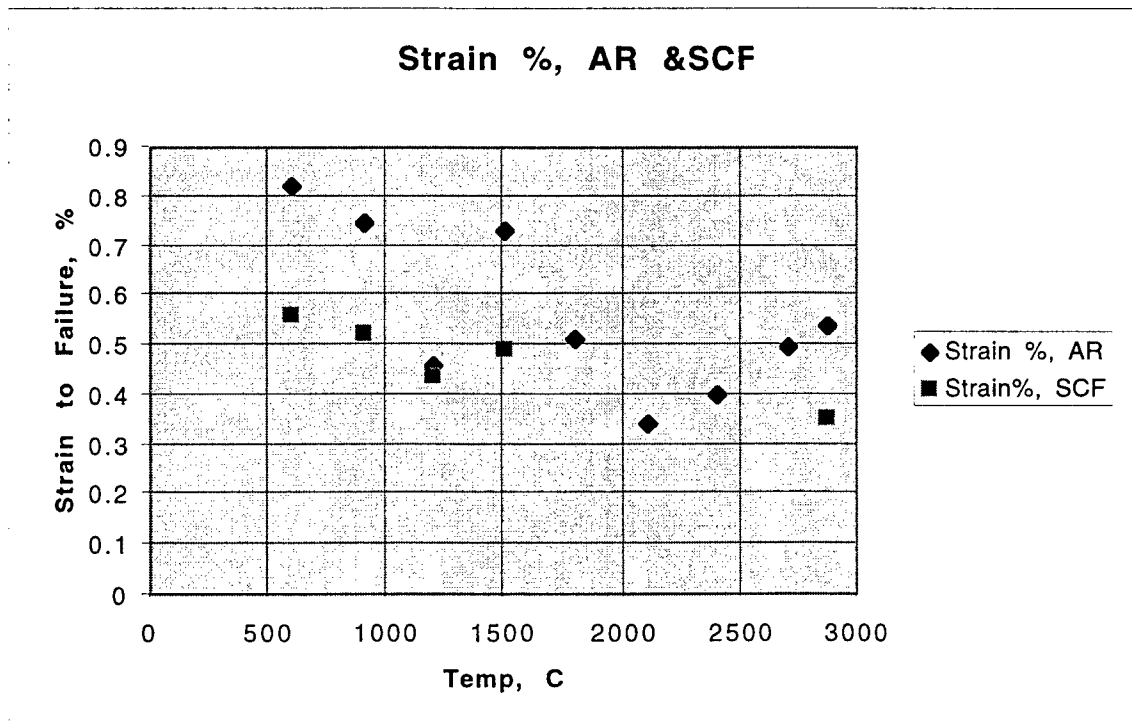


Figure 10. Strain to Failure of heat treated AR and SCF carbon fibers.

**Simulation of the Antenna Pattern of Arbitrarily Oriented
Very Large Phase/Time-Delay Scanned Antenna Arrays
With Systematic and Random Errors**

Carlos R. Ortiz, Ph.D.
Professor
Department of Electrical Engineering

Polytechnic University of Puerto Rico
377 Ponce de Leon Ave
Hato Rey, PR 00919

Final Report for:
Summer Faculty Research Program
Wright-Patterson Research Site

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC
and
Wright-Patterson Research Site

August 1998

**Simulation of the Antenna Pattern of Arbitrarily Oriented
Very Large Phase/Time-Delay Scanned Antenna Arrays
With Systematic and Random Errors**

Carlos R. Ortiz, Ph.D.

Professor

Department of Electrical Engineering

Abstract

A computer simulation to determine the field-pattern of arbitrarily oriented very large phase/time delay scanned antennas was developed. The simulation takes into account errors present in antenna arrays. These are systematic and random errors. The systematic errors considered here are the finite quantization of the phase produced by the use of digital N-bits phase/time-delay shifters and the flexing of the array aperture due to its large size and weight. The random errors considered are those caused by variations on the amplitude and phase of the elements current, variations on the radiation pattern of the elements, and missing elements (due to catastrophic failure), and variations in the location of the elements. To validate the simulation a number of patterns were computed. These included patterns of linear arrays, array panels, and arrays of panels. Experience as well as specific examples validated the ideal patterns. The "random-error-patterns" were compared to specific trends noted in earlier studies. The behavior of the computed patterns confirmed such trends.

**Simulation of the Antenna Pattern of Arbitrarily Oriented
Very Large Phase/Time-Delay Scanned Antenna Arrays
With Systematic and Random Errors**

Carlos R. Ortiz, Ph.D.

1.0 Introduction:

The purpose of this effort was to write a simulation to compute the field of arbitrarily oriented, very large phase/time-delay scanned antenna arrays with systematic and random errors. As explained in [1], errors in antenna arrays can be divided into two categories depending on whether they are predictable (systematic) or random. Among systematic errors is the finite quantization of the phase produced by the use of digital N-bits phase/time-delay shifter. Another error considered here is the flexing of the array aperture due to its large size and weight. Random errors are caused by variations on the amplitude and phase of the elements current, variations on the radiation pattern of the elements, missing elements (due to catastrophic failure), and others not considered here. Random errors may alter antenna parameters such as, increasing sidelobe level, reducing power gain, lowering directivity, etc.

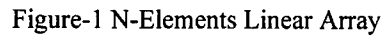
2.0 Theory:

The development of the necessary equations to simulate the field pattern of arbitrarily oriented, very large planar phase/time-delay scanned antenna arrays with systematic and random errors is presented in this section. Such development is based on the known principle of pattern multiplication [1].

2.1 Linear Arrays:

Consider the N-elements linear array shown in Figure 1. The array axis is in the $\hat{t}$ direction. The center of the array coincides with the origin of the t-axis and it is specified, with respect to a Cartesian (global) coordinate system, by the position vector $\vec{r}_c$. For an even number of elements, the elements are distributed equally on each side of the array. For an odd number of elements, the center element is located at the origin of the t-axis with an equal amount of elements on each side. The distance between any two consecutive elements is constant and it is represented by " d_t ." The elements are considered to be identical, with identical magnitude and progressive phase shift between consecutive elements. Such array can be referred to as a uniform array.

According to the method of pattern multiplication the far-field pattern of an array of identical elements is equal to the product of the element pattern and the array factor of the array. The element pattern is the field of a single element computed at a reference point -usually the origin of the array. The array factor is a function of the number of elements, their geometrical arrangement, their relative magnitude and phases, and their spacing.


$$F_N(\theta, \phi) = e_N^*(\theta, \phi) f_N(\theta, \phi). \quad (1)$$
$$e_N^* = e_N (1 + \delta e_N) \kappa.$$
$$F_N(\theta, \phi) = \sum_{n=1}^N I'_n e^{j\beta(\vec{u} \cdot \vec{r}'_n)}. \quad (2)$$
$$\beta = 2\pi / \lambda \quad (3)$$
$$I'_n = (1 + \delta I_n) |I_n| e^{j(\alpha_n + \delta \alpha_n)} \quad (4)$$

$|I_n|$ and α_n are the no-error magnitude and phase of the excitation. The δI_n as well as the $\delta \alpha_n$ are samples of random variables assumed to have a normal distribution with zero mean and variance $\sigma^2_{\delta I_n}$ and $\sigma^2_{\delta \alpha_n}$, respectively. The direction of transmission is given by,

$$\hat{u} = \cos\theta_i \sin\phi_i \hat{x} + \sin\theta_i \sin\phi_i \hat{y} + \cos\theta_i \hat{z}. \quad (5)$$

The distance vector from the n^{th} element to the observation point P is denoted by $\vec{r}_n$ which can be expressed as,

$$\vec{r}_n = \vec{r}_0 - \vec{r}'_{ng}. \quad (6)$$

$\vec{r}_0$ is the position vector of the observation point P, while $\vec{r}'_{ng}$ is the position vector of the n^{th} element with respect to the origin of the Cartesian coordinate system and can be expressed as,

$$\vec{r}'_{ng} = \vec{r}_c + \vec{r}'_{nl}. \quad (7)$$

Where,

$$\vec{r}_c = x_c \hat{x} + y_c \hat{y} + z_c \hat{z} \quad (8)$$

and $\vec{r}'_{nl}$ is the position vector of the n^{th} element with respect to the local coordinate system $\hat{n} = \hat{t} \times \hat{v}$ with origin at the center of the array. In terms of their scalar components along the Cartesian unit vectors, the unit vectors along the t, v and n-axis are $\hat{t} = t_x \hat{x} + t_y \hat{y} + t_z \hat{z}$, $\hat{v} = v_x \hat{x} + v_y \hat{y} + v_z \hat{z}$, and, $\hat{n} = n_x \hat{x} + n_y \hat{y} + n_z \hat{z}$, respectively.

Using the equations described above, and making the usual far-field approximations, Equation (2) becomes,

$$F_N = \sum_{n=1}^N |I'_n| e^{j(\alpha_n + \delta \alpha_n)} e^{j\beta(x'_n \sin \theta_i \cos \phi_i + y'_n \sin \theta_i \sin \phi_i + z'_n \cos \theta_i)}. \quad (9)$$

$$\text{Where, } |I'_n| = (1 + \delta I_n) |I_n|, \quad (10)$$

$$x'_n = x_c + \left(-\frac{L_t}{2} + (n-1) d_t \right) t_x + \delta_t t_x + \delta_v v_x + f(t) n_x, \quad (11)$$

$$y'_n = y_c + \left(-\frac{L_t}{2} + (n-1) d_t \right) t_y + \delta_t t_y + \delta_v v_y + f(t) n_y, \quad (12)$$

$$z'_n = z_c + \left(-\frac{L_t}{2} + (n-1) d_t \right) t_z + \delta_t t_z + \delta_v v_z + f(t) n_z. \quad (13)$$

In Equations (11) – (13), L_t is the length of the linear array along the t-axis. In arriving at these equations, it was assumed that the location of the elements is a random variable with independent distributions along the transversal and vertical directions. Both distributions are assumed to be normal with mean equal to the “real” location of the elements and with a respective standard deviation σ_{δ_t} and σ_{δ_v} . The δ_t and δ_v represent samples of such

distributions. In the code, σ_{δ_t} and σ_{δ_v} , are specified as inputs. $f(t)$ is the flexing of the array aperture along the t-axis. Here, it is assumed that $f(t)$ varies linearly from the center to the edge of the array. The slope of this variation is an input to the code also.

2.1.1 Beam Scanning:

Two different options for beam scanning were incorporated into this simulation. These are phase scanning and true-time delay scanning. These may be used for scanning the main beam of the field radiated by either the individual panels as well as that of the array of panels. Both options are discussed presently.

2.1.1.1 Phase Scanning:

By prescribing the elements with a progressive phase shift given by,

$$\alpha_n = \beta_t = -\beta \hat{u}_0 \bullet \vec{r}'_{ng}, \quad (14)$$

it is possible to point the main beam in a given direction, say $\hat{u}_0$. Equation (9) can be written now as,

$$F_N(\theta, \phi) = \sum_{n=1}^N |I'_n| e^{j\delta\alpha_n} e^{j(\beta \hat{u} \bullet \vec{r}'_{ng} + \beta_t)}. \quad (15)$$

The above expression produces a maximum radiation in a direction given by,

$$\hat{u}_0 = \sin \theta_0 \cos \phi_0 \hat{x} + \sin \theta_0 \sin \phi_0 \hat{y} + \cos \theta_0 \hat{z}. \quad (16)$$

2.1.1.2 True Time-Delay Scanning:

The following description is a generalization of the description found in [3]. In the true time-delay method, the progressive phase-shift is achieved by using N-bit "time-shifters" as opposed to phase shifters. For the n^{th} element, this is achieved by propagating the excitation of the element through a delay line whose length is designed to provide a time delay given by,

$$t_n(\theta_0, \phi_0) = (\hat{u}_0 \bullet \vec{r}'_{ng}) / c \quad (17)$$

where c = speed of light in free-space. For all frequencies $\omega = 2\pi f$, where f is the frequency of the excitation in Hertz, the phase of the excitation is now given by,

$$\beta_t = -\omega t_n(\theta_0, \phi_0). \quad (18)$$

Substituting Equation (18) into Equation (15), we obtain,

$$F_N(\theta, \phi) = \sum_{n=1}^N |I'_n| e^{j(\beta \hat{u} \bullet \vec{r}'_{ng} - \omega t_n(\theta_0, \phi_0))} \quad (19)$$

which also produces a maximum radiation in the direction θ_0, ϕ_0 . In this case, however, the time-delay factor is independent of frequency.

2.1.1.3 Phase/Time-Delay Quantization Error:

The quantized phase/time-delay factor results from the use of N-bits phase/time-delay shifters. The number of bits N, is an input to the computer program. Due to the quantization process, the actual phase/time-delay factor realized may not be exactly equal to the specified value. The resulting quantization error may be expressed as,

$$\delta\beta_i = \beta_i - \bar{\beta}_i \quad (20)$$

Where, $\delta\beta_i$ is the error, β_i is the specified phase/time-delay factor as given by either Equation (14) or Equation (18), and $\bar{\beta}_i$ is the quantized phase time/delay shift factor. Note that $\delta\beta_i$ can be either positive or negative depending on whether $\bar{\beta}_i$ is smaller or larger than β_i . Considering the phase/time-delay quantizing error, Equation (9) becomes,

$$F_N = \sum_{n=1}^N |I'_n| e^{j\delta\alpha_n} e^{j[\beta(x'_n \sin \theta_i \cos \phi_i + y'_n \sin \theta_i \sin \phi_i + \cos \theta_i) + \bar{\beta}_i]} \quad (21)$$

2.1.2 Pattern Synthesis:

It is often of interest to achieve a narrow main beam, accompanied by a low side lobe level. One of the most important methods is the Taylor Line-Source Design Method [4]. Taylor perfected this method for continuous line source antennas. Various authors, including Villeneuve [5] have developed several approaches of applying Taylor's method to arrays of discrete elements. Villeneuve showed that for discrete, uniformly spaced arrays, exciting the array elements with samples of the continuous distribution produces little differences in the excitations. Thus, the pattern realized by exciting the elements with samples of the continuous distribution, is undistinguishable from the pattern realized by exciting the elements with the corresponding distribution for discrete arrays. The agreement improves for larger arrays. Here we use the sampling method to compute the excitation of the elements.

2.2 Array Panels:

In this section, the pattern of a planar N by M elements array is discussed. The planar array is shown in Figure-2. Such array may be constructed by aligning M linear arrays-of the type described in section 2.1- along the v-axis. The spacing between the M linear arrays along the v-axis is d_v . Recalling the principle of pattern multiplication, the field pattern of the planar array can be expressed as,

$$F_{NM} = e_M(\theta, \phi) f_M(\theta, \phi) \quad (22)$$

In this case, the element pattern is the field pattern of a linear array given by Equation (1) and re-stated here as:

$$e_M(\theta, \phi) = e_N^*(\theta, \phi) f_N(\theta, \phi). \quad (23)$$

Thus,

$$F_{NM} = e_N^*(\theta, \phi) f_N(\theta, \phi) f_M(\theta, \phi). \quad (24)$$

Where, f_N is given by Equation (2) (for isotropic elements) and

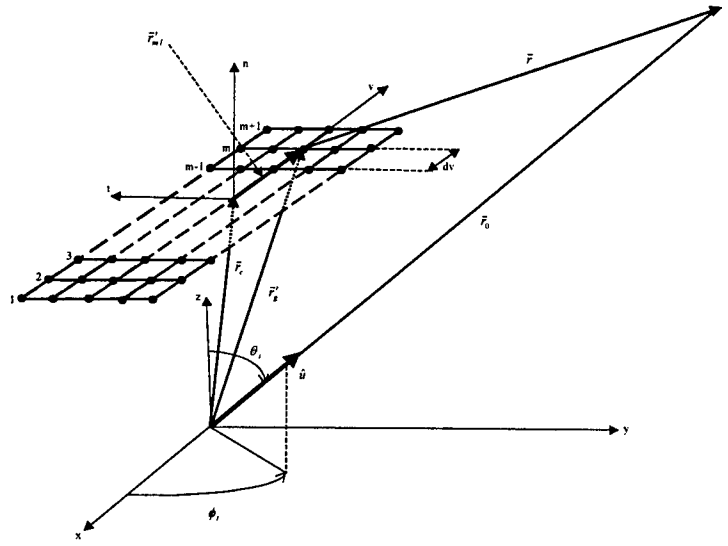


Figure-2 N by M Elements Planar Array

$$f_M(\theta, \phi) = \sum_{m=1}^M |I'_m| e^{j\delta\alpha_m} e^{j\beta(x'_m \sin \theta_i \cos \phi_i + y'_m \sin \theta_i \sin \phi_i + z'_m \cos \theta_i + \bar{\beta}_v)} \quad (25)$$

In Equation (25), $|I'_m|$ is the magnitude of the excitation current with error at the m^{th} element given by,

$$|I'_m| = (1 + \delta I_m) |I_m| \quad (26)$$

Where $|I_m|$ is the magnitude of the no-error excitation and δI_m are samples of a normal distribution with zero mean and variance $\sigma_{\delta I_m}^2$. Also, $\delta\alpha_m$ is the random error in the phase of the excitation. Here, $\bar{\beta}_v$ is the quantized phase/time-delay factor, which can be computed in a manner similar to the one described in section 2.1.1. The phases $\bar{\beta}_i$ and $\bar{\beta}_v$ are independent, but usually adjusted so that the main beam of is directed in the same direction. Finally,

$$x'_m = x_c + \left(-\frac{L_v}{2} + (m-1) d_v \right) v_x + \delta_i t_x + \delta_v v_x + f(v) n_x, \quad (27)$$

$$y'_m = y_c + \left(-\frac{L_v}{2} + (m-1) d_v \right) v_y + \delta_i t_y + \delta_v v_y + f(v) n_y, \quad (28)$$

$$z'_m = z_c + \left(-\frac{L_v}{2} + (m-1) d_v \right) v_z + \delta_i t_z + \delta_v v_z + f(v) n_z. \quad (29)$$

L_v is the length of the array along the v-axis and $f(v)$ is the flexing of the array aperture along the v-axis. It is assumed that $f(v)$ varies linearly. The remaining variables were defined in previous sections. We can consider

elements which are not isotropic in which case $e(\theta, \phi)$ is directive. Included in the code are three different options for the element pattern. These are isotropic, cosine squared and cardioid element patterns.

2.3 Array Of Panels:

The procedure described in section 2.2 can be used to compute the pattern of an array of N by M panels. In this case, the individual array elements shown as dots in Figure-2 are panels. The element pattern of the array of panels is the field pattern of a reference panel. The pattern of a panel may be computed using the procedure described in section 2.2 also. This can be expressed simply as,

$$F'_{NM} = e'_N(\theta, \phi) F_{NM}(\theta, \phi) \quad (30)$$

F'_{NM} is the field of the array of panels, $e'_N(\theta, \phi)$ is the element pattern of a reference panel given by Equation (25) and $F_{NM}(\theta, \phi)$ is the array factor of the array of panels which can also be computed as in Equation (25). These ideas were used here to write a code capable of computing the field-pattern of an N by M panels antenna array.

3.0 Results:

The purpose of this section is twofold. First, it serves as a validation of the results obtained with the code described here. Second, it serves as an implicit demonstration of the potential of the code. For the sake of completeness, patterns for linear arrays, array panels and, an array of panels are included here. Some results for directivity, gain and Sidelobe level are also included. All patterns are computed at increments of half a degree in elevation and a degree in azimuth to insure convergence. The linear arrays are oriented along the x-axis so that the t-axis coincides with the x-axis. The array panels are oriented in the x-y plane so that the t-axis and the v-axis coincide with the x-axis and y-axis, respectively. All patterns are computed in the x-z plane. The random-error patterns shown here were computed using the following data: RMS amplitude error = 0.002 units. RMS phase error = 10.00 degrees, RMS error in the elements location in both x and y directions = 0.002 cm, and RMS error in element pattern = 0.0. Unless stated otherwise, the element spacing is .5 cm, which corresponds to a $\lambda/2$ element spacing at 30 GHz. These values were chosen accordingly to examples found in the literature [6].

3.1 Linear Arrays:

Patterns shown in Figure-3 correspond to the normalized field-pattern of a 20-elements linear array. Two graphs are shown in Figure-3. The solid line corresponds to the normalized pattern of the array with uniform illumination. Experience as well as results found in the literature, validate the field pattern of this uniformly excited array. For instance, the directivity, gain, and sidelobe level were computed for this array also with the code described here. The directivity is 10, the gain 10 dB, and the sidelobe level -13.3 dB. These quantities further validate the results. The dotted line corresponds to a Taylor synthesized pattern with -25 dB sidelobe level ($\bar{n} = 5$). The Taylor weightings were validated with results found in the literature [4]. The gain, directivity, and sidelobe level for this case array are 9.10, 9.59 dB, -25.1 dB as expected. Figure-4 shows the pattern for the same

20-elements Taylor synthesized array scanned to 45° . In this figure, the solid line represents the no-error while the dotted line corresponds to the error-pattern. In computing this pattern only random errors were considered.

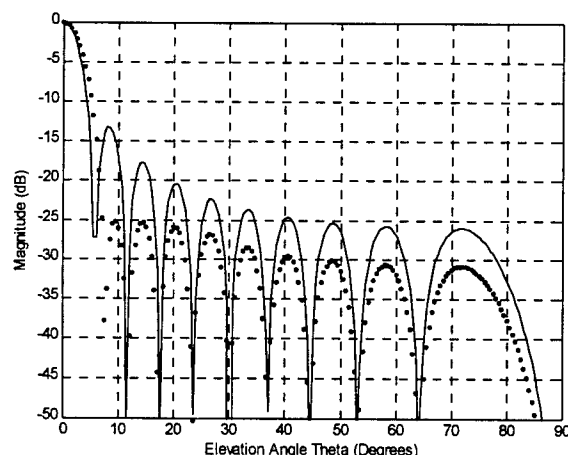


Figure 3. Pattern for a 20-elements linear array with uniform illumination (solid), -25 dB design sidelobe level (dotted)

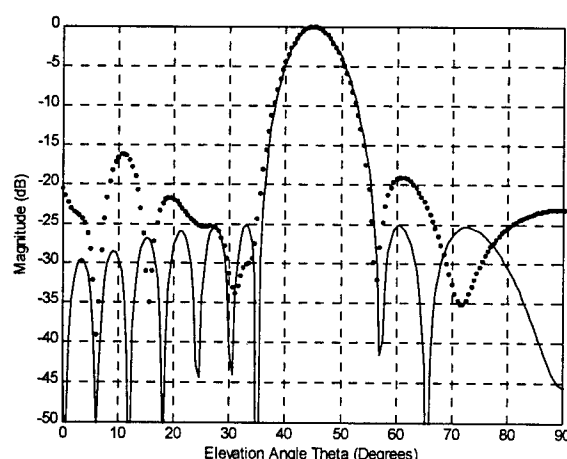


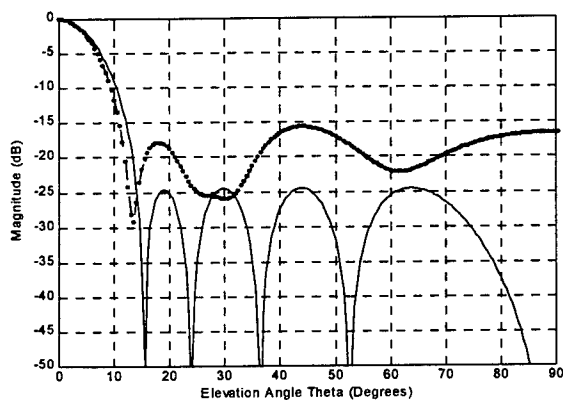
Figure 4. Pattern for a scanned 20-elements linear array with uniform illumination (solid), -25 dB design sidelobe level (dotted)

3.2 Array Panels:

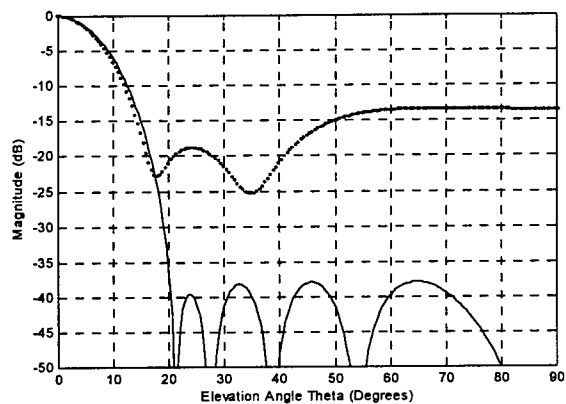
In this section, a number of no-random-error as well as random-error- patterns are shown. The error-patterns are validated against examples found in the literature [6]. However, these examples correspond to average patterns in a statistical sense. Meanwhile, the patterns shown here are computed considering actual random variations. The validation process consists of confirming various trends in tolerance analysis reported in the literature [6]. For a given set of tolerances, these trends are: (1) the rise in sidelobe level due to random error increases as design sidelobe level decreases; (2) the pattern deterioration decreases as the array is enlarged; (3) the sidelobe level increase due to random error does not depend on scan angle; (4) pattern deterioration is larger for an L elements linear array than for an $L \times L$ elements array; and (5) the pattern deterioration is mostly a result of translational errors in the position of the elements. A number of patterns were plotted to examine these trends. These are shown in Figures (4) to Figure (8). In these figures, the no-error patterns are plotted with solid line while the error-patterns are plotted with a dotted line.

Figure-4 (a) shows the pattern for a 10×10 elements planar array with -25 dB design sidelobe level ($\bar{n} = 5$), while Figure-4 (b) show the pattern for the same planar array but with a -40 dB ($\bar{n} = 10$) design sidelobe level. The rise in the sidelobe level is readily noticed. Thus, confirming the first trend.

The second trend may be examined through Figure-5. Figure-5 (a) shows the pattern for a 12×12 elements array, while Figure-5 (b) shows the pattern for a 20×20 elements array. Both arrays have a -30 dB ($\bar{n} = 7$) Taylor distribution. It is obvious that the rise in the sidelobe level is smaller for the larger array.

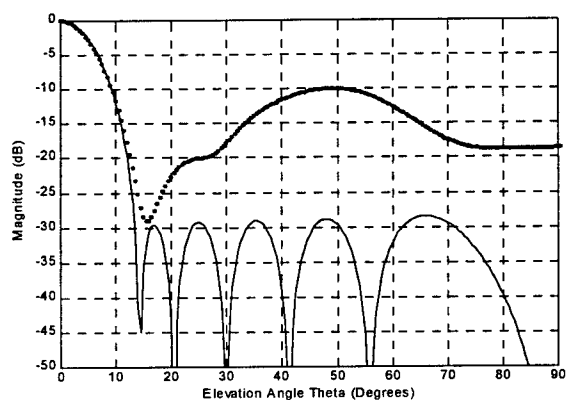


(a)

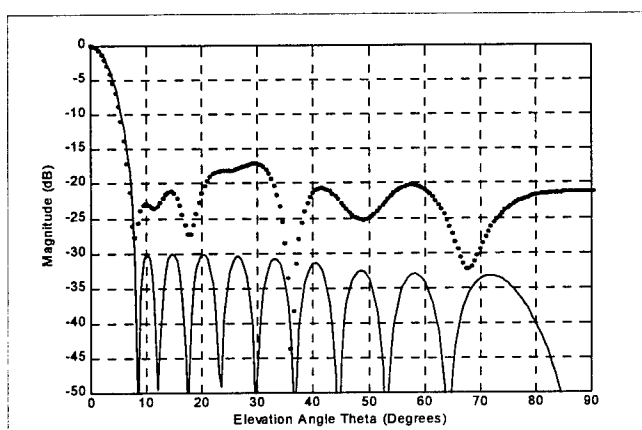


(b)

Figure-4. Error (dotted) vs. No-Error (solid) Pattern for 10 by 10 elements planar array with (a) -25 dB Taylor Distribution (b) -40 dB Taylor Distribution



(a)



(b)

Figure-5. Error (dotted) vs. No-Error (solid) pattern for a (a) 12 by 12 elements array (b) 20 by 20 elements array with -30 dB Taylor Distribution

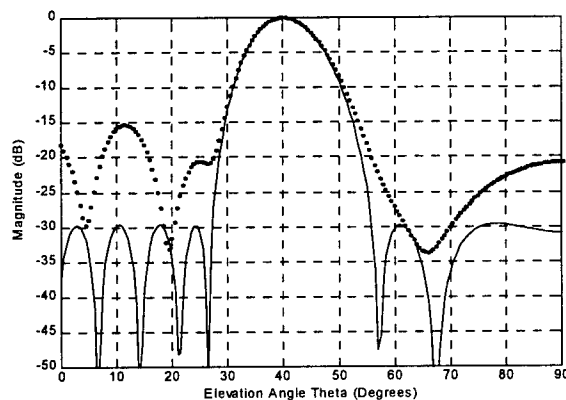
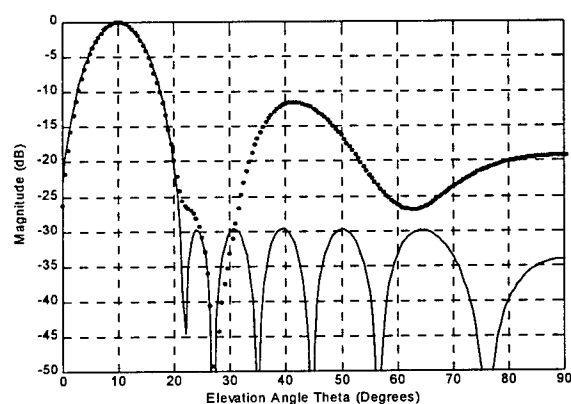


Figure-6 Error (dotted) vs. No-Error (solid) pattern for a 15X15 elements array scanned (a) scanned 10 degrees (b) 40 degrees from broadside

Figure-6 considers the third trend. In this case, the pattern of two 15 X 15 elements arrays are shown. Each array has a -25 dB ($\bar{n}=5$) Taylor distribution. A close look at the figure might suggest that in the average the sidelobe level variation be about the same in both cases. To examine the fourth trend, we consider Figure-2 and Figure-7. Figure-2 shows the error vs. no-error pattern for a 20 elements linear array with a -25 dB ($\bar{n}=5$) Taylor distribution, while Figure-7 shows the error vs. no-error pattern for a 20X20 elements array with the same illumination. Examination of the two figures confirms the trend. Finally, in Figure-8, the error pattern considering translational errors only and the error pattern considering excitation errors only are both plotted vs the no-error pattern. The array considered here is a 10 X10 elements array with a -25 dB ($\bar{n}=5$) Taylor Distribution. The solid line corresponds to the no-error pattern while the dotted line corresponds to the translational errors only pattern. The dashed line corresponds to the excitation errors only pattern. This plot superimposes precisely over the no error pattern. Thus, confirming the fifth trend.

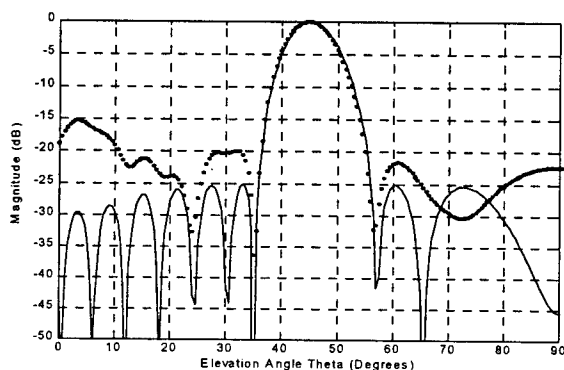


Figure 7. Error (dotted) vs. No-Error (solid) pattern for 20X20 planar array

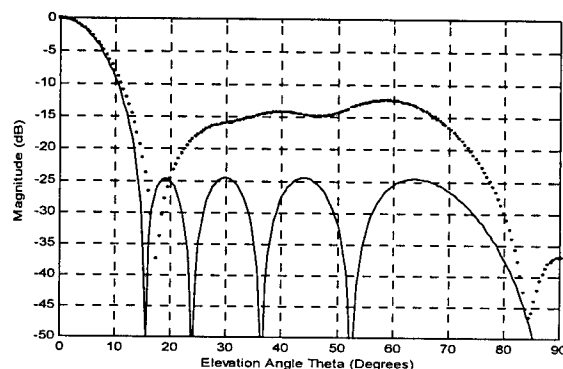


Figure-8 Error (dotted/dashed) vs. No-Error pattern for 10X10 planar array

For completeness, this section is concluded with a figure showing the “systematic-error-pattern” vs no-error of a 20X20 elements array with -25 dB Taylor distribution. The systematic errors considered here are flexing of the face of the array and N-bit quantization phase/time-delay shift error. The flexing of the array is assumed to vary linearly from the center to the edge of the array. In the results shown in Figure-9 the slope of the linear variation was chosen so that the maximum flexing would be 0.002 cm. Also, 3-bit phase-shifters are used to steer the main beam. It appears as if systematic errors had a slightly less impact on the array pattern than random errors.

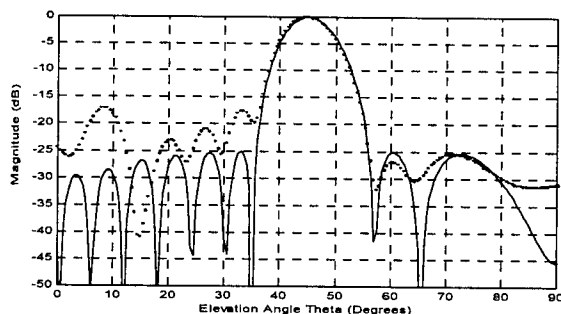


Figure-9 Systematic-Errors vs. No-Error Pattern for a 20 X 20 elements array

3.3 Array of Panels

In this case, we consider the field pattern of an array of panels. The specific array considered here is based on the low-altitude space-based radar (LASBR) described in the literature [7]. The LASBR is a 13.8m X 63.3m planar array. It consists of 49,152 elements distributed over 32 array panels. The specific arrangement of the panels as well as that of the elements within a panel is unknown. The gain for LASBR is 53 dB. It operates at a center frequency of 1.275 GHz. The array of panels considered here is a 13.1m X 63.3m array. It consists of 48,960 elements distributed over 32 panels. The panels are arranged in an array of 4 X 32 panels, with 90 X 17 elements per panel. The gain for this array is 38.8 dB. It is considerably lower than the LASBR, but these two arrays are not exactly the same. The array considered here also operates at 1.275 GHz. Figure-9 shows the no-error pattern for the array this array. Figure-9 shows the "random-error-pattern." The patterns are computed in the y-z plane. A considerable increase in sidelobe level may be observed in the error case. The RMS error in element location is 2mm, which may account for the considerably large increase in sidelobe level.

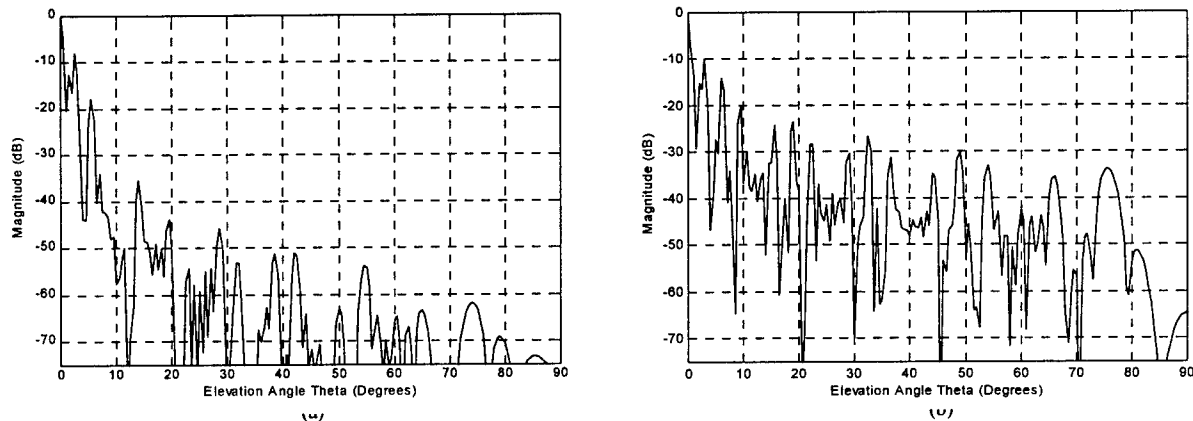


Figure-10. (a) No-Error Pattern (b) Error Pattern for a 4X32 panel array

4.0 Concluding Remarks:

A computer simulation to determine the field-pattern of arbitrarily oriented very large phase/time delay scanned antennas was developed. The simulation takes into account errors present in antenna arrays. These are systematic and random errors. The systematic errors considered here are the finite quantization of the phase produced by the use of digital N-bits phase/time-delay shifters and the flexing of the array aperture due to its large size and weight. The random errors considered are those caused by variations on the amplitude and phase of the elements current, variations on the radiation pattern of the elements, and missing elements (due to catastrophic failure), and variations in the location of the elements. To validate the simulation a number of patterns were computed. These included patterns of linear arrays, array panels, and arrays of panels. Experience as well as specific examples validated the ideal patterns. The "random-error-patterns" were compared to specific trends noted in earlier studies. The behavior of the computed patterns confirmed such trends.

A simulation like this is a useful tool for assessing the effect of tolerances on the performance of an antenna array. It can also be used as a design tool to determine required design parameters necessary to achieve a desired sidelobe level. There are some areas where the code could be improved. For instance, other types of element patterns could be considered. Moreover, the simulation could be modified to accept measured element patterns data.

5.0 References:

1. Collin, R.E., and Zucker, F.J. "Antenna Theory Part 1," McGraw-Hill , NY, 1969
2. Allen J.L. et.al. "Phased Array Antenna Studies: Tech Report NO. 236, 1 July 60 – 1 July 61, Lexington MA, Massachusetts Institute of Technology, Lincoln Labs, 13 Nov 1961
3. Wille Ng, et.al. "The First Demonstration of an Optically Steered Microwave Phased Array Antenna Using True-Time-Delay," Jou. Ligth.Tech.,Vol.9.NO.9,Sep.1991
4. Taylor, T.T. "Design of Line-Source Antennas for Narrow Beamwidth and Low Sidelobes," I.R.E. Transactions on Antennas and Propagation, AP-3. NO. 1: 16-28, January 1955
5. Villeneuve, A.T. "Taylor Patterns for Discrete Arrays," IEEE Trans. on Antennas and Propagations, AP-32: 1089-1093, October 1984
6. Chrisman, B.P. "Planar Array Antenna Design Analysis Volume I," MS Thesis, Air Force Institute of Technology, Wright-Patterson AFB, Dayton, OH, Dec 1989 (AD-A215 537)
7. Cantafio, L.J. Spaced-Based Radar, Artech House, 1989.

6.0 Acknowledgements:

The author wishes to thank Dr. Krishna Pasala for many helpful discussions. The support and interest of Dr. Stephen W. Schneider are acknowledged. Mr. John Mehr and Mr. Jim Mudd offered many helpful suggestions. The constant help received from Ms. Donna Gurnick at the AFIT Library is greatly appreciated.

FLUTTER PREDICTION METHODS FOR AEROELASTIC DESIGN OPTIMIZATION

**Ramana M. Pidaparti
Associate Professor
Department of Mechanical Engineering**

**Purdue School of Engineering and Technology, IUPUI
723 W. Michigan Street
Indianapolis, IN 46202-5132**

**Final Report for:
Summer Faculty Research Program
Wright Research Site**

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

and

Wright Patterson Air Force Base, Ohio

August 1998

FLUTTER PREDICTION METHODS FOR AEROELASTIC DESIGN OPTIMIZATION

Ramana M. Pidaparti
Associate Professor
Department of Mechanical Engineering
Purdue University at Indianapolis (IUPUI)

Abstract

In the design of flight vehicles, dynamic flutter instability is a critical parameter that must be considered in various design phases. Due to the numerically intensive nature of design optimization, accurate and efficient method of flutter prediction is needed. The objective of the report is to review and investigate the flutter prediction methods for realistic examples used in aircraft preliminary design. Specifically, demonstrate the flutter predictions with recently developed method using EigenVector Orientations (EVO). Comparisons of flutter predictions between the popular V-g method and Eigenvector Orientation method are presented. The examples studied include a sweptback untapered wing, jet transport wing/aileron, and intermediate complexity wing. Results presented illustrate that the EVO method can predict the onset of flutter for aircraft wings used in design optimization studies. Based on the results obtained, the EVO method shows the feasibility of automating the flutter prediction process and achieving real time control.

FLUTTER PREDICTION METHODS FOR AEROELASTIC DESIGN OPTIMIZATION

Ramana M. Pidaparti

Introduction

Design of modern flight vehicles requires multidisciplinary design optimization, MDO (integration of structures, aerodynamics, control and propulsion disciplines) to achieve an optimum design based on certain performance criteria and constraints. In the design of flight vehicles, flutter is a critical parameter that must be considered in the early stage of design cycle. Also, with the aging aircraft, certain restrictions need to be imposed to avoid flutter instability during the flight regime of the aircraft. A large number of parameter changes must be investigated in the flutter design of military as well as commercial aircraft due to complexity of modern composite/metallic structures. Due to the numerical nature of search optimization in MDO, accurate and efficient methods for flutter prediction are needed for design applications.

There are several aeroelastic design optimization codes being used for aircraft and spacecraft structures such as Automated Structural Optimization System (ASTROS) [1], and MSC-NASTRAN [2]. In aeroelastic design optimization, predominant dynamic flutter instability needs to be investigated for most flight vehicles. Several methods are being used to predict the onset of flutter instability in aeroelasticity design optimization. Most of these methods are based on complex eigenvalues (p-k method, K-method or KE method) [2]. Recently, Afolabi, Pidaparti and Yang [3] developed a method based on eigenvector orientations (EVO) to predict the onset of flutter instability. In their work, the method of eigenvector orientations was demonstrated for a limited examples of panel flutter. To gain a better understanding of the eigenvector orientation method of flutter prediction, further investigations need to be carried out on realistic aircraft structures. This understanding will help to devise a prediction methodology that may be implemented in MDO codes for preliminary design of flight vehicles.

The objective of this report is to review and investigate the flutter prediction methods for realistic examples used in aircraft preliminary design and further validate the eigenvector orientation method [3]. A brief review of commonly used flutter prediction methods is described in the next section. Three different examples are studied and the results of flutter prediction are compared between V-g method and EVO method. These comparisons may serve as a benchmark for predicting the onset of flutter condition for design applications.

Flutter Prediction Methods

The aeroelastic system of equations for an aircraft structural component can be written as

$$[M]\{\ddot{x}\} + [C]\{\dot{x}\} + ([K] + [K_A])\{x\} = \{0\} \quad (1)$$

where $[M]$ is the mass matrix, $[K]$ is the stiffness matrix, $[C]$ is the damping matrix due to aerodynamic or structural sources, and $[K_A]$ is the unsteady aerodynamic stiffness matrix. The above system of equations can be solved as a complex eigenvalue problem by assuming the unknowns $\{x\}$ to follow a harmonic motion. For a large system, normal modes approach can be used to reduce the aeroelastic system of equations. Usually, a free vibration analysis (by neglecting the damping and aerodynamic stiffness) is performed to obtain the natural frequencies and mode shapes. Then, using a sufficient number of lowest modes (usually 6-12 modes), the original aeroelastic system (Eq. 1) can be reduced as follows,

$$[m]\{\ddot{q}\} + [c]\{\dot{q}\} + ([k] + [k_A])\{q\} = \{0\} \quad (2)$$

where $[m]$, $[c]$, $[k]$, and $[k_A]$ are the modal mass, damping, structural stiffness and aerodynamic stiffness matrices, respectively. The system of equations in Eq. (2) results in a complex eigenvalue problem corresponding to a given value of reduced frequency (k) or aerodynamic parameter (q). The flutter boundary is obtained when the lowest two eigenvalues coalesce at a critical value of k or q (assuming no aerodynamic damping), or the damping (g) of one of the modes is zero when crossing from negative (stable) to positive (unstable) values. Different solution techniques are being used in the literature to obtain the critical flutter boundary (flutter speed and frequency) which are briefly described below.

K- Method (V-g analysis)

In the K-method of flutter solution, the complex eigenvalue problem for V-g aeroelastic analysis can be written as,

$$\{(1 + ig)[k] - \omega^2([m] + [A])\}\{\phi\} = \{0\} \quad (3)$$

where $[k]$ and $[m]$ are the modal stiffness and mass matrices, and $[A]$ is the complex aerodynamic matrix derived from some aerodynamic theory, for example, doublet lattice method for subsonic flight regime [5]; constant pressure panel theory [6] or piston theory [7] for supersonic flight regime. In equation (3), the parameter g is the artificial structural damping, and ω is the frequency. Rewriting the above equation as,

$$\{[k] - \lambda([m] + [A])\} \{\phi\} = \{0\} \quad (4)$$

where $\lambda = \omega^2/(1+ig)$ is the complex eigenvalue and $\{\phi\}$ is the complex eigenvector. The complex eigenvalue can be written as,

$$\lambda = a + ib \quad (5)$$

where the damping, g is defined as $(-b/a)$ and the velocity V is given by

$$V = \sqrt{-(a^2 + b^2)/a} \quad (6)$$

The frequency, f (Hz) is given in terms of reduced frequency, k by

$$f = \frac{kV}{\pi c} \quad (7)$$

In the K-method flutter solution, the complex eigenvalue problem is solved for each reduced frequency (k) or velocity (V) until the flutter condition is obtained (damping becomes zero). The damping values (g) for each mode is plotted against the airspeed (V) in a V-g plot. The flutter reduced frequency (k_f) is obtained when the damping of one of the modes is zero when crossing from a negative value to a positive value. Damping may not be a single valued function of velocity (V) in the K-method when multiple eigenvalues exist in the system. When reduced frequency or velocity $1/k$ is used, the damping becomes single valued function. The flutter speed (V_f) is obtained from k_f , and then the flutter frequency is obtained using Eq. (7). Misidentification of flutter stability may be obtained if mode crossings are not properly tracked during flutter analysis [4].

The K-method has been extended to KE method in MSC/NASTRAN by neglecting the damping and solving the resulting eigenvalue problem as an efficient way to find the flutter condition. The modes are extracted in a specific order and tracked to determine the flutter instability when damping becomes zero for one of the modes similar to the K-method. This method does not extract eigenvectors from the flutter analysis.

p-k Method

In the p-k method of flutter solution, the complex eigenvalue problem for V-g aeroelastic analysis can be written as,

$$\{[k] - q[A] + \lambda[m]\} \{\phi\} = \{0\} \quad (8)$$

where

$$\lambda = (V/b)^2 p^2 = \omega^2 (\gamma + i)^2 \quad (9)$$

In the above equation, p is the root of the p-k aeroelastic analysis equation (8) which is defined as $p=k(\gamma+i)$, with γ being the true damping in the system. The aerodynamic matrix $[A]$ in the p-k method differs from that of the K- method by a factor which is given as $2k^2/\rho b^2$. In the p-k method flutter solution, the complex eigenvalue problem (Eq.8) is solved for each velocity (V) until the flutter condition is obtained. The damping values (γ) for each mode is plotted against the airspeed (V) in a V-g plot to determine the flutter velocity.

An iterative loop must be carried out to match the p root for the $[A]$ matrix in terms of assumed k until a convergence is achieved. Convergence in terms of under- or overrelaxation on k values may not be obtained at or near the flutter boundaries. A more robust method of convergence for k is needed for use in commercial software for design applications. Also, there is a possibility that more than n modes at a given airspeed (V) may match the p root. These additional modes represent aerodynamic poles rather than true roots of the p-k aeroelastic equation. These extraneous roots may cause problems in mode tracking algorithms [4].

Both the above two methods have been implemented in MSC/NASTRAN software, and only p-k method has been implemented in ASTROS software. A systematic study of flutter problems by Tischler et al. [9] revealed a good agreement between NASTRAN and ASTROS softwares.

Eigenvector Orientation (EVO) Method

Both K-method and p-k method flutter solutions are based on the complex eigenvalues (tracking zero values of modal damping) to predict the onset of flutter instability. In other words, when the damping is ignored, the flutter instability is found when the lowest two eigenvalues coalesce and become complex conjugate pairs.

The aeroelastic system of equations in terms of airspeed or reduced frequency parameter (Eqs. 4 or 8) when cast into a complex eigenvalue problem can be solved to get eigenvalues and eigenvectors. Previous studies in the literature focussed mostly on the complex eigenvalues but not on the complex eigenvectors. Recently, Afolabi, Pidaparti and Yang [3] used the idea of eigenvector orientation based on complex eigenvectors to predict the onset of flutter instability. The eigenvectors initially real and orthogonal to each other, lose their orthogonality at the flutter instability. So, when the angle between any two eigenvectors deviates from 90 degrees indicates the extent to which the aeroelastic system is close to the flutter boundary. Therefore, in the EVO method, the angle between any two complex eigenvectors (any modes), or its deviation from 90 degrees or when the rate of change of EVO angle is zero, can be used to track the flutter condition. For real time flutter control process, one can use the EVO method as it can be easily monitored the EVO angle close to zero indicating the flutter condition. In addition, the EVO method serves as a complement to the existing methods based on eigenvalues.

The angle between two complex vectors is calculated by mapping a complex vector into a real vector by grouping the imaginary components after the real components. For example, the angle between two eigenvectors (ϕ_1 and ϕ_2) can be found as,

$$\theta_{12} = \cos^{-1} \left(\frac{\phi_1 \cdot \phi_2}{\|\phi_1\| \|\phi_2\|} \right) \quad (10)$$

The angle between any two eigenvectors is 90 degrees initially (orthogonal modes) and approaches zero at the flutter instability (modes are no longer orthogonal) when the airspeed (V) or reduced frequency (k) is varied. This method of tracking flutter condition is completely different from V-g plot using K and p-k methods. The EVO method of predicting the onset of flutter instability has never been implemented in any general purpose aeroelastic design optimization software.

Applications

Three examples were selected (see Fig. 1) to demonstrate the flutter prediction based on EVO analysis as discussed in the previous section. For each of the examples, MSC/NASTRAN software was used to calculate the flutter condition based on V-g plot using PK method. A post processing program was developed to extract the complex eigenvectors, and determine the angle between them by the EVO method. Results of flutter boundaries are presented to illustrate the relative merits of EVO method compared to V-g method for tracking the onset of flutter condition. The examples studied range from subsonic to supersonic flight regimes, and transport aircraft wing and sweptback fighter type wing to demonstrate the further applicability of the EVO method for design applications.

Example 1: A 15-degree Sweptback Wing

A 15° swept untapered wing model (see Fig. 1) made from magnesium which was tested in a wind tunnel is analyzed for supersonic flutter analysis. The wing is modeled using 28 thin shell QUAD4 elements (40 nodes) and is fixed at the root. The elements representing leading and trailing edge taper in thickness to zero whereas the rest of the elements have a thickness of 0.041 inches. For the aerodynamic model, the wing is divided into four chordwise regions and six equal width strips. Piston theory is used in the analysis. More details about structure and aerodynamic model can be found in Ref. [2]. The flutter analysis results are presented in Figure 2. The results of eigenvalues and damping for the lowest three modes are given in Figs. 2 (a) and (b). The angle between the lowest three eigenvectors (1,2 and 3) is presented in Fig. 2c. The V-g plot gives a flutter velocity of 2077 ft/s compared to an experimental value of 2030 ft/s [2]. The flutter velocity predicted from EVO method (Fig. 2c) is about 2067 ft/s which compares well with experiments as well as V-g analysis.

The same 15° swept wing is analyzed for low supersonic conditions ($M=1.3$) using Mach Box lifting surface aerodynamic theory. Twenty control points were specified for the aerodynamic model. An internal surface spline in Nastran interpolates between aerodynamic grid points and Mach Box centers. The calculated aerodynamic forces are transferred to structural nodal points with a linear surface spline. The flutter analysis results are compared in Fig. 3. It can be seen the trend for EVO angle is similar to the previous example. The flutter velocity predicted from EVO method (Fig. 3c) is about 1630 ft/s which compares well with V-g analysis (1582 ft/s) using NASTRAN and 1683 ft/s using ASTROS softwares. In both the above cases, the angle between the two eigenvectors corresponding to lowest modes reaches a minimum (about 12°) and increases again. This is a completely different trend from the panel flutter examples showed by Afolabi, Pidaparti and Yang [3]. In all their examples, the angle between the eigenvectors reached approximately zero condition at the flutter instability. However, in the present case, the flutter condition is reached when the two modal eigenvectors are not completely orthogonal (close to losing orthogonality). Therefore, the flutter criteria may corresponds to a situation when the EVO angle reaches a minimum and increases again (rate of change of EVO angle is zero, as in Fig. 3(b)).

Example 2: Jet Transport Wing/Aileron

This example is taken from MSC/NASTRAN Aeroelastic Analysis manual [2]. The structural model consists of inputting the flexibility influence coefficients at 10 grid points. The grid point 11 corresponds to the rigid fuselage at the root of the wing elastic axis. Strip theory aerodynamics with an approximation to circulation function was used. The wing is idealized into 58 aerodynamic boxes (five for fuselage and six aileron) for aerodynamic model in the flutter analysis. Figure 4 shows the results of flutter speed using V-g analysis and EVO method. The rate of change of EVO angle and damping is presented in Fig. 4(b). The EVO method predicts the flutter speed of 1183 ft/s whereas V-g method predicts the flutter speed to be 1133 ft/s. A good agreement is seen in this case too.

Example 3: Intermediate Complexity Wing

The intermediate complexity wing (ICW) example is taken from ASTROS software and has been used in many applications in MDO. The wing geometry is given in Fig. 1. The geometry of ICW is 30° sweptback wing with upper and lower skin surfaces consisting of balanced composite lay-up of $(0^\circ/\pm 45^\circ/90^\circ)$ with 0° fibers aligned with the midspar of the wing. The structural model of the wing (total 88 nodes) consists of 62 quadrilateral and 2 triangular membrane elements (skins), 55 shear panels (ribs and spars), and 39 rod elements (posts). The wing is cantilevered at the root. For aerodynamics, the wing is modeled as a flat plate lifting surface with 72 boxes (9 spanwise and 8 chordwise with unequal spacing). The aero structural interconnection is defined by two surface splines. Flutter analysis is carried out at $M = 0.7$ using PK method. Figure 5 shows the results of flutter speed using V-g analysis and EVO method. The EVO method predicts the flutter speed around 800 ft/s whereas V-g method predicts the flutter speed to be 939 ft/s and 926 ft/s using NASTRAN and ASTROS softwares, respectively. The difference in flutter speed between EVO and V-g analyses may be attributed to the fact that the present finite element model is not accurate enough to give good mode shapes. It will be interesting to remesh ICW and see how the EVO angle predicts the flutter speed.

The three different examples studied in this investigation illustrate that EVO method of flutter prediction indeed predicts results comparable to V-g method and experiments. The question now is to address the possible benefits of the EVO method as compared to the V-g method. In the EVO method, the angle reduces from 90° monotonically, definitely reaches a minimum and increases again. If one is to use real time automatic tracking of flutter instability, tracking the EVO angle from 90° seems an easy task compared to tracking damping (which could be multiple valued) for any general case.

Concluding Remarks

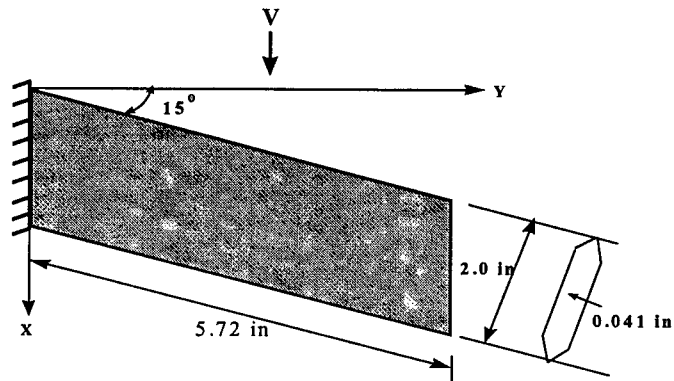
Flutter prediction methods based on complex eigenvalues is reviewed along with recently developed method based on complex eigenvectors. Comparisons of flutter predictions between the popular V-g method and the new Eigenvector Orientation method were made for three different examples. Good comparison of the flutter speeds between EVO method and V-g method illustrate the applicability of EVO method for realistic aircraft wings. One limitation of EVO method of flutter prediction may be the mesh dependancy to get accurate mode shapes. Based on the examples studied, the EVO method shows the feasibility of automating the flutter prediction process and achieving real time control.

Acknowledgements

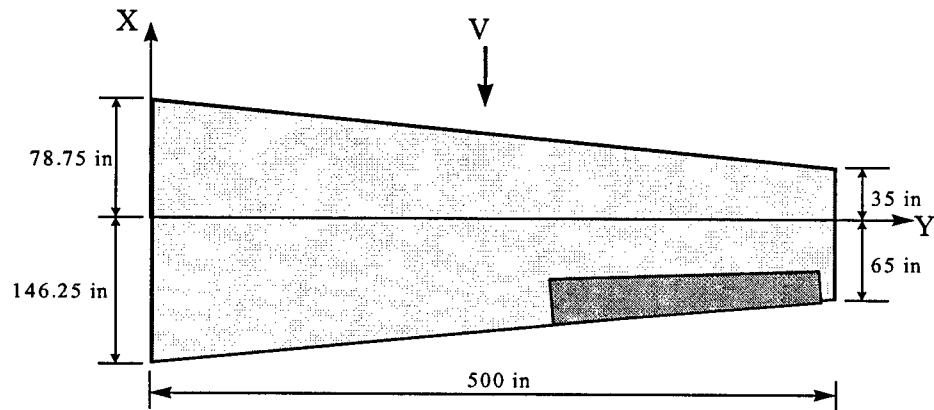
The author thanks Dr. Venkayya for his guidance and support, and to Ms. Tischler for help with NASTRAN runs. Support from the AFOSR at the Wright Patterson Air Force Base, Ohio is acknowledged.

References

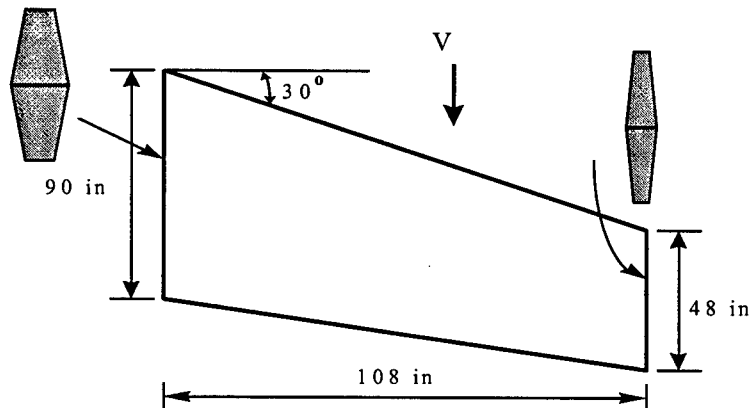
1. Rodden W.P. and E.H. Johnson. (ed.) "MSC/NASTRAN Handbook for Aeroelastic Analysis, Version 68," MacNeal-Schwendler Corp., 1994.
2. Johnson, E.H. and Venkayya, V.B., "Automated STRuctural Optimization System (ASTROS), Volume I, Theoretical Manual, Airforce Wright Patterson AFB, AFWAL-TR-88-3028, Wright Patterson Air Force Base, Ohio, 1988.
3. Afolabi, D., Pidaparti, R.M.V. and Yang, H.T.Y., "Flutter Prediction using Eigenvector Orientation Approach," AIAA Journal, Vol. 36, No.1, pp. 69-74, 1998.
4. Eldred, M.S., Venkayya, V.B. and Anderson, W.J., "New Mode Tracking Methods in Aeroelastic Analysis," AIAA Journal, Vol. 33, No. 7, pp. 1292-1299, 1995.
5. Albano, E., and Rodden, W.P., "A Doublet-Lattice Method for Calculating Lift Distributions on Oscillating Surfaces in Subsonic Flows," AIAA Journal, Vol. 7, No.2, pp. 279-285, 1969.
6. Blair, M., "A Compilation of the Mathematics Leading to the Doublet Lattice Method," WL-TR-92-3028, 1992.
7. Appa, K., "Constant Pressure Panel Method for Supersonic Unsteady Airloads Analysis," Journal of Aircraft, Vol. 24, No. 10, pp. 696-702, 1987.
8. Lighthill, M.J., "Oscillating Airfoils at High Mach Number," Journal of Aerospace Sciences, Vol. 20, No. 6, pp. 402-406, 1953.
9. Tischler, V.A., Kolonay, R.M., and Venkayya, V.B., "Unsteady Aeroelastic Analysis Capabilities in NASTRAN", Presented at the NASTRAN User's Conference, 1997.



Example 1: 15-degree Sweptback Wing



Example 2: Jet Transport Wing/Aileron



Example 3: Intermediate Complexity Wing

Fig. 1 Three Different Examples Studied for Flutter Predictions

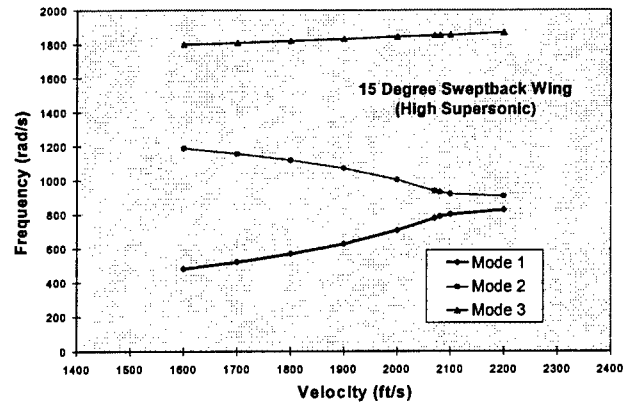


Fig. 2(a) Frequency Versus Velocity of Three Lowest Modes for 15 deg. Sweptback Wing

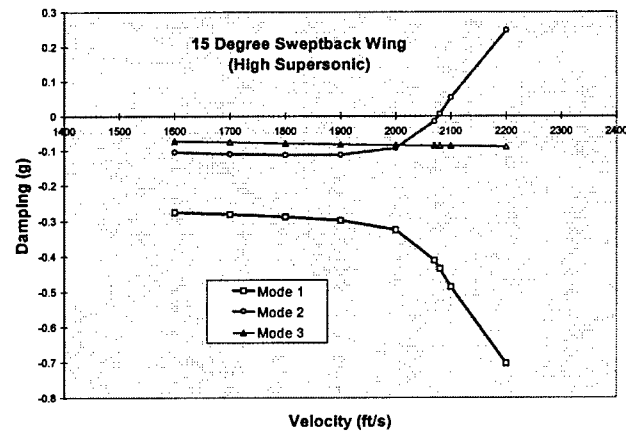


Fig. 2(b) Damping Plot for 15 deg. Sweptback Wing under High Mach Number

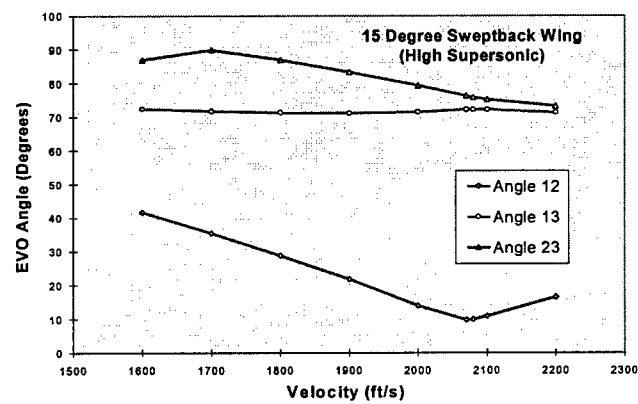


Fig. 2c EVO Angle versus Velocity Plot for 15 deg. Sweptback Wing under High Mach Number

Fig. 2 Comparison of Eigenvalue Coalescence, Damping and EVO angle for 15 degree Sweptback Wing (High Supersonic)

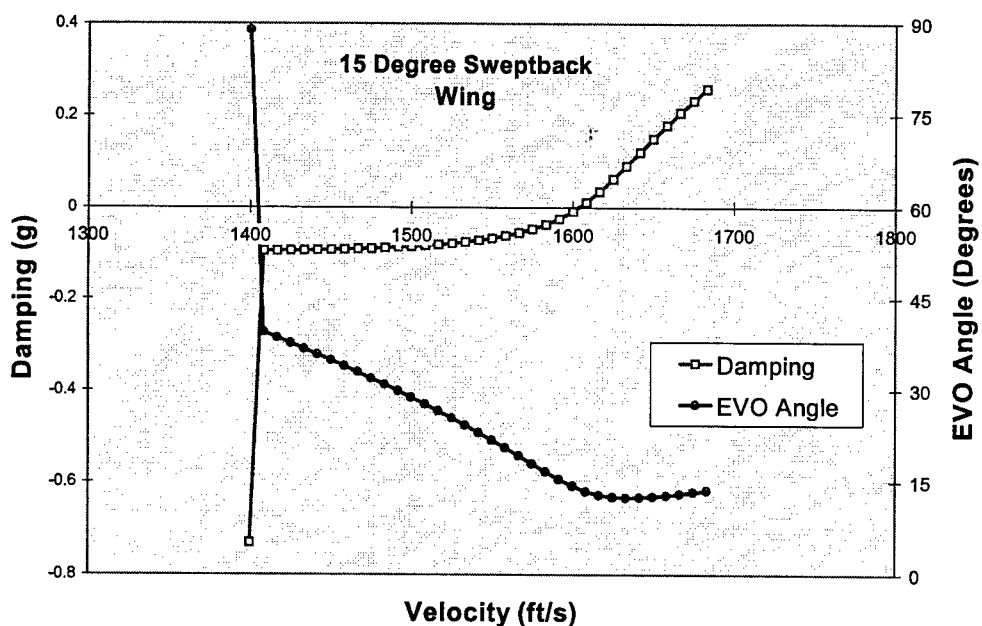


Fig. 3(a) Comparison of V-g versus V-EVO angle plots for 15 degree Sweptback Wing (Low Supersonic)

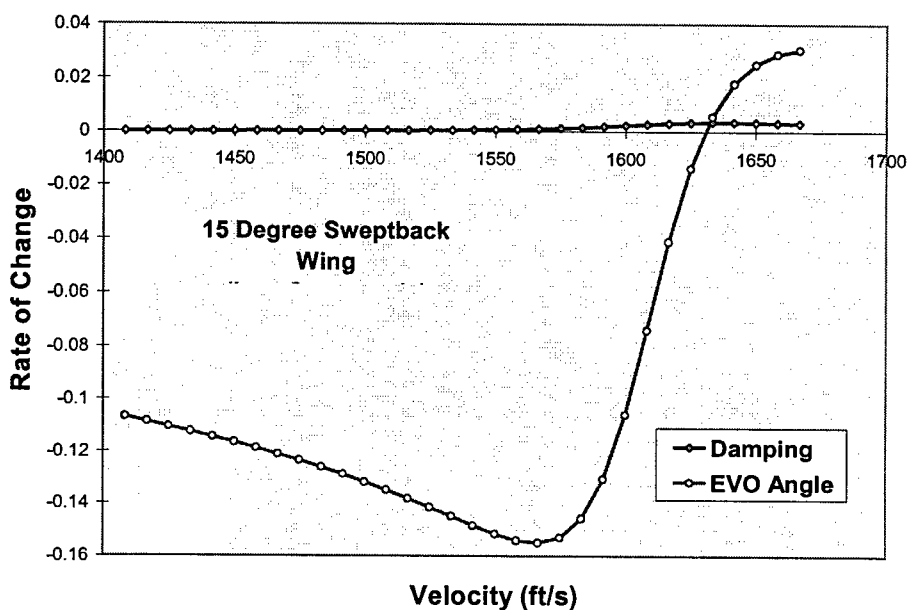


Fig. 3(b) Rate of Change of Damping and EVO Angle with Velocity for Low Supersonic 15 Deg. Sweptback Wing

Fig. 3 Comparison of Damping and EVO angle and their rate using V-g method and V-EVO method for 15 degree Sweptback Wing (Low Supersonic)

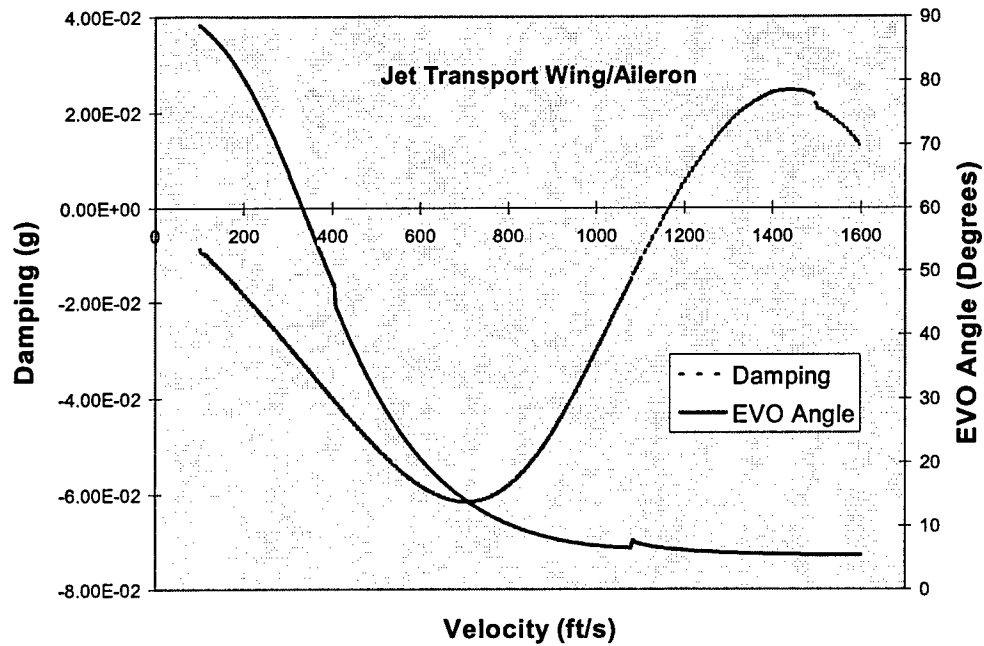


Fig. 4(a) Comparison of V-g and V-Evo Angle plots for Jet Transport Wing/Aileron

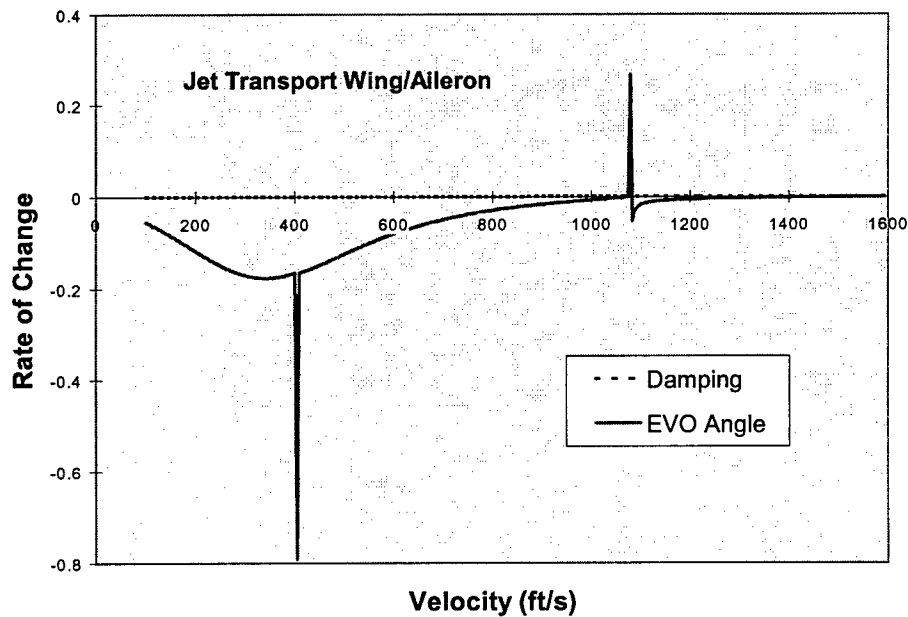


Fig. 4(b) Rate of Change of Damping and EVO Angle for Jet Transport Wing/Aileron

Fig. 4 Comparison of Damping and EVO angle and their rate using V-g method and V-EVO method for Jet Transport Wing/Aileron

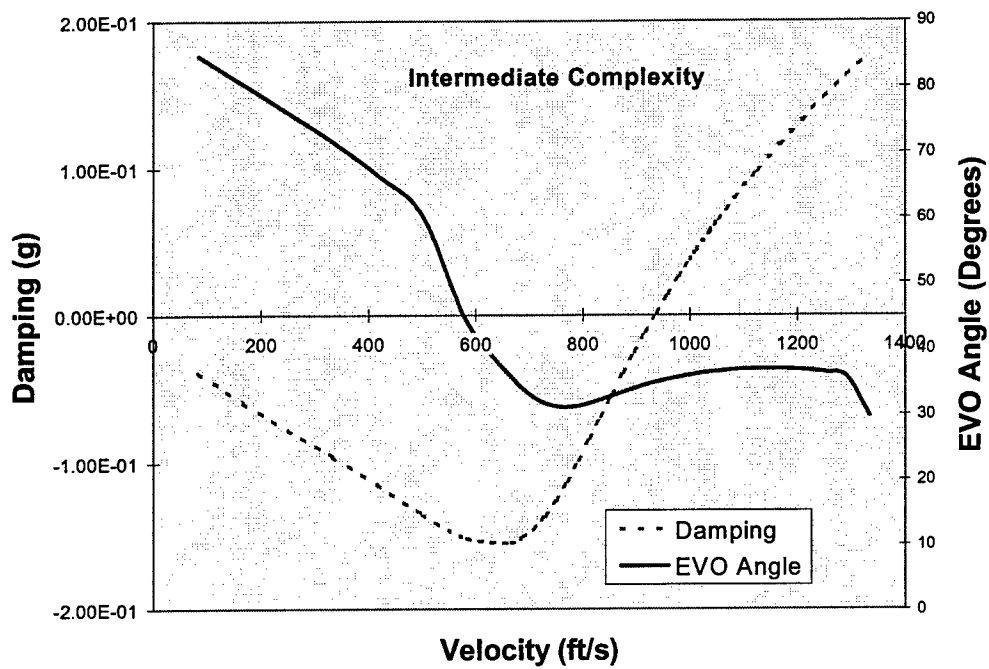


Fig. 5(a) Comparison of V-g versus V - Evo angle plots for Intermediate Complexity Wing

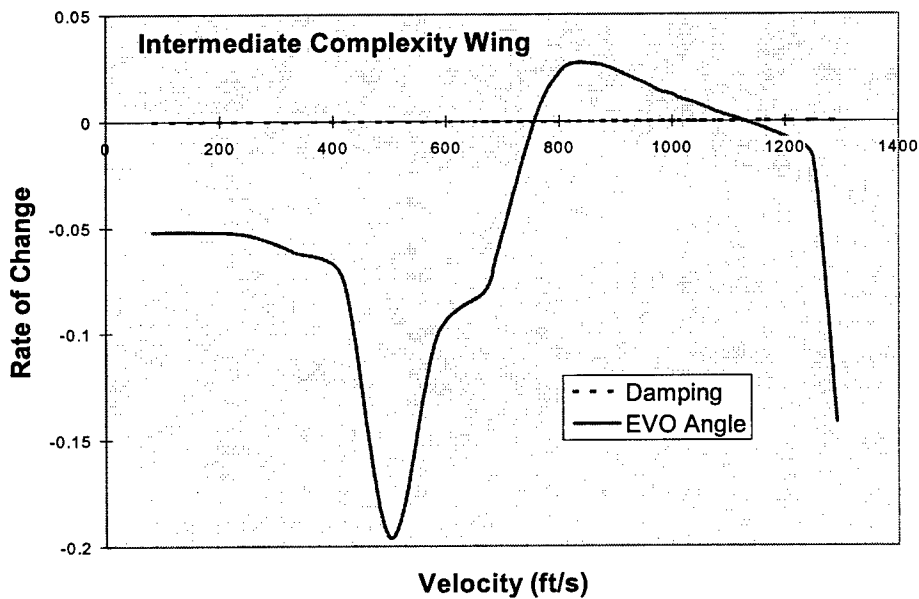


Fig. 5(b) Rate of Change of Damping and EVO Angle for Intermediate Complexity Wing

Fig. 5 Comparison of Damping and EVO angle and their rate using V-g method and V-EVO method for Intermediate Complexity Wing

Characterization of BN-Doped SiC Epitaxial Layers

Stephen E. Sadow
Assistant Professor
Emerging Materials Research Laboratory
Department of Electrical & Computer Engineering

Mississippi State University
Box 9571
Mississippi State, MS 39762

Final Report for:
Summer Faculty Research Program
Air Force Research Laboratories
Wright Patterson AFB

Sponsored by:
Air Force Office of Scientific Research
Bolling Green Air Force base, DC

And

Air Force Research Laboratories
Wright Patterson AFB

August 1998

Characterization of BN-Doped SiC Epitaxial Layers

Stephen E. Sadow
Assistant Professor
Emerging Materials Research Laboratory
Department of Electrical & Computer Engineering
Mississippi State University

Abstract

The electrical properties of 4H-SiC epitaxial layers on semi-insulating substrates were studied. The epitaxial layers were grown by the Emerging Materials Research Laboratory (EMRL) using an existing cold-wall chemical vapor deposition (CVD) system. These layers were doped during growth using a solid-phase boron-nitride source, which had previously been shown to dope the layers with boron, nitrogen and aluminum. Capacitance-Voltage (C-V), cathodoluminescence (CL), Hall effect, thermal admittance spectroscopy (TAS) and optical admittance spectroscopy (OAS) were used to characterize these layers. CL data indicated the presence of the boron-related D-center ($E_A = E_V + 0.60 \pm 0.2$ eV), which has a characteristic green luminescence when donor to D-center recombination occurs. TAS showed evidence of a deep level but sample instabilities prevented complete characterization from being performed. OAS data was gathered and compared with semi-insulating substrate material from Cree Research Inc. which permitted a direct comparison between observed optically-active defects in both materials. Finally Hall data indicate the possibility of hopping conduction in slightly p-type layers, while in slightly n-type layers there was a measurable mobility. However, characterization of thin films on insulating substrates using the Hall Effect was inconclusive and will require further study.

Characterization of BN-Doped SiC Epitaxial Layers

Stephen E. Saddow

1. Introduction

Silicon carbide, despite on-going efforts to develop a reliable implantation doping technology, remains an epitaxial technology. Chemical vapor deposition (CVD) remains the principal means by which device quality SiC semiconducting films are obtained with specific doping concentrations. There are three principal dopants for "shallow" donor and acceptor doping, they are nitrogen, aluminum, and boron. Nitrogen doping for donor formation is a relatively satisfactory technology. Acceptor doping favors aluminum because aluminum forms a shallower acceptor level than does boron. However, interest remains for boron acceptor doping. More recently, suggestions for using the boron-related D-center deep-level defect (reported to lie between $N_V + 0.58$ eV and $N_V + 0.73$ eV) to compensate nitrogen donors to produce semi-insulating SiC epitaxial layers have been published.¹ This paper provides experimental results to suggest that codoping SiC with boron and nitrogen from a boron-nitride solid source may provide an economical means for achieving semi-insulating epitaxial layers.

The Emerging Materials Research Laboratory (EMRL) at Mississippi State University (MSU) has been developing a process for growing semi-insulating (SI) epitaxial layer. This technique involves the use of a boron-nitride (BN) solid-source dopant scheme whereby radiant heating of a BN disk is sufficient to out gas boron, nitrogen and aluminum from the BN disk. These elements are then transported to the SiC growth region via a hydrogen carrier gas, the result being the doping of the resulting epitaxial layer with all three species. Numerous applications for SI SiC epitaxial layers are envisioned, including intrinsic (i.e., "i") layers for p-i-n diodes, field termination structures for high-voltage power devices² and buffer layers for high-frequency devices.

Experimental results have been reported for boron-doping 6H silicon carbide epitaxial layers from a solid boron-nitride source. Boron doping concentrations from more than 10^{18} cm⁻³ to less than 10^{15} cm⁻³ were observed.³ Doping concentrations were correlated with the temperature of the BN source during CVD growth of the 6H SiC. Donor co-doping, most likely from nitrogen, was also observed, in some cases leading to substantial compensation. Unfortunately, the exact physical mechanism responsible for high resistivity of this material cannot be extracted when these layers are grown on conducting substrates. Since semi-insulating (SI) substrates in 4H-SiC are commercially available, the EMRL BN-doping process was transitioned to SI substrate material, which should permit a more accurate assessment of the BN-doped film to be made.

2. Epitaxial Layer Samples

A standard silane/propane based precursor growth technology was used to grow several 4H-SiC epitaxial layers on 4H-SiC substrates. A 50-mm diameter quartz reaction tube with an outer water jacket to permit "cold wall" processing was used. This system is substantially similar to

that developed by the group at the NASA Lewis Research Center. A 30-mm silicon-carbide coated graphite susceptor is supported on a quartz boat that is inserted into the reaction tube. One or more 1 cm square pieces of 4H silicon carbide rest on the susceptor. A radio-frequency induction coil surrounds the quartz reaction tube, which heats the susceptor to the growth temperature. For the results reported here, the growth temperature at the substrate was typically maintained around 1550 C, as measured with an optical pyrometer and corrected for absorption through the quartz. The 4H substrate pieces were cut from SI Cree wafers polished on one side (Si face) which were provided by Dr. W. Mitchel of the AFRL Materials Laboratory. Prior to insertion into the reactor, the substrate pieces were cleaned ultrasonically with solvents, rinsed in deionized water, and blown dry with nitrogen.

Co-doping with boron and nitrogen from a solid source was provided by a single piece of boron nitride cut from a two-inch wafer specifically manufactured as a boron diffusion source for use in silicon processing. The trailing edge of the BN was placed a measured distance upstream of both the susceptor and the silicon-carbide substrate (see Fig. 1). The substrate was typically between 10 and 15 mm downstream of the leading edge of the susceptor. While the BN was never placed on the heated susceptor, on one occasion it was allowed to touch the upstream ("leading") edge of the susceptor (the "0-mm" case). In the samples prepared for this SFRP effort, the BN was placed 3 mm from the susceptor, since growth runs on conductive substrates indicated this situation resulted in highly-resistive material

The typical growth run proceeded as follows. After loading and purging with ultra-high purity (UHP) argon followed by UHP hydrogen, the reaction tube was evacuated with a mechanical pump several times. Each time, the reaction tube was back filled with UHP hydrogen, and then pumped back down. Following the pump-purge cycles, a steady atmospheric-pressure UHP hydrogen flow was established at a rate of 3 l/m. The RF power source was then engaged and the susceptor rapidly heated to the growth temperature (typically requiring between 1 and 2 minutes). As the growth temperature was reached, flows of 3% silane (SiH_4) in hydrogen and 3% (by volume) propane (C_3H_8) in hydrogen commenced. The flow rates were metered continuously with mass-flow controllers, and were set to 15 sccm for silane while the propane flow was varied to set a desired silicon-to-carbon atomic ratio in the growth tube. (Experience with epitaxial layer growth on conductive substrates indicated that a silicon to carbon ration on the order of 0.16-0.18 should yield SI epi.) Growth proceeded for 75 minutes and was terminated by the following purge sequence implemented by a programmable logic controller: The C_3H_8 and SiH_4 flows were terminated, followed by 30 seconds of UHP H_2 flow at 3 l/m and at growth temperature, followed by 4 minutes of UHP argon flow at 3 l/m and at growth temperature, followed by RF power shutoff and susceptor cool down for 10 additional minutes in UHP argon. Upon removal, both the BN pieces and the SiC pieces were inspected. The surfaces of the SiC pieces invariably had a specular or near specular reflection, indicating good quality homoepitaxial layers. Generally, this was confirmed by optical microscopy, which did reveal the usual epitaxial-layer morphological imperfections, some of which were clearly related to substrate issues. Except in the cases previously noted, the BN pieces were largely unchanged. The SiC pieces were then transported to the AFRL by the principal investigator for characterization, the focus of this SFRP effort.

Epitaxial Sample Summary

Various epitaxial layers were grown by EMRL for characterization during the SFRP at the AFRL. The details of the epitaxial layers are listed in Table 1 for reference. Each sample was grown to achieve a specific objective to help develop a full understanding of the electrical properties of EMRL BN-doped epitaxial layers.

Table 1 EMRL SiC Epitaxial Layer Growth Summary

Sample No.	Si/C	Net Doping Density [cm ⁻³] ^{A)}	Net Doping Density [cm ⁻³] ^{E)}	Characterization Planned ^{F)}
MSU-98-4H-51	0.16	$2-6 \times 10^{14}$ (p-type)	—	Hall, CV
MSU-98-4H-52	0.18	$\sim 10^{14}$	$\sim 10^{14}$	CV, CL, OAS, TAS
MSU-98-4H-53 ^{B)}	0.18	$\sim 10^{14C)}$	—	Hall
MSU-98-4H-56	0.18	$10^{14}-10^{15}$	$10^{14}-10^{15}$	Hall, CV, TAS
MSU-98-4H-57 ^{D)}	0.20	$\sim 10^{14}$	$\sim 10^{14}$	Hall, CV, TAS
MSU-98-4H-69a	0.20	See note G	See note H	Hall
MSU-98-4H-69b	0.20	“	“	CV, OAS, TAS
MSU-98-4H-70a	0.18	“	“	Hall
MSU-98-4H-70b	0.18	“	“	CV, OAS, TAS

^{A)} as determined by Hg-probe C-V at EMRL.

^{B)} a degenerate cap layer was grown in-situ after BN-doped epi growth.

^{C)} cap layer conc. only measured. BN-doped epi conc. should be identical to sample 52.

^{D)} BN doped with x = 3.5 mm.

^{E)} as determined by Schottky contact C-V at AFRL

^{F)} estimated epi thickness $\sim 2 \mu\text{m}$ for all samples except MSU-98-4H-57 which is $\sim 3 \mu\text{m}$

^{G)} Hg-probe data too noisy for concentration estimation

^{H)} CV not measured at AFRL at time of this report

Sample Preparation

Hall Effect and Admittance Spectroscopy characterization require that metal contacts be placed onto the sample. In the first case, Hall Effect, these contacts must display “ohmic-like” behavior, which can be achieved in one of two ways. The classical approach is to select a metal whose work function is such that ohmic behavior is achieved. For semiconductors where such a suitable metal is not available, ohmic contact is achieved via the use of a heavily-doped degenerate layer which ensures that a Schottky Tunneling contact is realized. Admittance Spectroscopy requires

only a Schottky contact, and thus a degenerate layer is not required. Cathodoluminescence (CL) does not require well-defined contacts, but rather contact between the sample surface and the low-temperature cryostat such that excess charge can be bled off of the sample surface. Indium solder was used to achieve this goal, after which the In was removed using a HCL and Nitric Acid dip.

For sample MSU-98-4H-53 electrical contact was formed to the N^+ degenerate layer via the deposition of $1,000\text{\AA}$ of Ni which was subsequently annealed at 950°C in a rapid thermal annealing (RTA) system. The sample was then loaded into a reactive ion etching (RIE) system⁴ to remove the N^+ epitaxial layer between the Ni contacts (see Figure 1). A total of $0.5\mu\text{m}$ was removed, ensuring that the electrical conduction between the Ni contacts was due to the underlying BN-doped epitaxial layer only. After RIE, the electrical behavior between the contacts were observed to be that of a pn junction, indicating that the underlying epitaxial layer is not purely SI but slightly p-type.

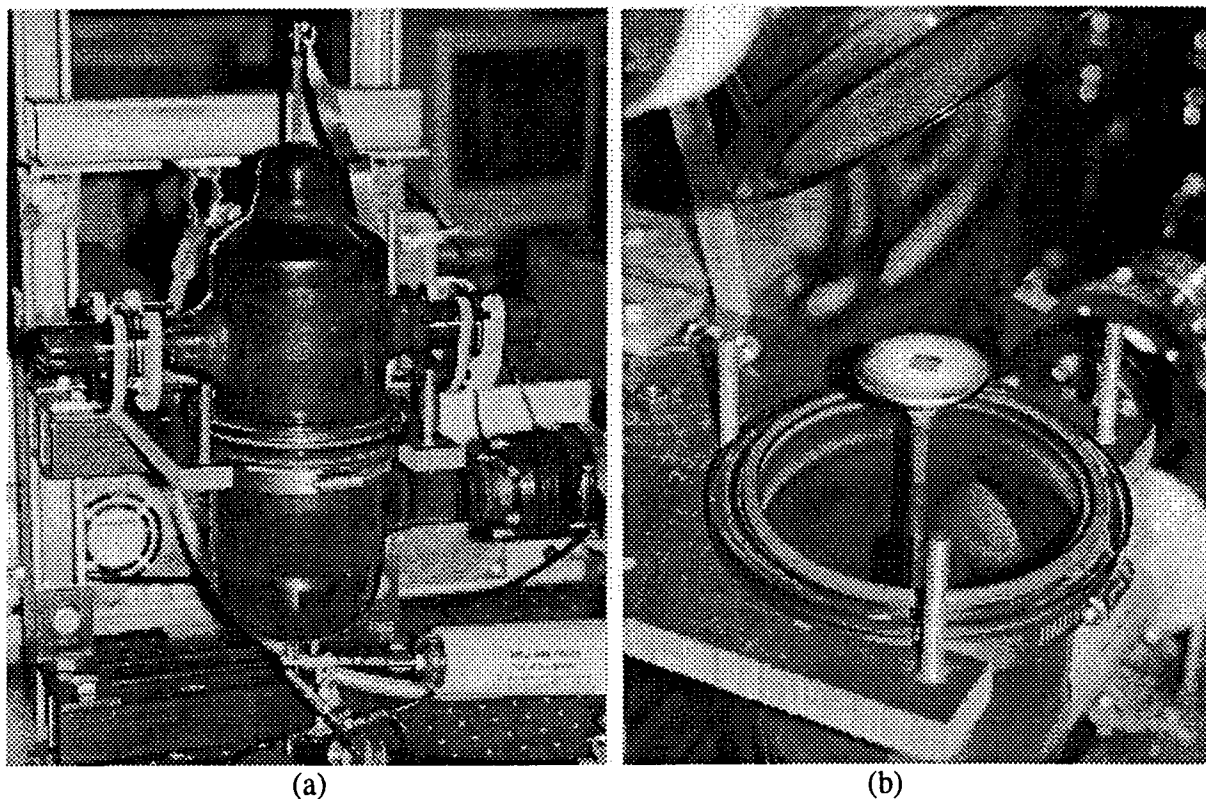


Figure 1 AFIT RIE system used to etch MSU SiC cap layers. (a) Photo of system with sample loaded. (b) MSU-98-4H-53 sitting on parallel plate electrode inside RIE system after etch. The Van der Pauw contacts are visible in the image.

Using a standard ohmic contact recipe developed by Scofield et al,⁵ contact was made to sample MSU-98-4H-56 via the DC sputtering of Al/Ni/W/Au of $2,000\text{\AA}/2,000\text{\AA}/2,000\text{\AA}/5,000\text{\AA}$, respectively. A shadow mask was used to pattern a van der Pauw structure on the sample. The

contacts were then annealed at 925 °C in an 4% N₂ in H₂ forming gas. Ohmic contact was observed via the measurement of the current-voltage (I-V) behavior between the contacts. A similar process was used to form van der Pauw contacts on sample MSU-98-4H-57, 69a and 70a. To prepare Schottky contacts for Admittance Spectroscopy, Al contacts were sputtered and patterned using photolithography.

Except for sample MSU-98-4H-53, all other samples (MSU-98-4H-56, 57, 69, 70) did not contain a degenerate layer for ohmic-like contact synthesis, but rather relied upon deposition of the proper metalization to achieve ohmic contacts. Metal deposition was performed at AFRL on these samples by Gerry Landis of University of Dayton Research Institute (UDRI) and consisted of the metal layers listed in Table 2.

Table 2 Ohmic Contacts on n- and p-type SiC

Material	Metal 1	Metal 2	Metal 3	Metal 4	Anneal
n-type	Al	Ni	W	Au	925 °C*
p-type	NiCr [§]	Cr	Au	—	925 °C*

* Anneal performed in 4% N₂ in H₂ forming gas for 10 minutes

§ Anneal performed after NiCr deposition

3. Epitaxial Layer Characterization

It is well known that accurate characterization assessment of epitaxial layer properties requires that a multitude of characterization techniques be employed. The Air Force Research Laboratories (AFRL) has considerable characterization resources, specifically intended to assess the material properties of wide band gap (WBG) semiconductors, like 4H-SiC. As a consequence, several characterization experiments were performed during the SFRP visit to AFRL. These were Cathodoluminescence (CL) to assess optically active defects, Hall Effect to measure the film resistivity and carrier mobility, and Admittance Spectroscopy to determine the Fermi Level (E_F) position in the band gap. Each of these experiments are now described in detail.

Capacitance-Voltage (CV)

Using standard Capacitance-Voltage (CV) methods, the doping density and carrier type of each epitaxial layer were measured. The results are listed in Table 1 for reference. For each epitaxial layer the conductivity was determined to assist with subsequent data analysis. In addition, the well-known fact that SiC dopants are not "classically shallow," (i.e., within a few kT of the band edges) requires the assessment of the conductivity type to be performed at elevated temperatures. As a consequence, C-V profiles were measured at elevated temperatures to ensure an accurate assessment of both the net doping density as well as the conductivity type.

Cathodoluminescence (CL)

Using the luminescence system (both CL and photoluminescence, PL) of the Air Force Institute of Technology (AFIT), a 4H-SiC sample containing what was believed to be a semi-insulating (or at least highly compensated) epitaxial layer was characterized by the PI and Dr. J. Scofield of AFRL. The CL system is capable of detecting luminescence from centers located throughout the 4H-SiC band gap.⁶ One defect of considerable interest is the boron-related D-center ($E_A = E_V + 0.60 \pm 0.2$ eV), which displays a characteristic green luminescence in 4H-SiC with a peak energy of approximately 2.3-2.4 eV.⁷ Observation of this characteristic defect signature would indicate that a sufficient D-center concentration was present in the material which is the critical deep-level required to achieve close compensation in these films.

A second center that is expected to be present from earlier characterization of these films is the aluminum acceptor³ ($E_A = E_V + 0.15$ eV). A classic donor-acceptor pair (DAP) recombination in the violet portion of the visible spectrum is associated with this recombination event, and should be clearly observable if sufficient aluminum acceptors are present.

Both of these luminescent signatures were observed during the CL experiments, as the low-temperature (9K) experimental data of Figure 1 indicates. The violet transition was clearly visible to the naked eye when the sample was cooled to less than 50K, otherwise the green D-center luminescence was dominant. Comparison of this data to the established literature on 4H-SiC indicates that these spectra are indeed the nitrogen to aluminum and nitrogen to D-center DAP recombination signatures, respectively.⁸

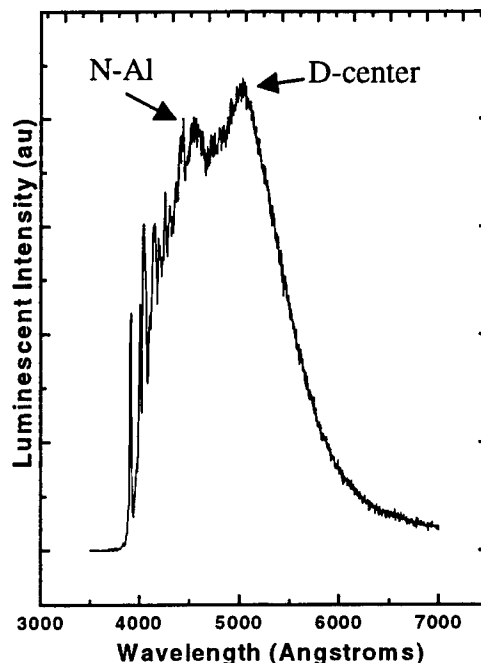


Figure 1. CL spectra from epitaxial layer no. MSU-98-4H-52. Spectra taken at 9 K. Note the nitrogen to aluminum and nitrogen to D-center DAP signatures at $\sim 4,500$ Å and $\sim 5,000$ Å, respectively. The beam current was $80 \mu\text{A}$

In order to fully characterize the observed luminescence, the temperature dependence of the various peaks must be measured (see Fig. 2). We see that the violet transition indicative of the nitrogen to aluminum DAP recombination quenches for temperatures greater than about 50 K, and the nitrogen to D-center DAP dominates, and is clearly observable for temperatures as high as room temperature (see Fig. 2(a)). In addition, we took a finer look at the luminescence structure near the band edge ($\sim 4,000$ Å) to see if any exciton peaks were observable (see Fig. 2(b)). We did not observe any excitonic transitions in this material which, given the extra doping of this material as compared to undoped 4H-SiC,² was not a surprise.

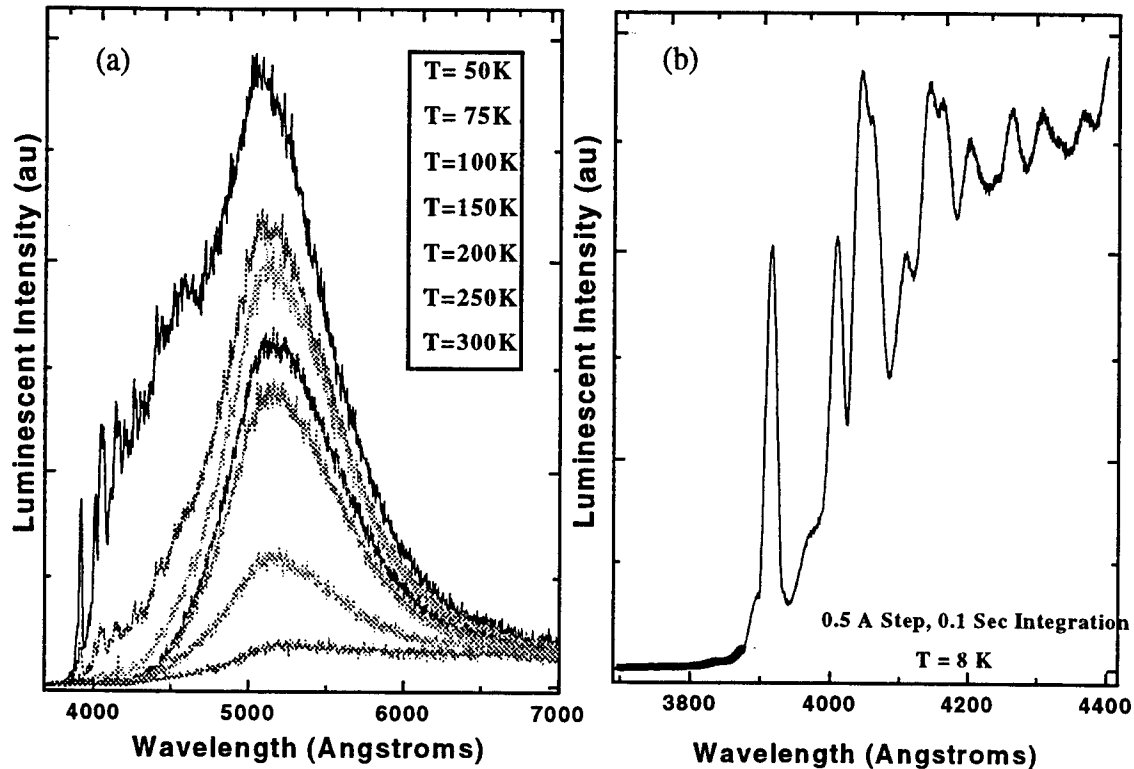


Figure 2. (a) CL spectra taken from 50K to 300K, showing nitrogen to D center DAP recombination peak. Note spectra does not shift with temperature.
(b) Band-edge CL spectra taken at 8K.

Admittance Spectroscopy

Admittance Spectroscopy is one technique that is useful for determining the position of the Fermi Energy Level, E_F , in the semiconductor bandgap. As is the case for all spectroscopic methods, the semiconductor conductivity is measured while a control parameter is varied, such as electrical bias, sample temperature or incident optical radiation. Using the extensive characterization resources available within the Materials Laboratory and staffed by the University of Dayton Research Institute (UDRI), Optical Admittance Spectroscopy (OAS)⁹ and

Thermal Admittance Spectroscopy (TAS)¹⁰ measurements were performed on EMRL epitaxial layers. Samples MSU-98-4H-52 and MSU-98-4H-56 were investigated using OAS and TAS, the details of which are now described.

Optical Admittance Spectroscopy (OAS)

Sample MSU-98-52 was investigated using OAS and the results compared with vanadium-doped Cree substrate material. Since the Cree material is semi-insulating (SI), this comparison should reveal interesting similarities and/or differences between EMRL compensated epi and V-doped SI material of the same polytype. Figure 3 shows the OAS data taken during the SFRP visit along with the OAS spectrum for the Cree V-doped SI substrate.

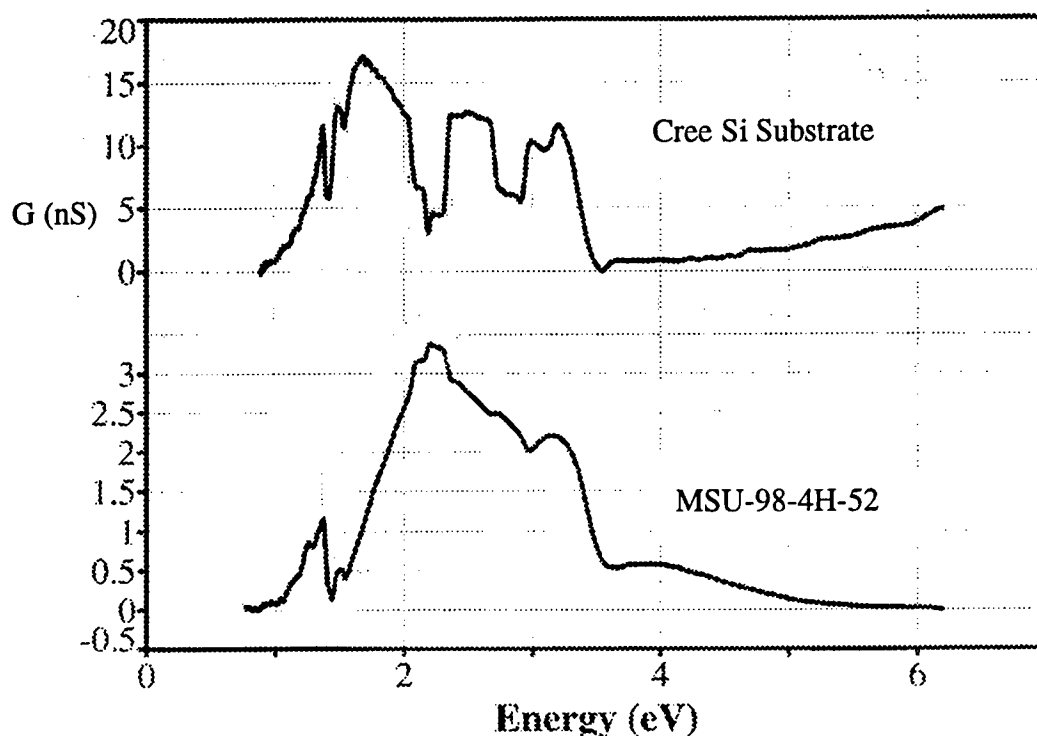


Figure 3 OAS spectrum of sample MSU-98-4H-52 compared with Cree V-doped SI substrate.

Inspection of Figure 4 reveals the following differences between the V-doped Cree material and the EMRL epitaxial layer: Both samples contain peaks related to Ti, as well as band-edge transitions. The primary difference is the large peak centered around 2.5 eV that is only present in the EMRL material. We believe this peak is related to the boron-related D-center, which is located approximately 2.5 eV below the conduction-band edge (i.e., $E_A = E_C - 2.5$ eV). This result is consistent with the CL data discussed in the last section, where the green luminescence was associated with donor to acceptor (DAP) recombination between the nitrogen donor and D-center. The difference with respect to the OAS data is that this DAP transition cannot be observed; rather, only conduction band to recombination center transitions are observable.

Thermal Admittance Spectroscopy (TAS)

Using measurement techniques developed at AFRL and described in the literature,⁹ thermal admittance spectroscopy (TAS) measurements were performed on Sample MSU-98-4H-52 at measurement frequencies of 15, 30, 45, 60 and 75 kHz. A broad TAS peak was observed during preliminary measurement of this sample in a low-temperature cryostat (sample measurement conducted from 50 K to 380 K, with a peak determined to be greater than 380 K). Due to instabilities in the device properties, TAS peaks could only be recorded for the lowest three frequencies. As a consequence, an accurate determination of the activation energy of the peak could not be made; however, upper and lower bounds for the observed level could be estimated and are believed to be between 50 and 120 meV, which is consistent with the known nitrogen donor in n-type 4H-SiC.

Hall Effect

A sample, identical to MSU-98-4H-52 with the addition of a n^+ cap layer, was grown by EMRL for the specific intention of providing a suitable device structure for Hall Effect measurement. This sample, MSU-98-4H-53, was prepared for Hall measurement using the RIE sample preparation procedure outlined earlier. Unfortunately, this sample proved to be slightly p-type, and thus pn junctions were formed under the RIE patterned Ni contacts. As a consequence, Hall measurement of this sample was not possible using the Ni contacts.

Four additional epitaxial layers were grown and these samples delivered to AFRL to support the SFRP effort. A summary of the Hall samples investigated during the SFRP has been provided in Table 1. Each sample will now be discussed in detail.

Hall data taken during the SFRP on sample MSU-98-4H-56 indicated that the epitaxial layer is not semi-insulating, with a room-temperature sheet resistivity of $8 \times 10^5 \Omega\text{-cm/square}$. Given an epi thickness estimated to be 2 μm , the bulk resistivity of the epitaxial film is thus approximately 160 $\Omega\text{-cm}$, which is far from semi-insulating. What is interesting about this Hall data is that the geometrical ratio between the contacts is strongly temperature dependent, which is an indication of highly anisotropic behavior. In addition, the epitaxial layer concentration versus temperature profile indicates that hopping conduction may be occurring in this sample, which is typically the case for highly-doped material.¹¹ Yet the net doping concentration, as determined by CV, is in the low 10^{15} cm^{-3} range. It may therefore be the case that close compensation is responsible for this peculiar result. High temperature Hall measurement should reveal the presence of hopping conduction as a mode of carrier transport, which would be manifested in a double-slope behavior evident in the resistivity versus reciprocal temperature profile.

Given the relatively low resistivity of sample MSU-98-4H-56, another sample was prepared with an attempt to push the sample into the semi-insulating doping regime. This was accomplished in two ways. First, the ratio of silicon-to-carbon (Si/C) was increased from $\text{Si/C} = 0.18$ to $\text{Si/C} = 0.2$. Second, in an effort to reduce the incorporation of nitrogen from the BN solid-phase doping source, the BN doping source was placed 0.5 mm farther away from the graphite susceptor to

decrease the N flux in the reactor during the growth run.³ Hg-probe C-I-V indicated that the stated objectives of these two epitaxial layer growth process changes did indeed result in a film that is likely highly compensated. Van der Pauw contacts were deposited as described earlier, and Hall measurements performed. Figure 4 shows representative set of Hall data from this sample for reference.

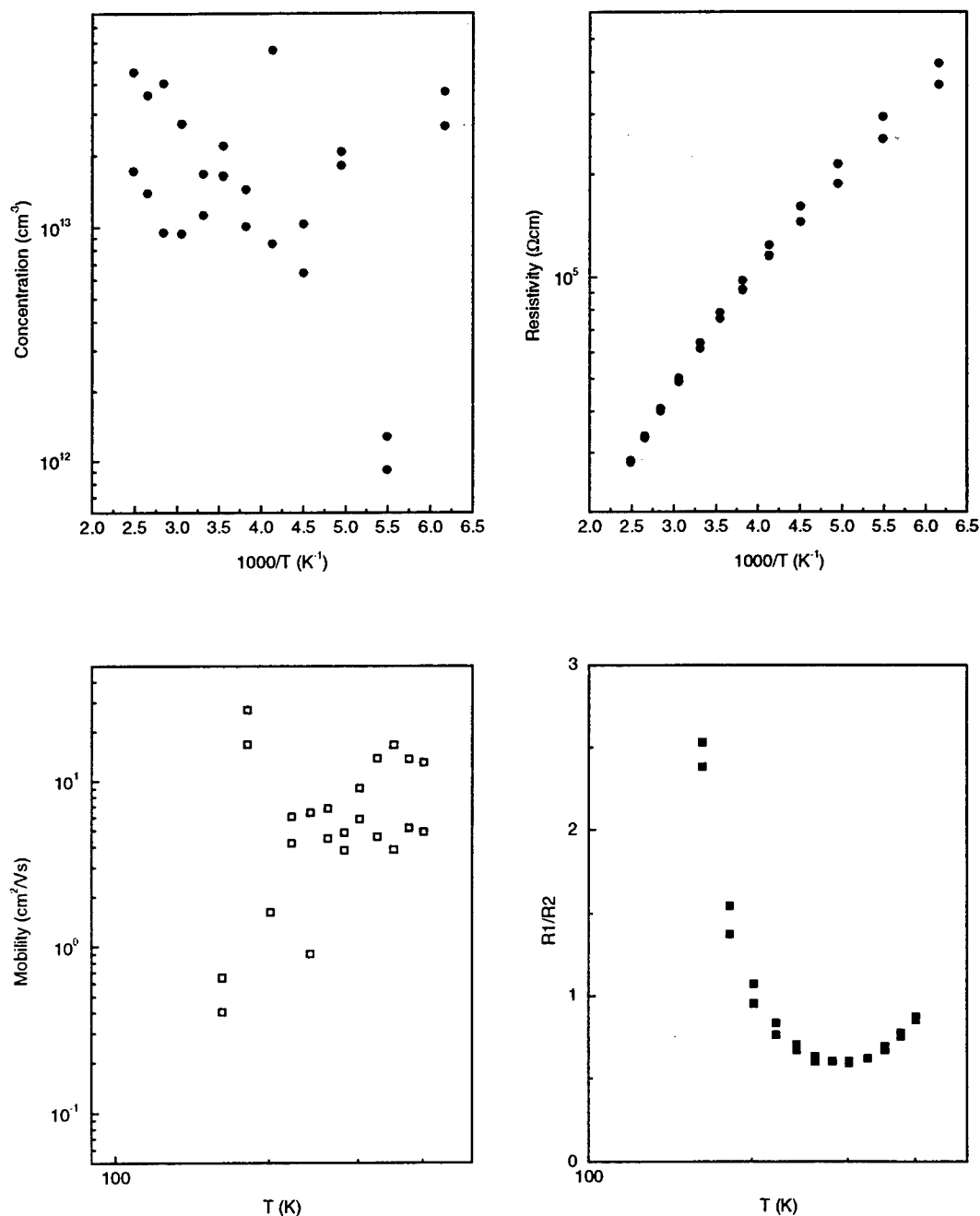


Figure 4 Low-temperature Hall data from sample MSU-98-4H-57 showing evidence of hopping conduction. Note the sample has a room temperature resistivity of ~160 Ω-cm. The scatter in the mobility data does not permit the concentration to be estimated.

Since neither of these layers displayed semi-insulating performance, a second set of Hall samples were prepared, whose parameters are listed in Table 1. For these last two samples, high-resistivity material was again not observed, as shown in Figure 5. However, the sample did have a measurable resistivity, mobility and concentration, although there is concern that the interpretation of these results is suspect due to the nature of the sample (thin film on SI substrate). The primary difference is that the material displayed n-type conductivity whereas the epi of Fig. 4 was p-type. This subtle difference could be significant, as will be described in the next section of this report.

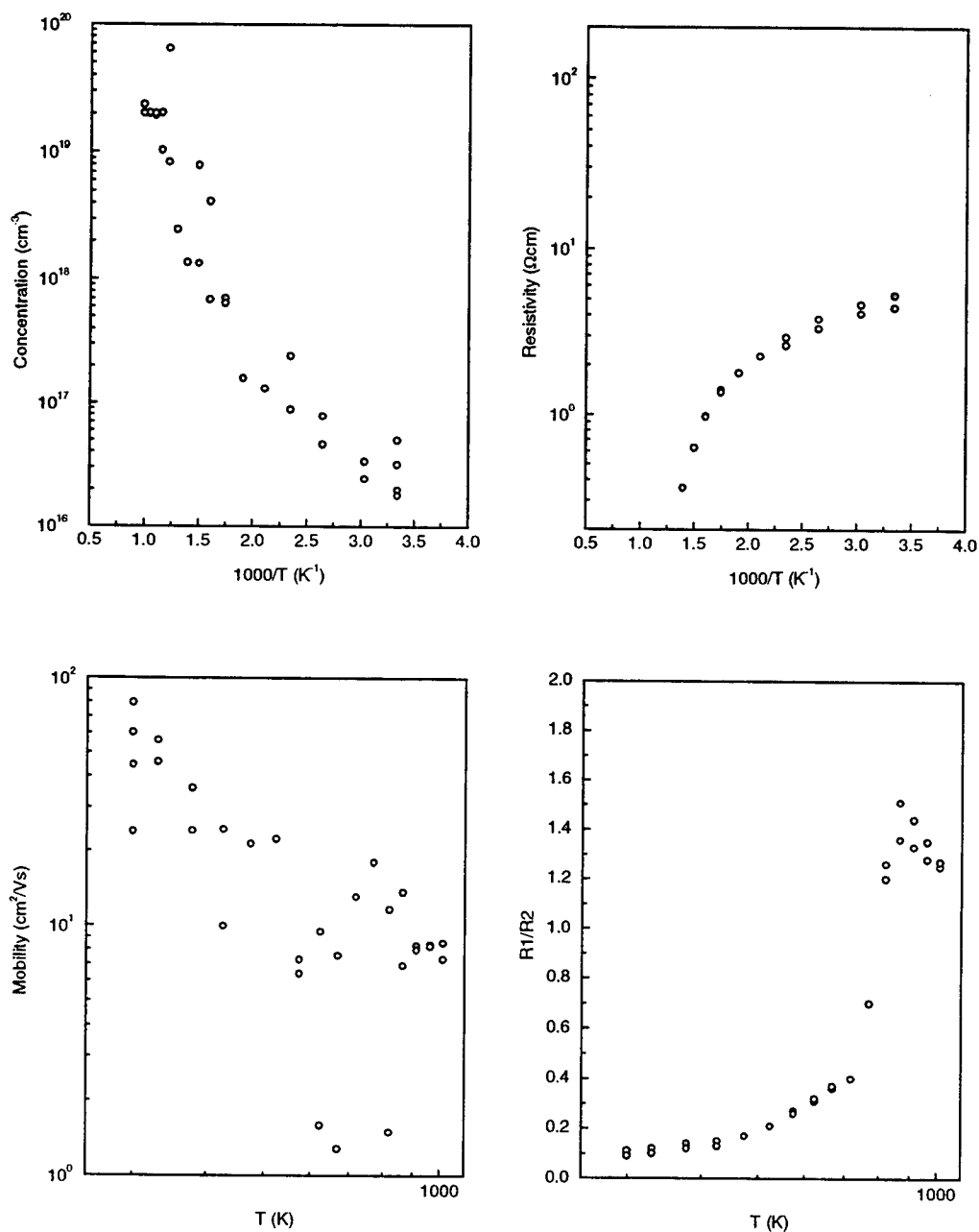


Figure 5 Hall data from sample MSU-98-4H-69a, showing a room temperature mobility, resistivity and concentration of $\sim 50 \text{ cm}^2/\text{V-s}$, $4 \text{ } \Omega\text{-cm}$ and $2 \times 10^{16} \text{ cm}^{-3}$.

4. Discussion

The experimental data gathered during the SFRP tour of duty at AFRL at first appears to be inconsistent, but after looking at all of the data the outcome is quite encouraging. For material that was somewhat p-type the capacitance was very low (1-2 pF as determined by Hg-probe CV), while for the material that was slightly n-type the capacitance was even lower (see Appendix A for C-I-V data on samples 57 and 69a). However, in both cases the BN-doped epi films appeared to be fully depleted, which is either a sign of very lightly doped (net doping density of less than 10^{14} cm^{-3}) or compensated material.

For the cases of slightly n-type material, the D-center was observed via cathodoluminescence (CL) indicating that not only is there boron incorporated in these films (this had been established earlier via SIMS characterization³) but that an observable quantity of boron is forming the deep-level needed to compensate these films based on the EMRL technique: the D-center.

The Hall Data is, in some ways, supportive of these results. First, for slightly n-type material free carrier mobility and concentrations were observed, indicating that transport was primarily due to free-carrier motion and not limited by carrier trapping and detrapping. In the case of slightly p-type material, evidence of Hopping Conduction, which is conduction mediated by traps, was observed. This is fully consistent with the model of the D-center serving as an acceptor defect whereby the material would be compensated on the p-side of the type conversion boundary.

The data that is perhaps the most interesting is the admittance data, and in particular the optical admittance spectroscopy (OAS) results. Here we see little similarity to the OAS spectra observed from Cree V-doped SI substrates, in particular the V-lines are absent. What is common to both spectra is the observation of Ti peaks and a peak that appears to be originating from the D-center. At the time of this report Dr. Steve Smith of UDRI is continuing to analyze this data and compare it with other known samples to develop a more detailed analysis of these results.

What is clear at this juncture is the need to continue to both refine the BN-doping scheme and to continue to perform both growth and characterization experiments to fill in the missing data necessary to interpret these results. The most obvious result of the Hall experiments is the need to ensure that this method is even suitable for analysis of such thin films on insulating substrates. The obvious question in this regard is "what is the actual epi volume that charge conducts through" since the estimate of this volume greatly affects the data interpretation. Clearly experiments need to be performed on non-compensated (i.e., non-BN-doped) films on the same substrate material. If classic Hall results are achieved for this very simple situation, then perhaps more attention can be paid to the results gathered during this SFRP on BN-doped epi layers.

The most important outcome of the SFRP visit to AFRL is the solid working relationship that has been formed between the EMRL and AFRL SiC researchers. Although the principal investigator has been familiar with the AFRL SiC group for some time, and a modest collaboration had been started prior to the undertaking of this program, there is no substitute for working together to formulate new ideas and concepts. This is, in the opinion of the author, the most significant

outcome of the summer visit and one that will hopefully lead to an ever expanding collaboration in the future.

5. Future Work

Based on the Experimental Discussion of the last section, it is viewed by the PI that two primary tasks should be undertaken to continue this collaborative effort. The first task is simply a continuation of the experimental investigation undertaken during the SFRP tour of duty. Indeed, at the time of this report writing, the PI has already delivered to AFRL a n-type, conventionally-doped epi layer on a Cree SI substrate for hall investigation. As mentioned in the last section, if classic Hall data results from this sample, then it may become possible to interpret the results of Hall experiments conducted on EMRL BN-doped epi layers. In addition, AFRL has continued to gather data after the PI returned to MSU, and this data continues to be analyzed by both groups. In addition, TAS and OAS experiments have recently begun at AFRL on the latest samples grown during the SFRP visit. These samples will then be shipped to EMRL where DLTS experiments will be performed to both confirm the presence of the D-center in this material and to determine its concentration. Knowing the D-center concentration as a function of growth parameters is critical to achieving SI epi since the concentration of donors, acceptors and deep-level impurities must be in balance for close compensation to occur.

The second task was jointly discussed between Dr. William Mitchel and the PI extensively during the visit and is really two subtasks, both having to do with EMRL performing epi experiments to support AFRL efforts to improve SiC substrates and ion implants. Considerable work has already been conducted by groups either working at AFRL or under contract to AFRL¹² in the area of chemical mechanical polishing (CMP). The ultimate proof of improved substrate material is the performance of device layers on these substrates. EMRL has agreed to assist AFRL in this fashion by growing epi layers on both CMP and etched substrate material and delivering this material to AFRL for device fab and analysis. EMRL will also work on its own CMP methods and will provide identical material to AFRL for comparison. The second subtask involves the anneal of ion implants and is a collaboration between AFRL, Purdue University (Dr. M. Capano) and EMRL. EMRL, using its silane-based CVD reactor, will perform annealing experiments in an attempt to develop a method of fully activating both Al and N implants while maintaining surface morphology. Indeed, EMRL has already received several implanted samples from Dr. Capano to begin this effort and has ordered a dedicated quartz tube to serve as an annealing furnace.

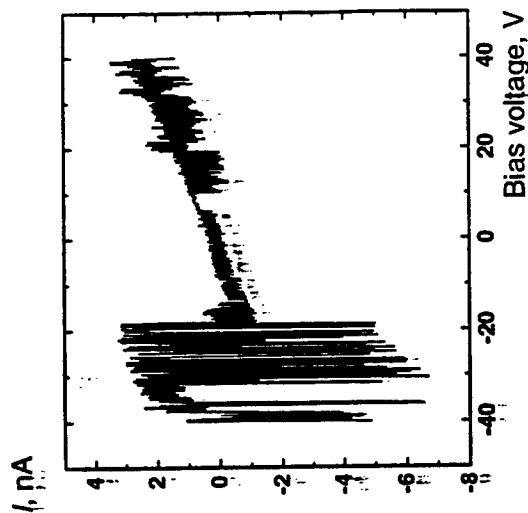
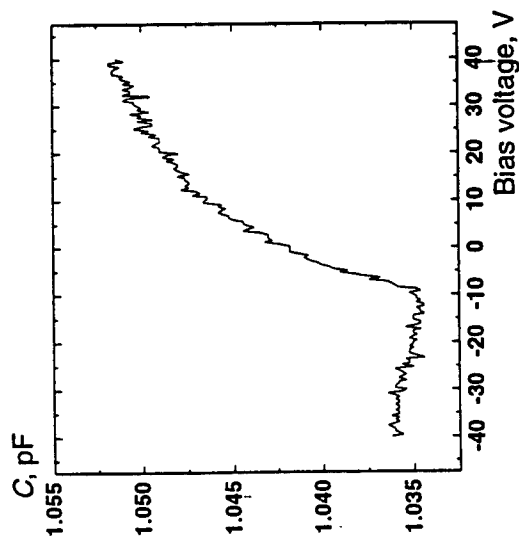
Both of these activities are the subject of a proposal to AFOSR under the Summer Faculty Research Extension Program (SFREP) which will be submitted in the near future.

6. References

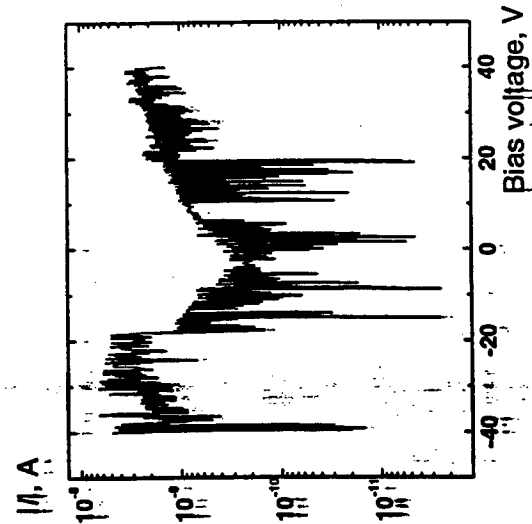
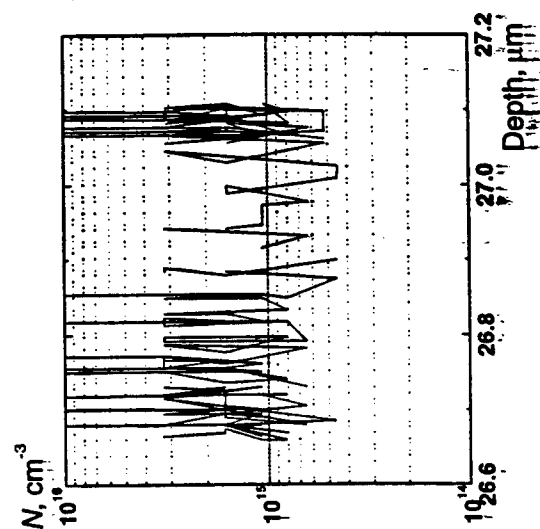
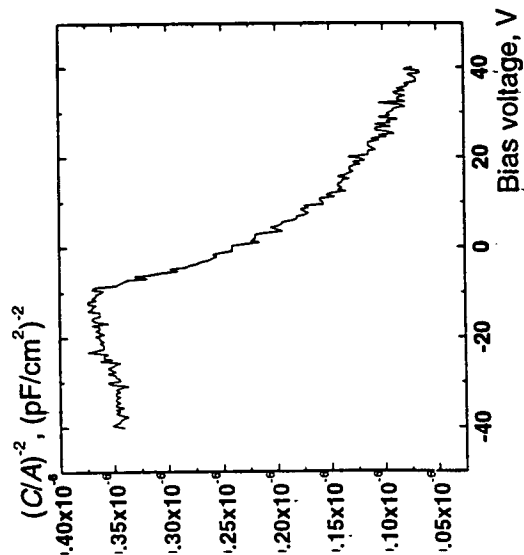
- ¹ M. S. Mazzola, P. G. Neudeck, D. J. Larkin, C. W. Tipton, and S. E. Sadow, Government Microelectronics Applications Conference (GOMAC'96), Orlando, FL, 18-21 March, 1996.
- ² S. E. Sadow and M. S. Mazzola, *International Semiconductor Device Research Symposium (ISDRS'97)*, Charlottesville, VA, 11-13 Dec. 1997.
- ³ M. S. Mazzola, S. E. Sadow, Adolf Schoener, *Materials Science Forum* Vols. 264-268 (1998) pp. 119-122, Trans Tech Publications, Switzerland. (Ref. conf. proc. of the *Int'l Conf. on SiC and III-N'97*, Stockholm, Sweden, 1-5 September 1997).
- ⁴ J. D. Scofield, P. Bletzinger, and B. N. Ganguly, accepted for publication.
- ⁵ S. Liu, S. Scofield, High Temperature Electronics Conference, Albuquerque, NM, June 1998.
- ⁶ J. Scofield CL ref.
- ⁷ M. Ikeda, H. Matsunami, and T. Tanaka, *Phys. Rev. B*, Vol. 22, No. 6, 2642-2854, 15 Sep. 1980.
- ⁸ W. A. Davis, *Photoluminescence Spectroscopy of 4H and 6H-SiC*, Master's Thesis, AFIT, Dec. 1994.
- ⁹ A. O. Evwaraye, S. R. Smith, and W. C. Mitchel, *J. Appl. Phys.*, Vol. 79, No. 1, 1 January 1996.
- ¹⁰ A. O. Evwaraye, S. R. Smith, and W. C. Mitchel, *J. Appl. Phys.*, Vol. 75, No. 7, 1 April 1997.
- ¹¹ W. C. Mitchel, A. O. Evwaraye, S. R. Smith and M. D. Roth, *J. of Elec. Mater.*, Vol. 26, No. 3, 1997.
- ¹² L. Zhou, V. Audurier and P. Pirouz, "Chemomechanical Polishing of Silicon Carbide," *J. Electrochem. Soc.*, Vol. 144, No. 6, June 1997.

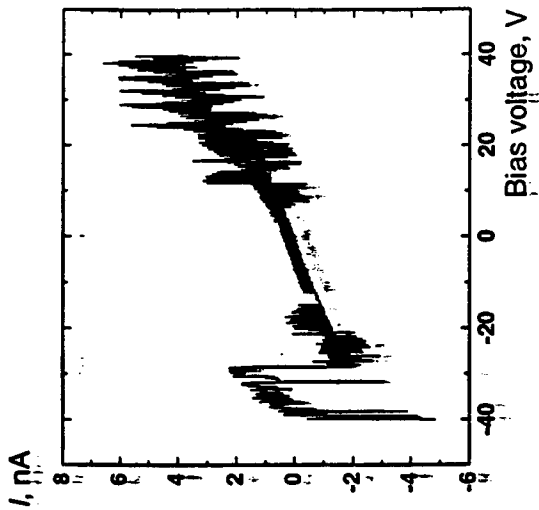
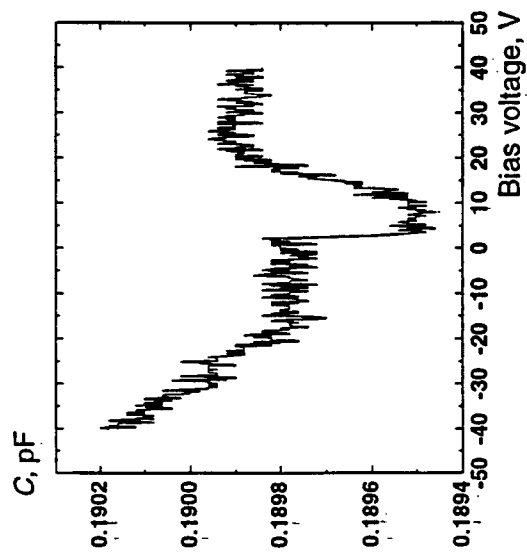
Appendix A

Capacitance-Voltage (CV) and Current-Voltage (IV) data from samples MSU-98-4H-57 and MSU-98-4H-69a, taken with a Hg-probe by the Emerging Materials Research Laboratory (EMRL) to support the characterization effort undertaken at AFRL under the SFRP. Note that sample 57 is slightly p-type and has a zero bias capacitance of approx. 1.5 pF, whereas sample 69a is slightly n-type with a capacitance of less than 1 pF.



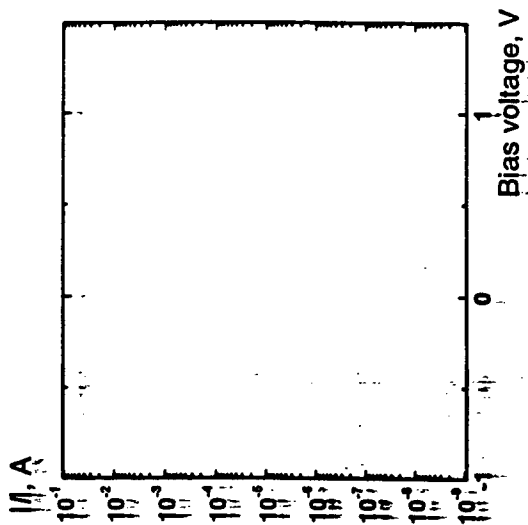
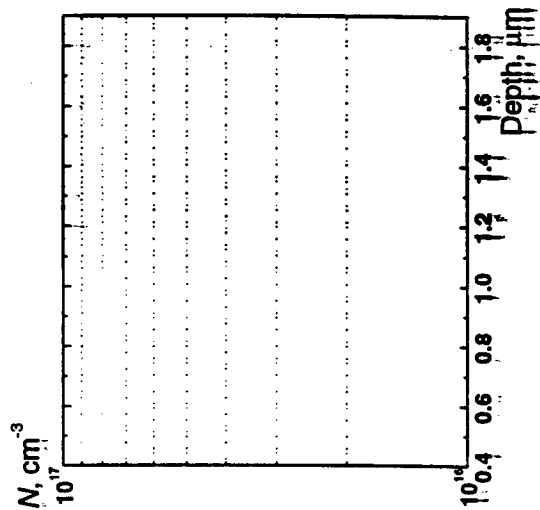
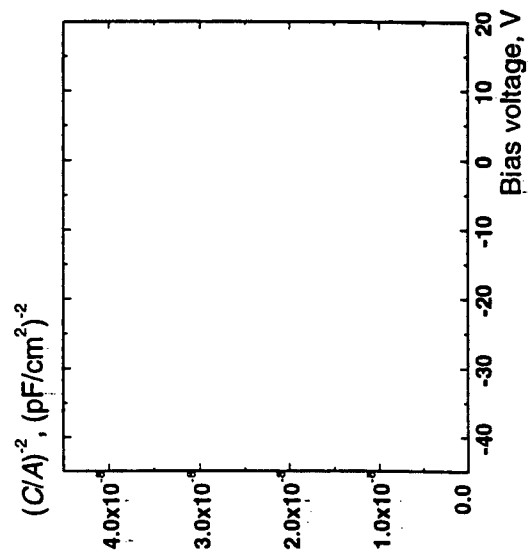
EMRL sample 98-4H-57
 Substrate: Si-type
 Contacts: Hi : Schottky Hg
 Lo: Hg
 Schottky contact $\phi=635 \mu\text{m}$
 C-V: AC frequency 1MHz
 AC level 30 mV





EMRL sample 98-4H-69a
 Substrate: Si-type
 Contacts: Hi : Schottky Hg
 Lo: Hg return
 Schottky contact $\phi=635 \mu\text{m}$
 C-V: AC frequency 1MHz
 AC level 30 mV

26-19



COMPUTER MODELLING OF NONLINEAR VISCOUS PANEL FLUTTER

R. Panneer Selvam
Associate Professor
Department of Civil Engineering

BELL 4190 University of Arkansas
Fayetteville, AR 72701
Email: rps@engr.uark.edu

Final Report for:
Summer Faculty Research Program
Wright Research Site

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

and

Wright Research Site

August 1998

COMPUTER MODELLING OF NONLINEAR VISCOUS PANEL FLUTTER

R. Panneer Selvam
Associate Professor
Department of Civil Engineering
University of Arkansas

Abstract

Nonlinear viscous panel flutter is reported for the first time in Selvam , Visbal and Morton [1]. Further study on Euler and viscous panel flutter is reported in this work. Divergence occurred when the Mach number M is less than one. There are two equilibrium positions during divergence depending upon the initial perturbations is reported in [3] and similar two equilibrium positions are simulated in this work for both Euler and viscous flows. The viscous flow features are much more complex than Euler flow. Many small vortices are observed during viscous negative divergence. The flow details and pressure contours are reported. Computer model to study the three-dimensional panel flutter is implemented. At this time the nonlinear plate equation is solved by finite difference method in space and Newmark- β method in time. Verification is underway.

COMPUTER MODELLING OF NONLINEAR VISCOUS PANEL FLUTTER

R. Panneer Selvam

1. Introduction

In the previous work [1] Selvam et. al. has shown the importance of coupled solution procedure to that of lagged approach in the aeroelastic or fluid-structure problems. By the coupled approach one could achieve second order accuracy by using fluid and structure solvers of second order. The computed stability boundary for flutter using Euler flow for two-dimensional panel is in excellent agreement with Dowell [2] and Bendiksen & Davis [3]. Stability boundary for viscous flow at $Re=1 \times 10^6$ is also reported. This is the first time one could simulate the panel flutter for viscous flow. In this work further investigation is done for 2D panel flutter to understand the viscous effect and the results are reported. Attempt is made to model the 3D panel flutter using the fluid code reported in [4] and by developing a dynamic nonlinear plate program using finite difference method.

2. Method of Solution

The flow equations are solved by Beam-Warming, alternate-direction, implicit scheme. The structural equations are solved by the implicit Newmark- β method as discussed in [1]. The equations are approximated in space by finite difference method. The details of the 2D equations are given in [1].

2.1. Nonlinear Dynamic Analysis of Plate

For 3D also the flow equations are solved similar to 2D as given in [4]. The three dimensional (3D) plate equations can be solved by either finite element method (FEM) or finite difference method (FDM) for simple regions. For complex regions like cantilever plate or wing finite element method is preferred. In this work as a start, rectangular panels are considered and hence FDM is used. The governing equations for the nonlinear dynamic analysis of plate as given in [5-8] is as follows:

$$D\nabla^4 w = p + N_x w_{,xx} + N_y w_{,yy} + 2N_{xy} w_{,xy} - m w_{,tt} \quad (1)$$

$$N_{x,x} + N_{xy,y} = 0 \quad (2)$$

$$N_{xy,x} + N_{y,y} = 0 \quad (3)$$

$$\text{Where } \nabla^4 = w_{,xxxx} + 2w_{,xxyy} + w_{,yyyy},$$

$$N_x = B[u_{,x} + (w_{,x})^2/2 + \nu(v_{,y} + (w_{,y})^2/2)]$$

$$N_y = B[v_{,y} + (w_{,y})^2/2 + \nu(u_{,x} + (w_{,x})^2/2)]$$

$$N_{xy} = B((1-\nu)/2)[u_{,y} + v_{,x} + w_{,x}w_{,y}]$$

$$D = Eh^3/(12(1-\nu^2)) \text{ and } B = Eh/(1-\nu^2)$$

Here u, v and w are the displacements in the x, y and z directions, N_x and N_y are the stresses in the x and y direction, N_{xy} is the shear stresses, E is the modulus of elasticity, h is the thickness of the plate, ν is the Poisson's ratio, p is the pressure and m is the mass per unit area. Here the terms following a comma represents the differentiation with respect to that variable. Here it is assumed that the inertial effect in the axial direction is very small and hence it is neglected.

The equations are highly nonlinear and also the equation in the z direction is fourth order. The solution procedure used by Vallabhan and Selvam [9] is used in this work for equation 1. The FDM stencils for equal spacing is given in [5]. At this time the plate is assumed to be simply supported at all edges and the displacements u and v along the edges are also assumed to be zero. The computed displacement for static problem reported in [7] is in excellent agreement with FEM solution in [7]. Then the program is modified to incorporate the Newmark- β method for dynamic problems. Currently it is in the process of verifying the fluid-structure interaction code. The plate equations are rewritten in a non-dimensional form as follows:

$$(\mu/\lambda) \nabla^4 w = \mu (p + N_x w_{,xx} + N_y w_{,yy} + 2N_{xy} w_{,xy}) - w_{,tt} \quad (1)$$

$$N_{x,x} + N_{xy,y} = 0 \quad (2)$$

$$N_{xy,x} + N_{y,y} = 0 \quad (3)$$

$$\text{Where } \nabla^4 = w_{,xxxx} + 2w_{,xxyy} + w_{,yyyy},$$

$$N_x = t_1 [u_{,x} + (w_{,x})^2/2 + \nu(v_{,y} + (w_{,y})^2/2)]$$

$$N_y = t_1 [v_{,y} + (w_{,y})^2/2 + \nu(u_{,x} + (w_{,x})^2/2)]$$

$$N_{xy} = t_1 ((1-\nu)/2)[u_{,y} + v_{,x} + w_{,x}w_{,y}]$$

$$\mu = \rho_\alpha l / (\rho_s h), \quad \lambda = \rho_\alpha u_\alpha^2 l^3 / D \text{ and } t_1 = B / (\rho_\alpha u_\alpha^2 l)$$

Here u, v and w are non-dimensional displacements with respect to the length of the plate l , μ is the mass ratio, λ is the dynamic pressure and t_1 is the non-dimensional term for the membrane forces. Further work is underway to solve the plate equations using FEM similar to Xue and Mei [10].

3. Results

3.1 Divergence of 2D Panels

Panel response in the stable and flutter regions are investigated at $M=1.2$ for Euler and viscous flows in Selvam et. al. [1]. When M is less than 1 and λ is greater than critical λ , rather than flutter, divergence occurs. Davis and Bendiksen [3] found that for Euler flow there are two equilibrium positions depending upon the initial perturbation. In this work also it is observed that if the initial velocity is positive (negative) then the divergence is in the positive (negative) direction as shown in Fig. 1 and 2. For the illustration $\mu=0.1$ and $\lambda=200$ is considered. The plate is divided into 50 equal spacing and the time step is considered to be 0.05. The nonlinear coefficient t_1 for 2D used in [1] is 125. To study the transonic features in detail t_1 is considered to be 25. When the deflection is around 0.021, shock is noticed at the front and back of the plate for negative divergence as shown in Fig. 3. For positive divergence the shock appears at the end of the plate the Mach number ranges from 0.7 to 1.2.

In the case of viscous flows also divergence occurred similar to Euler flow. But the flow features are much complicated as shown in Fig. 5 to Fig. 10. For negative divergence, rather than coming to a stand still position as in Euler flow, the displacements at the midpoint and quarter point of the plate are oscillating as in Fig. 5. Even when the time step is reduced to 0.01 and ran for longer time this oscillation is noticed. For positive divergence, the oscillation is very small as in Fig. 6. When small time step is used and ran for longer time the oscillating amplitude approached close to zero. These oscillations may be due to the vortices created close to the plate as shown in Fig. 10. The effect is more for negative divergence because many vortices are created as in Fig. 10. The pressure coefficients and the deflected shape of the plate is plotted in Fig. 7 and 8 for negative and positive divergence. The pressure changes sign more than once on the plate because of the complex flow over the plate during viscous effect. Experiments were conducted by computing flow over cavity and considering the structure to be rigid. Then these oscillations were not noticed as much. Hence this may be essentially due to the fluid-structure interaction. The pressure contours are plotted in Fig. 9 for both cases. For the negative divergence pressure waves are noticed. This may lead to aero-acoustic problems. These pressure waves disappeared when the flow is computed considering the plate to be rigid after it diverged. Hence it is clearly due to fluid-structure interaction.

3.2 Three-dimensional panel flutter

The nonlinear dynamic analysis program is verified and combined with flow code. Currently the fluid structure code is verified. Further work is under way to implement FEM procedure to solve the response of the plate.

4. Conclusions

Nonlinear viscous panel flutter is reported for the first time in Selvam et. al. [1]. Further study on Euler and viscous panel flutter is reported in this work. Divergence occurred

when the Mach number M is less than one. There are two equilibrium positions during divergence depending upon the initial perturbations is reported in [3] and similar two equilibrium positions are simulated in this work for both Euler and viscous flows. The viscous flow features are much more complex than Euler flow. Many small vortices are observed during viscous negative divergence. The flow details and pressure contours are reported. Computer model to study the three-dimensional panel flutter is implemented. At this time the nonlinear plate equation is solved by finite difference method in space and Newmark- β method in time. Verification is underway.

5. Acknowledgments

The author acknowledges the partial summer support provided by the Air Force Office of Scientific Research through the summer faculty research program. My sincere thanks to Dr. Miguel Visbal for giving me the opportunity to work on this problem and helping me during the process of this research. Thanks also to Dr. Ray Gordnier and Dr. Reid Melville to put the 3D fluid and solid code together.

References

- [1] Selvam, R.P., Visbal, M.R. and Morton, S.A., Computation of Nonlinear Viscous Panel Flutter Using a Fully-Implicit Aeroelastic Solver, *AIAA 39th Structures, Structural Dynamics, and Materials Conference, Long Beach, CA, April 20-23, 1998*, AIAA-98-1844
- [2] Dowell, E.H., *Aeroelasticity of Plates and Shells*, Noordhoff International Publishing, The Netherlands, 1975
- [3] Davis, G.A., and Bendiksen, O.O., Transonic Panel Flutter, AIAA -93-1476, April 1993
- [4] Melville, R. and Gordnier, R., Numerical Simulation of Large Amplitude Aeroelastic Wing Response, AIAA 98-2657, June 1998.
- [5] Szilard, R., *Theory and Analysis of Plates*, Prentice Hall, Inc., New Jersey, 1974
- [6] Murray, D.W., *Large Deflection Analysis of Plates*, Ph.D. thesis, University of California at Berkely, California, 1967
- [7] Murray, D.W. and Wilson, E.L., Finite-Element Large Deflection Analysis of Plates, *Journal of the Engineering Mechanics Division*, Vol. 95, pp. 143-165, 1969
- [8] Brebbia, C. and Connor, J., Geometrically Nonlinear Finite-Element Analysis, *Journal of the Engineering Mechanics Division*, Vol. 95, pp. 463-483, 1969
- [9] Vallabhan, C.V.G. and Selvam, R.P., Nonlinear Dynamic Response of Window Glass Plates Using Finite Difference Method, in *Dynamic Response of Structures*, Edited by: G.C. Hart and R.B. Nelson, ASCE, New York, pp. 373-381, 1986
- [10] Xue, D.Y. and Mei, C., Finite Element Nonlinear Panel Flutter with Arbitrary Temperatures in Supersonic Flow, *AIAA Journal*, Vol. 31, pp. 154-162, 1993

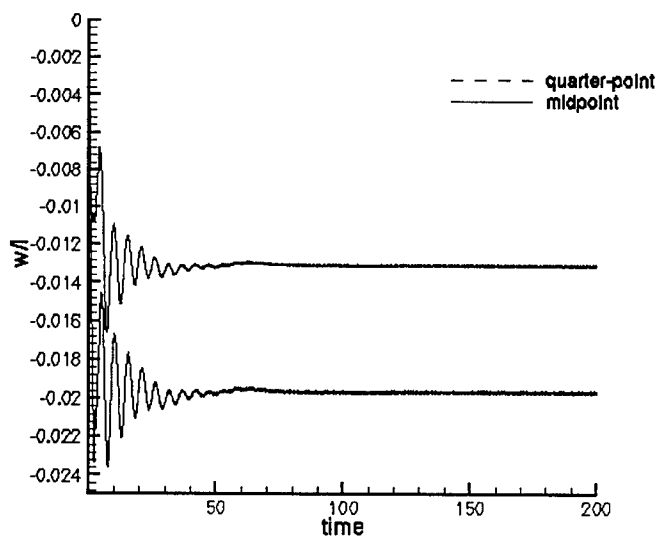


Fig.1. Panel response during transonic divergence in negative direction for $M=0.9$. Using Euler flow. Simply supported panel, $\mu=.1$ and $\lambda=200$.

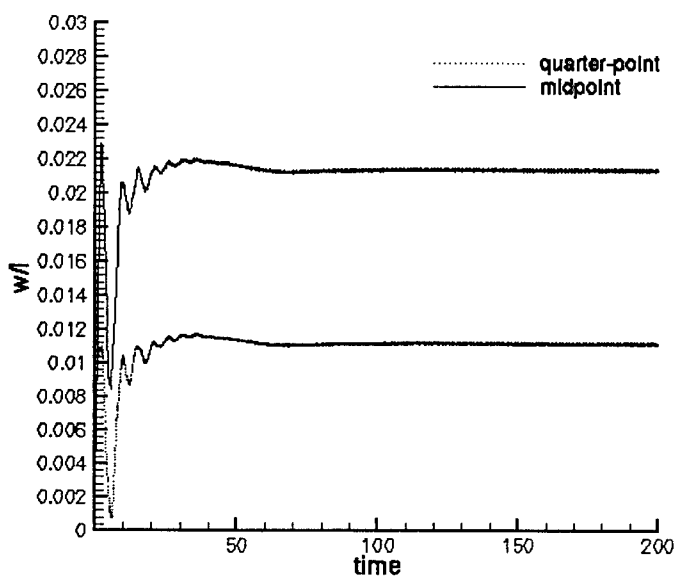


Fig. 2. Panel response during transonic divergence in positive direction for $M=0.9$. Using Euler flow. Simply supported panel, $\mu=.1$ and $\lambda=200$.

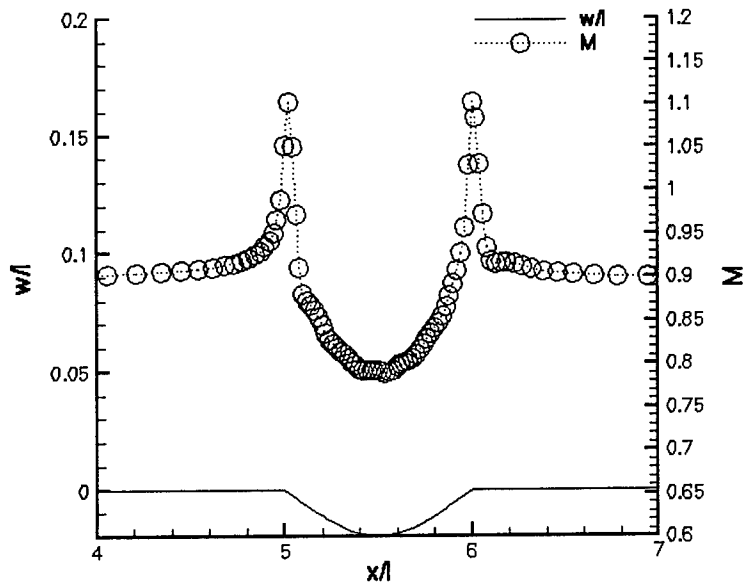


Fig. 3. Panel surface Mach number and deflected shape at negative divergence for $M=0.9$. Using Euler flow. Simply supported panel, $\mu=.1$ and $\lambda=200$.

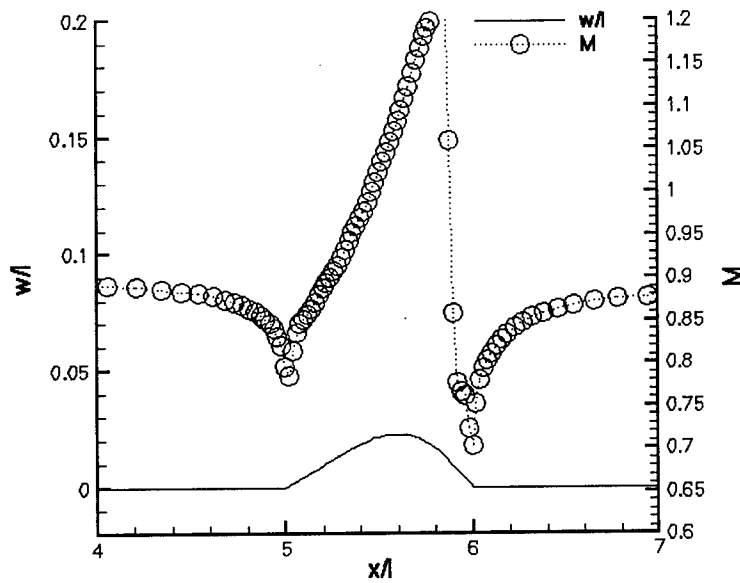


Fig. 4. Panel surface Mach number and deflected shape at positive divergence for $M=0.9$. Using Euler flow. Simply supported panel, $\mu=.1$ and $\lambda=200$.

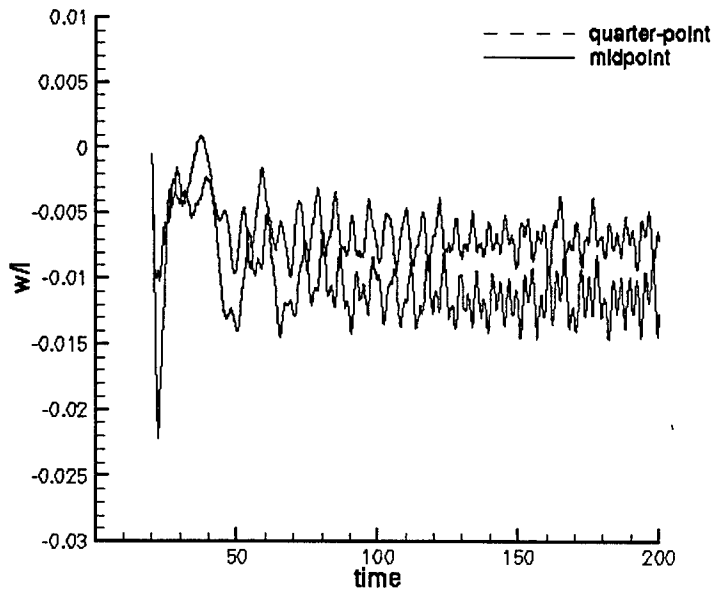


Fig.5. Panel response during transonic divergence in negative direction for $M=0.9$. Using viscous flow. Simply supported panel, $\mu=.1$ and $\lambda=200$.

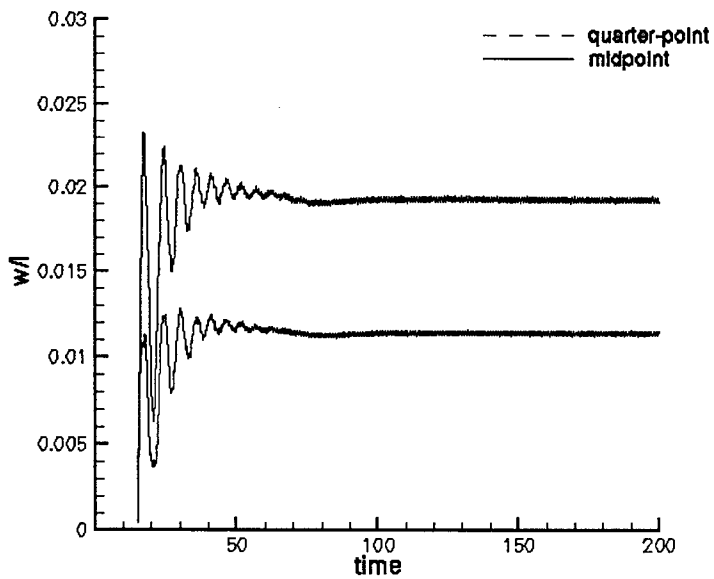


Fig. 6. Panel response during transonic divergence in positive direction for $M=0.9$. Using viscous flow. Simply supported panel, $\mu=.1$ and $\lambda=200$.

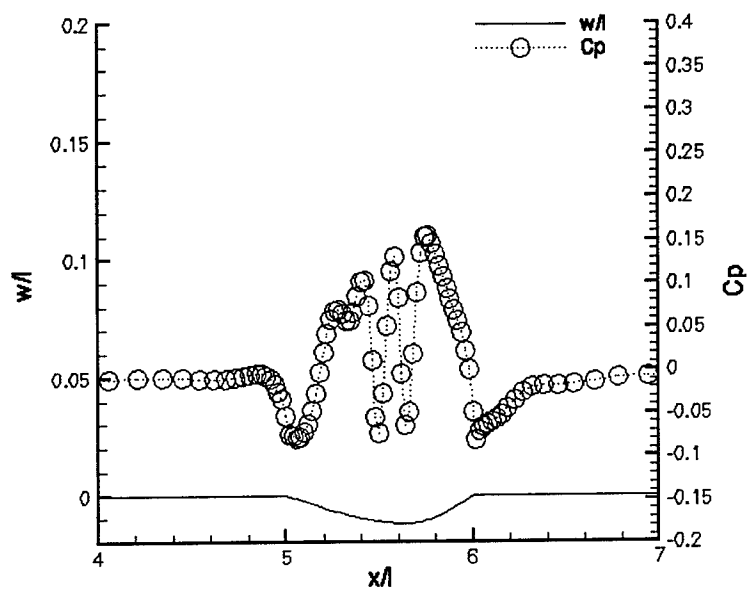


Fig. 7. Panel surface pressure coefficient and deflected shape at negative divergence for $M=0.9$. Using viscous flow. Simply supported panel, $\mu=.1$ and $\lambda=200$.

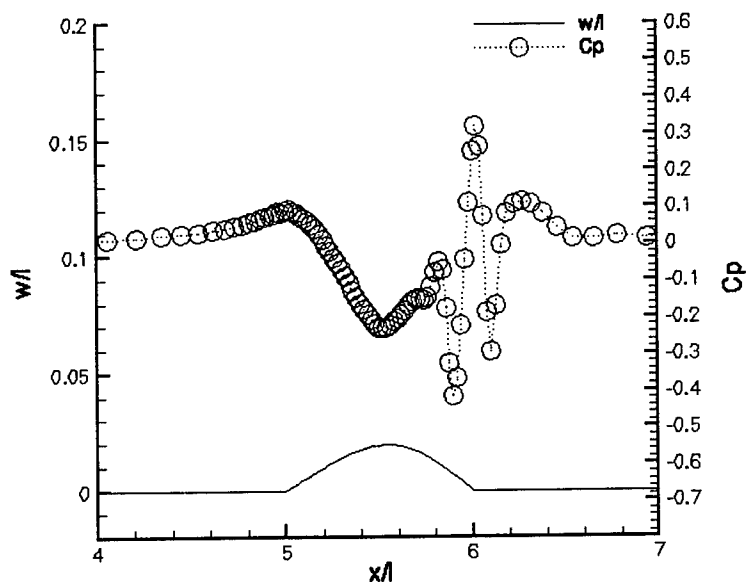


Fig. 8. Panel surface pressure coefficient and deflected shape at positive divergence for $M=0.9$. Using viscous flow. Simply supported panel, $\mu=.1$ and $\lambda=200$.

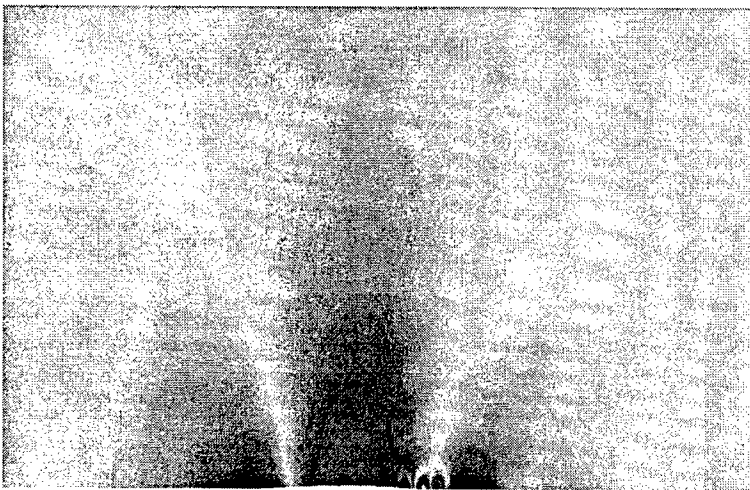
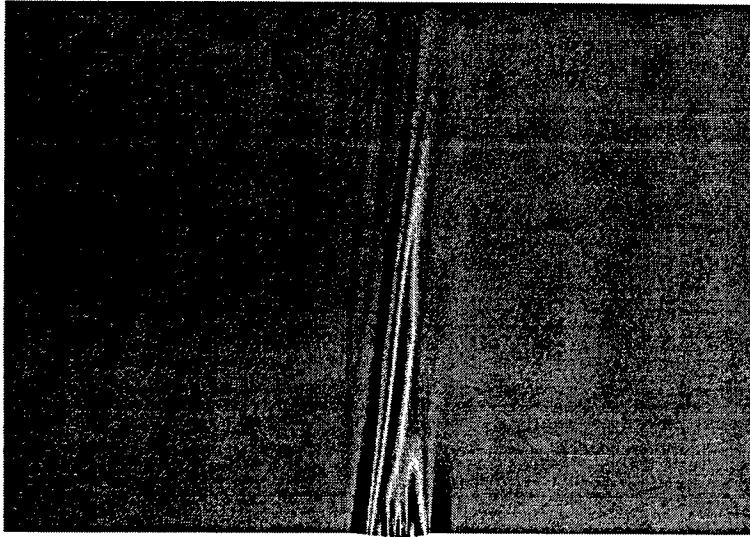


Fig. 9. Pressure contours for negative and positive divergence for $M=0.9$. Using viscous flow. Simply supported panel, $\mu=.1$ and $\lambda=200$.

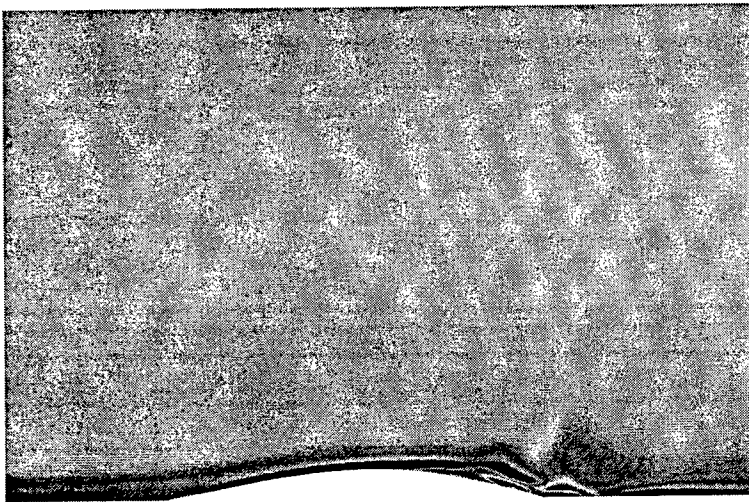
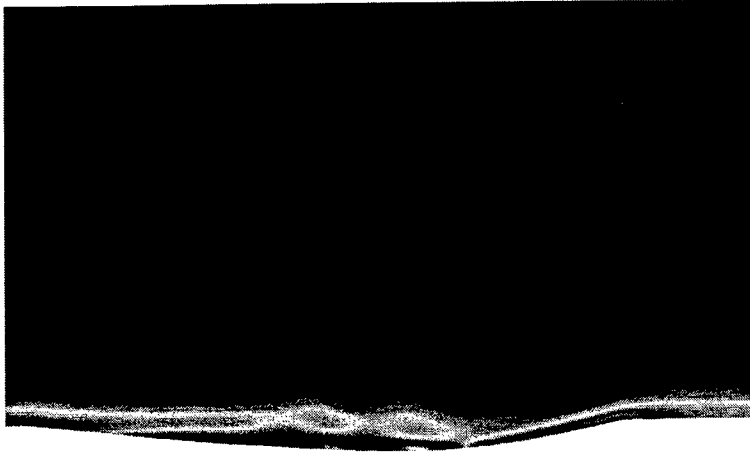


Fig. 10. Vorticity contours for negative and positive divergence for $M=0.9$. Using viscous flow. Simply supported panel, $\mu=.1$ and $\lambda=200$.

**A COMPUTATIONAL STUDY OF TURBINE BLADE INTERACTIONS
WITH CYLINDER WAKES AT VARIOUS REYNOLDS NUMBERS**

**Paavo Sepri
Associate Professor
Aerospace Engineering Program
Division of Engineering Sciences**

**Florida Institute of Technology
150 West University Boulevard
Melbourne, FL 32901-6975**

**Final Report for:
Summer Faculty Research Program
Wright Patterson Air Force Base
Turbines Branch**

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

and

Wright Patterson Research Site

September 1998

**A Computational Study of Turbine Blade Interactions
with Cylinder Wakes at Various Reynolds Numbers**

**Paavo Sepri
Associate Professor
Aerospace Engineering Program
Florida Institute of Technology**

ABSTRACT

The cyclic passages of wakes from upstream engine components introduce severe fluctuations into the aerothermal environment of turbine blades. The associated adverse effects are amplified at higher flight altitudes, as the attendant reduction in Reynolds number results in unsteady flow separation on the blades and a corresponding decrease in engine performance. In the present study, computations simulate the unsteady interactions of small cylinder wakes with a cascade of Langston turbine blades in a sequence of Reynolds number cases (660 K - 16.5 K). Vortex shedding from the cylinders produces an unsteady environment for the downstream boundary layers that may have adjustable content in frequency and length scale. Results from the existing VBI2D code are presented concerning the flowfield, pressure distributions, and total pressure loss coefficient. Flow structures are seen that are rich in oscillations and separations, especially at the lower Reynolds numbers. A somewhat surprising structure is seen in the generation of total pressure, which indicates the need for further study. A provisional explanation is offered for the corresponding behavior in pressure loss coefficient by means of a basic analytical model.

A Computational Study of Turbine Blade Interactions with Cylinder Wakes at Various Reynolds Numbers

Paavo Sepri

INTRODUCTION

Objectives

One objective of the present study has been to provide further computational characterizations of two-dimensional unsteady flow in turbine stages that are progressively influenced by flow separations associated with decreasing Reynolds number. As modern aircraft are designed to operate in an ever increasing range of altitudes, with the goal of not compromising engine performance, it becomes challenging to design and match components that remain efficient under varying operating conditions. In particular, at sea level, the Reynolds number (based on axial chord) of flow around a low pressure turbine blade is approximately 500,000. As the altitude increases to 50 Kft, the ambient density decreases by an order of magnitude, resulting in a corresponding decrease in Reynolds number. This change in flow condition is often accompanied by a dramatic flow separation on the suction side of the blade, which results in a large loss in turbine stage performance.

Another issue which influences turbine blade performance is related to the unsteadiness inherent in the flow around the blade, which is composed of a high level of turbulence as well as the periodic fluctuations produced by wake passages from stator vanes upstream of the turbine blades. These large fluctuations cause increased heat transfer rates to the turbine blades, and also produce variations in the location and extent of transitional boundary layers. Furthermore, at lower Reynolds numbers, boundary layers may re-laminarize in regions of favorable pressure gradients. In the present study, the effects of wake passages have been introduced by placing a row of cylinders upstream of the blades. Vortex shedding from the cylinders provides an influence on the turbine blade boundary layers that may be varied both in scale size and frequency by adjusting the cylinder diameter and flow rate. The objective here has been to utilize an existing computer code, termed Vane-Blade Interaction [1,2], to provide two-dimensional unsteady simulations of these effects.

Background

The Vane-Blade Interaction (VBI) code has been developed by several investigators [1-4] at the Allison Engine Company under contract to the United States Air Force. The earlier two-dimensional version (VBI2D) has recently been extended to include 3-D unsteady computational capability, although this latter version is still under evaluation. In a previous study, Rivir et al. [5] utilized the VBI2D code to calculate flow fields for Langston [6,7] turbine blade cascades in the steady state mode and without wake effects for a variety of Reynolds numbers and cascade solidities. The code incorporates the Baldwin-Lomax [8] two-layer algebraic turbulence model. Among other results, it was observed that the Langston cascade (at the original pitch to chord spacing of 0.93) demonstrated separation at $Re=50K$, oscillating transition at $100K$, and attached flow for higher Reynolds numbers.

Experimental investigations of flow in turbine passages have recently been reported by Blair [9], Baughn et al. [10], and Murawski et al. [11]. Blair has noted the several regions of high heat transfer produced by three-dimensional flow effects within the rotor passage, and has measured the effects of wall roughness on heat transfer. Furthermore, he discusses the sensitivity to changes in Reynolds number and inlet flow angle. Baughn et al. present comparisons of linear cascade experiments performed at UC Davis and the USAF Academy. The cascade geometries are similar to those of earlier tests performed by Langston et al. [6] and Graziani et al. [12] in linear cascades, and also to those of Dring et al. [13] and Blair et al. [14] in rotating tests. In the Reynolds number range of 67K to 144K, Baughn et al. note that an elevated level of free-stream turbulence influences the flow in augmenting heat transfer, advancing the location of transition, and removing observed flow striations from the blade pressure side (presumably caused by Görtler vortices). Furthermore, at the lower Reynolds numbers, evidence is presented of boundary layer separation on the blade suction side. Murawski et al. performed experiments in the linear, 2-D airfoil cascade at AFIT at Wright Patterson Air Force Base. The Reynolds number (based on exit velocity and suction surface length) was varied in the range of 50K to 300K, and the level of free-stream turbulence (FST) was varied in the range of 1.1% to 8.1 %. Separation on the suction surface was observed for all Reynolds numbers, with a shrinkage of the separation zone with increasing Reynolds number and increasing turbulence level. Wake surveys with a hot film have indicated that the width of the wake decreases with an increase of FST, and that it also decreases slightly with increasing Reynolds number. These trends are demonstrated to correlate well with the measured decreases in pressure loss coefficient as Reynolds number is increased. Calculations using the VBI2D code also indicate this trend of decreasing loss coefficient, although the effect of elevated FST is not included in the calculations, and the numerical results appear to underpredict the experimental values in some cases.

Among the many other studies of flow in turbine blade passages are those of Hodson [15], Zaccaria and Lakshminarayana [16], and Doorly and Oldfield [17]. At Cambridge University, Hodson has measured velocities and surface pressure distributions by means of instrumentation fixed to the rotor system, so that time-resolved influences of stator wake passages could be directly observed. It is concluded that the main effects of unsteadiness can be associated with the periodic convection of the wakes through the blade passage. An interpretation of the origin of blade surface velocity fluctuations is offered, although it is also stated that the phase variations of the pressure distributions along the blade surfaces are not understood. At the Axial Flow Turbine Research Facility at Pennsylvania State University, Zaccaria and Lakshminarayana have utilized a two-dimensional LDV system to measure both time-resolved and cycle-averaged wake flow properties in a rotor passage. Among other observations, it is noted that wake interactions with the rotor leading edge cause considerable downstream influences as the wake introduces significant swings in the effective angle of attack. Doorly and Oldfield have utilized a moving cylindrical wake generator upstream of a stationary cascade of rotor blades at the Isentropic Light Piston Tunnel at Oxford University. By means of Schlieren photography, heat transfer gauges, and surface pressure measurements, they show that the passage of wakes and shocks initiate leading edge separation bubbles on the turbine blades. The bubbles subsequently convect a short

distance downstream and then collapse. Unsteady heat transfer measurements indicate that a collapsing bubble generates a turbulent boundary layer patch which also is swept downstream, causing a locally elevated rate of heat transfer.

A related topic of study is that of the effects of elevated levels of free-stream turbulence, which are generated by components upstream of the turbine stage. Although the periodic unsteadiness caused by wake passages may dominate downstream flow performance, it has long been known that high levels of FST also influence heat transfer, transition, and boundary layer development. Recent studies aimed at increasing the level of generated FST have utilized upstream grids that include active blowing through holes in hollow rods. By these means, Young et al. [18] have extended heat transfer measurements on a flat plate for turbulence intensities ranging up to 25%. Welsh et al. [19] have applied the same technique at the U. S. Air Force Academy to influence flow in a Langston turbine blade cascade. In both cases, higher levels of FST advance the location of transition and enhance heat transfer, although it appears that these effects may reach a limiting form at the highest levels tested. Funazaki [20] and Funazaki et al. [21] have measured the effects of FST and wake passages on heat transfer in the region of a cylindrical forebody that includes film cooling from discrete holes. Information concerning film cooling effectiveness was obtained through liquid crystal thermography, which demonstrates that elevated levels of FST and increased wake Strouhal numbers both have adverse effects on film-cooling effectiveness. Fridman [22] has recently presented an algebraic relaxation-length model of turbulence to simulate FST effects in flat plate turbulent boundary layers, which apparently may be extended to turbine blade flows.

Computational simulations of flow in turbine blade passages have included two-dimensional models, and more recently, three-dimensional models. Abhari [23] has simulated rotor-stator interactions involving both steady and unsteady film cooling by means of a 2-D unsteady code that couples an explicit inviscid algorithm with a thin shear layer Navier-Stokes implicit scheme solved on a body-fitted boundary layer grid. The Baldwin-Lomax turbulence model is used to specify boundary layer transition in these computations. Comparisons with existing experimental results indicate good representations of film cooling on the surface heat flux, except in regions immediately downstream of the injection rows. Large scale pulsations of static pressure in the flow cause pulsations of coolant flow from the film holes, causing choking and unchoking in some cases. Arnone and Pacciani [24] have developed a time-accurate 2-D code to analyze rotor-stator interactions that utilizes a fully implicit time discretization to remove stability limitations. Accelerating strategies include local time stepping, residual smoothing, and multigrid computations. The code has been applied to a highly loaded gas turbine stage, for various combinations of stator-turbine pitch ratios.

Three-dimensional simulations of flow in engine components have been performed by Koya and Kotake [25], Boyle and Jackson [26], and Kang and Hirsch [27]. Koya and Kotake calculate periodic, transonic flows in highly twisted turbine stages using a finite-volume integral method to solve the Euler equations. Attention has been given to the treatment of boundary conditions involving periodicity and to phase relationships between time dependent stator

and rotor flow fields. Boyle and Jackson have compared the results of two different steady-state 3-D Navier-Stokes codes applied to a turbine vane with end walls, with the objective of calculating heat transfer and pressure distributions. Several algebraic models for turbulence transition have been compared in the study, and the Baldwin-Lomax model is among these. Heat transfer predictions on the pressure side of the vane have agreed well with experimental data, and a transition model formulated by Chima has apparently performed the best on the suction side. Kang and Hirsch have applied a 3-D time-dependent Reynolds-averaged Navier-Stokes code to simulate conditions in a linear compressor cascade including tip clearance with respect to a stationary wall. The Baldwin-Lomax turbulence model is also utilized. The computations appear to capture the development of secondary flows and tip leakage flows. However the approach to separation, leading to the tip leakage vortex, appears to be underpredicted, which the authors speculate to result from a limitation of the Baldwin-Lomax turbulence model.

Lengthier and more detailed accounts of experimental studies of flow within multi-stage low pressure turbines have been given in the dissertation by Halstead [28] at Iowa State University and in the report by Dorney et al. [29] at United Technologies Research Center. Halstead has taken extensive boundary layer measurements near both pressure and suction surfaces by means of densely packed hot-film sensors. He has observed regions of laminar, transitional, and turbulent flows that are influenced by the passages of wakes through the blade rows. At low Reynolds numbers, Halstead notes that the suction side boundary layer remains laminar over more than seventy percent of the surface. Dorney et al. have utilized the numerical scheme developed by Rai [30] to simulate experiments in the Large Scale Rotating Rig at UTRC. Both 2-D and 3-D versions have been implemented to study the migration of combustor hot streaks through stator and rotor stages. The 2-D code indicates that the blade count ratio could have a substantial effect on unsteady flow characteristics. Results from the 3-D code agree with the experimental observation that hot streaks tend to accumulate on the pressure side of rotor blades, thereby creating time-averaged surface hot spots.

Loss mechanisms in turbomachines are described in some detail in the *1993 IGTI Scholar Lecture* by Denton [31]. He describes the traditional three categories of profile loss, end-wall loss, and leakage loss. In the present study of 2-D flow, the latter two categories do not apply directly. Also discussed are several definitions of loss coefficient, of which the stagnation pressure loss coefficient for the turbine blade is utilized herein. Denton describes the connection of the various loss coefficients to the increase of entropy experienced in turbine stages. Of relevance to the pressure loss suffered through flow separation, especially at low Reynolds numbers, is the recent computational study by Muti Lin and Pauley [32]. A 2-D unsteady incompressible scheme is applied to flow around an Eppler 387 airfoil in the Reynolds number range 60K to 200K. A detailed time history of suction side separation, vortex development, vortex pairing, and subsequent convection is presented for several angles of attack of the airfoil. The mechanism for instability is attributed to be the inviscid Kelvin-Helmholtz one, with 2-D large scale vortex shedding to be the result.

RESEARCH DESCRIPTION

Computational Methodology

The computer code used for the present study has been developed earlier by several investigators [1-4] at the Allison Engine Company under contract to the United States Air Force. The version used here is designated VBI2D, as it calculates vane-blade flow interactions in a two-dimensional unsteady manner, utilizing a Reynolds-averaged compressible viscous flow model. Specifics of the computational methodology and operating procedures are described by Rao et al. [1,2]. Briefly, an O-grid system surrounds each vane or blade element, and the O-grid is embedded into an outer H-grid, which is contoured to conform to the main passages between the elements constituting each row of vanes or blades. The computational grid used for the present study is shown in Fig. 1.

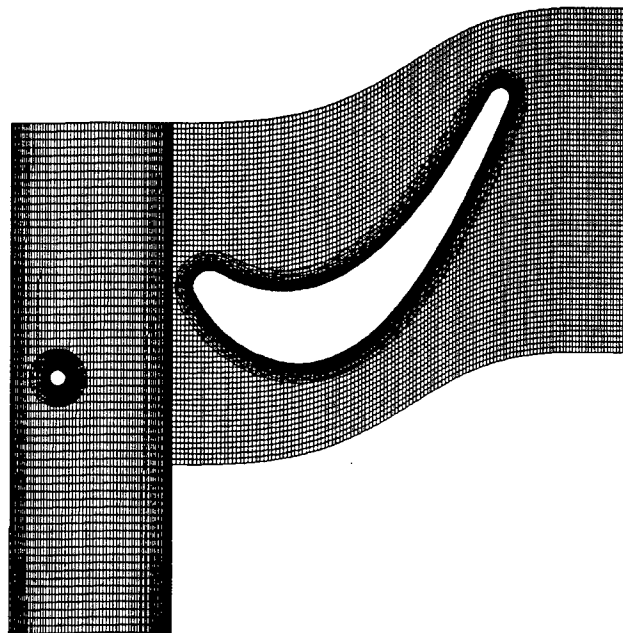


Figure 1: *Cylinder/Langston Blade Geometry and Computational Grids.*

The geometry is taken to be that of an axisymmetric body of revolution, along which the fluid flows tangent to stream-surfaces. Specific vane and blade geometries are defined through further transformations, while the flow is restricted to lie within the stream-surfaces, thereby eliminating the possibility of 3-D flow. The vane and blade grid systems may be stationary, or may be in relative rotation. Interpolation procedures are used to transfer data between the overlapping O- and H- grids, as well as the overlapping vane and blade H-grids. The numerical method utilizes a five-stage Runge-Kutta scheme with implicit residual smoothing. Turbulence is modeled with the Baldwin-Lomax [8] two-layer algebraic eddy viscosity model. The computational procedure has consisted of an initial run in the steady state mode to set up an approximate flow condition (typically 2000 iterations were used), followed by a run in unsteady mode

to set up the time dependent flow (40,140 iterations), and finally followed by an unsteady run describing one periodic passage of wakes with respect to the downstream configuration. During this last period, output has been generated at twenty equal time intervals for each Re case, in order to resolve flow variations within each cycle.

Computational Cases

The geometry has consisted of an upstream row of circular cylinders ($R = 0.1$ in) and a downstream row of Langston turbine blades (axial chord 4.0 in). The solidity (ratio of pitch to axial chord) was chosen to be 1.075 to conform with a recent experiment [10] at the Air Force Academy. Each spatial period consists of two cylinders upstream of three turbine blades, as may be deduced from Fig. 1. In all cases, the blade row was held stationary, while the upstream cylinders were set in motion, in order to produce the passage of wakes. The upstream flow incidence, with respect to the horizontal, was chosen to be 45° , close to the conditions of previous experiments. The axial and circumferential inlet Mach number components were taken to be 0.2, producing a total Mach number of 0.2828. The circumferential Mach number of the cylinders was taken to be 0.1, while for the last series it was set at -0.1. The upstream stagnation temperature was specified to be 520 °R. In order to simulate the change in Reynolds number (Re) that occurs with increasing altitude, the upstream total pressure was varied in seven increments from a maximum of 14.7 psia to a minimum of 0.3675 psia. The corresponding change in density results in a variation of axial chord Re from 660K to 16.5K, as indicated in Table 1. The beginning of a temporal period may be chosen arbitrarily in the cycle. Here, the VBI2D code has indicated 40,140 iterations to comprise one cycle in the present configuration. The center of an upstream wake passes the top blade nosetip at approximately iteration 66,231, as indicated in Table 1.

Iteration (Time)	Reynolds Number (Thousands)				Reynolds Number (Thousands)			
	16.5	33.0	66.0	99.0	132.0	264.0	660.0	99.0*
40140	100	200	300	400	500	600	700	800
42147	101	201	301	401	501	601	701	801
44154	102	202	302	402	502	602	702	802
46161	103	203	303	403	503	603	703	803
48168	104	204	304	404	504	604	704	804
50175	105	205	305	405	505	605	705	805
52182	106	206	306	406	506	606	706	806
54189	107	207	307	407	507	607	707	807
56196	108	208	308	408	508	608	708	808
58203	109	209	309	409	509	609	709	809
60210	110	210	310	410	510	610	710	810
62217	111	211	311	411	511	611	711	811
64224	112	212	312	412	512	612	712	812
66231	113	213	313	413	513	613	713	813
68238	114	214	314	414	514	614	714	814
70245	115	215	315	415	515	615	715	815
72252	116	216	316	416	516	616	716	816
74259	117	217	317	417	517	617	717	817
76266	118	218	318	418	518	618	718	818
78273	119	219	319	419	519	619	719	819
80280	120	220	320	420	520	620	720	820

* cylinders in downward motion.

Table 1: Identification of Output Cases with Axial Chord Reynolds Number and Time Interval (Denoted by Iteration Number).

Simplified Analytical Model

As an overall description to aid in the interpretation of the numerical results, it is convenient to consider the following idealized model of steady 2-D flow across a turbine stage. The model is derived from considerations presented by Hawthorne [33] and private communications with Dr. Robert Gray [34].

Steady flow across an axisymmetric stator/rotor stage may be represented by the usual equations of mass conservation, momentum, energy, and fluid properties. Here, the geometry is taken to have the possibility of streamsurface flaring, as represented by the inlet radial thickness, b_1 , and the exit radial thickness, b_2 . The vanes and blades may either be stationary or in rotation, but it is assumed that the inlet and exit flows are circumferentially uniform in an averaged sense. It is assumed that radial flow components are negligible. In a volume averaged sense, in a stationary coordinate frame, the basic equations may be listed as follows in the commonly used notation for an arbitrary segment of the stage.

$$\rho_1 v_1 A_1 \cos \beta_1 = \rho_2 v_2 A_2 \cos \beta_2 = \dot{m}, \quad (1)$$

$$\dot{m} [v_2 \sin \beta_2 - v_1 \sin \beta_1] = -F_y, \quad (2)$$

$$\dot{m} [v_2 \cos \beta_2 - v_1 \cos \beta_1] = p_1 A_1 - p_2 A_2 + F_x, \quad (3)$$

$$\dot{m} \left[c_p (T_2 - T_1) + \frac{1}{2} (v_2^2 - v_1^2) \right] = \dot{Q} - \dot{W}, \quad (4)$$

$$p = \rho R T. \quad (5)$$

Here, the cross-sectional area, A , may vary with streamsurface height, denoted by b . F_y and F_x represent net forces acting on the fluid in vertical and horizontal directions, respectively, as exerted by vanes and blades within the volume. The angle, β , represents the local flow angle relative to the axial direction. It is assumed that hub and end-wall effects are not included. The rotational rate of the turbine blades is represented by: $\bar{\Omega} = \Omega \bar{e}_x$. Accordingly, turbine power extracted from the flow is given by: $\dot{W} = \Omega R F_y$, and a possible heat transfer rate into the flow is represented by $\dot{Q}$.

For any stage, the basic equations may be transformed into a frame moving with the rotating system according to the usual vector decomposition, as follows:

$$v \cos \beta = u \cos \alpha \quad (6)$$

$$v \sin \alpha = u \sin \alpha + \Omega R \quad (7)$$

With appropriate subscripts, these equations may be applied at either the inlet or the exit. Here, u and α represent the flow speed and relative flow angle as seen in the rotating frame.

Upon squaring these expressions, combining at each station, and subtracting the totals at the two stations, one obtains:

$$u_1^2 - u_2^2 = v_1^2 - v_2^2 - 2\Omega R (v_1 \sin \beta_1 - v_2 \sin \beta_2), \quad (8)$$

which may be combined with Eq. (2) to yield the following:

$$u_1^2 - u_2^2 = v_1^2 - v_2^2 - 2\Omega \frac{RF}{\dot{m}}. \quad (9)$$

Further substitution into Eq. (4) results in the energy equation as viewed in the rotor frame:

$$\dot{m} \left[c_p (T_2 - T_1) + \frac{1}{2} (u_2^2 - u_1^2) \right] = \dot{Q}. \quad (10)$$

Here, the turbine power term has disappeared since the blades are stationary in this frame. From Eqs. (1) - (5), after transformation to the rotor frame, one may obtain the explicit relation:

$$\frac{u_2}{u_1} = \frac{B - (B^2 - 4AC)^{1/2}}{2A}, \quad (11)$$

where: $A = \gamma M_1^2 \cos^2 \alpha_2 - \frac{\gamma - 1}{2} M_1^2$; $B = \frac{\cos \alpha_2}{\cos \alpha_1} + \gamma M_1^2 \cos \alpha_2 (\cos \alpha_1 - d)$; $C = 1 + \frac{\gamma - 1}{2} M_1^2 + q$.

Here, d and q represent dimensionless drag and heating applied to the stage, respectively. The dimensionless lift alters the flow angle through the stage. Therefore, it is evident that the mean flow across the turbine stage depends on blade geometry (which influences flow deflection), drag (d), and heating (q), in addition to γ and M_1 . Other flow properties may be explicitly obtained from the above representation. In particular, the pressure loss coefficient is expressed as follows:

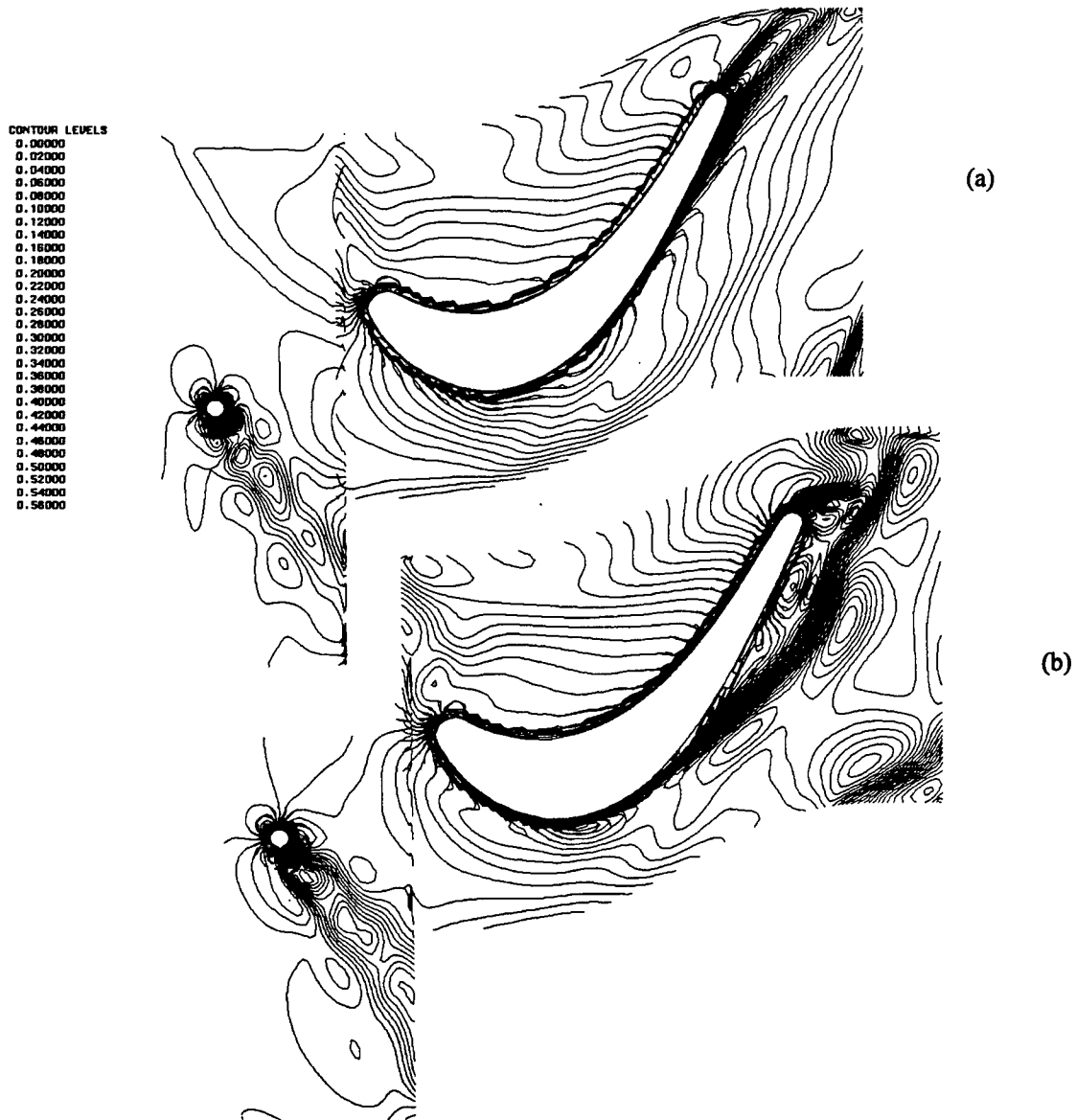
$$Y_p = \frac{p_{t1} - p_{t2}}{p_2 - p_2} = \frac{\frac{p_1}{p_2} \left[1 + \frac{\gamma - 1}{2} M_1^2 \right]^{\frac{\gamma}{\gamma-1}} - \left[1 + \frac{\gamma - 1}{2} M_2^2 \right]^{\frac{\gamma}{\gamma-1}}}{\left[1 + \frac{\gamma - 1}{2} M_2^2 \right]^{\frac{\gamma}{\gamma-1}} - 1}. \quad (12)$$

It is noted that the static pressure ratio is frame independent, whereas the Mach numbers depend on frame of reference. Therefore, the pressure loss coefficient expressed in a stationary frame varies with the rotational speed of the rotor. A numerical case is presented later, as it applies to the computational research considered herein.

RESULTS

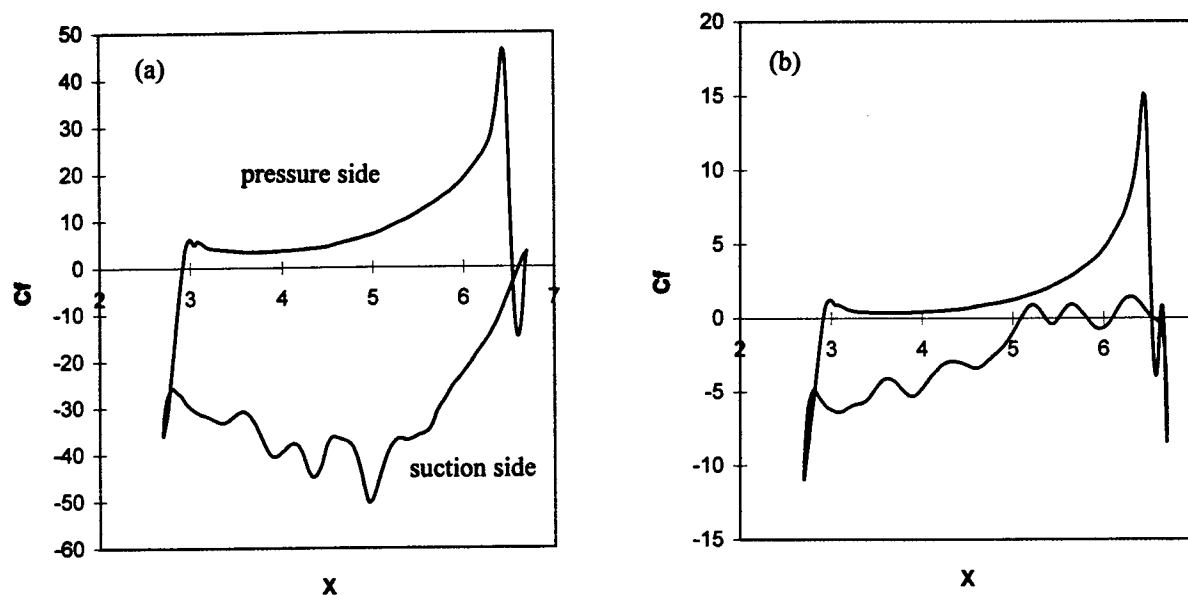
Flow Description

Typical velocity flowfields are shown in Figs. 2a and 2b for high (660K) and low (16.5K) Reynolds numbers, respectively, as computed using VBI2D (cases 713 and 113 in Table 1). It is evident that the suction side boundary layer remains largely attached for the high Re case, while it separates approximately midway along the axial chord for the other case. A comparison of the various runs reveals that the effective flow incidence at the blade nosetip may shift by ± 20 deg. during the passage of an upstream wake.



Figures 2a and 2b: *Constant Velocity Contours for Cases (a) 713 and (b) 113.*

Corresponding skin friction distributions are shown in Figs. 3a and 3b (note the different scales used for increased clarity). In the present convention, reverse flow on the suction side is indicated by positive values of skin friction, while streamwise flow is indicated by negative values. These indications are reversed in sign on the pressure side. Some of the oscillations occurring on the suction side are perhaps attributable to the influence of cylinder induced vortices, while the aft flow reversals at low Re (Case 113) are connected with vortices originating from the turbine blade itself.



Figures 3a and 3b: *Turbine Blade Skin Friction for Cases (a) 713 and (b) 113.*

The locations of transition to turbulence, as set by the Baldwin-Lomax model, may be observed in Fig. 4 for Case 713, in which the turbulent coefficient of viscosity is plotted at one grid point away from the surface along the periphery of the turbine blade. It is somewhat curious that the model is triggered at two isolated locations near the leading edge of the blade and briefly along the aft pressure side. The low Re cases remain totally laminar, whereas for Case 713, the main transition occurs at two locations on the suction side near mid-chord.

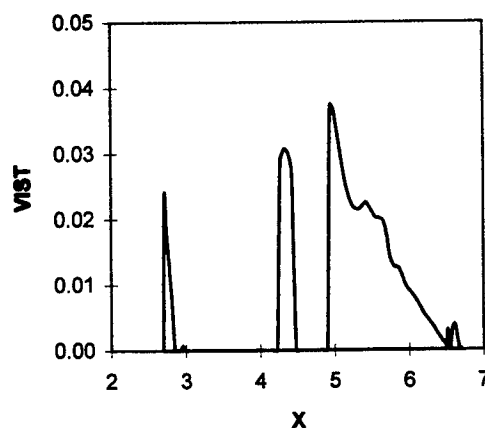
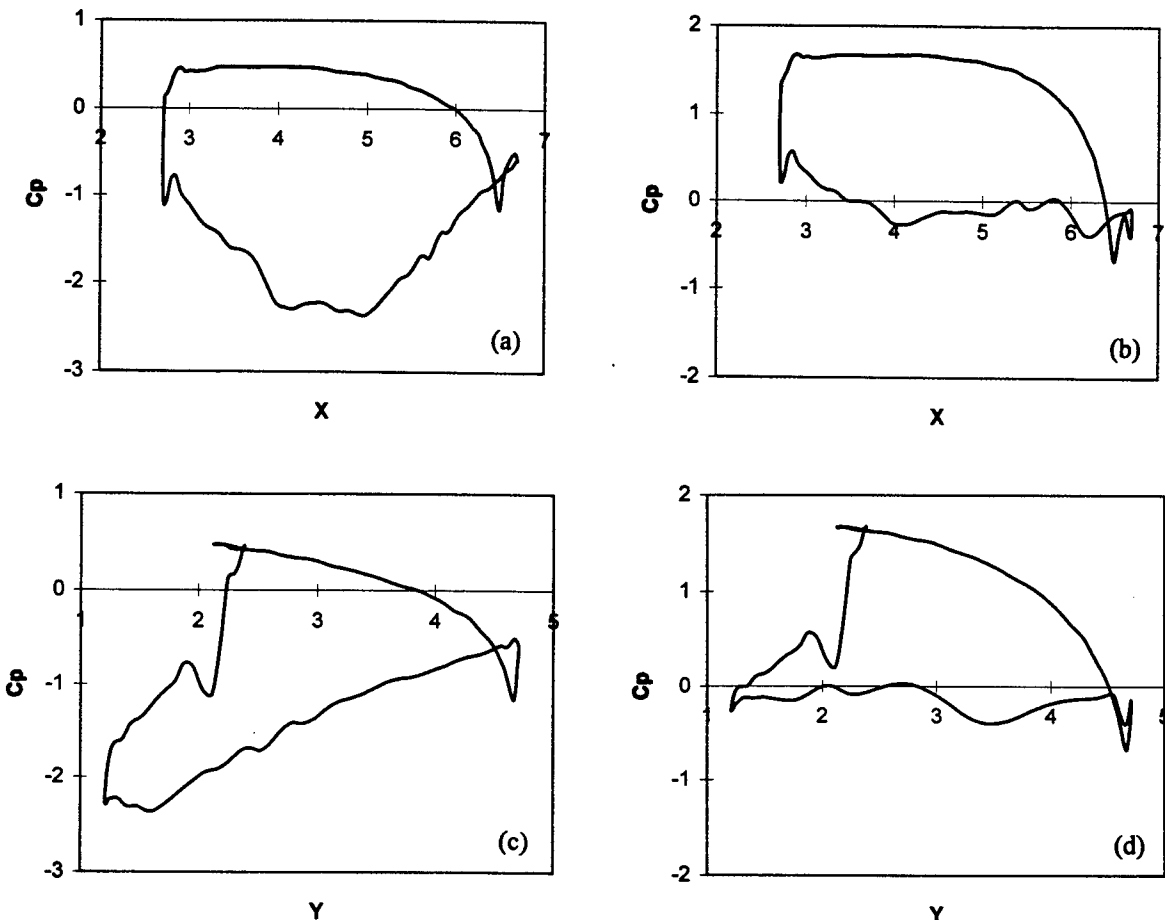


Figure 4: *Turbulent Viscosity Near Blade Surface (Case 713). Baldwin-Lomax Model.*

Pressure Coefficient

Typical pressure coefficient distributions on the turbine blade surface are shown in Figs. 5a - 5d. Here, the reference pressure is taken to be the constant static pressure imposed upstream of the cylinders at the *start* of each Re case. As each run progresses, the ambient static pressure level changes with time as power is either added or removed through the rotation of the cylinder stage. Presumably, the pressure level adjusts itself according to the emanation of acoustic waves through upstream and downstream boundaries. The choice of a constant reference pressure for each Re case proves useful in comparing pressure distributions at different times, although the C_p distributions shown in the figures exhibit uncharacteristically high or low levels owing to the shift in actual ambient pressure with time. Pressure differentials, both vertical and axial, decrease significantly at lower Re, as the degree of effective suction is reduced.

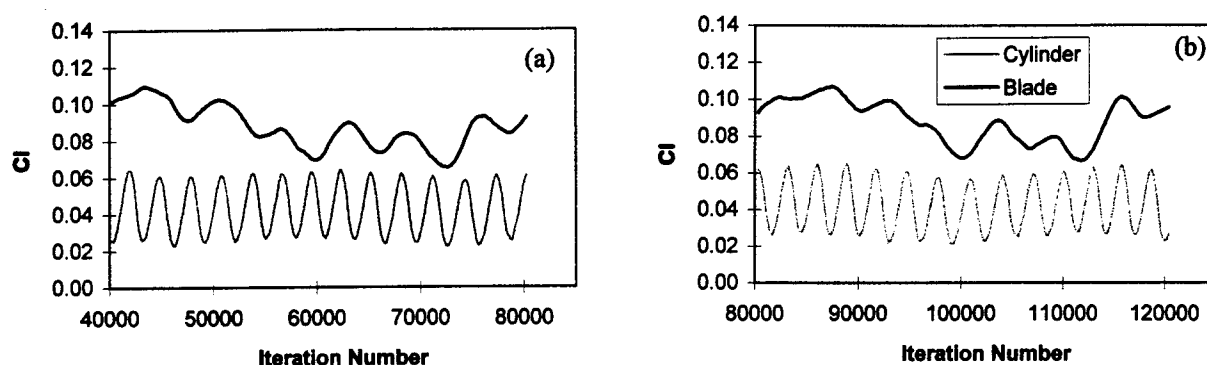
Normally, one associates with separated flow an increase of airfoil drag. However, since the turbine blade tail section is highly inclined to the axial direction, a loss of relative suction corresponds to a decrease in both axial pressure drag and torque.



Figures 5a - 5d: C_p vs. Axial Distance for (a) Case 713 and (b) Case 113.
 C_p vs. Circumferential Distance for (c) Case 713 and (d) Case 113.

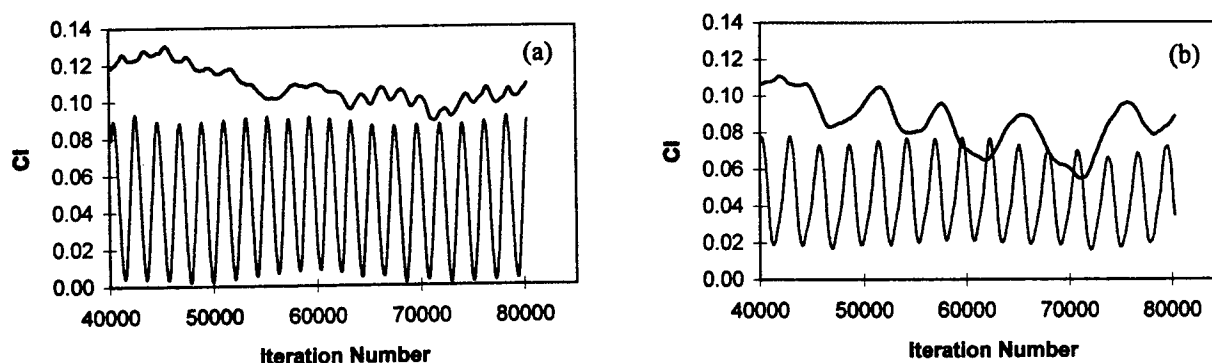
Flow Periodicity

One objective of the present study has been to observe the time dependent effects of cylinder/blade interactions. For this purpose, output from the VBI2D code has been extracted at twenty time intervals within one flow cycle, as indicated in Table 1, after the unsteady computations appeared to have reached an approximately periodic condition. This periodicity was checked for a low Reynolds number case (16.5K) by continuing the computations for an additional period. The results shown in Figs. 6a and 6b demonstrate the high degree of similarity for two neighboring cycles. It may be expected that such histories would not be identical owing to random differences of vortex shedding, although further computations over several periods have not yet been attempted.



Figures 6a and 6b: Cylinder and Blade Lift Histories ($Re = 16.5K$) for Two Cycles
Iterations: (a) 40140-80280 (b) 80280-120420.

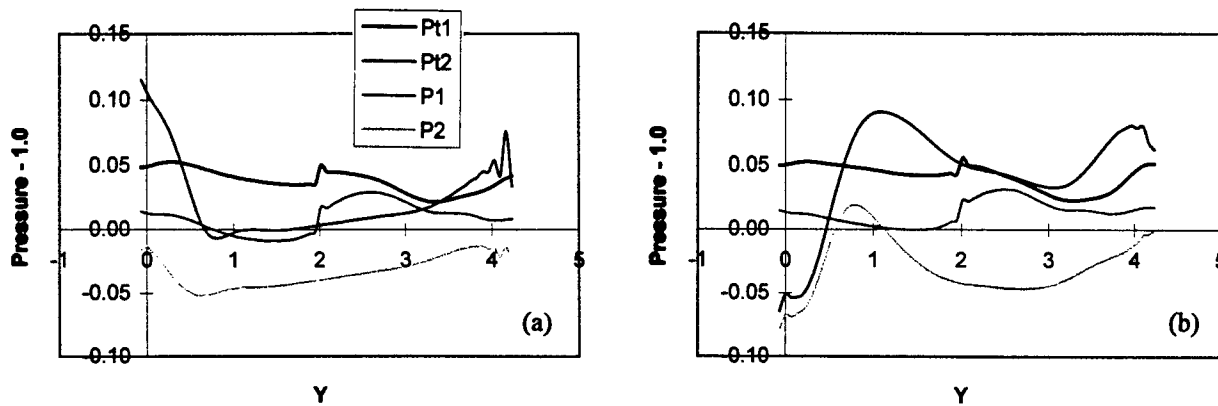
Variations in lift histories at two Reynolds numbers are shown in Figs. 7a and 7b. At the higher Re (660K), the blade lift oscillates at a frequency that is approximately the same as that of the cylinder, whereas in Fig. 7b (33K), the blade frequency is lower. The cylinder shedding frequency evidently decreases with decreasing Re , which implies a change in Strouhal number, since the flow velocity is not varied here.



Figures 7a and 7b: Lift Histories for Cases (a) $Re = 660K$ and (b) $Re = 33K$.

Pressure Profiles

The stagnation pressure loss coefficient, as defined by Eq. 12, is computed for the VBI2D cases by performing a numerical quadrature in respective vertical cuts in order to find averaged values for the pressures. The inlet location was chosen to be the first column of the blade H-grid, and the exit location was the last column. Sample vertical distributions of the pressures are shown in Figs. 8a and 8b. Here, fluctuations are seen both upstream and downstream of the turbine blade. Presumably, some of these variations may be due to influences of vortices swept downstream from the cylinders, and by the modification of total pressure in the unsteady flow around the blades. In particular, it is interesting to note the distribution of total pressure downstream of the turbine blade. In some locations, the total pressure rises significantly above its upstream value, whereas in other regions of the wake, the level is markedly lower. The reduction of total pressure by viscous losses is to be expected, but its generation may be somewhat surprising.



Figures 8a and 8b: *Vertical Pressure Profiles for Cases (a) 713 and (b) 113.*

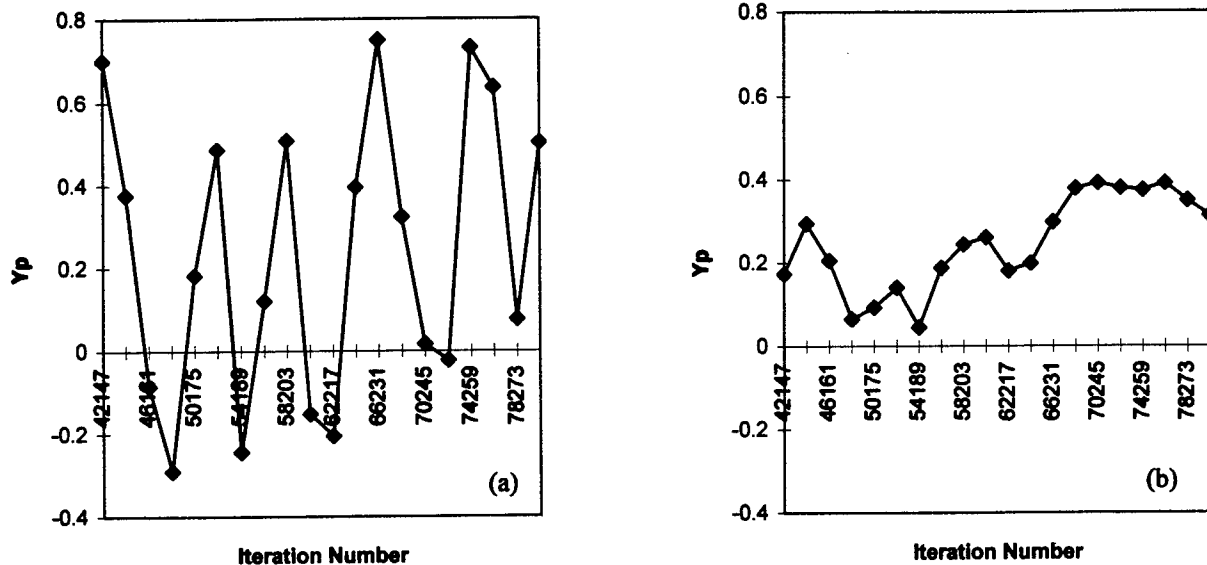
Alternating regions of high and low stagnation pressure also occur in the wakes of the cylinders. In Fig. 9, near peak locations in the wake, the stagnation pressure exceeds its level attained upstream of the cylinder. While such a circumstance is not to be expected for steady flows around stationary objects, it apparently can occur for unsteady flow about stationary objects, and certainly for flow about moving objects which impart mechanical energy to the flow. A preliminary literature search by the author has not yielded prior mention of this basic phenomenon in regard to wake flows past stationary cylinders.



Figure 9: *P_T Contours for Case 714.*

Pressure Loss Coefficient

Variations in total pressure loss coefficient within one cycle are shown in Figs. 10a and 10b for two cases. Large oscillations occur within the cycle for the low Re case, so that Y_p is even negative at times. It should be noted that flow structures, which are convected streamwise through the blade passages, would involve a phase lag between exit and inlet locations, whereas Y_p provides an instantaneous comparison between these locations. The higher Re case exhibits no negative excursions.



Figures 10a and 10b: Total Pressure Loss Coefficient Variation Over One Cycle for Re of (a) 33K and (b) 660K.

The results of averaging Y_p over ten equally spaced time intervals during each cycle are shown in Fig. 11 as Re is varied. These results are somewhat contrary to earlier expectations. From a previous study [11], it was anticipated that Y_p would have small values at higher Re; that there would be a marked break in the trend near 100K; and that Y_p would increase dramatically at lower Re owing to the large losses associated with suction side separation. The present results do not support this view in that the calculated Y_p remains fairly constant (range of 0.2 – 0.3), with a peak occurring in the distribution near $Re = 132K$.

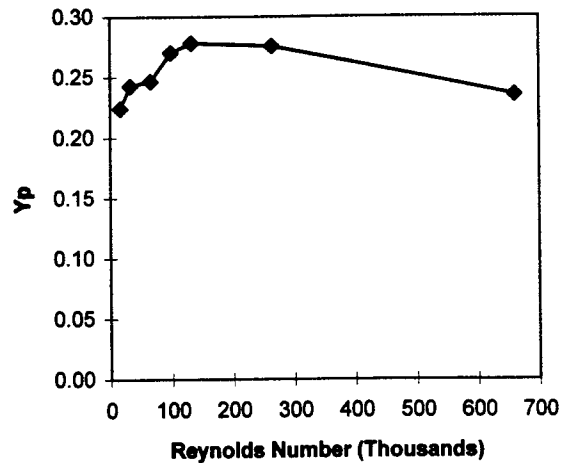


Figure 11: Y_p Averaged Over One Cycle vs Re.

Comparison with Analytical Model

In an effort to reconcile this unexpected result, computations have been applied to the simple analytical model presented earlier. A plot of Y_p as a function of exit angle and dimensionless drag is presented in Fig. 12. The inlet flow angle is taken to be 45 deg. and the inlet Mach number is 0.3. As expected, the pressure loss coefficient increases with increasing drag, but it is of interest to note that with increasing turbine blade camber (resulting in a larger flow deflection), Y_p decreases and may even become negative. Evidently, within this model, the compressive effect on the pressure side of the blade can overcome losses due to drag, provided that the flow deflection can be maintained to be large. A possible explanation for the earlier VBI2D results may then be as follows. At higher Reynolds numbers, with minimal flow separation, the exiting flow angle is relatively large (approximately -65 deg. from Fig. 2a), while the turbine blade drag is relatively high (refer to Fig. 5c). As the Reynolds number decreases, the suction side boundary layer separates, thereby decreasing turbine blade effectiveness, and reducing the flow deflection angle. However, owing to the relative loss of suction, both lift and drag are reduced on the blade. According to Fig. 12, it is possible for the pressure loss coefficient to remain relatively constant in this process, provided the drag decreases as the deflection angle decreases. It had been anticipated earlier that with flow separation, the flow angle would decrease but that the drag would increase, resulting in a large increase of Y_p . This explanation is to be regarded as being provisional, with further investigation recommended.

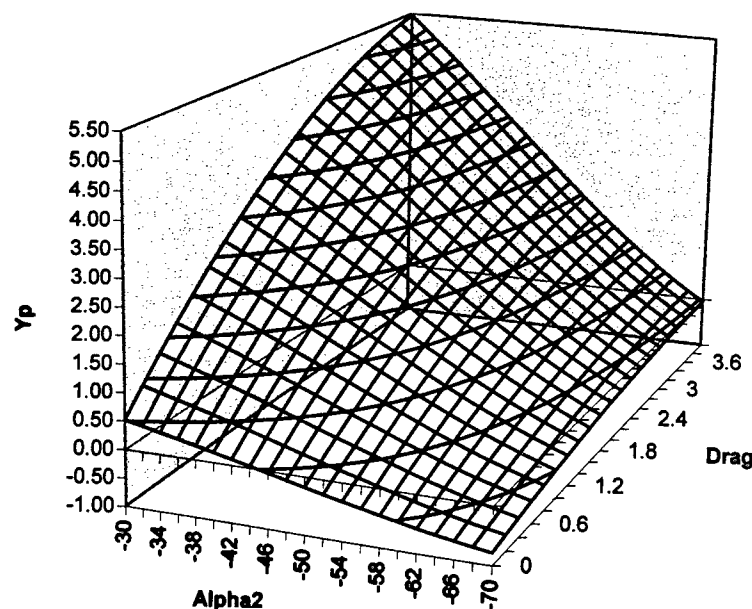


Figure 12: Total Pressure Loss Coefficient (Y_p) vs Drag and Flow Deflection (α_2) Using Eq. (12).
($\alpha_1 = 45^\circ$; $M_1 = 0.3$)

CONCLUSIONS

The unsteady VBI2D code has been applied to a sequence of seven different Re flows around a stationary cascade of Langston turbine blades influenced by the wakes of moving upstream cylinders. Vortex shedding and wake deficits cause significant shifts (± 20 deg.) in the effective flow incidence at the blade nosetip. At the lower Reynolds numbers (16.5K-66K), extensive suction side separations have been computed and quantified, whereas in the higher Re range (132K-660K), flow separation is non-existent or of minimal extent. Computations of the total pressure loss coefficient are presented at twenty equal time steps within a flow interaction cycle, and cycle averaged results are presented as a function of Re. A simplified analytical model is formulated to aid in the interpretation of the Yp results. A provisional conclusion is that while flow separation causes a dramatic loss of turbine power, the competing effects of decreased flow deflection and decreased drag combine so that Yp is not altered as much as had been anticipated initially. The identification of these unsteady phenomena, along with unanticipated swings in total pressure, provide substance for continued investigation.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the support received from the AFOSR Summer Research Program through RDL. The research was conducted at Wright-Patterson AFB in the Turbines Branch, directed by Dr. Charles D. MacArthur. The work was made possible through the guidance of, and collaboration with, Dr. Richard B. Rivir, who served as the Laboratory Focal Point. The author also wishes to thank Dr. Robert Gray for his many insightful discussions during this period, and Dr. Rolf Sondergaard for his invaluable assistance with VBI2D and its implementation. Appreciation is also expressed for the many stimulating discussions held with Mr. Gregory C. Harding during a parallel research topic, appearing under his authorship and titled, "Interactions Between Weakly Ionized Gas Plasmas And Shock Waves, A Review".

REFERENCES

1. Rao, K.V., Delaney, R.A., and Topp, D.A., 1994, "Turbine Vane-Blade Interaction: Vol. I, 2-D Euler/Navier-Stokes Aerodynamic and Grid Generation Developments," Wright Patterson AFB, WL-TR-94-2073.
2. Rao, K.V., Griffin, J.W., and Delaney, R.A., 1994, "Turbine Vane-Blade Interaction: Vol. II, 2-D Computer Program User's Manual," Wright Patterson AFB, WL-TR-94-2074.
3. Lewis, J.P., Delaney, R.A., and Hall, E.J., 1989, "Numerical Prediction of Turbine Vane-Blade Aerodynamic Interaction," *ASME Journal of Turbomachinery*, Vol. 111, pp. 387-393.
4. Rao, K.V., Delaney, R.A., and Dunn, M.G., 1994, "Vane-Blade Interaction in a Transonic Turbine, Part I: Aerodynamics," *Journal of Propulsion and Power*, Vol. 10, No. 3, pp. 305-317.

5. Rivir, R., Sondergaard, R., Dahlstrom, M., and Ervin, E., 1996, "Low Reynolds Number Turbine Blade Cascade Calculations," ISROMAC-6 The 6th International Symposium on Transport Phenomena and Dynamics of Rotating Machinery, Honolulu, HI.
6. Langston, L.S., Nice, M.L., and Hooper, R.M., 1977, "Three-Dimensional Flow Within a Turbine Cascade Passage," *ASME Journal of Engineering for Power*, Vol. 99, pp. 21-28.
7. Langston, L.S., 1989, "Research on Cascade Secondary and Tip-Leakage Flows – Periodicity and Surface Flow Visualization," AGARD Conference Proceedings No. 469, pp. 1-19.
8. Baldwin, B.S., and Lomax, H., 1978, "Thin Layer Approximation and Algebraic Model for Separated Turbulent Flows," AIAA Paper 78-257, pp. 1-8.
9. Blair, M.F., 1994, "An Experimental Study of Heat Transfer in a Large-Scale Turbine Rotor Passage," *ASME Journal of Turbomachinery*, Vol. 116, pp. 1-13.
10. Baughn, J.W., Butler, R.J., Byerley, A.R., and Rivir, R.B., 1995, "An Experimental Investigation of Heat Transfer, Transition, and Separation on Turbine Blades at Low Reynolds Number and High Turbulence Intensity," ASME International Mechanical Engineering Congress and Exposition.
11. Murawski, C.S., Sondergaard, R., Rivir, R.B., Kambiz, V., Simon, T.W., and Volino, R.J., 1997, "Experimental Study of the Unsteady Aerodynamics in a Linear Cascade with Low Reynolds Number Low Pressure Turbine Blades," ASME Paper 97-GT-95.
12. Graziani, R.A., Blair, M.F., Taylor, J.R., and Mayle, R.E., 1980, "An Experimental Study of Endwall and Airfoil Surface Heat Transfer in a Large Scale Turbine Blade Cascade," *Journal of Engineering for Power*, Vol. 102, pp. 257-267.
13. Dring, R.P., Joslyn, H.D., and Hardin, L.W., 1982, "An Investigation of Axial Compressor Rotor Aerodynamics," *ASME Journal of Engineering for Power*, Vol. 104, pp. 84-96.
14. Blair, M.F., Dring, R.P., and Joslyn, H.D., 1989, "The Effects of Turbulence and Stator/Rotor Interactions on Turbine Heat Transfer: Part 2 - Effects of Reynolds Number and Incidence," *ASME Journal of Turbomachinery*, Vol. 111, pp. 91-103.
15. Hodson, H.P., 1984, "Measurements of Wake-Generated Unsteadiness in the Rotor Passages of Axial Flow Turbines," *Journal of Engineering for Gas Turbines and Power*, Vol. 107, pp. 467-476.
16. Zaccaria, M.A., and Lakshminarayana, B., 1997, "Unsteady Flow Field Due to Nozzle Wake Interaction With the Rotor in an Axial Flow Turbine: Part I – Rotor Passage Flow Field," *ASME Journal of Turbomachinery*, Vol. 119, pp. 201-213.
17. Doorly, D.J., and Oldfield, M.L.G., 1985, "Simulation of the Effects of Shock Wave Passing on a Turbine Rotor Blade," *Journal of Engineering for Gas Turbines and Power*, Vol. 107, pp. 998-1006.
18. Young, C.D., Han, J.C., Huang, Y., and Rivir, R.B., 1992, "Influence of Jet-Grid Turbulence on Flat Plate Turbulent Boundary Layer Flow and Heat Transfer," *ASME Journal of Heat Transfer*, Vol. 114, pp. 65-72.

19. Welsh, S.T., Barlow, D.N., Butler, R.J., Van Treuren, K.W., Byerley, A.R., Baughn, J.W., and Rivir, R.B., 1997, "Effect of Passive and Active Air-Jet Turbulence on Turbine Blade Heat Transfer," ASME Paper No. 97-GT-131, 1997, pp. 1-11.
20. Funazaki, K., 1996, "Studies on Wake-Affected Heat Transfer Around the Circular Leading Edge of Blunt Body," *ASME Journal of Turbomachinery*, Vol. 118, pp.452-460.
21. Funazaki, K., Yokota, M., and Yamawaki, S., 1993, "Effect of Periodic Wake Passing on Film Effectiveness of Discrete Cooling Holes Around the Leading Edge of a Blunt Body," *ASME Journal of Turbomachinery*, Vol. 119, pp. 292-301.
22. Fridman, E., 1997, "Simulation of Heat Transfer From Flow With High Free-Stream Turbulence to Turbine Blades," *ASME Journal of Turbomachinery*, Vol. 119, pp. 284-291.
23. Abhari, R.S., 1996, "Impact of Rotor-Stator Interaction on Turbine Blade Film Cooling," *ASME Journal of Turbomachinery*, Vol. 118, pp. 123-133.
24. Arnone, A., and Pacciani, R., 1996, "Rotor-Stator Interaction Analysis Using the Navier-Stokes Equations and a Multigrid Method," *ASME Journal of Turbomachinery*, Vol. 118, pp. 679-689.
25. Koya, M., and Kotake, S., 1985, "Numerical Analysis of Fully Three-Dimensional Periodic Flows Through a Turbine Stage," *Journal of Engineering for Gas Turbines and Power*, Vol. 107, pp. 945-952.
26. Boyle, R.J., and Jackson, R., 1997, "Heat Transfer Predictions for Two Turbine Nozzle Geometries at High Reynolds and Mach Numbers," *ASME Journal of Turbomachinery*, Vol. 119, pp. 270-283.
27. Kang, S., and Hirsch, C., 1996, "Numerical Simulation of Three-Dimensional Viscous Flow in a Linear Compressor Cascade With Tip Clearance," *ASME Journal of Turbomachinery*, Vol. 118, pp. 492-505.
28. Halstead, D.E., 1996, "Boundary Layer Development in Multi-Stage Low-Pressure Turbines," Doctoral Dissertation, Iowa State University.
29. Dorney, D.J., Davis, R.L., and Edwards, D.E., 1992, "Investigation of Hot Streak Migration and Film Cooling Effects on Heat Transfer in Rotor/Stator Interacting Flows," United Technology Research Center Report 91-29, N00140-88-C-0677.
30. Rai, M.M., 1989, "Three-Dimensional Navier-Stokes Simulations of Turbine-Stator Interaction," *ALAA Journal of Propulsion and Power*, Vol. 5, pp. 307-319.
31. Denton, J.D., 1993, "Loss Mechanisms in Turbomachines," *ASME Journal of Turbomachinery*, Vol. 115, pp. 621-656.
32. Muti Lin, J.C., and Pauley, L.L., 1996, "Low-Reynolds Number Separation on an Airfoil," *AIAA Journal*, Vol. 34, No. 8, pp. 1570-1577.
33. Hawthorne, W.R., 1964, *Aerodynamics of Turbines and Compressors*, Princeton University Press.
34. Gray, R.E., 1998, Private Communication.

DEVELOPMENT OF A PROBABILISTIC
ASSESSMENT FRAMEWORK FOR HIGH CYCLE FATIGUE
FAILURES OF GAS TURBINE ENGINE BLADES

M.-H. Herman Shen
Associate Professor
Dept. of Aerospace Engineering,
Applied Mechanics, & Aviation

The Ohio State University
328 Bolz Hall, 2036 Neil Avenue
Columbus, Ohio 43210-1276

Final Report for:
Summer Faculty Research Program
AFRL Materials Directorate

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

and

AFRL Materials Directorate

August 1998

DEVELOPMENT OF A PROBABILISTIC ASSESSMENT FRAMEWORK FOR HIGH CYCLE FATIGUE FAILURES OF GAS TURBINE ENGINE BLADES

M.-H. Herman Shen
Associate Professor
Dept. of Aerospace Engineering,
Applied Mechanics, & Aviation
The Ohio State University

Abstract

A framework for the probabilistic analysis of high cycle fatigue is developed. The framework will be useful to U.S. Air Force and aeroengine manufacturers in the design of high cycle fatigue in disk or compressor components fabricated from Ti-6Al-4V under a range of loading conditions that might be encountered during service. The main idea of the framework is to characterize vibratory stresses from random input variables (uncertainties such as initial crack size, crack location, loading, material properties, and manufacturing variability). The characteristics of such vibratory stresses will be portrayed graphically as histograms, or probability density function (PDF). The outcome of the probability measures associated with all the values of a random variable exceeding the material capability is achieved by a failure function $g(\mathbf{X})$ defined by the difference between the vibratory stress and Goodman line or surface such that the probability of HCF failure is $P_f = P[g(\mathbf{X}) \leq 0]$. The framework can be used to facilitate the development design tools for the prediction of inspection schedules and reliability in aeroengine components. Such tools could lead ultimately to improved life extension schemes in aging aircraft, and more reliable methods for the design and inspection of critical components.

DEVELOPMENT OF A PROBABILISTIC ASSESSMENT FRAMEWORK FOR HIGH CYCLE FATIGUE FAILURES OF GAS TURBINE ENGINE BLADES

M.-H. Herman Shen

Introduction

The gas turbine engines operating environment of nonuniform flow field repeated loading creates conditions that are favorable to high cycle fatigue (HCF) degradation of metallic rotating components. This type of component demands durability, high reliability, light weight, and high performance. Therefore, lifetime failure-free design criteria based on Goodman Diagram has been adopted by the aircraft engine design community for ensuring safety of critical structural components.

For example, a traditional HCF turbine blading system design procedure is shown schematically in Figure 1. This design process usually consists of a structural dynamics analysis to determine natural frequencies and mode shapes at certain operating speed ranges and a stress analysis to calculate dynamic stress distribution for identifying maximum vibratory stress location or area under a series of given excitations. Once the maximum stresses for each vibration mode are determined, high cycle fatigue assessment can be achieved by measuring the margin between the maximum vibratory stress and the material fatigue capability which is a straight line (so called Goodman line in this study) drawn between the mean ultimate strength at zero vibratory stress and mean fatigue strength at 10^7 cycles (or infinite life).

Therefore, the traditional HCF design procedure is usually established by using the results of a simple deterministic stress analysis without take into account the information

such as degradation of material properties, scatter in fatigue life testing, and uncertainties inherent in the operating conditions in the real world.

In turn, as it has been reported, a number of structural failures have occurred on aircraft engines during development testing and operational service. These engines were optimally designed and assumed to be valid during their service life. These incidents triggered an awareness of the fact that current aircraft engine critical structural components, blading system in particular, although they satisfy the lifetime failure-free design criteria, sometimes fail.

Failure Analysis of Gas Turbine Blades

Gas turbine blade failures are generally resulting from high vibratory stress due to excessive vibration in blading during forced response. Forced response usually occurs, under a nonuniform flow field operating environment, when the excitation frequency from unsteady aerodynamic load is coincident with any blade's natural frequency. Hence, an adequate blading design will be the case if all the resonant conditions could be avoided. Realistically, it is impossible due to the fact that insufficient real world loading information provided at the time of the analysis and design.

Since the effects of randomness in real-world operating environment cannot be quantified accurately by a traditional deterministic design approach as shown in Figure 1, it is beyond the capability of such approach to provide the important information regarding risk of failure, or the sensitivities of the design variables to risk. Hence, in this work, a probabilistic prediction procedure as presented in Figure 2, which accounts for variability in geometry and aerodynamic loading has been developed to calculate the important design and maintenance information of sensitivity, (high cycle) fatigue reliability, and associated risk. In addition, in this study, the reliability due to the variability in materials (e.g. damage, cracks, degradation, etc.) will be discussed.

Blade Vibration Model

An analytical procedure has been obtained by Shen [1] for modeling the behavior of rotating blades with taking into account the effects, so-called "centrifugal-softening and centrifugal-stiffing", of the displacement-dependent centrifugal forces. The procedure provides close form equations of motion for a rotating flat uniform blade in flapping motion and lead-lag motion as

$$EI_z(w'''' + \beta_x''') + m\ddot{w} + m\Omega^2 xw' - \frac{1}{2}m\Omega^2(l^2 - x^2)w'' = 0, \quad (1)$$

$$EI_z(w''' + \beta_x'') - A\kappa G\beta_x = 0, \quad (2)$$

$$EI_y(v'''' + \beta_y''') - m\ddot{v} + m\Omega^2 xv' - \frac{1}{2}m\Omega^2(l^2 - x^2)v'' = 0, \quad (3)$$

$$EI_y(v''' + \beta_y'') - A\kappa G\beta_y = 0, \quad (4)$$

Here, v , w are the displacements, β_x , β_y are the shears angles, κ is the shear correction factor, and E , G are the effective blade bending and shear moduli, respectively.

For high cycle fatigue applications, in accordance with the above analytical procedure, a 3-D finite element model (Figure 3) was developed [2] through classical variational principles. The finite element model can be used in developing the HCF reliability model.

HCF Reliability Model

HCF is a result of repeated load, but well below the yield strength of the material cycling, and the consequential damage, places restrictions on the usable lifetime of gas turbine engines experience these loads. The cycling can be mechanical (e.g. vibration), aeroelastic (e.g. flutter, resonance) or environmental (e.g. thermal), and in its early stages the accumulating damage is often very difficult to detect. The mechanical or aeroelastic

cycling due to variability in geometry, aerodynamics, and materials cannot be predicted with certainty, and must be considered a random process. The effects of this randomness cannot be quantified accurately by a deterministic analysis, and loading must therefore be treated as a random variable.

During the design phase or service life, these uncertainties or random variables can be combined with other operating experience from tests and flights to calculate reliability and risk using a probabilistic analysis procedure. The main idea of the probabilistic procedure is to characterize vibratory stresses from random input variables (uncertainties such as initial crack size, crack location, loading, material properties, and manufacturing variability). The characteristics of such vibratory stresses will be portrayed graphically as histograms, or probability density function (PDF).

Failure State Function

The outcome of the probability measures associated with all the values of a random variable below a target value is a cumulate distribution function (CDF). For example, in structural reliability analysis, a PDF or a failure state function or limit state function $g(\mathbf{X})$ needs to be defined by the difference between the vibratory stress (σ) and Goodman line or surface (ϕ) in terms of a vector of basic random variables (or uncertainties), $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$,

$$g(\mathbf{X}) = \phi - \sigma(\mathbf{X}) \quad (5)$$

such that the CDF or the probability of failure is

$$P_f = P[g(\mathbf{X}) \leq 0] \quad (6)$$

Example: Structural Reliability of Turbine Blades

This example illustrates a general procedure for predicting gas turbine blade's structural reliability and associate risk when high vibratory stress fluctuation develop due to uncertainties such as manufacturing variability and other stress raisers that arise when the excitation frequency from unsteady aerodynamic load is coincident with any blade's natural frequency.

Suppose that a high vibratory stress fluctuation is developed in a rotating gas turbine blade and progressively increases along a particular stress ratio R . The first step in avoiding the catastrophic failure of the entire blading system is estimating the reliability or the probability of failure of the blade when high vibratory stress appears.

As it has been discussed above, the reliability estimation procedure consists of determining the probability density function (PDF) and calculating the cumulative probability of failure (CDF) for crack propagation through FPI process.

If we choose the vibratory stress σ as the probability measure, the performance function of FPI can be written as

$$g(\mathbf{X}) = \phi - \sigma(\mathbf{X}) \quad (7)$$

where ϕ is the Goodman line or Goodman surface. In this study, the vibratory stress fluctuation is assumed to be controlled by the uncertainties that associate with the manufacturing variability and characterized by structural damping variation. The vibratory stress at the various high frequency modes is calculated using the 3-D finite element model [2].

An example of a gas turbine blade that was tested is made of Ti-6Al-4V having dimensions: 3.0 inch wide $\times$ 7.0 inch long. The Haigh or Goodman diagram of the Ti-6Al-4V as shown in Figure 4 is obtained from the study by Maxwell and Nicholas [3]. These data will be used to form the failure criteria or failure boundary ϕ in the failure state function g for the prediction of HCF reliability and risk.

A finite element analysis has been conducted for the damped free and force vibrations analysis of the non-rotating and rotating blades using the blade model as shown in Figure 3. In accordance with the testing data available in the public domain [4], highest vibratory stresses are usually occurred on the high frequency stripe vibration modes. Hence, in this study, the proposed reliability prediction procedure is demonstrated on the third stripe modes only.

As shown in Figure 5, the third (11925.1 Hz) free stripe vibration mode (3s) of the blade are calculated with damping loss factor (ξ) 0.0077 (inverse of the damping capability). Next, let's consider the cases of the blade under a concentrated external harmonic excitation force and a static load applying at center portion of the blade. The magnitudes of the excitation force and the static load are allowed to be adjusted for achieving appropriate stress ratios and vibratory stress ranges.

The 3rd (11925 Hz) stripe mode forced response was computed for various loss factors ξ obtained from the normally distributed testing data set with 0.0077 mean value and the 0.0025 standard deviation. The mode shape and the corresponding principle stress distribution are shown in Figures 6 and 7. One should note that the mode shape is in close agreement with the free vibration mode shape in Figure 5, and the maximum von Mises vibratory stress, as expected, occurs near the trailing edge of the blade.

Figure 8 shows the von Mises vibratory stress distributions $\sigma(\xi)$ calculated through finite element analysis under various values of damping at stress ratio (R) equation to - 1. The distributions are presented in the form of probabilistic density function (PDF), the probability of stress occurrence, for reliability prediction. With the vibratory stress distributions $\sigma(\xi)$ the probability of failure, p_f , reliability prediction has been accomplished via the formulation

$$p_f = \int_{-\infty}^{\infty} \sigma(\xi) \phi d\sigma_v \quad (8)$$

and qualitatively represented by the interference between the PDF curve and the Goodman surface (due to the scatter in testing data). In other words, the larger the overlapped area indicates the higher the failure probability.

However, in general, such formulation is very difficult to evaluate not only because of the multiple integration, but the joint probability distribution may not even be available. Therefore, a numerical method (FPI) [5] is employed in this study to calculate an approximation of the probability of failure p_f distribution (a cumulative distribution function CDF) as shown in Figure 9.

Probabilistic Goodman Diagram

With the vibratory stress distributions $\sigma(\xi)$ calculated through finite element analysis for various values of stress ratio R and alternating stress, the probability of failure, p_f is qualitatively represented by the intersection of the vibratory stress distribution and the Goodman line as shown in the 3-dimensional plot on Figure 10. In turn, a new reliability design diagram, which is named the probabilistic Goodman Diagram, is constructed by plotting the probability of failure, p_f locus on the original Goodman diagram. Furthermore, since the reliability of the blade can be defined as the probability of non-failure, and is given by $R_s = 1 - p_f$, the probabilistic Goodman Diagram is then extended to include the R_s locus as shown in Figure 10.

Conclusion

A probability-based risk assessment framework has been developed to predict reliability of gas turbine engine blades subjected to high cycle fatigue. The procedures provides a systemic approach for predicting and designing turbomachinery blading reliability against various potential fatigue problems for all relevant vibratory modes and taking into account variability in geometry (e.g. dimensional variation, mistuning, etc.), unsteady aerodynamics, structural damping, and thermal loading. The variability in materials (e.g. damage,

cracks, degradation, etc.) can be also considered in this approach. A reliability prediction was performed on gas turbine blades at high frequency modes (e.g. third stripe mode) using a probabilistic vibratory stress distribution in conjunction with the Haigh or modified Goodman Diagrams. The cumulative reliability (R_s) and probability of failures (p_f) are then calculated using the fast probability integration (FPI) technique to construct a novel probabilistic Goodman (or Haigh) diagram which provides blade's lifetime design guide line and an optimal maintenance strategy in management in decision-making relating to the PM/inspection scheduling, replacement, spare parts requirements.

References

- [1]. M.-H. H. Shen, "Passive Vibration Suppression of Rotating Blades Using Piezoelectric Devices", under review, *Journal of Smart Materials and Structures*.
- [2]. M.-H. H. Shen, "Reliability High Cycle Fatigue Design of Gas Turbine Blades Using the Probabilistic Goodman Diagram", under review, *International Journal of Fatigue*.
- [3]. D.C. Maxwell and T. Nicholas, "A Rapid Method for Generation of a Haigh Diagram for High Cycle Fatigue", Submitted to ASTM for review, 1997.
- [4]. A.J.A. Mom and M.D. Raizenne, "AGARD Engine Disc Cooperative Test Program", *AGARD-R-766*, 1988.
- [5]. M.-H. H. Shen and M. S. Shen, 1994, "Probability of Failure and Risk Assessment of Structures with Fatigue Cracks", *AIAA Journal*, Vol. 32, No. 12, pp. 2447-2455.

Acknowledgment

The author is grateful to Dr. Ted Nicholas for his valuable discussions and expert suggestions.

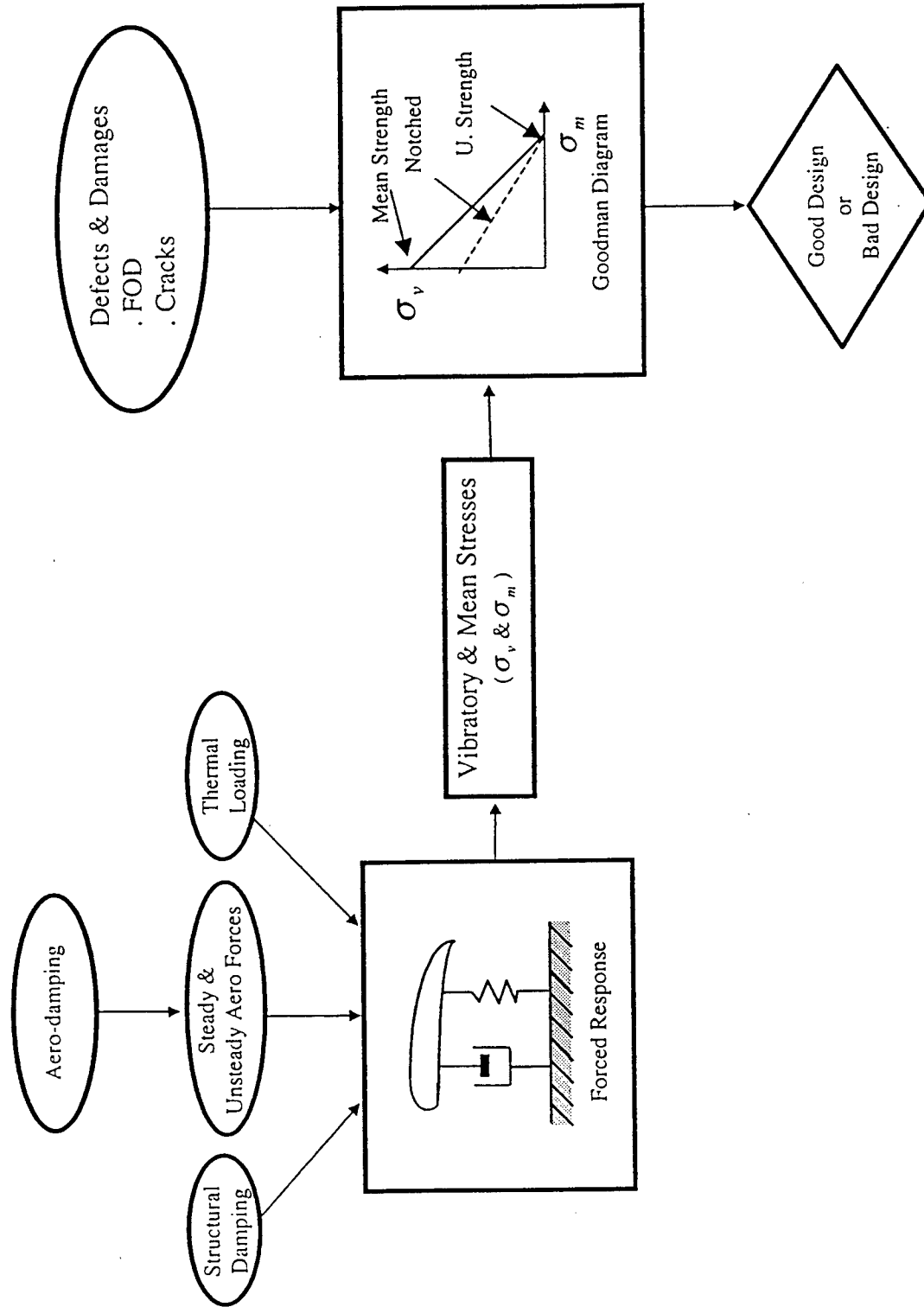


Figure 1. Traditional HCF Design Process

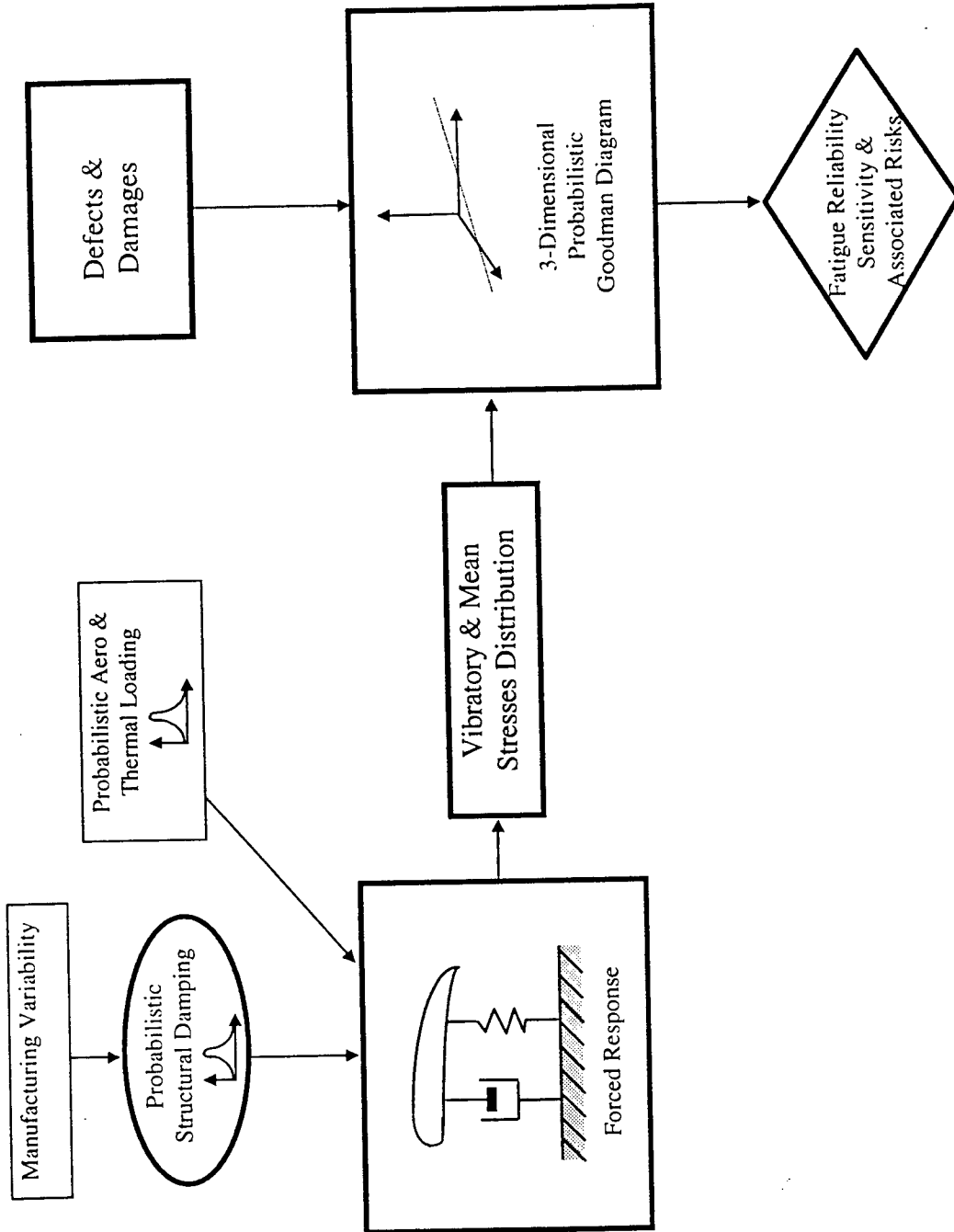


Figure 2. Proposed Reliability HCF Design Process

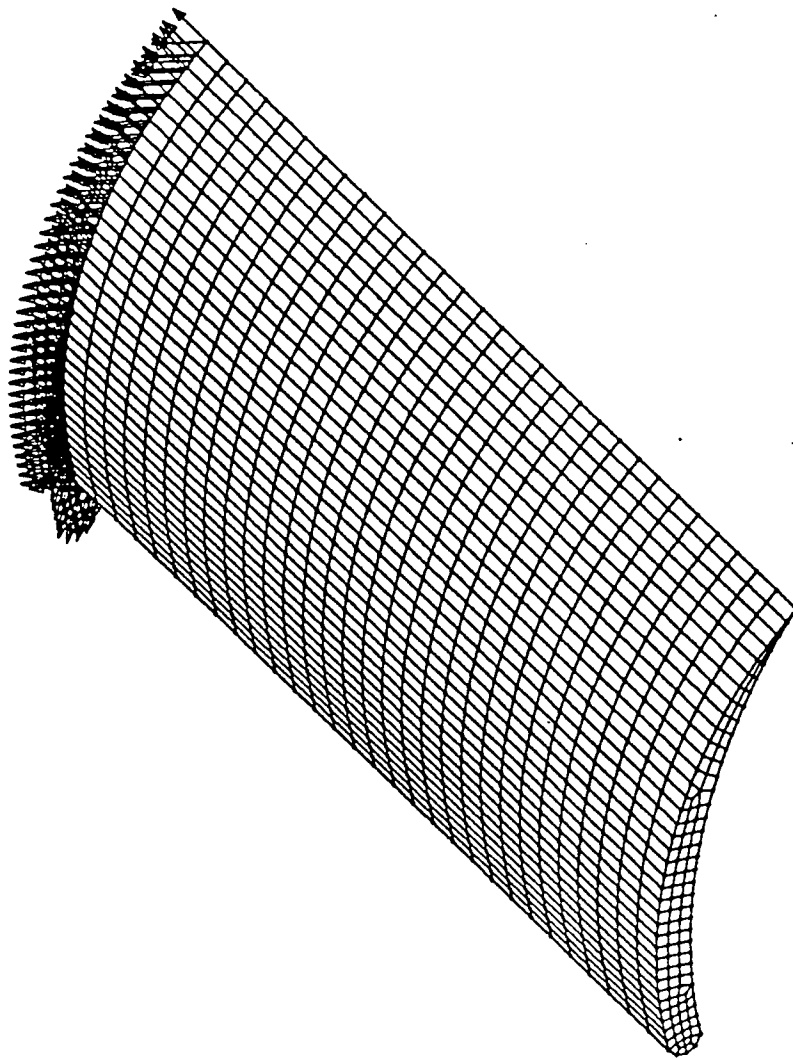


Figure 3: The Finite Element Mesh of OSU-Shen Yen-Blade#2

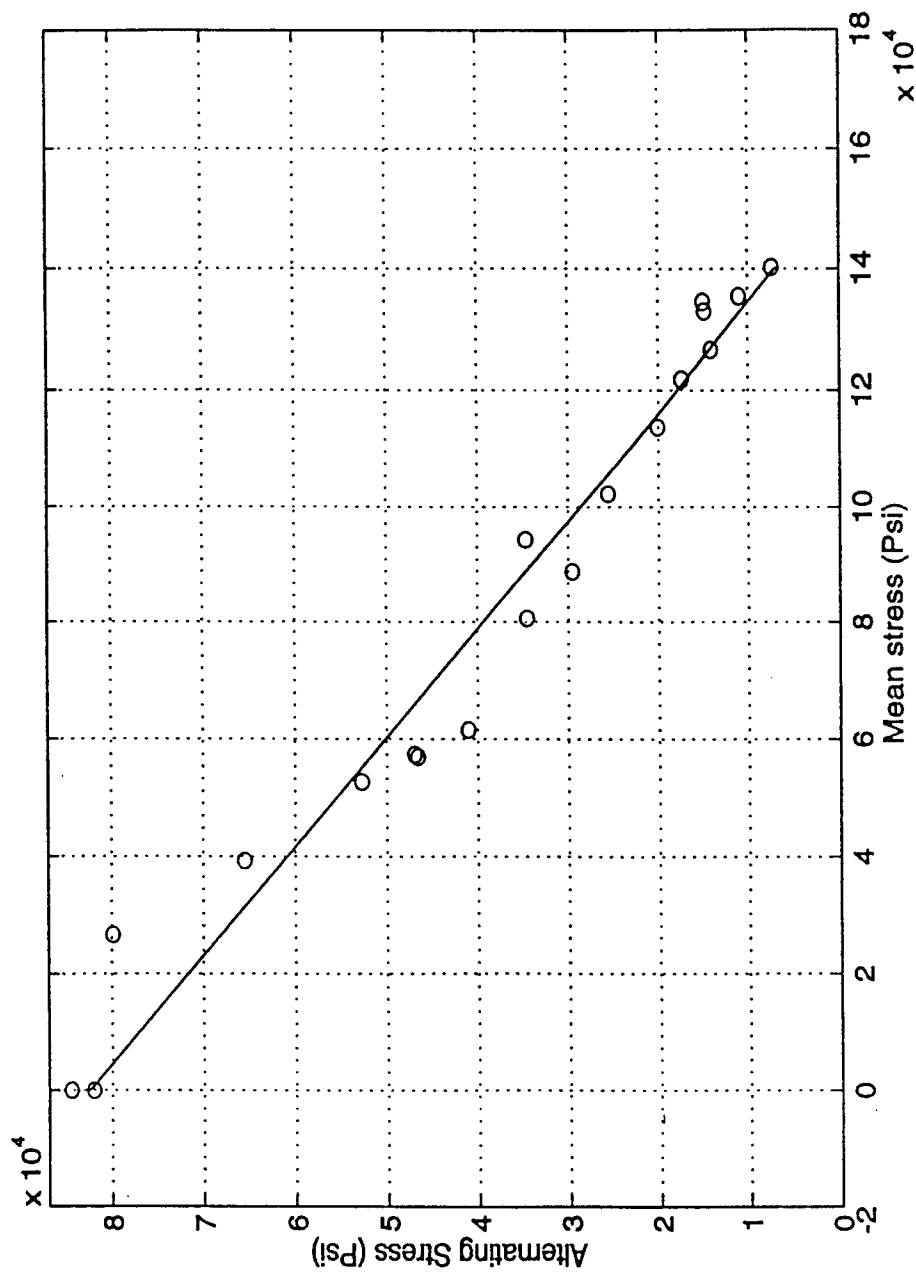


Figure 4: Hiigh (or Goodman) Diagram for Ti-6Al-4V

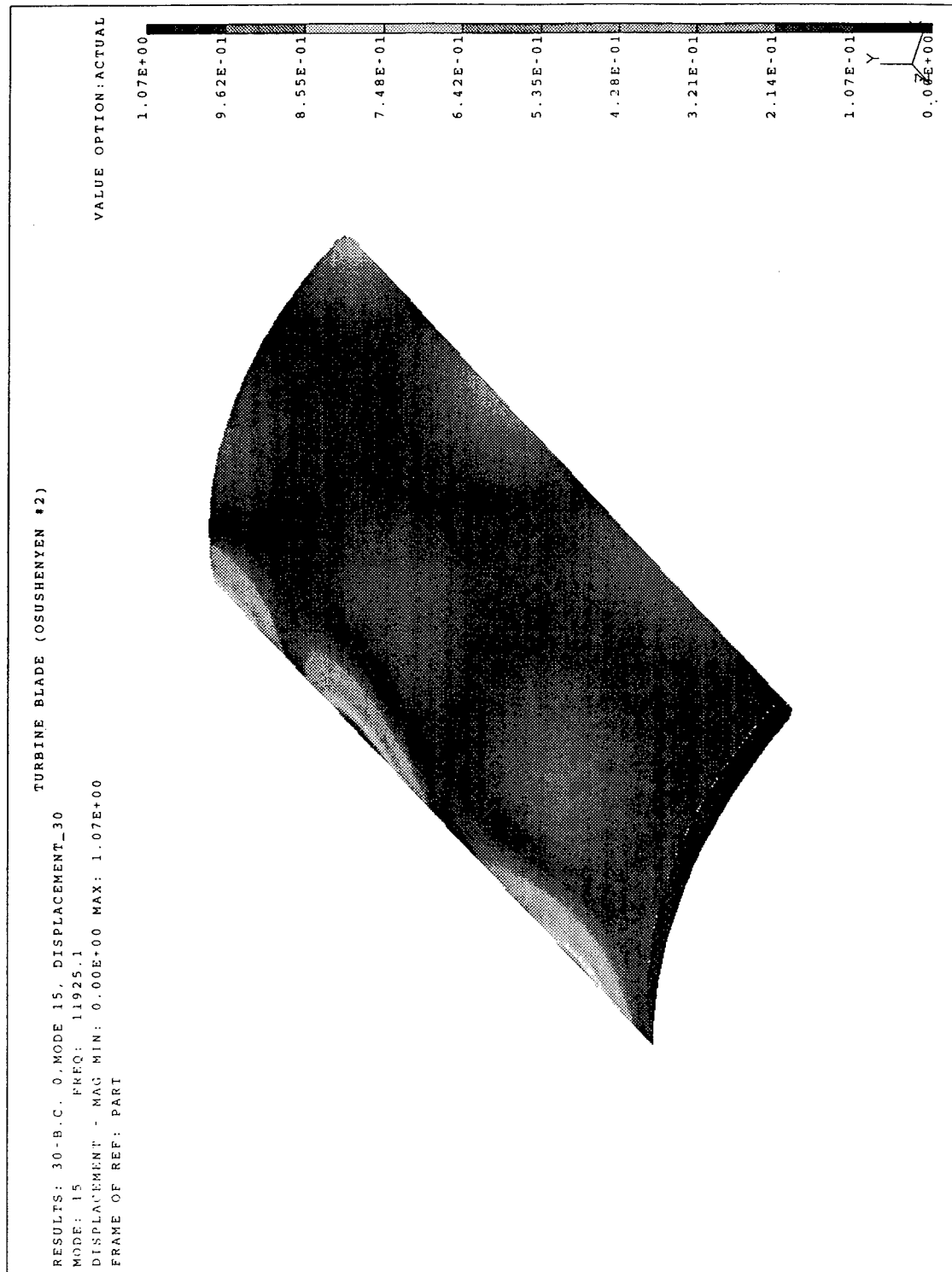


Figure 5. Free Third Stripe Vibration Mode Shape

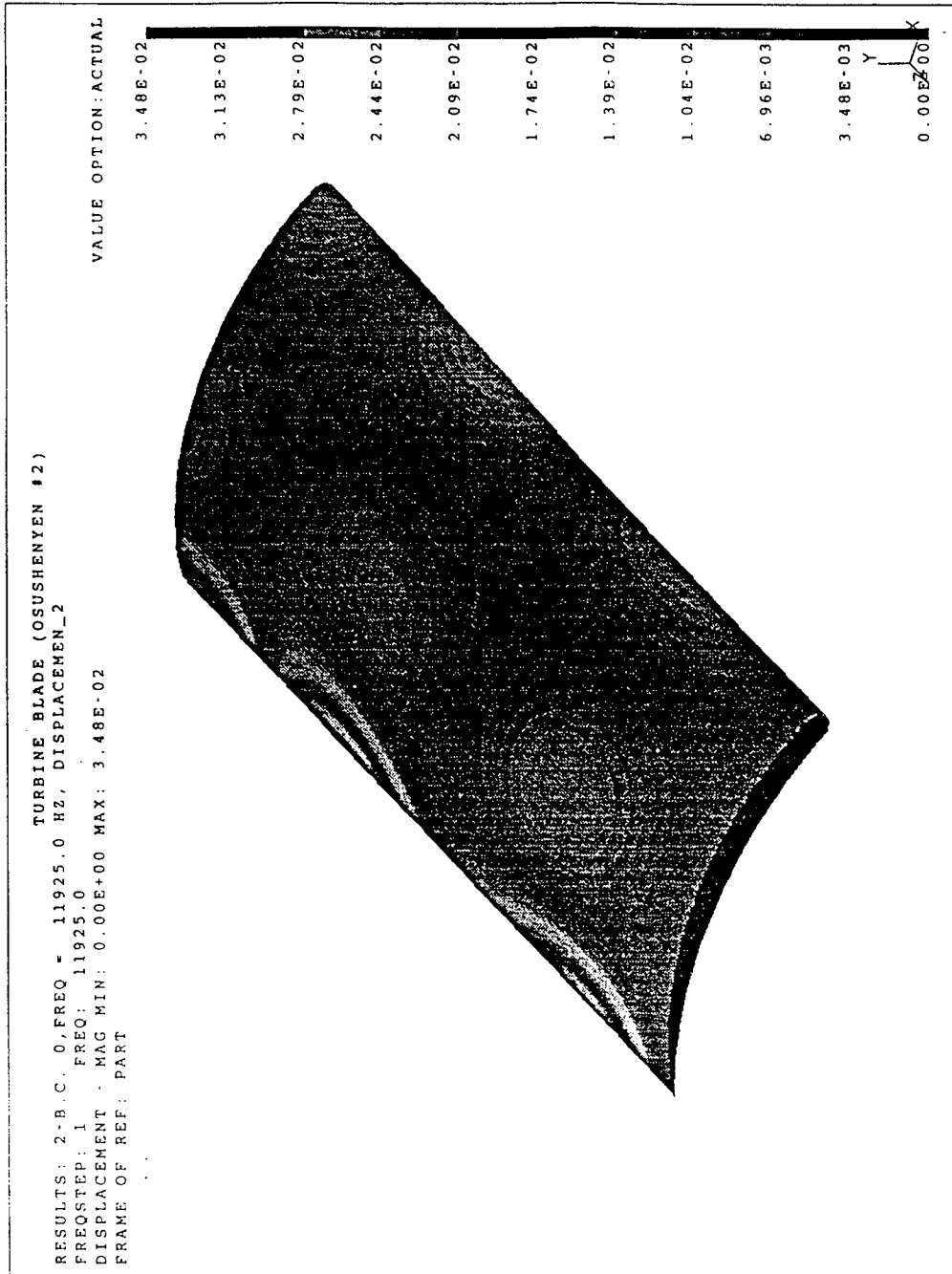


Figure 6. Forced Third Stripe Vibration Mode Shape

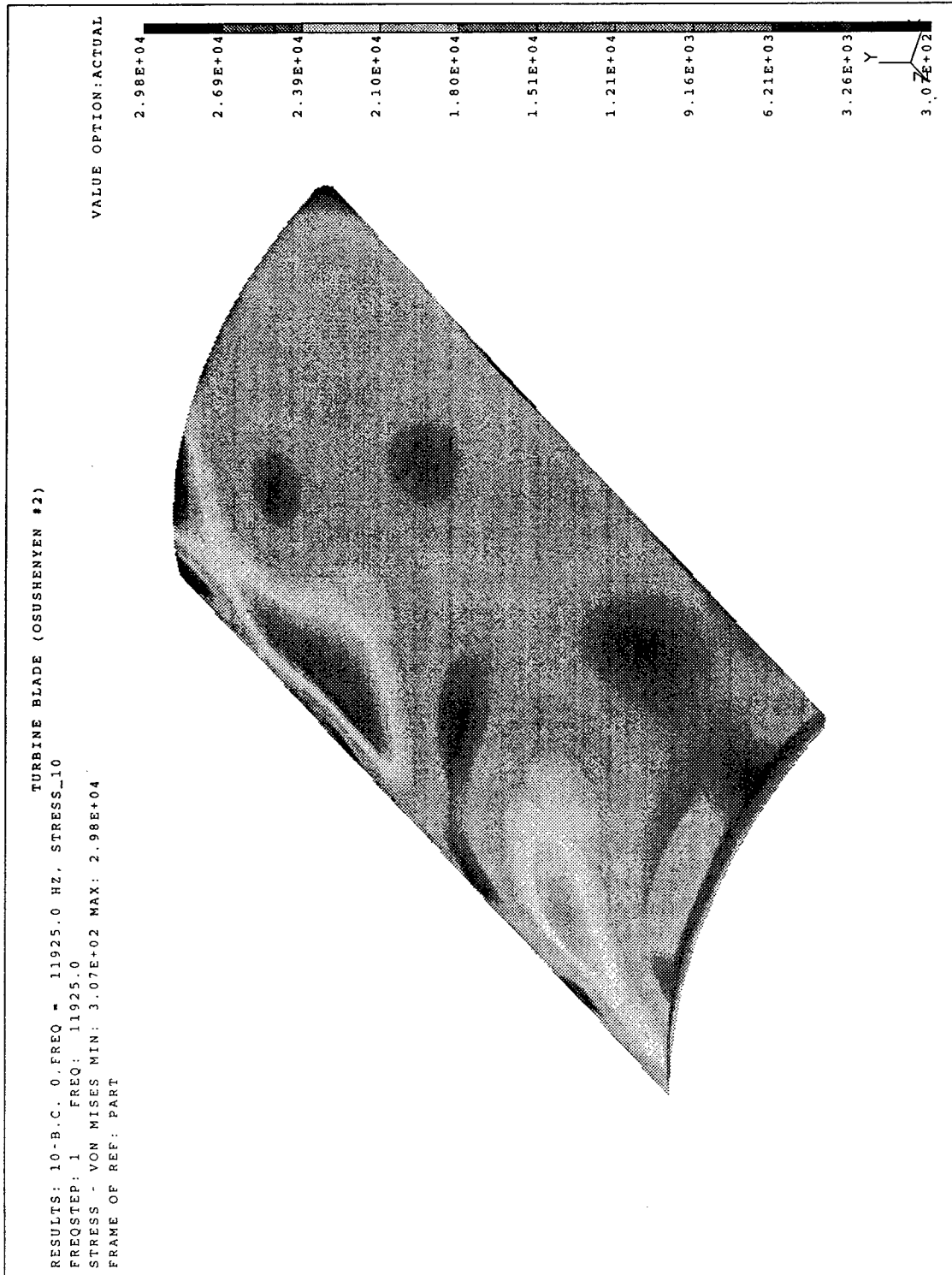


Figure 7. Vibratory Stress (von Mises) at 3rd Stripe Mode

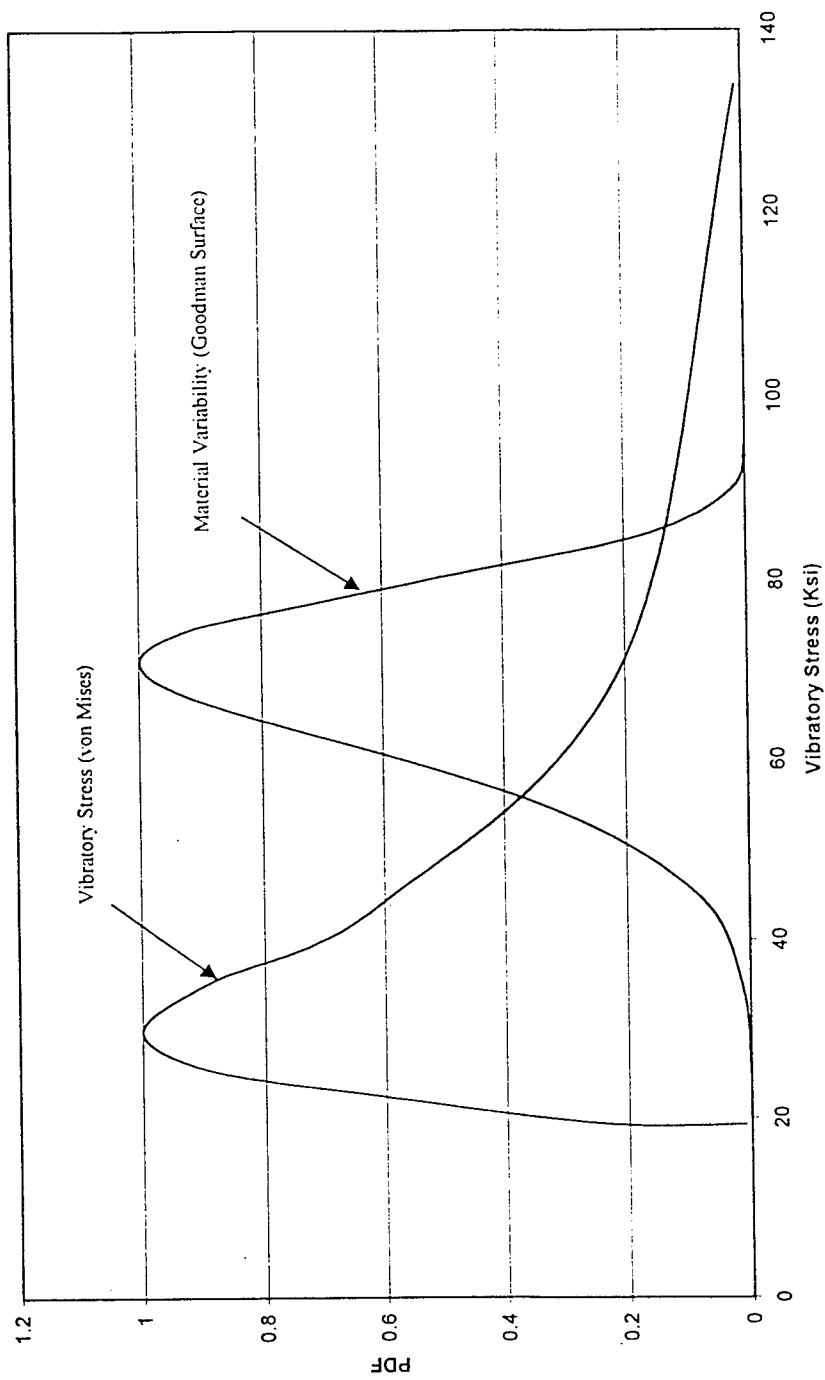
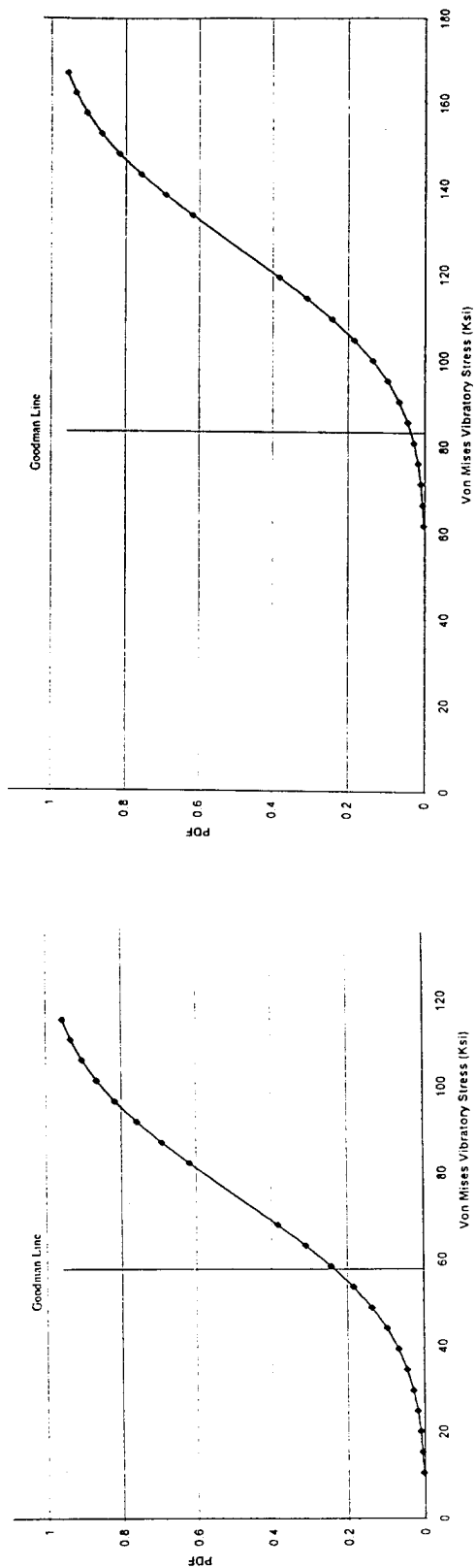
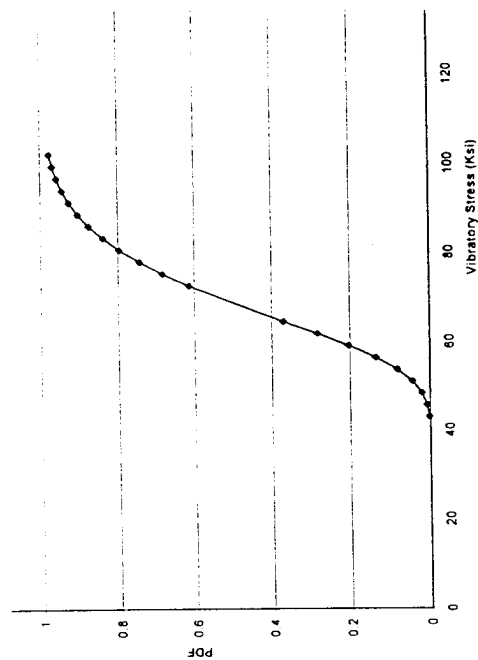


Figure 8. Vibratory Stress and Material Variability Distributions ($R=-1$)



(a) Material capacity = 57 Ksi ($R=-1$)

(b) Material capacity = 83 Ksi ($R=-1$)



(c) Material capacity is normally distributed between 57-83 Ksi ($R=-1$)

Figure 9. Cumulative Probability that Vibratory Stress > Material Capability.

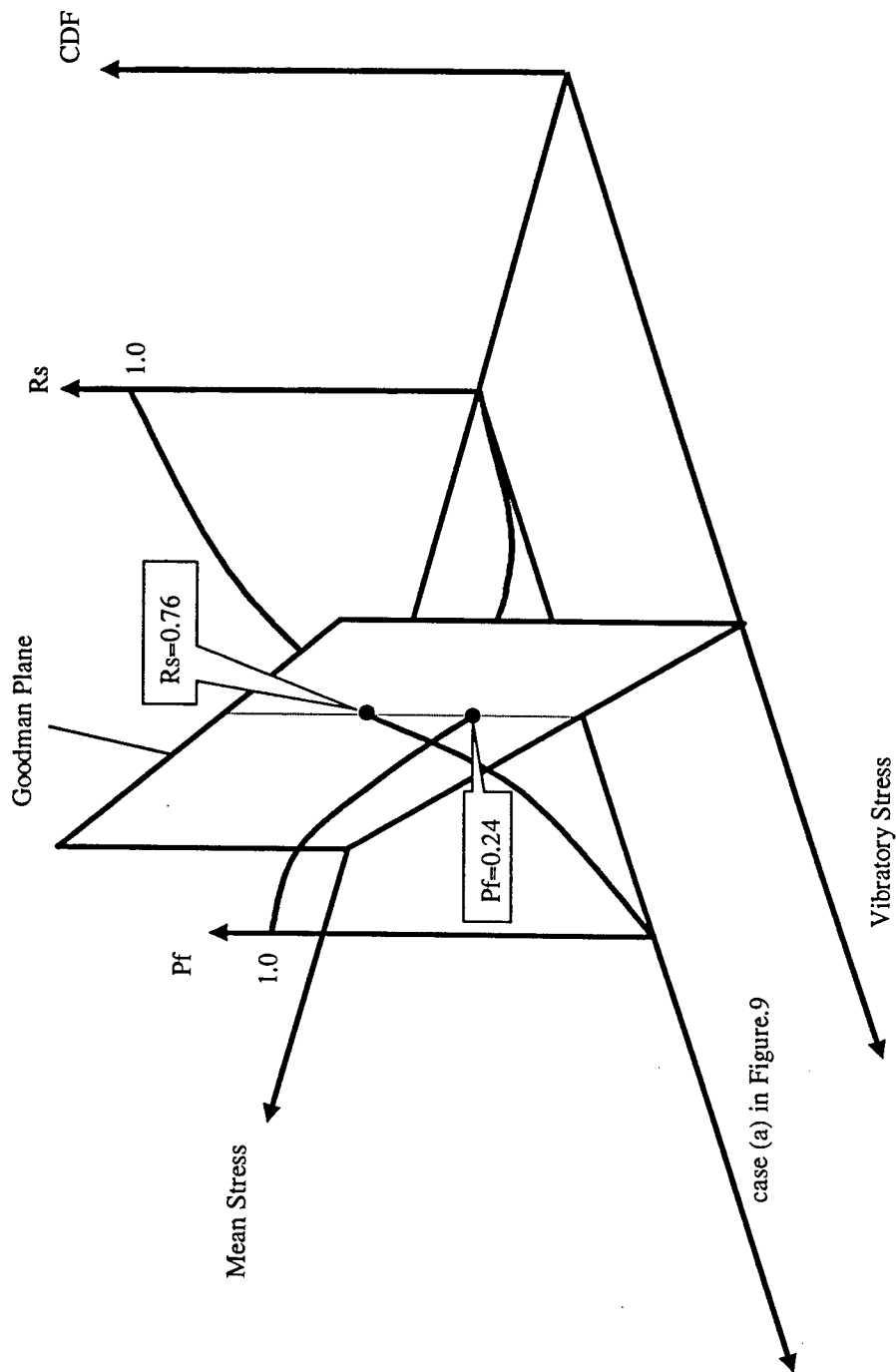


Figure 10: Proposed 3-dimensional probabilistic Goodman diagram

A STUDY OF MODELS AND TOOLS FOR
PROGRAMMING THE VGI PARALLEL COMPUTER

Hongchi Shi
Assistant Professor
Department of Computer Engineering & Computer Science
201 Engineering Building West
University of Missouri-Columbia
Columbia, MO 65211

Final Report for:
Summer Faculty Research Program
Air Force Research Laboratory / Munitions Directorate

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

and

Air Force Research Laboratory / Munitions Directorate
Eglin Air Force Base, FL

August 18, 1998

A STUDY OF MODELS AND TOOLS FOR PROGRAMMING THE VGI PARALLEL COMPUTER

Hongchi Shi

Assistant Professor

Department of Computer Engineering & Computer Science
University of Missouri-Columbia

Abstract

A recent trend in dataflow architecture design is incorporation of dataflow concept with multithreading concept. The VGI parallel computer is a multithreaded static dataflow computer intended for processing stream data present in video, graphics, and image processing applications. The massively parallel VGI computer consists of a large number of processors connected through reconfigurable communication networks. Two connected processors use a simple handshake protocol for communication and synchronization.

This report focuses on the issues of programming models, tools, and compilation technology for programming the VGI parallel computer. The issues are discussed at three programming levels: machine level, signal flow graph level, and SISAL language level. Future research work is also discussed in the report.

A STUDY OF MODELS AND TOOLS FOR PROGRAMMING THE VGI PARALLEL COMPUTER

Hongchi Shi

Introduction

The dataflow approach to parallel computing has been attractive to many researchers due to its simplicity and elegance in specifying parallelism and data dependencies [2, 8, 7, 10, 6]. The pure dataflow model allows parallel and pipelined execution at the level of individual instructions. Pure dataflow machines are theoretically able to exploit all the available program parallelism. However, scheduling each instruction separately and synchronizing on each data value lead to a large amount of overhead [10, 1, 11]. A recent trend in the design of dataflow machines is incorporation of von Neumann architectural features and multithreading concept into dataflow architectures to reduce the overhead incurred from synchronization and instruction scheduling at the instruction level [4, 7, 10, 6]. In such hybrid dataflow machines that integrate dataflow and control flow models, the basic scheduling and synchronization unit is a thread which is a sequence of instructions.

Following the ideas of dataflow and multithreading, Srimi and Rabaey have developed a multiprocessor stream data processing computer [14, 15, 16, 13]. The parallel stream data processing computer is named VGI and designed for video, graphics, and image processing applications. A VGI computer may contain several VGI chips connected using the I/O processors in the chips. Each chip contains 16 clusters of processors. Each cluster has four computational processors, one memory processor, and one I/O processor. The processors are interconnected through two levels of segmented buses. A handshake protocol is used for synchronization between two communicating processors. Each computational processor has its own instruction memory for instructions processing streams of data present in video, graphics, and image processing applications. Each memory processor can be statically configured as a delay line, FIFO, lookup table, or RAM. The I/O processors are designed for external communication. From a more abstract perspective, a VGI computer contains a set of interconnected processors that each repeatedly execute a thread. The synchronization between two processors is implemented such that it happens only when the source processor is ready to send and the destination processor is ready to receive. The handshake technique for synchronization is much simpler than the traditional token matching technique and reduces the overhead significantly. Once the VGI computer is configured, the connections among processors and the threads on the processors are fixed until the computer is reconfigured. Threads are statically scheduled to processors during configuration for execution. Each thread usually receives inputs from its predecessors; performs some operations on the inputs; and outputs the results to its successors. The predecessors and successors are statically determined by the connections.

The eventual success of the VGI computer will depend on its programmability. In general, writing parallel programs is currently much more difficult than writing sequential programs [1]. The main difficulties are how to determine which tasks execute and when, how tasks communicate and synchronize, and how to configure

the computer to accomplish the work efficiently. Despite the VGI architecture supports well parallelism at thread level and provides simple synchronization, it is still a challenge to efficiently map algorithms onto the underlying VGI computer. The users have to worry about how to assign tasks to each processor and how to connect processors through switch settings. Writing a parallel program by explicitly specifying how to handle these issues is extremely complicated. To relieve the programmers from the demanding undertaking, abstract programming models and tools have to be developed for parallel programming.

This report focuses on the issues of programming models, tools, and compilation technology for programming the VGI parallel computer. We organize the rest of the report as follows. In the next section, we give an overview of the VGI parallel computer. In the following sections, we discuss programming models and tools for the VGI computer. We also use a simple example to illustrate programming at different levels. We conclude the report with brief

conclusions and future work in the last section.

VGI Parallel Computer

The VGI computer is based on the concept of processing data streams as they are generated from sensors for applications involved with video, graphics, and image processing. It consists of VGI chips, bus interface chips, and SRAM chips on a single board or in a multiple board box and usually works as a co-processor controlled by a host such as a workstation, PC, or PDA. The VGI chips can be connected in many different ways through the I/O processors in each chip and interconnection networks.

The core of the VGI computer is the stream data processing chips referred to as VGI chips. The VGI chip is a multiple instruction multiple data stream (MIMD) parallel processing chip and operates at 66 MHz clock rate. It is composed of 16 clusters, each consisting of four VGI processors, one memory processor, and one I/O processor which are connected through a local communication network referred to as level-1 network. The level-1 networks are connected through the global level-2 network. Thus, the VGI chip has 64 VGI processors, 16 memory processors for supporting on-chip data memory, 16 input/output (I/O) processors for communication with external memory and other VGI chips, and a crossbar-like interconnection network which interfaces between the VGI processors, memory processors, and I/O processors, as shown in Figure 1. The chip can be rapidly configured through a scan chain that reaches all the processors. The scan chain is used to initialize the local instruction memory in each VGI processor, the data memory in each memory processor, the I/O connections in each I/O processor, and the switches in the communication networks which, when configured using the scan chain, allow both point-to-point and simple broadcast types of communication.

VGI Processor

The VGI processor is simple in architecture as shown in Figure 2. There is a small instruction memory that can store 16 single-word instructions. Thus, a VGI processor program can be at most 16 instructions long. The VGI processor has a controller, a 16-bit data path, and an ALU. It supports arithmetic and logical operations (including multiplication), branching, and byte manipulation.

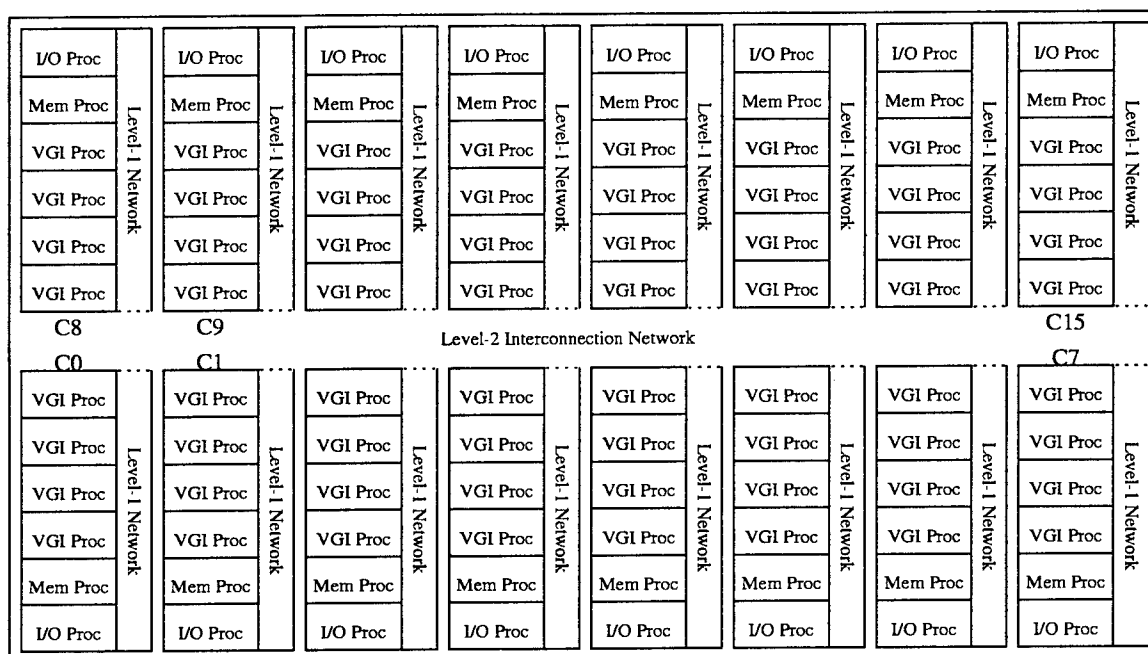


Figure 1: The major components of the VGI chip

There are 6 programmable registers in each VGI processor which can also be configured as 3 queues as shown in Figure 3. The six registers are divided into three groups of two registers each. Each group of two registers can be individually configured as either two random access registers or a two-element first-in first-out queue. Queues are used for communication with other processors. As queues they obtain inputs through the communication networks. As registers they can be initialized through the scan chain during configuration.

Memory Processor

It is known that off-chip communication is expensive in terms of power, time delay, and chip area. With on-chip memory, the memory bandwidth can be increased and off-chip traffic can be reduced. Each cluster of the VGI chip has one memory processor which contains a 256×16 memory. Like the VGI processors, the memory processor is connected to the level-1 network and has a three-stage pipeline comprising fetch, memory access, and communicate. The fetch and communicate stages are used to synchronize the memory processor with the VGI processors. The architecture of the memory processor is as shown in Figure 4.

The memory processor is the interface between other processors and the on-chip memory. It is designed to implement four types of commonly used memory accesses in signal processing: RAM, lookup table, FIFO, and delay line. When configured as a RAM, the memory processor can execute any read or write operation. A read or write request is sent through the corresponding handshake line. The lookup table mode is similar to the RAM mode except that write is not allowed during execution and the table in the memory array is loaded through the scan chain during configuration. In the FIFO mode, addresses for read and write

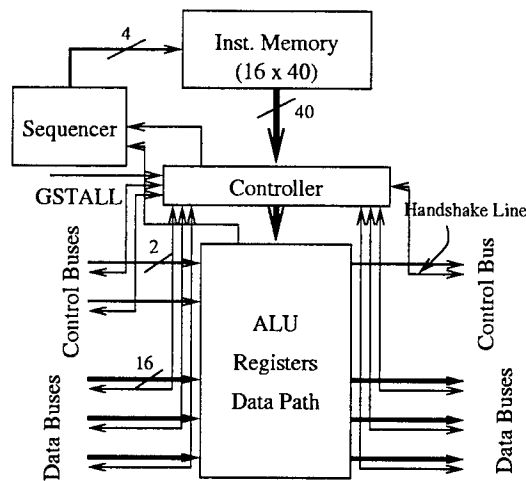


Figure 2: The VGI processor

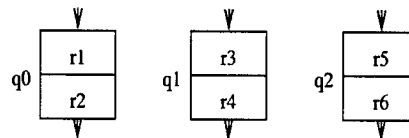


Figure 3: The registers/queues in the VGI processor

come from the read and write pointers. The write data is from the communication networks. Read requests are communicated using read handshake lines, and write requests are communicated using write handshake lines. The delay line mode is similar to the FIFO mode except that read is executed automatically once the delay line is full.

I/O Processor

The I/O processor handles communication with off-chip devices such as external memory and other VGI chips. To accommodate commercial SRAM chips and other VGI chips, the I/O processor uses two cycles for data and control communication. Figure 5 shows the block diagram of the I/O processor.

The data buses in the I/O processor can be connected to the level-1 network and the I/O pad. The I/O processor, configured through the scan chain, either takes data from an on-chip VGI processor and outputs to a pad, or the other way around. If the I/O processor is supposed to take data from the pad, it outputs the data over the output control bus and one of the last two output data buses. If the I/O processor is set up to output data to the pad, then it takes the data from the input control and data buses.

Communication Networks

The VGI processor has three data output ports connected to the three output buses and three data input ports connected to the three queues. It also has two input ports and one output port for control data used for communicating conditions such as carries for pipelining some complicated operations such as multiplication.

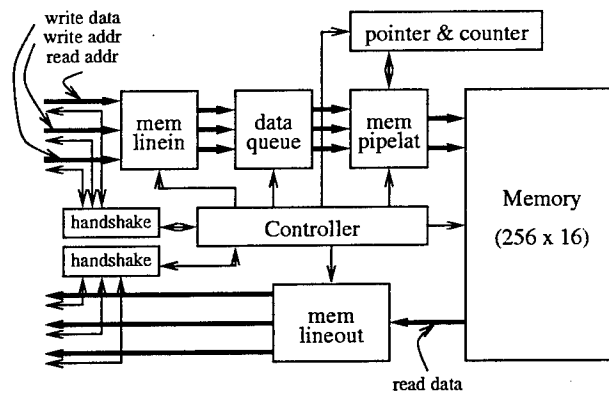


Figure 4: The memory processor

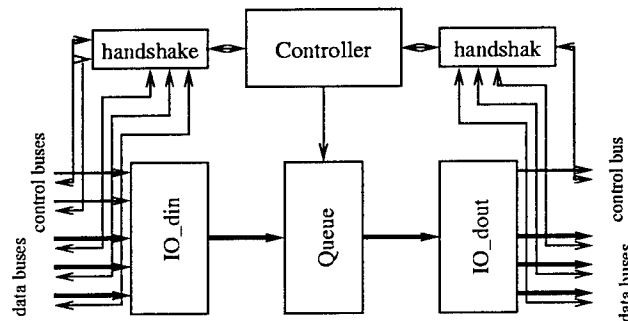


Figure 5: The I/O processor

The memory processor also has three data input ports and three data output ports, but does not have control input/output ports. The I/O processor has three data input ports and three data output ports. It also has two control input ports and one control output port. All the ports are associated with corresponding handshake lines for synchronization. The processors can be abstracted as shown in Figure 6 for the purpose of connecting processors to implement an algorithm.

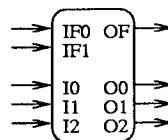


Figure 6: A processor with input/output ports

Each of the VGI and I/O processors in a cluster has 17 programmable switches for connecting data and control ports to the level-1 network. The memory processors have fewer switches since they do not have control ports. There are 6 data buses and 4 control buses in the level-1 network. There are also 3 internal bypass buses connecting neighboring processors. The switches are numbered in Figure 7.

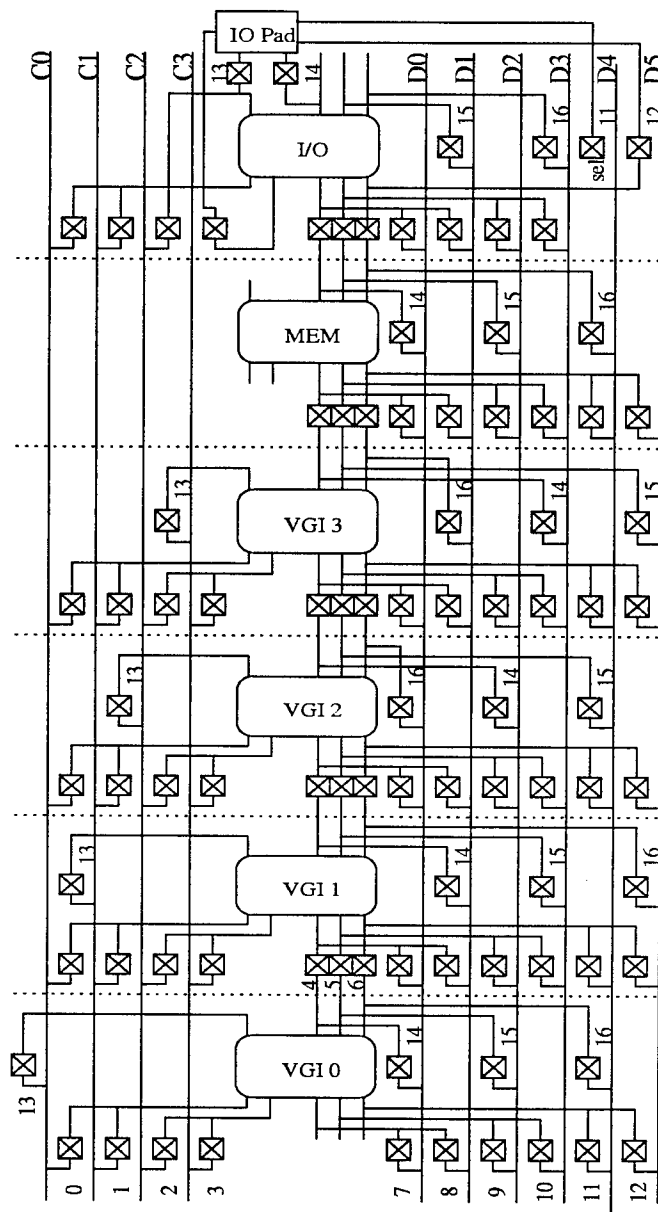


Figure 7: The level-1 switches in each cluster

For processors in one cluster to communicate with processors in another cluster, the level-2 communication network is provided. The level-2 network consists of 16 16-bit data buses and 8 2-bit control buses with their associated handshake lines. The level-1 and level-2 buses form a crossbar with programmable switches at the crossings of the level-1 and level-2 buses. The switches allow connections of the level-1 and level-2 buses. The 50 switches for each cluster are set through the scan chain. Figure 8 shows the switches for the connection of an upper cluster and a lower cluster to the level-2 buses.

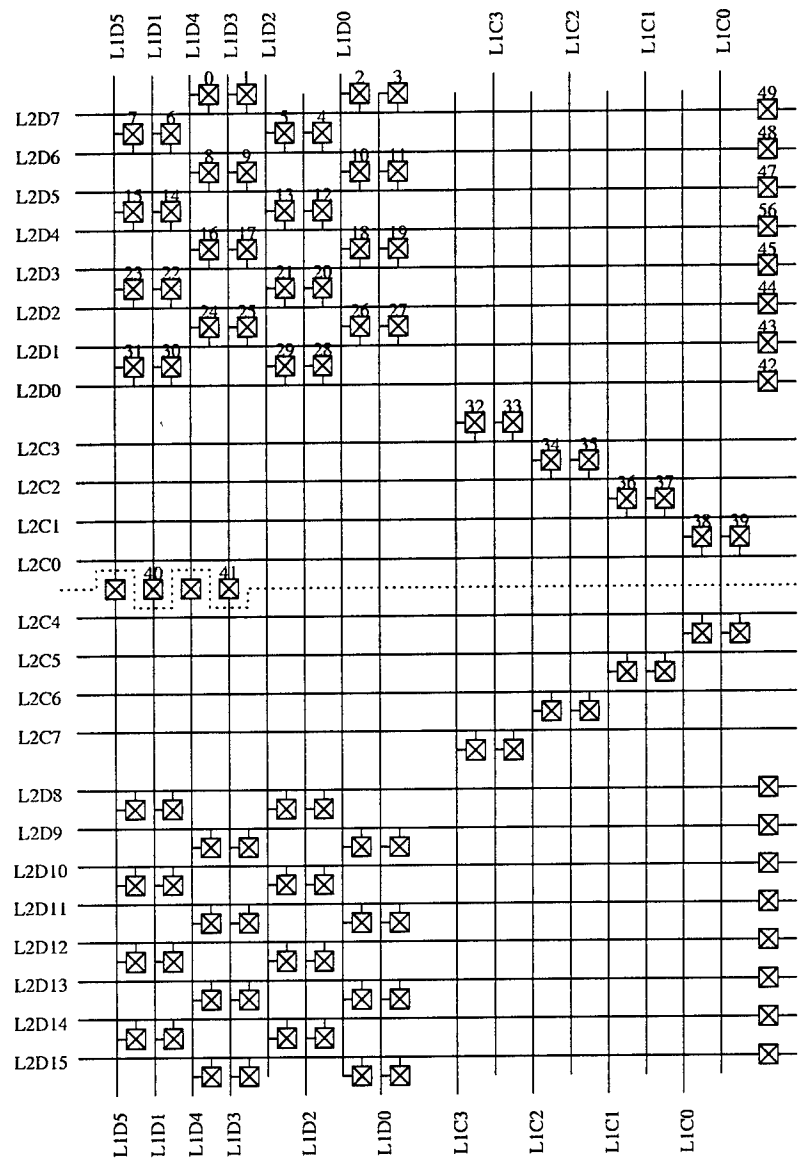


Figure 8: The level-2 switches for each upper and lower cluster pair

Machine Level Programming

The lowest level programming is at the machine level. At this level, the programmer is aware of all the physical processors and the switches that can be set to form paths for connecting the processors.

Programming Model

The machine level programming model abstracts the VGI computer for the purpose of programming the computer.

Each VGI processor can implement some function that transforms every set of data items from its predecessors into a set of data items for its successors. The function is fixed for every set of inputs and is usually repeated in a loop. The function should be simple enough to be implemented in at most 16 VGI instructions. Unlike the VGI processors, the memory and I/O processors work in some mode among a few predefined modes. The specific mode is chosen during configuration and fixed during execution.

A processor can send data items to other processors through its three output ports and can receive data items from other processors through its three input ports. An output port of a processor can be connected to an input port of another processor by setting switches to form a path between the output port and the input port. The connection, after configured, is fixed during execution. Two communicating processors have to wait until the data is transferred from the sender to the receiver. The transfer only takes one cycle when both the sender and the receiver are ready. There are also two input ports and one output port in the VGI and I/O processors for communicating control data such as carries.

To implement a solution to a stream data processing problem, the programmer has to break up the solution into a sequence of tasks, each transforming some input streams into some output streams. Each task is assigned to a physical processors. If an output stream of task A assigned to processor A is an input stream of task B assigned to processor B, there should be a connection between the output port of processor A for the output stream of task A and the input port of processor B for the input stream of task B. The tasks are then implemented using the VGI instructions. The switch settings for processor connections and the assignment of tasks to processors are stored in a file as the VGI program for the stream data processing problem.

Three tedious processes are involved in the above programming paradigm after the programmer divides the algorithm into tasks and figures out the relationships among tasks in terms of inputs and outputs. The first process is to assign the tasks to the processors. The second process is to determine which data stream goes through which port. The third process is to find switches that form a path to connect an output port of a processor to an input port of another processor. Since assignment of processors and input/output ports affects the availability of paths between communicating processors, the above processes may have to be repeated with different processor and input/output port assignments to find a valid assignment. These three processes affect each other, making the machine level programming very difficult. There is a need for tools to keep track of resources and assist the routing process. Figure 9 shows the tools and information flow in machine level programming.

Tools

Since processors in each cluster have more connections among themselves, a good heuristic for processor assignment is to assign tasks with more connections among themselves to the same cluster. Finding paths can be assisted with a routing tool. The routing tool searches the currently available switches to form a path between an output port of a processor and an input port of another processor. It shows the possible paths for the programmer to choose. Since the paths with switch numbers are hard to imagine, it is better to display all the switches in the level-1 and level-2 networks and mark the paths already taken for the programmer.

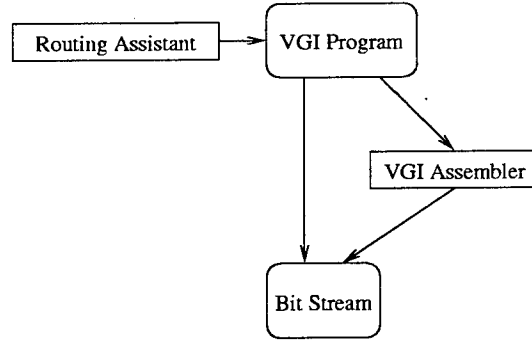


Figure 9: Machine level programming tools and information flow

This will also allow the programmer to interactively select switches to form paths.

After the processors, input/output ports, paths are determined, the programmer then writes the assembly code for each processor. The code for each processor can be assembled using the VGI assembler.

The switch settings for the networks and the assembled code for each processor are put together to generate a scan chain bit stream for configuration of the VGI computer.

A VGI simulator has been under development for VGI program execution and verification. The functional simulator is designed to give fast feedback to the programmer in the trial-and-error machine level programming.

A Programming Example

To illustrate further the machine level programming paradigm, we describe in details the development of a 1D convolution on the VGI computer. The 1D convolver computes the output signal stream, $b_i, i = 0, 1, 2, \dots$, from the input signal stream, $a_i, i = 0, 1, 2, \dots$, by

$$b_i = k_{-1}a_{i-1} + k_0a_i + k_{+1}a_{i+1}.$$

We can divide the computation of b_i into four tasks. Task 1 computes $k_{-1}a_{i-1}$; task 2 computes k_0a_i ; task 3 computes $k_{+1}a_{i+1}$; and task 4 accumulates the results from tasks 1, 2, and 3. The input stream to task 1 is $a_0, a_1, a_2, \dots$ and the output stream is $0, k_{-1}a_0, k_{-1}a_1, k_{-1}a_2, \dots$. The input stream to task 2 is $a_0, a_1, a_2, \dots$ and the output stream is $k_0a_0, k_0a_1, k_0a_2, \dots$. The input stream to task 3 is $a_0, a_1, a_2, \dots$ and the output stream is $k_{+1}a_1, k_{+1}a_2, \dots$. Let task 3 takes the input stream from the signal input. Task 3 can forward a copy of its input to task 2, and task 2 can forward a copy of its input to task 1. Task 4 accumulates the output streams of tasks 1, 2, and 3, and outputs $k_0a_0 + k_{+1}a_1, k_{-1}a_0 + k_0a_1 + k_{+1}a_2, k_{-1}a_1 + k_0a_2 + k_{+1}a_3, \dots$. To synchronize the tasks, task 1 needs to output one 0 before it reads any input, and task 3 must not generate any output on the first input. The tasks and their relationship are represented with the signal flow graph in Figure 10.

We can assign the tasks to one cluster, say cluster 3. We then assign task 3 to VGI processor 0 (P0.C03), task 2 to VGI processor 1 (P1.C03), task 1 to VGI processor 2 (P2.C03), and task 4 to VGI processor 3 (P3.C03). We can use the I/O processor of cluster (I0.C03) to handle the output. We need the I/O processor

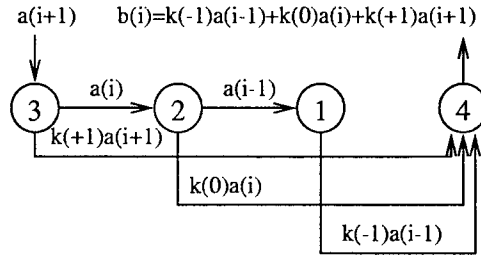


Figure 10: A signal flow graph for the 1D convolution

of another cluster, say cluster 4, to handle the input. We also need to assign input/output ports for each connection. The processor and input/output port assignment is shown in Figure 11.

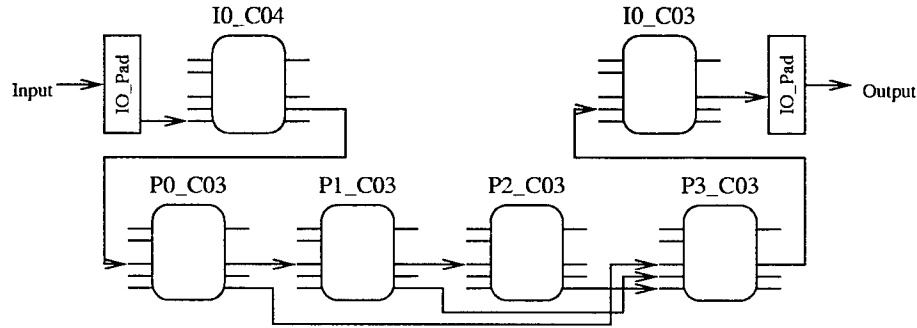


Figure 11: Assignment of processors and input/output ports for the 1D convolution

Now we need to find a path connecting the output port of the processor to the input port of the other processor for each communicating processor pair. This process can be done with help of the routing tool. For the connection between O1 of P0_C03 and I0 of P1_C03, the routing tool finds two possible paths (using at most 4 switches): L1S4P1_C03 (level-1 switch 4 of VGI processor 1 in cluster 3) and L1S14P0_C3 - L1S7P1_C03. We choose the first path (which uses the bypass bus between P0_C03 and P1_C03). Note that switch L1S4P1_C03 and its associated bus cannot be used for other connections. We can similarly choose paths for other processor pairs. The paths are summarized in Table 1.

The easiest process is to write the VGI code for each processor. The three input ports I0, I1, and I2 corresponds to the three queues q1, q2, and q3, respectively. The three output ports O0, O1, and O2 corresponds to the three output buses c0, c1, and c2, respectively. When the queues are not for inputs, they can be used as registers. Here are the VGI code for each task and the VGI main program for the 1D convolution.

```

/*****
Program for task 1
Name: k_m1.s
Input: a(0) a(1) a(2) ...
Output: 0 k(-1)a(0) k(-1)a(1) k(-1)a(2) ...
*****/

```

Table 1: The switch settings for the 1D convolution

Connection		Switches
Output	Input	
O0 (P0_C03)	I0 (P1_C03)	L1S4P1_C03
O2 (P0_C03)	I0 (P3_C03)	L1S16P0_C03 - L2S3C03 - L2S0C03 - L1S7P3_C03
O0 (P1_C03)	I0 (P2_C03)	L1S4P2_C03
O2 (P1_C03)	I1 (P3_C03)	L1S16P1_C03 - L2S7C03 - L2S4C04 - L1S9P3_C03
O2 (P2_C03)	I2 (P3_C03)	L1S6P3_C03
O0 (P3_C03)	I1 (I0_C03)	L1S14P3_C03 - L1S10I0_C03
O0 (I0_C03)	IO_Pad (I0_C03)	L1S14I0_C03
O1 (I0_C04)	I0 (P0_C03)	L1S15I0_C04 - L2S15C04 - L2S46C03 - L2S14C03 - L1S8P0_C03
IO_Pad (I0_C04)	I2 (I0_C04)	L1S12I0_C04

```

.init q0 = q
.init q1 = r
.init q2 = r

.init r3 = 16384 /* k(-1): 0.25 (16384 X 2^-16) */
.init pc = 0

0: r0 = xor(ra, ra),
   c2 = ra, /* output 0 to 02 */
   next = 1;

1: r4 = or(q0, q0), /* input from I0 */
   next = 2;
2: r0 = xor(ra, ra), /* clear accumulator */
   next = 3;
3: r0 = mstartu(ra, r3, r4), /* start multiplication */
   next = 4;
4: r0 = mstepu(ra, r3, rc),
   next = 5;
5: r0 = mstepu(ra, r3, rc),
   next = 6;
6: r0 = mstepu(ra, r3, rc),
   next = 7;
7: r0 = mstepu(ra, r3, rc),
   next = 8;
8: r0 = mstepu(ra, r3, rc),
   next = 9;
9: r0 = mstepu(ra, r3, rc),
   next = 10;
10: r0 = mstepu(ra, r3, rc),
    c2 = ra, /* output product to 02 */
    next = 1;

/*****
Program for task 2
Name: k_0.s

```

```

    Input: a(0) a(1) a(2) ...
    Output: k(0)a(0) k(0)a(1) k(0)a(2) ...
    *****/

.init q0 = q
.init q1 = r
.init q2 = r

.init r3 = 32768 /* k(0): 0.5 (32768 X 2-16) */
.init pc = 0

0: r4 = or(q0, q0), /* input from I0 */
   c0 = ra, /* output to 00 */
   next = 1;
1: r0 = xor(ra, ra), /* clear accumulator */
   next = 2;
2: r0 = mstartu(ra, r3, r4), /* start multiplication */
   next = 3;
3: r0 = mstepu(ra, r3, rc),
   next = 4;
4: r0 = mstepu(ra, r3, rc),
   next = 5;
5: r0 = mstepu(ra, r3, rc),
   next = 6;
6: r0 = mstepu(ra, r3, rc),
   next = 7;
7: r0 = mstepu(ra, r3, rc),
   next = 8;
8: r0 = mstepu(ra, r3, rc),
   next = 9;
9: r0 = mstepu(ra, r3, rc),
   c2 = ra, /* output product to 02 */
   next = 0;

/*****
Program for task 3
Name: k_p1.s
Input: a(0) a(1) a(2) ...
Output: k(+1)a(1) k(+1)a(2) ...
*****/

.init q0 = q
.init q1 = r
.init q2 = r

.init r3 = 16384 /* k(+1): 0.25 (16384 X 2-16) */
.init pc = 0

0: r4 = or(q0, q0), /* input from I0 */
   c0 = ra, /* output to 00 */
   next = 1;

1: r4 = or(q0, q0), /* input from I0 */
   c0 = ra, /* output to 00 */

```

```

    next = 2;
2:  r0 = xor(ra, ra), /* clear accumulator */
    next = 3;
3:  r0 = mstartu(ra, r3, r4), /* start multiplication */
    next = 4;
4:  r0 = mstepu(ra, r3, rc),
    next = 5;
5:  r0 = mstepu(ra, r3, rc),
    next = 6;
6:  r0 = mstepu(ra, r3, rc),
    next = 7;
7:  r0 = mstepu(ra, r3, rc),
    next = 8;
8:  r0 = mstepu(ra, r3, rc),
    next = 9;
9:  r0 = mstepu(ra, r3, rc),
    next = 10;
10: r0 = mstepu(ra, r3, rc),
    c2 = ra, /* output to 01 */
    next = 1;

/*****
Program for task 4
Name: sum.s
Input: 0 k(-1)a(0) k(-1)a(1) ...
       k(0)a(0) k(0)a(1) k(0)a(2) ...
       k(+1)a(1) k(+1)a(2) ...
Output: k(0)a(0)+k(+1)a(1) k(-1)a(0)+k(0)a(1)+k(+1)a(2) ...
*****/

.init q0 = q
.init q1 = q
.init q2 = q

.init pc = 0

0:  r0 = add(q0, q1), /* add values from I0 and I1 */
    next = 1;
1:  r0 = add(q2, ra), /* add value from I2 */
    c0 = ra,          /* output sum to 00 */
    next = 0;

/*****
This is the main program that contains the configuration
information for the 1D convolution.
Name: conv1d.set
*****/

# ----- Nanoprocessor allocation: -----

P0_C03  k_p1
P1_C03  k_0
P2_C03  k_m1

```



```

P3_C03  sum
IO_C03
IO_C04

# ----- Switch settings: -----
#   CN   PN   MN   IN   SN

# IO_Pad --> I2-IO_C04
S   4   -1   -1   0   12

# 01-IO_C04 --> IO-P0_C03
S   4   -1   -1   0   15
S   4   -1   -1   -1  14
S   3   -1   -1   -1  46
S   3   -1   -1   -1  14
S   3   0   -1   -1   8

# 00-P0_C03 --> IO-P1_C03
S   3   1   -1   -1   4

# 02-P0_C03 --> IO-P3_C03
S   3   0   -1   -1  16
S   3   -1   -1   -1   0
S   3   -1   -1   -1   3
S   3   3   -1   -1   7

# 00-P1_C03 --> IO-P2_C03
S   3   1   -1   -1   4

# 02-P1_C03 --> I1-P3_C03
S   3   1   -1   -1  16
S   3   -1   -1   -1   4
S   3   -1   -1   -1   7
S   3   3   -1   -1   9

# 02-P2_C03 --> I2-P3_C03
S   3   -1   -1   -1   6

# 00-P3_C03 --> I1-IO_C03
S   3   3   -1   -1  14
S   3   -1   -1   0  10

# 00-IO_C03 --> IO-Pad
S   3   -1   -1   0  14

/*****/

```

Programming with Signal Flow Graphs

Stream data processing algorithms can often be described naturally by block diagrams in which computational nodes are interconnected by links that represent data streams [3, 13]. The development of stream data algorithms on the VGI computer can be better supported using signal flow graphs.

Programming Model

The signal flow graph (SFG) programming model allows the programmer to represent any stream data parallel algorithm as a directed graph with nodes denoting tasks and edges denoting data flow. The SFG model supports hierarchical graph structure in the sense that a node of a graph may be a subgraph representing a sub-algorithm. In the SFG programming paradigm, the programmer divides the algorithm into small tasks. Each task takes inputs from some other tasks and/or the external input, and generates outputs for other tasks and/or the external output. If one task outputs to another task, there should be an edge from the node representing the first task to the node representing the second task. Each task can be independently coded with a function that transforms the input streams into the desired output streams.

In the 1D convolution example, the programmer only needs to develop an SFG shown in Figure 10. We can abstract the SFG as a node as shown in Figure 12 for constructing an SFG for a 2D convolution.

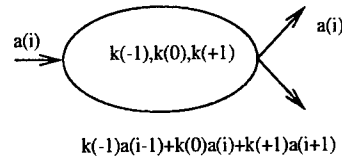


Figure 12: An abstract node representing the 1D convolution

We then implement the 2D convolution

$$\begin{aligned}
 b_{i,j} = & k_{-1,-1}a_{i-1,j-1} + k_{-1,0}a_{i-1,j} + k_{-1,+1}a_{i-1,j+1} \\
 & + k_{0,-1}a_{i,j-1} + k_{0,0}a_{i,j} + k_{0,+1}a_{i,j+1} \\
 & + k_{+1,-1}a_{i+1,j-1} + k_{+1,0}a_{i+1,j} + k_{+1,+1}a_{i+1,j+1}
 \end{aligned}$$

with the SFG as shown in Figure 13.

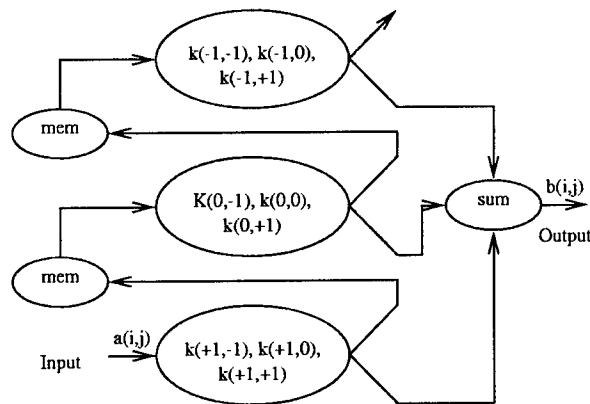


Figure 13: An SFG for the 2D convolution

SFG GUI and SFG-to-VGI Mapping

To support the SFG programming paradigm, we need a graphical user interface (GUI) for the programmer to edit signal flow graphs. In addition to the graph editing capabilities, The GUI should provide a way for the programmer to specify the function of each node as an attribute of the node.

The major problem with SFG programming is that we need an effective algorithm to assign each task to a physical processor, to assign each end of each edge to a port of the processor, and to find switches to form a path for each edge. To solve this difficult problem, we need to model the VGI computer as a graph and devise a mapping algorithm that map SFGs to the VGI graph. More research is needed to accomplish this.

Programming in Functional Language SISAL

The SISAL language is a functional programming language intended for high performance scientific computing on highly parallel computers [9, 12]. Streams of integers and streams of integer arrays are natural representations for the data processed in signal, image, and video processing applications [5]. The SISAL language naturally supports development of stream data algorithms.

Programming Model

The SISAL programming model allows the programmer to specify naturally a stream processing algorithm as a recursive function that defines a result stream as the result of concatenating a new element with the stream produced by a recursive application of the function. In this programming paradigm, the programmer develops stream processing algorithms without dealing with individual small tasks as in the SFG and machine level programming models.

A stream is a data type that may be created for any type. For example, `+type T2 = stream [T1]+` defines stream type T2 for type T1. The values of T2 are streams (i.e., sequences of indefinite length) of elements of type T1. Three basic operations, `stream_first` (for first element of a stream), `stream_rest` (for the stream of the remaining elements), and `||` (for concatenating two streams), are provided for stream manipulation. A stream of a single element `x` of type `T` can be constructed by `stream T [x]`. Elements of a stream can also be accessed using subscripts. As an example, our 1D convolution can be expressed in SISAL as follows.

```
type Signal = stream [integer];
function conv1d (a: Signal returns Signal)
  stream integer [ (km1*a[0]+k0*a[1]+kp1*a[2] )
    || conv1d (stream_rest (a))
end function
```

SISAL Compiler

To support SISAL programming, we need an SISAL compiler to translate SISAL programs into efficient VGI programs. One important issue in compiling SISAL programs is the recognition of recursive functions that can be transformed into non-recursive signal flow graphs. More research is needed to construct an efficient SISAL compiler for the VGI computer.

Conclusions and Future Work

The VGI parallel computer has been proposed by Srini and Rabaey for processing stream data present in video, graphics, and image processing applications. It is a hybrid architecture with features of von Neumann and dataflow architectures. It implements the static dataflow model with multithreading concept. It provides a simple synchronization mechanism, for any two communicating processors. So far, the research on the VGI computer has been focused on its hardware. However, the success of the VGI computer lies in its programmability. This report has discussed some issues related to programming models and tools on top of the VGI hardware. More research efforts are needed to have a more user-friendly environment for solving stream data processing problems on the VGI computer.

First of all, research is needed to provide and improve several tools for assisting machine level programming. A functional VGI simulator has been under development which incorporates the VGI assembler and the routing assistant. The routing assistant should be an interactive tool that displays all the switches with already taken paths clearly marked on the screen. In addition to the capability to suggest paths between an output port of a processor and an input port of another processor, it should allow the programmer to choose switches interactively to form a path and to modify previously found paths. Furthermore, a timing tool is needed for estimating the execution time of VGI programs. Since the VGI computer executes multiple threads communicating asynchronously, it is hard for the programmer to trace the execution to estimate the execution time. The timing tool should compute the latency (the time between the first input and the first output) and the per output time (the time between an output and the next output).

Secondly, research is needed on modeling the VGI computer as a graph and mapping SFGs to the VGI graph to support SFG programming. The processors can be modeled as nodes. The bus segments can also be modeled as nodes. The switches can be modeled as edges. Finding paths for SFG edges is then converted to a problem of finding disjoint paths in the VGI graph.

Thirdly, research is needed on compilation techniques for translating SISAL programs into SFGs. There are algorithms to recognize simple recursive SISAL functions and translate them into SFGs. More research should be performed to handle more general cases.

Acknowledgements

This work was supported by the 1998 Air Force Office of Scientific Research Summer Research Program. The author wishes to thank Dr. Patrick Coffield, focal point at the Air Force Research Laboratory / Munitions Directorate at Eglin AFB. The author also wishes to thank Dr. Dennis Goldstein at Eglin AFB and the staff of Research & Development Laboratories (RDL) for their assistance.

References

- [1] B. S. Ang, Arvind, and D. Chiou. StarT the next generation: Integrating global caches and dataflow architecture. In G. R. Gao, L. Bic, and J.-L. Gaudiot, editors, *Advanced Topics in Dataflow Computing and Multithreading*. IEEE Computer Society Press, Los Alamitos, CA, 1995.

- [2] Arvind and D. E. Culler. Dataflow architectures. *Annual Review in Computer Science*, 1:225–253, 1986.
- [3] S. S. Battacharyya, P. K. Murthy, and E. A. Lee. *Software Synthesis from Dataflow Graphs*. Kluwer Academic Publishers, Boston, MA, 1996.
- [4] D. E. Culler, et al. TAM – a compiler controlled threaded abstract machine. *Journal of Parallel and Distributed Computing*, 18(3):347–370, 1993.
- [5] J. B. Dennis. Stream data types for signal processing. In G. R. Gao, L. Bic, and J.-L. Gaudiot, editors, *Advanced Topics in Dataflow Computing and Multithreading*, pages 87–101. IEEE Computer Society Press, Los Alamitos, CA, 1995.
- [6] G. R. Gao, L. Bic, and J.-L. Gaudiot, editors. *Advanced Topics in Dataflow Computing and Multithreading*. IEEE Computer Society Press, Los Alamitos, CA, 1995.
- [7] G. R. Gao, J.-L. Gaudiot, and L. Bic. Special issue on dataflow and multithreaded architectures: Guest editors' introduction. *Journal of Parallel and Distributed Computing*, 18(3):271–172, July 1993.
- [8] J.-L. Gaudiot and L. Bic, editors. *Advanced Topics in Dataflow Computing*. Prentice Hall, Englewood Cliffs, NJ, 1991.
- [9] J. McGraw, et al. SISAL: Streams and iteration in a single assignment language: Reference manual version 1.2, 1985. Technical Report M-146, Rev. 1.
- [10] B. Lee and A. R. Huson. Dataflow architectures and multithreading. *Computer*, pages 27–39, August 1994.
- [11] O. C. Maquelin. Load balancing and resource management in the ADAM machine. In G. R. Gao, L. Bic, and J.-L. Gaudiot, editors, *Advanced Topics in Dataflow Computing and Multithreading*. IEEE Computer Society Press, Los Alamitos, CA, 1995.
- [12] R.R. Oldehoeft, et al. SISAL reference manual: Lanuage version 2.0, 1992. Technical Report.
- [13] V. P. Srimi. Low-power parallel DSP computer, 1998. Final report to the Air Force Research Laboratory / Munitions Directorate, Eglin AFB, FL.
- [14] V. P. Srimi, N. Chow, R. A. Sutton, and J. M. Rabaey. MultiPADDI-2 board for image processing. In *Proceedings of the SPIE Conference on Parallel and Distributed Methods for Image Processing*, San Diego, CA, 1997.
- [15] V. P. Srimi, Spartan-Team, and J. M. Rabaey. An architecture for Web-based image processing. In *Proceedings of the SPIE Conference on Parallel and Distributed Methods for Image Processing*, San Diego, CA, 1997.
- [16] V. P. Srimi, J. Thendean, S.-Z. Ueng, and J. M. Rabaey. A parallel DSP with memory and I/O processors. In *Proceedings of the SPIE Conference on Parallel and Distributed Methods for Image Processing II*, San Diego, CA, 1998.

Joule Heating Simulation of Poly-Silicon Thermal Micro-Actuators

Donald J. Silversmith

Professor

Department of Electrical and Computer Engineering

**Wayne State University
5050 Anthony Wayne Drive
Detroit, Michigan 48202**

**Final Report for:
Summer Faculty Research Program
Air Force Research Laboratory
Hanscom AFB, MA**

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

And

**Air Force Research Laboratory
Hanscom AFB, MA**

September 1998

Joule Heating Simulation of Poly-Silicon Thermal Micro-Actuators

**Donald J. Silversmith
Professor**

Department of Electrical and Computer Engineering

Wayne State University
5050 Anthony Wayne Drive
Detroit, Michigan 48202

Professor
Department of Electrical and Computer Engineering
Wayne State University
5050 Anthony Wayne Drive
Detroit, Michigan 48202

ABSTRACT

Previous studies of surface micro-machined polycrystalline silicon MEMS thermal micro-actuators have established that these simple and compact devices can provide deflections on the order of 10 micrometers at CMOS compatible drive voltages. These thermo-mechanical devices operate by differential thermal expansion caused by ohmic heating in higher resistance regions of the double beam device. Because of thermal conductivity and temperature dependent resistivity in polycrystalline silicon, the temperature profile along the "pusher" section of the beam is not uniform, and motion simulation of the thermal actuator can be complex. This work represents an initial attempt at correlation of experimental near-IR observations of thermal actuator operation to a new commercial Joule-heating MEMS simulation package that integrates pertinent mechanisms into a no parameter fit model.

Introduction

Comtois et al ¹ have presented poly-silicon in-plane horizontal double beam thermal actuators that demonstrate large deflections ($>10\mu\text{m}$) and operate at CMOS compatible drive voltages and at current levels below 5mA. This actuator concept has been applied to a wide variety of applications, including scanning and rotating micro-mirrors, Fresnel lenses, movable gratings, linear micro-motors, and self-assembly systems.^{2, 3, 4} A representative diagram of this beam design, used both in the experimental and simulations studies for report, is indicated in Fig. 1. Differential ohmic heating effects in a double beam structure drive these thermal actuators which consist of two parallel beams, coupled at the free end, with one beam significantly more electrically and thermally conductive than the second beam. As the more resistive beam (with a smaller cross section) heats up relative to the attached less resistive beam (with a larger cross section), the effect of differential thermal expansion causes the narrower beam to push on the wider beam at the free end. The experimentally optimized design in Fig. 1 has been used in number of other studies, with limited theoretical evaluation.^{5, 6}

Functional modeling of these devices is complex because localized ohmic heating depends on the temperature distribution in the expanding beam elements. This temperature distribution depends on a number of criteria, primarily the absolute thermal conductivity of the heated elements, and, to a lesser extent, the temperature dependence the resistance of the doped poly-crystalline silicon elements, which heat up. Simulation studies in this work utilize a new MemCad 4.0 MEMS simulation module available from Microcosm, Inc.⁷ This "MemECad" Joule heating module, in conjunction with other parts of this commercial finite element design and simulation package, computes the thermal and electrical potential field distributions resulting from an applied voltage or current flow through a resistive material. The localized Joule heating effects are coupled to the MemCad mechanical simulation package to determine overall displacement and stresses related to thermal expansion. This iterative approach accounts for the voltage potential distribution throughout the volume of the double beam structure, as well as the impact of localized differential thermal conductivity, which is related to the different geometrical cross sections in different parts of the structure. The accumulated thermal deformations

of the double beam structure are determined from the thermal expansion effects in different finite elements of the structure.

Devices

The devices used for the experimental observations were originally fabricated at the MCNC/MUMPS processing facility as part of an array of test structures for related work.⁸ Test chips as received from MCNC were released, mounted and bonded in an open package. A SEM micrograph of a typical device is indicated in Fig. 2. The length of the 2.5 μ m wide, 2.0 μ m thick, narrow "pusher" beam is 235 μ m. The nominal resistivity of the polycrystalline silicon is $10 \pm 5 \text{ } \Omega/\square$, resulting in a room temperature device resistivity of $1300 \pm 650 \Omega$. Because of this resistivity variability, device current is a more pertinent independent variable in terms of calibrating thermal actuator functions.

These devices were connected individually to a stable variable voltage DC power supply, and the current was monitored with a digital milliamp meter. Beam deflection was observed through a microscope coupled to a PULNIX TM 840 TV camera, which has significant relative sensitivity below $\lambda=1.0\mu$ m wavelength. Pictures were captured in a bit map format. A typical captured image of an operational device is given in Fig. 3, which was taken at a temperature just before the luminosity of near-IR radiation can be observed. Deflections can be measured using the vernier structure at the coupled free end of the beam. Images were taken both in bright optical field illumination, and in dim optical illumination, in order to highlight the IR luminosity of the hot part of the narrow beam. The IR luminosity as seen with the PULNIX camera is not seen for low applied voltages, and increases in intensity with voltage. For example, no IR luminosity is seen for the device of Fig. 3 below 8 volts applied to the two-anchored beam-ends. The sharp focus position of the microscope for bright optical illumination and sharply defined IR radiance are not coincident, most likely, because of the wave length dependence of the optics. Fig. 4 indicates typical behavior of an actuator indicated in Fig. 2 for both current and displacement. Precision displacement measurements are limited by the use of vernier indicators and by "stiction" of the dimple spacers in contact with the substrate. These

“dimples” act as spacers to prevent contact of the double beam structure with the substrate surface, nominally $2\mu\text{m}$ below the beam.

Experimental Results

The current and displacement dependence with applied voltage for a typical actuator is indicated in Fig. 4. The displacement measurements are based on observation of the vernier at the free end of the beam, and are also subject to “stiction” of the dimple spacers on the substrate. A displacement of over $12\mu\text{m}$ is achieved at less than 5mA , in agreement with prior observations.¹ The current dependence with voltage is sub-ohmic, most likely because the resistance of the poly-silicon declines with temperature. No IR luminosity is observed for this device below 8 volts applied to the beam ends, but above this value, IR luminosity is quite apparent, even with moderate white light optical illumination, as seen in Fig. 5.

The temperature at the center of the IR radiance is estimated to be in the range of $850 - 1000\text{ }^\circ\text{K}$, while the edge of the IR luminosity is estimated as $800\text{ }^\circ\text{K}$ from blackbody spectral exitance considerations.⁹ Clearly, the ohmic temperature distribution along the narrow beam is far from uniform, and the highest temperature region—corresponding to the elements of largest thermal expansion—extend no more than 30% of the length of this beam.

The wide, cold side of the coupled beam structure has a flexure section $40\mu\text{m}$ long with a $2.0 \times 2.0\mu\text{m}^2$ cross section that serves as a hinge. Without this flexure element, little thermally activated deformation would be seen at the free end of the coupled beam structure. Because the area cross section of this flexure region is 20 pct. smaller than that of the “hot” beam, there could be a concern that Joule heating in the flexure will mitigate the deformation at the free end of the “hot” beam. The IR observation indicates that both the large thermal conductivity and the small voltage drop in the short flexure is sufficient to prevent the appearance of IR luminosity in this part of the structure. This observation has also been confirmed by the simulation studies.

The range of IR luminosity never extends more than 50 μm on the 235 μm long pusher beam, and never appears on the 40 μm flexure. Because of thermal heat flow effects, one might presume that the IR luminescence should be localized about the center of the pusher beam, but the actual position varies from device to device, even on the same chip. One possibility is that local variations of the effective doping in the pusher beam could shift the "hot spot." Another source for this phenomena could be local etch fabrication variations in the cross section of the pusher beam.

MemCad Simulations

MemCad is an integrated set of MEMS simulation tools enabling the design, specification, modeling, and engineering of accurate and performance predictive devices and systems with a wide variety of functional properties and applications. The complete suite of MemCad modules, used in this work, resides on a HP workstation at the AFRL Information Technology Division at Griffith AFB, Rome, NY, and is accessed remotely at Hanscom AFB. The recent availability of an integrated Joule heating module provided an opportunity to compare experimental observations of local heating and thermal conductivity for thermal actuators to the new simulation code, without the need to assume or infer a temperature profile on the narrow "pusher" beam.

Recently, Allen et al.¹⁰ have modeled a very similar ohmic heating structure using the competitive IntelliCad simulation program,¹¹ getting very good correlation to prototype experimental devices. Because the code used by Allen et al. did not have a Joule heating module at the time of their work, a temperature profile of the narrow beam had to be devised. In the current work, using the MemCad MemETherm module, there are no fitable parameters, other than specification of poly-silicon properties, including the poly-silicon resistivity, and the voltages applied to fixed ends of the double beam structure. Because of the variability in Poly-Si resistivity, device current is, generally, a more pertinent independent variable than applied voltage.

Simulation Results

Simulated results using MemCad are indicated in Fig 6. Because MemCad does not incorporate a model for material plastic flow or melting, calculations in the range of temperatures above 1500 °K are not meaningful. As clearly evidenced in the graph, the simulated displacement and reaction force at the anchors increases in a nominally linear manner with temperature. These calculations also demonstrate that only the middle sections of the pusher beam get "hot," and that the free end of the double beam structure gets no hotter than 400 °K, although the center of the pusher beam can reach 1500°K with only 3.5 volts applied to the fixed ends of the beams, as indicated in Fig. 6. The simulated calculation also indicates that virtually all the voltage drop is across the pusher beam, and only 10 pct. across the flexure. Since, in addition, both ends of the flexure are constrained to close to room temperature, no IR "hot spot" can develop in this part of the structure.

Summary

Using Near-IR microscopy on MEMS Thermal Actuators, reasonable agreement between experiment and a new Joule-heating finite element simulation module has been established. This work demonstrates the need to develop optimized design criteria in this class of devices to mitigate the thermal conduction effects in the pusher beam, which degrades displacement performance. This work also indicates the need to incorporate the temperature variation of material resistance in the calculations to enhance accuracy of the overall model.

Acknowledgments

The author acknowledges the support and assistance of his supervisor, Captain J. Robert Reid. He also acknowledges the assistance of George Roberts, Stephen Spaziani, William Waters, and Pearl Yip at AFRL, Steven Bart and Mark DaSilva at Microcosm, Inc., and Alfredo Coralles at Florida International University.

Bibliography

- 1) "Thermal microactuators for surface-micromachining processes", J. H. Comtois, V. M. Bright, and M. W. Phipps, Proc. SPIE 2642, pp. 10-21, (1995); "Applications for Surface-Micromachined Polysilicon Thermal Actuators and Arrays," J. H. Comtois and V. M. Bright, Sensors and Actuators A58, 19-25, (1997).
- 2) "Automated Assembly of Flip-Up Micromirrors," J. R. Reid, V. M. Bright, J. H. Comtois, Technical Digest of the 1997 Int. Conf. On Solid-State Sensors and Actuators(Transducers '97), Chicago, IL., June 16-19, Vol.1 pp. 347-350; "Automated Assembly of Flip-Up Micromirrors," J. R. Reid, V. M. Bright, and J. T. Butler, Sensors and Actuators, A66, pp. 292-298, (1998).
- 3) "Automated Assembly of MEMS," V.M. Bright, J. T. Butler, W. D. Cowan, D. M. Burns, and J. Robert Reid, Int. J. of Adv. Manufac. Syst. 1, in press,(1998).
- 4) "Arrays of thermal micro-actuators coupled to micro-optical components," J. R. Reid, V. M. Bright and J. H. Comtois, Proc. SPIE 2865, pp. 74-82, (Aug. 1996).
- 5) "SPICE Modeling of Polysilicon Thermal Actuators," J. T. Butler, V. M. Bright and W.D. Cowan, Proc. SPIE 3224, pp284-292, (1997).
- 6) "Electrothermal responses of lineshape microstructures," Liwei Lin, Mu Chiao, Sensors and Actuators A55, pp. 35-41(1996).
- 7) Microcosm Technologies Inc., Development Center, 101 Rogers St., Suite 213, Cambridge, MA 02142
- 8) "Force measurements of polysilicon thermal micro-actuators," J. R. Reid, V. M. Bright, and J. H. Comtois, Proc. SPIE 2882, pp. 296-306, (Oct. 1996).
- 9) ORIEL Corporation, Stratford, CT. Light Source Catalog pp. 1-8, -9: Spectral Exitance Considerations.
- 10) "Design, Finite Element Analysis, and Experimental Performance Evaluation of a Thermally-Actuated Beam Used to Achieve Large in-Plane Mechanical Deflections," P. B. Allen, J. T. Howard, E. S. Kolesar, and J. M. Wilken, Proc. 1998 Solid State Sensor and Actuator Workshop, Hilton Head, N. C. MEMS Conf., June 8-11, 1998, Late News Poster Session, pg. 5-6 .
- 11) IntelliSence Corp., 16 Upton Dr., Wilmington, MA 01887

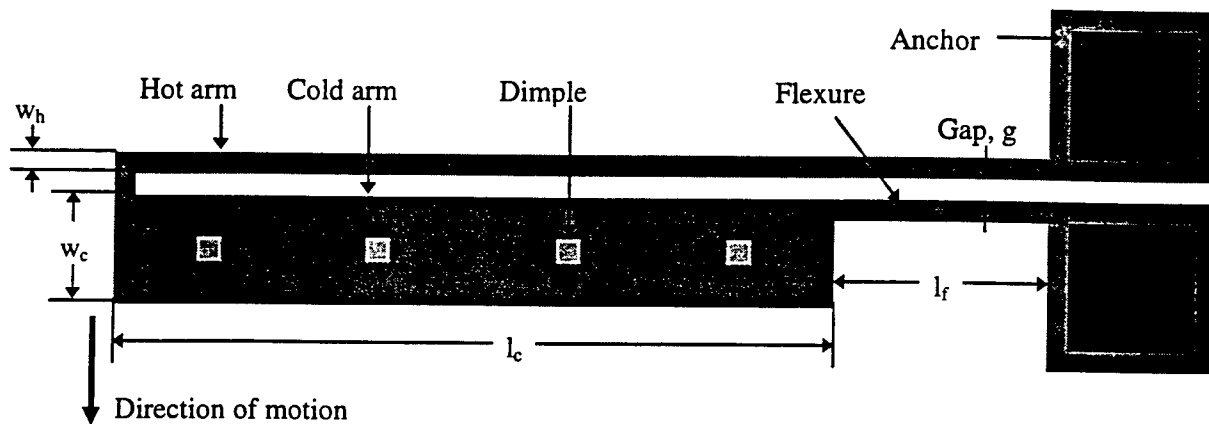


Figure 1. Basic layout of the lateral thermal actuator. Typical dimensions are: 'hot' arm $2.5\text{ }\mu\text{m}$ wide, $240\text{ }\mu\text{m}$ long; 'cold' arm $16\text{ }\mu\text{m}$ wide, $200\text{ }\mu\text{m}$ long; flexure $2.5\text{ }\mu\text{m}$ wide, $40\text{ }\mu\text{m}$ long, and gap $2\text{ }\mu\text{m}$ wide. The $4\text{ }\mu\text{m}$ square dimples prevent stiction. Deflections of $16\text{ }\mu\text{m}$ can be achieved at 3 V and 3.5 mA in the $2\text{ }\mu\text{m}$ thick first polysilicon layer of the MUMPS process.

Fig. 2 SEM Micrograph of a lateral thermal actuator. The fixed vernier grating has a period of $4\mu\text{m}$, while the vernier at the free end of the double beam has a period of $5\mu\text{m}$. The double beam structure is $2\mu\text{m}$ thick doped poly-Si, and is suspended $2\mu\text{m}$ above the substrate.

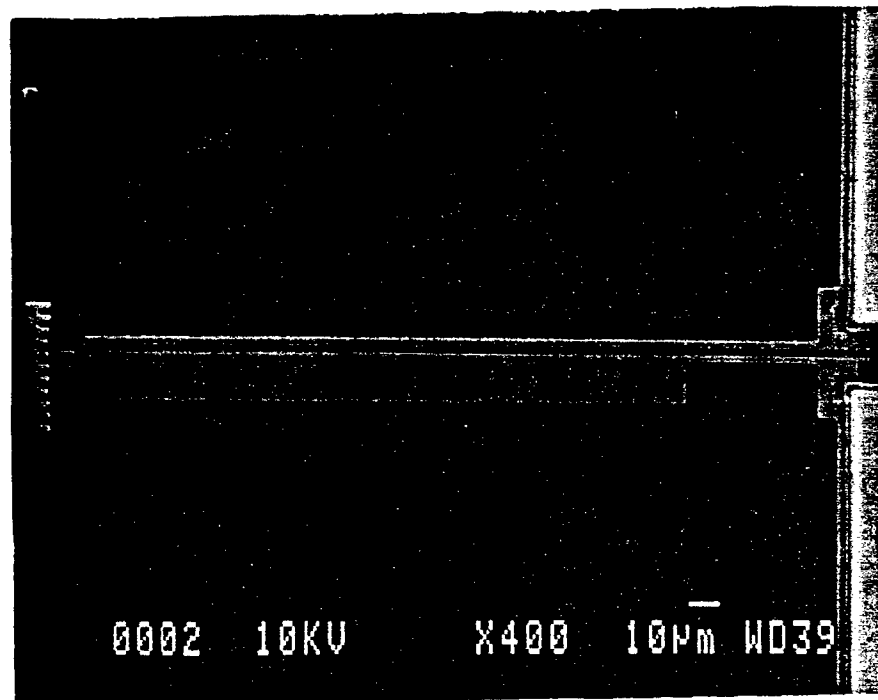
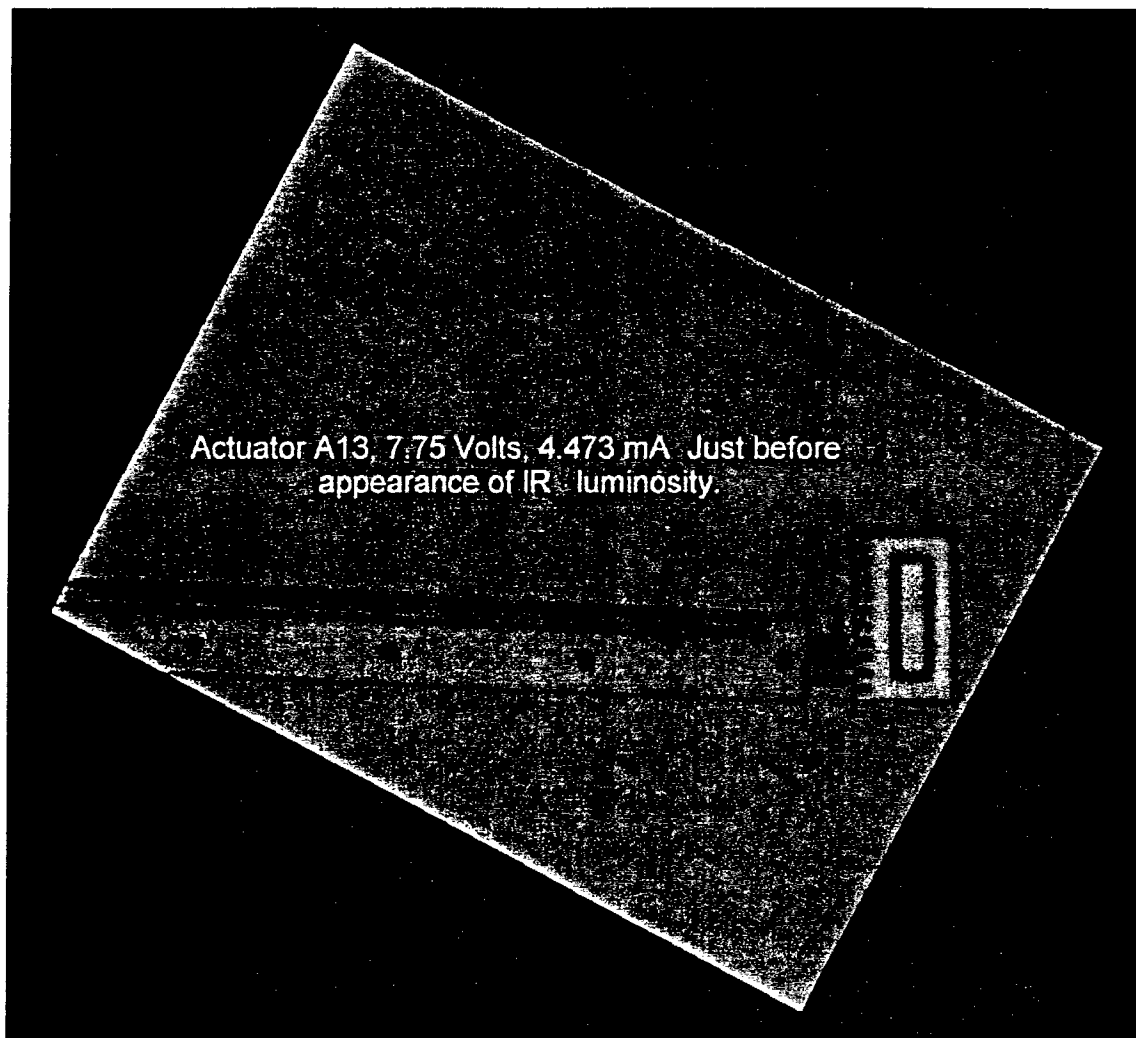


Fig. 3. Image of typical device on test chip with voltage applied to anchor pads. For this device no IR luminosity is seen for applied voltages below 8 volts.



A 13 Thermal Actuator

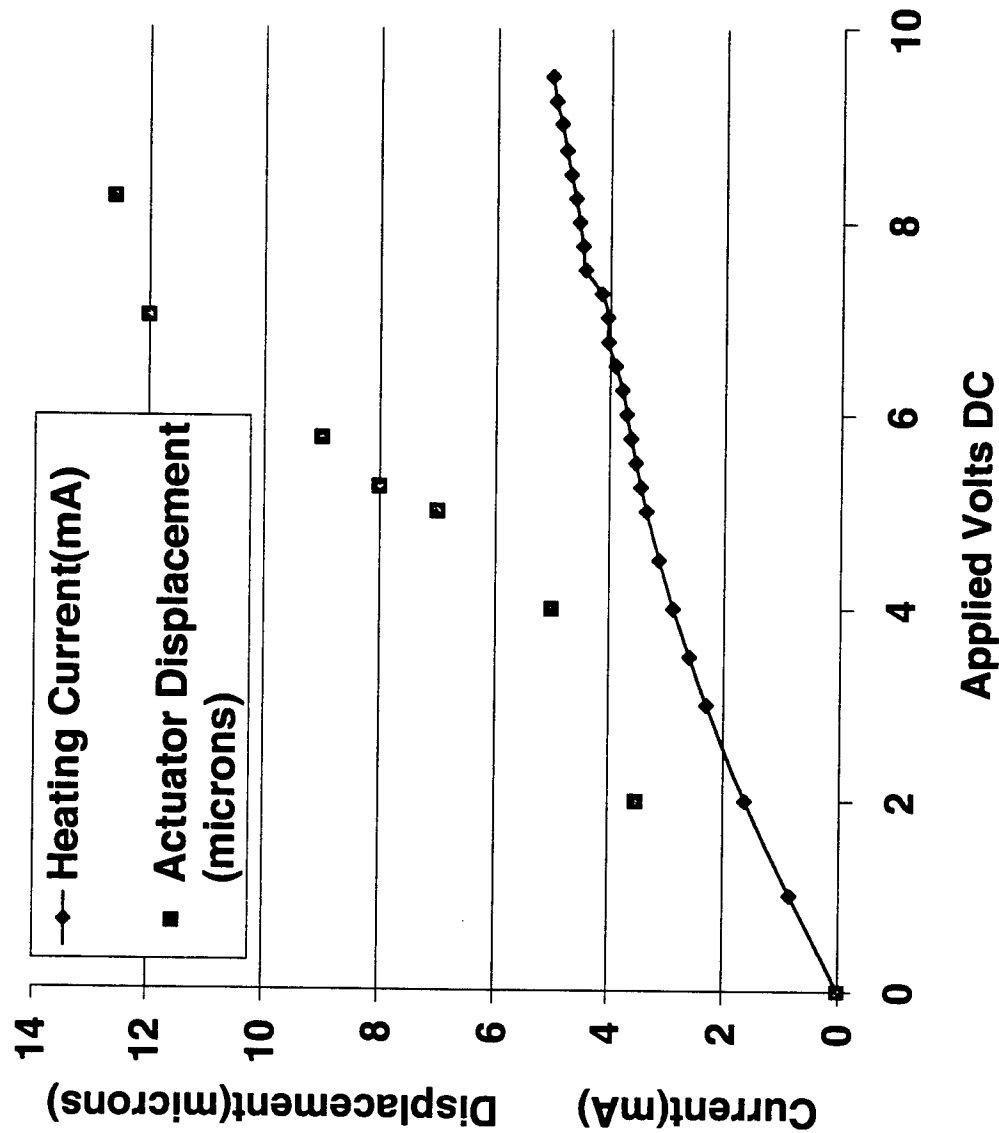
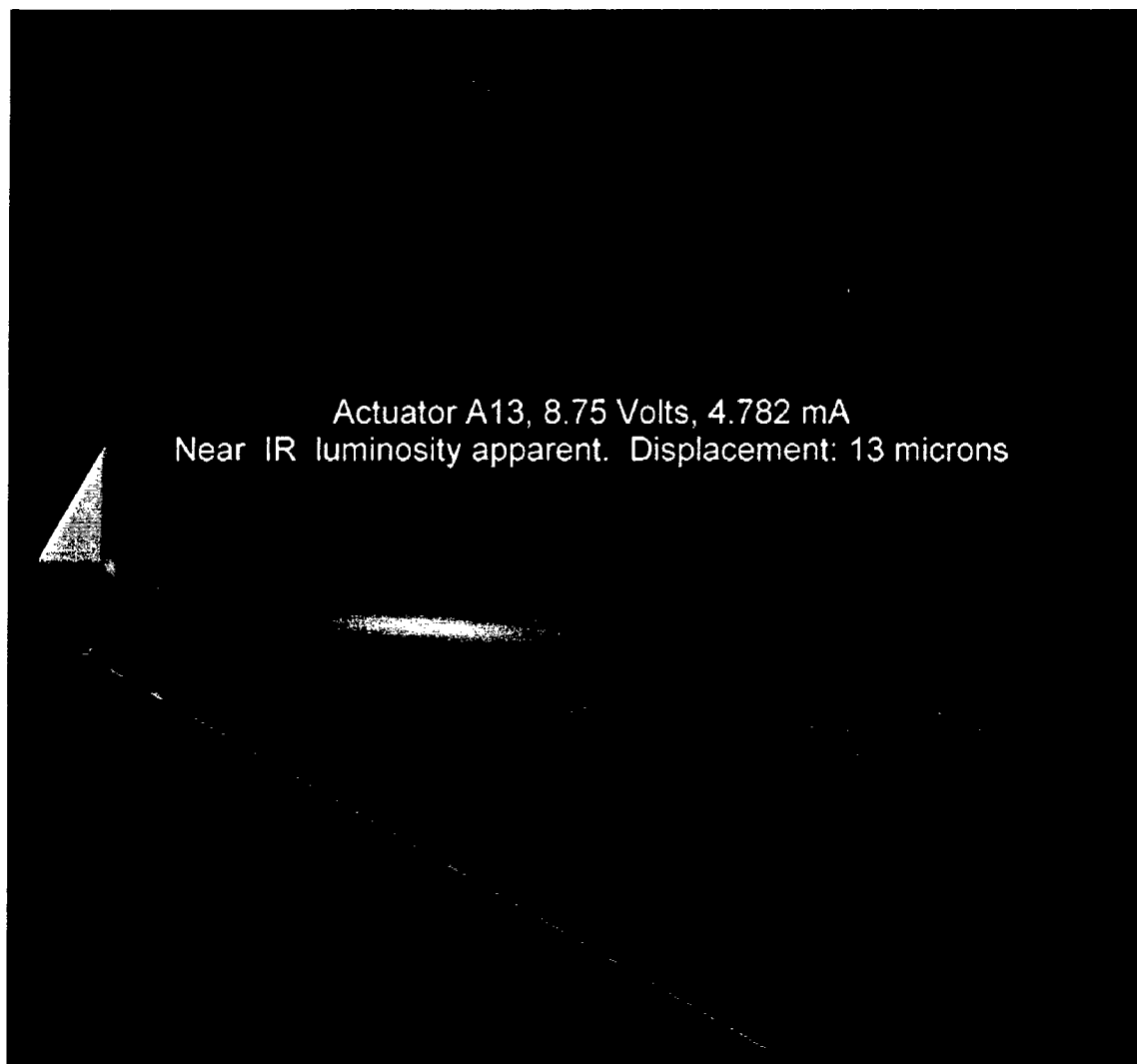


Fig. 4. Measured current and displacement as a function of applied voltage for the device indicated in Fig. 3, using the typical thermal actuator design from Fig. 1. Displacements are observed on a vernier indicator built on the device. Near IR luminosity is seen to commence when the applied voltage is > 8.0 volts.

Fig. 5 Image of same device indicated in Fig. 3 with sufficient voltage applied to indicate near-IR luminosity.



Thermal Actuator Characteristics

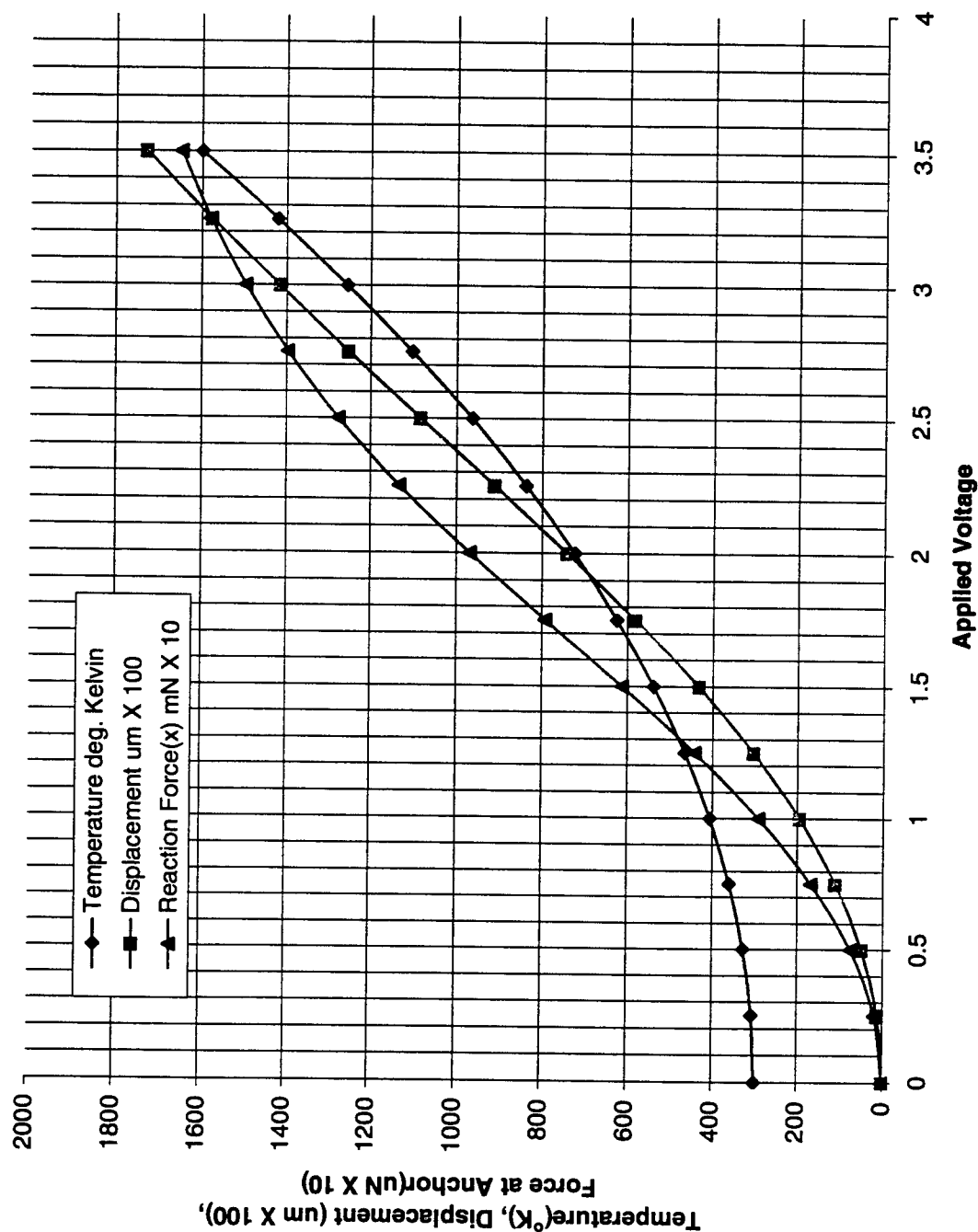


Fig. 6. Results of MemCad/MemETherm simulation indicating maximum temperature achieved in narrow beam, beam displacement, and reaction force at anchors as a function of applied voltage. Because of plastic flow of poly-Si above 1500 deg. K, useful simulations are limited to this temperature.

ALIAS-FREE PROCESSING OF P-3 SAR DATA

Mehrdad Soumekh
Department of Electrical Engineering

201 Bell Hall
State University of New York at Buffalo
Amherst, New York 14260

Final Report for:
Summer Faculty Research Program
Rome Research Site

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

and

Rome Research Site

September 1998

ALIAS-FREE PROCESSING OF P-3 SAR DATA

Mehrdad Soumekh

Department of Electrical Engineering

201 Bell Hall

State University of New York at Buffalo

Amherst, New York 14260

Telephone: (716) 645-2422, extension 2138

Email: msoum@eng.buffalo.edu

ABSTRACT

This paper is concerned with multidimensional signal processing and image formation with FOLIage PENetrating (FOPEN) airborne imaging radar data which were collected by a Navy P-3 ultra wideband (UWB) radar in 1995 [Raw]. A commonly-used assumption for the processing of the P-3 data is that the beamwidth angle of the radar is limited to 35 degrees [Bes], [Goo]; based on this assumption, the PRF of the P-3 SAR system yields alias-free data in the slow-time Doppler domain. However, controlled measurements with the P-3 radar have indicated a beamwidth which exceeds 35 degrees [Raw]; e.g., at the radar frequency of 300 MHz, the beamwidth angle is close to 100 degrees.

In this paper, we examine a method for processing of the P-3 data in which the incorrect assumption that its radar beamwidth angle is limited to 35 degrees is not imposed. In this approach, a SAR processing scheme which enables the user to extract the SAR signature of a specific target area (digital spotlighting) is used to ensure that the resultant reconstructed SAR image is not aliased [S94], [S95], [S99]. The images which are formed via this method with 8192 pulses are shown to be superior in quality to the images which are formed via the conventional P-3 processor with 16386 pulses which was developed at the MIT Lincoln Laboratory [Bes].

In the presentation, we also introduce a method for converting the P-3 deramped data into its alias-free baseband echoed data, and the resultant interpretation of Radio Frequency Interference (RFI) in the measured P-3 data. A two-dimensional spectral domain method for calibration from an in-scene target signature is discussed. Squint-mode image formation with the P-3 data is also examined. The original figures for this report may be obtained by contacting the author.

ALIAS-FREE PROCESSING OF P-3 SAR DATA

Mehrdad Soumekh

1. INTRODUCTION

Synthetic Aperture Radar (SAR) is an airborne or satellite-borne radar system which provides high-resolution maps of remote targets on a terrain, a planet, etc. The AWACS airplanes, which are used extensively in reconnaissance missions in the Persian Gulf region, and the NASA space shuttles are equipped with this radar system. SAR systems are a highly developed combination of precision hardware and electronic design for data acquisition, and advanced theoretical principles of mathematics and physics to convert the acquired data to high-resolution images.

National security and safety issues, such as the dangers from Scud missiles encountered in the Gulf War and drug interdiction problems, created the need for developing FOLiage PENetrating (FOPEN) SAR systems that utilize Ultra WideBand (UWB) UHF/VHF radars for detection of concealed targets. The SAR digital signal processing issues associated with these FOPEN reconnaissance SAR systems brought new complexities and misunderstandings for those familiar with the traditional SAR systems. This is mainly due to the wide-bandwidth and wide-beamwidth of these SAR systems which are desired to operate in near range (few hundred meters) as well as far range (several kilometers).

One of the first set of FOPEN SAR data were collected at Stanford Research Institute (SRI). A stripmap pulsed-type SAR was used in the frequency band of [200,400] MHz to image a target area with a near range value of approximately 500 meters. The imaging results indicated interesting features and capabilities of a FOPEN SAR system [S95]. However, to accurately retrieve the information in the SRI data and avoid Doppler aliasing, the user should utilize new signal processing tools for analyzing this SAR database, and forming images from it.

Imaging with FOPEN SAR data requires the use of a new SAR imaging method (for example see [Caf], [Car], [S90], [S91], [S92], [S94], [S95], [S99]) which is based on the Gabor's theory of *wavefront reconstruction* [Gab]. SAR wavefront reconstruction theory not only provides a tool for SAR image formation but also reveals functional properties of the SAR signal which contradict or were not predicted in the approximation-based theoretical foundation of either range-Doppler imaging or polar format processing [S95], [S99].

For instance, the beamwidth angle of the SRI radar was in excess of 90 degrees. It can be shown that the Pulse Repetition Frequency (PRF) which was used for the SRI

data collection was not sufficient to accommodate such a large beamwidth angle; as a result, the SRI data contained slow-time Doppler aliasing. The wavefront reconstruction theory reveals unique functional properties of and redundancies in the SAR signal which cannot be found in any other information processing modality. These redundancies could be exploited to recover uncorrupted SAR data from the collected aliased SRI data.

While slow-time Doppler aliasing is an issue in the processing of the SRI's FOPEN data, however, most of the users of the SRI data, who had background in either range-Doppler imaging or polar format processing, reported problems which stem from incorrect implementation of the SAR wavefront reconstruction algorithm. It turns out that an extensive theoretical and practical knowledge base has been shaped by the inherent approximations in the range-Doppler imaging and polar format processing methods over the past forty years; some of these concepts are either incorrect or not applicable when viewed in the framework of the modern high-resolution or FOPEN SAR systems, and the wavefront reconstruction theory.

A more extensive FOPEN stripmap SAR data collection was carried out in 1995 [Raw] with a Navy P-3 data by the Environmental Research Institute of Michigan (ERIM); the resultant database is referred to as P-3 data. The P-3 SAR system utilized a chirp radar within the frequency band of [215,730] MHz with a near range value of approximately 5400 meters. The P-3 data posed similar challenges and problems as the SRI did for the user.

The P-3 radar also possessed a relatively wide beamwidth angle [Raw]. Similar to the SRI data, the PRF for the P-3 radar was not sufficient enough to prevent slow-time Doppler aliasing within the large beamwidth angle of the radar. (The wide beamwidth angle for the P-3 data also posed other practical problems, for example, processing a relatively large number of slow-time samples.) The preliminary processing of the P-3 data, which was done at ERIM, was based on processing a 31.7 degree beamwidth angle. This was rationalized based on the following [Goo]:

For UWB/WB SAR systems which operate at low frequencies, the above (theoretical) definition (for the radar beamwidth angle) leads to an impractical wide coherent integration angle. As an example, the P3 UWB SAR would require a 73.3 degree integration angle based on this definition. Instead, we suggest a working definition of azimuth resolution based on the center wavelength. This definition yields a more practical integration angle of 31.7 degrees for the P3 UWB system at a center frequency of 469.5 MHz.

The final processing of the P-3 data was done by a group at the MIT Lincoln Laboratory (MIT-LL) [Bes]. The P-3 processor which was implemented at the MIT-LL mostly utilized the recommendations and procedures which were suggested by the ERIM group; for instance, the processing was performed on a 35 degree beamwidth angle. We will show later that the MIT-LL P-3 processor suffered from slow-time Doppler aliasing despite limiting the beamwidth angle to 35 degrees (Section 4). Moreover, there existed other problems with the MIT-LL P-3 processor which introduced other forms of artifacts in the reconstructed SAR image; these will be discussed.

Unfortunately, the other P-3 processors which are found in the literature suffer from various misunderstandings, misconceptions, and incorrect implementation of the SAR wavefront reconstruction. For instance, in [Kir], the authors mention a problem with the wavefront reconstruction which does not exist. Then, to solve the problem, they introduce an approximation which is attributed to the wavefront reconstruction though the approximation seems to be the one which is used in the polar format processing. It is likely that the problem which is observed by these authors in the P-3 wavefront reconstruction is due to the use of an incorrect reference fast-time point [S99], that is, an error in the implementation of the algorithm.

Recently, the use of the direct Time Domain Correlation (TDC) or backprojection algorithm for SAR image formation has become popular [Bar], [Cur], [S99]. The backprojection algorithm is a relatively simple approach for forming accurate SAR images though it carries a relatively heavy computational burden. We should point out that irrespective of the algorithm used for image formation (for example, wavefront or backprojection), an aliased SAR database results in artifacts and loss of contrast in the reconstructed image. Examples of this phenomenon with the backprojection reconstructions with the SRI data are provided in [S99].

This report provides a wavefront reconstruction theory framework for alias-free signal processing of P-3 data. The scheme is based on what is referred to as *digital spotlighting* of stripmap SAR data [S95], [S99]; this method was also utilized for the alias-free processing of the SRI data in [S99]. The resultant digitally-spotlighted P-3 data are then used to form the image of the target scene via the wavefront reconstruction algorithm.

The processing is also based on converting the deramped P-3 data into alias-free baseband echoed data of the target area [S99]. This operation has two desirable features. First, the deramped data suffers from what is called the Residual Video Phase (RVP) error which is commonly removed using certain approximations [Car]; baseband echoed data do not contain RVP error.

In addition, FOPEN SAR frequency band usually covers a part of the radio, television and/or communication channels; these produce what is referred to as Radio Frequency Interference (RFI) for FOPEN SAR systems. These RFI sources are fairly narrowband and commonly visible in the baseband echoed data. However, the deramping operation alters the spectral nature of the RFI sources (band broadening) which makes it difficult to excise the RFI from the desired SAR data. In fact, as we will see in Section 4, the RFI suppression which is performed in the MIT-LL P-3 processor results in degradations in the reconstructed image.

We will show that the baseband echoed data, which are restored from the deramped P-3 data, provide a two-dimensional fast-time and slow-time spectral domain in which the relatively strong direct path RFI sources appear as isolated signatures in the fast-time frequency and Doppler (aspect angle) domains (Sections 3 and 4). The user can utilize a two-dimensional nulling scheme in this domain to reduce the RFI effects with a minimal amount of distortion in the desired SAR data.

We will examine issues which are associated with calibration of the P-3 data with respect to the phase functions of the radar radiation pattern and the radar signal spectrum. A two-dimensional calibration phase function in the spectral domain is developed using the signature of an in-scene target (Section 5).

The MIT-LL processor is also known to yield artifacts at the azimuth points which are far from the aimpoint (that is, an image formation which is referred to as *squint*-mode SAR processing [Car], [S92], [S94], [S95], [S99]). While these degradations have been attributed to the errors in the GPS motion compensation [Bes], we will show that these are byproducts of processing aliased data (Section 4). We will examine alias-free processing of the P-3 data in the squint-mode, and provide reconstruction results for this SAR modality using P-3 data (Section 6).

We begin with an analysis of stripmap SAR system, and the P-3 SAR system data collection (Section 2).

2. STRIPMAP SAR SYSTEM

A. Narrow-Bandwidth and Narrow-Beamwidth Model

Consider the target region within the radar range swath which is composed of a set of point reflectors with reflectivity σ_ℓ that are located at the coordinates (x_ℓ, y_ℓ) ($\ell = 1, 2, \dots$) in the spatial (x, y) domain. (A discrete target model is used for notational simplicity; the principles which follow are also applicable for a continuous target region.) The variable x

is used for the range domain (also slant range or down range), and the variable y identifies the cross-range (also called azimuth or along the track).

The synthetic aperture domain is identified by the variable

$$u = v_r \tau,$$

where v_r is the speed of the radar-carrying aircraft in the along-track (azimuth) domain, and τ represents the *slow-time* domain. The sample spacing in the slow-time domain is dictated by the PRF of the radar system:

$$\Delta_\tau = \frac{1}{\text{PRF}};$$

thus, the sample spacing in the synthetic aperture domain is

$$\Delta_u = \frac{v_r}{\text{PRF}}.$$

The mean range swath of the radar is denoted with X_c ; see Figure 1. Let the length of the radar swath in the range domain be $2X_0$. Thus, the range values of the targets in the illuminated target area satisfy

$$x_\ell \in [X_c - X_0, X_c + X_0].$$

(In the case of P-3 data, $X_c \approx 5871.2$ m, and $X_0 \approx 483.7$ m.) In the conventional processor for a narrow-bandwidth and narrow-beamwidth stripmap SAR system, the user identifies the azimuth region for image formation

$$y_\ell \in [-Y_0, Y_0].$$

Under the assumption that the transmitting/receiving radar has a narrow-bandwidth and a narrow-beamwidth, a constant beamwidth angle (or beamwidth extent $\pm B$; see Figure 1) is used for slow-time coherent integration of the SAR data at all the desired range bins within the radar range swath. This corresponds to a synthetic aperture interval of (see Figure 1)

$$u \in [-L, L],$$

where

$$L = Y_0 + B.$$

If the frequency domain methods are used for this processing, the user should make sure that *circular convolution* aliasing is avoided. (This is a *spatial* domain-based aliasing which is different from the slow-time Doppler domain aliasing.)

B. Wide-Bandwidth and Wide-Beamwidth P-3 Model

The wide-bandwidth and wide-beamwidth of the P-3 SAR data poses new challenges which are not predicted in the conventional stripmap SAR processors. Figures 2a-b show examples of the beamwidth angle of the P-3 radar versus the radar frequency [Raw]. These figures show that the P-3 radar beamwidth angle is within 40 to 110 degrees in the radiated chirp radar signal ([215,730] MHz).

However, the results which are shown in Figures 2a-b represent the 3 dB beamwidth of the radar. For alias-free processing of the SAR data, the user should also incorporate the first side lobe of the radar in calculating the radar beamwidth [S99]. We should also point out that a 6:1 tapped delayed line (lowpass Doppler filter) in the slow-time domain was used to reduce the beamwidth of the P-3 radar [Bes], [Raw]. As we will show in our results, this lowpass Doppler filter was not successful in sufficiently reducing the beamwidth to prevent slow-time Doppler aliasing. In fact the minimum beamwidth angle for the P-3 radar after lowpass Doppler filtering turns out to be over 50 degrees in the [215-550] MHz radar frequency band; this will be shown with the real P-3 data.

Consider the P-3 data over $N = 16384$ slow-time samples; this approximately corresponds to a 35-degree beamwidth angle for the center of target area (called *aimpoint* in [Bes]). With the sample spacing of

$$\Delta_u = .4 \text{ m},$$

in the synthetic aperture domain, the half-size of the processed synthetic aperture is

$$L = \frac{N \Delta_u}{2} = 3276.8 \text{ m}.$$

Depending on the choice of Y_0 , the user faces different constraints for alias-free processing of the P-3 data.

An example is shown in Figure 3a. In this scenario, the desired target area is within the radar radiation pattern for all the available radar frequencies. from the SAR signal theory point of view, the resultant database should be treated as a *spotlight* SAR data. (Examples of this for the SRI data are provided in [S99].) In this case, the sampling in the synthetic aperture domain should satisfy [S92], [S94], [S99]:

$$\Delta_u \leq \frac{\lambda_{\min}}{4 \sin \theta_{\max}},$$

where $\lambda_{\min}$ is the wavelength at the highest frequency of the radar ($\lambda_{\min} = .41$ m at 730 MHz), and

$$\theta_{\max} = \arctan\left(\frac{Y_0 + L}{X_c - X_0}\right),$$

is the maximum aspect angle of the target area as seen by the radar within the synthetic aperture interval of $u \in [-L, L]$; see Figure 3b.

C. MIT Lincoln Laboratory P-3 Processor

The MIT-LL P-3 processor is based on coherent integration of 16384 slow-time pulses over 4096 fast-time samples. The resultant SAR image has $n_x = 4096$ range bins with sample spacing of $\Delta_x = .236$ m; this is based on the half-range swath of

$$X_0 = \frac{\Delta_x n_x}{2} = 483.7 \text{ m.}$$

(Due to the curvature of the radar radiation pattern in the FOPEN SAR systems, a finer range sample spacing should be used [S94], [S95], [S99].)

The SAR image of the MIT-LL processor produces 16384 cross-range bins with sample spacing of $\Delta_y = \Delta_u = .4$ m. However, only the middle $n_y = 4096$ cross-range bins are retained, and the remaining 75 percent of the cross-range bins are considered to be contaminated (circular convolution aliasing) and, thus, are discarded. In this case, the half-width of the formed image in the cross-range domain is

$$Y_0 = \frac{\Delta_y n_y}{2} = 819.2 \text{ m.}$$

3. RADIO FREQUENCY INTERFERENCE SUPPRESSION VIA CONVERSION OF DERAMPED SIGNAL TO ECHOED SIGNAL

The FOPEN SAR systems operate in the VHF and UHF frequency bands in which various television and communication channels reside. These sources, which are called Radio Frequency Interference (RFI), contaminate the collected SAR data. (There are other sources of RFI, e.g., Electronic Counter Measure, ECM, signals; for a discussion and treatment of ECM signals in SAR see [S99]).

Figure 4a shows the measured deramped P-3 data for four consecutive of the radar bursts versus the chirp radar instantaneous frequency. The resultant data can be interpreted as the SAR signal in the fast-time frequency ω domain [S99]. There are certain

inaccuracies which are involved in this assumption that are called Residual Video Phase (RVP). One can use an approximation to reduce the RVP error [Car]. The MIT-LL P-3 processor utilizes this scheme [Bes].

The main problem with this processing is not the approximation which is used to remove the RVP error. It turns out that the deramping of the measured signal results in *spectral spreading* of the RFI signals in the fast-time frequency domain. Figure 5a shows the two-dimensional spectrum (i.e., fast-time frequency and slow-time Doppler frequency) for a portion of the P-3 data. The horizontal lines which appear in this figure are the spectrally-spread RFI signals.

The MIT-LL processor utilized a scheme for RFI suppression in the P-3 data that is based on placing notches in the range equalization filter at the frequency bin locations that are occupied by the strong RFI signals. As we will show later, this would result in degradations in the Point Spread Function (PSF) of the SAR image, particularly in the range domain. In addition to degrading the PSF, there is no study on the effect of the adverse effects of the spectral spreading of the RFI signals in the reconstructed SAR image.

The processing which we used to encounter RFI is based on converting the deramped data into the baseband echoed signal [S99]. Figure 4b shows the converted echoed data for the deramped signals in Figure 4a. Note that the RFI signals appear as fairly isolated signatures (not spread) in Figure 4b; for example the audio and video signals for Channel 13 appear around 215 MHz. The two-dimensional spectrum of the echoed signal in Figure 5b exhibits the same phenomenon. Note that an RFI signal which correspond to a *direct* path propagation appears at specific slow-time Doppler frequency (or *angle of arrival* with respect to the synthetic aperture). The components seen at the other Doppler frequencies are due to multi-path effects as well as slow-time *incoherence* in the broadcasted TV signal.

This representation of the RFI signals allows the user to utilize a variety of the array signal processing methods for RFI suppression. However, one may also use simple methods for RFI suppression. For instance, since the direct path RFI signals pose the main distortion in the reconstructed SAR image, the user may apply power equalization on the spectral components of the SAR signal where the direct path signal is strong.

Note that the SAR wavefront reconstruction is based on a mapping of the spectral domain of the SAR signal into the target function spectrum; this is known as Stolt transformation. Thus, the user may apply spectral filtering for RFI suppression after forming the SAR image. This allows the user to determine whether such an RFI filtering, which is likely to degrade the PSF of the SAR system, is a necessary operation. In fact, our results indicate that the RFI strongest signal in the baseband echoed signal results in a

background noise which is below 60 dB of the strongest reflector in the scene. (This observation is based on coherent slow-time processing of 8192 P-3 pulses; the relative signal to RFI power ratio improves as the number of pulses is increased.)

4. DIGITAL SPOTLIGHTING AND SLOW-TIME (PRF) UPSAMPLING FOR ALIAS-FREE DOPPLER PROCESSING

A. MIT-LL P-3 Processor

As we mentioned earlier, the MIT-LL P-3 processor forms an image in the spatial domain region identified via

$$(x, y) \in [X_c \pm X_0, \pm Y_0],$$

where $X_c = 5871.2$ m, $X_0 = 483.7$ m, and $Y_0 = 819.2$ m. This area is formed by processing 16384 pulses in the synthetic aperture interval of

$$u \in [-L, L] = [-3276.8, 3276.8] \text{ m.}$$

Based on these parameters, the maximum aspect angle of the desired target area with respect to is

$$\theta_{\max} = \arctan\left(\frac{Y_0 + L}{X_c - X_0}\right) = 37^\circ.$$

Thus, at the radar frequencies where the beamwidth angle is less than 74 degrees ($\pm 37^\circ$), the desired target area is observable to the radar. If we use the constraint on the synthetic aperture sample spacing

$$\Delta_u \leq \frac{\lambda_{\min}}{4 \sin \theta_{\max}},$$

with $\Delta_u = .4$ m, and $\theta_{\max} = 37^\circ$, the following condition should be satisfied by the minimum wavelength:

$$\lambda_{\min} \geq .963 \text{ m.}$$

For a 74-degree beamwidth, this implies that the P-3 data contain aliasing for radar frequencies which are greater than $\frac{3 \times 10^8}{.963} = 312$ MHz.

However, this does not imply that the data from the targets outside the desired azimuth region (that is, $|y| > Y_0$) are not aliased. The aliased data from these targets not only results in shifted and smeared images of these targets (which is the result of aliasing

fold over in their own spectral band) but also noise-like artifacts (which is the result of aliasing fold over in the spectral band of the targets with $|y| < Y_0$) that appear in the image of the desired target area. The two-dimensional spectral data in Figures 5a-b show these Doppler aliased data; see the ramp-looking signatures. For instance, the signature of a target at an aspect angle of 40 degrees appears alias-free for radar frequencies which are less than 300 MHz. (The 40-degree ramp is cut at the highest slow-time Doppler frequency $k_u = 8$ rad/m, or 333 Hz, at around 300 MHz). The signature of this target at the higher radar frequencies is folded over, and is aliased.

As we stated earlier, the beamwidth angle of the P-3 radar is likely to be close to 50 degrees at the highest radar frequency. The coverage of the resultant data in the spatial frequency domain of the target image is shown in Figure 6 via two solid lines. The two dashed lines in this figure show the target spectral region when the radar beamwidth is assumed to be 35 degrees. The lightly shaded region represents the processed spectral data which are not aliased. The darker shaded area is the aliased spectral region of the 50-degree beamwidth data.

Although the MIT-LL P-3 processor applies a 35-degree Doppler filter on the data to remove slow-time Doppler aliasing. However, this filter cannot remove the entire *fold-over* aliasing data which are shown in Figure 6. Moreover, the 35-degree filter removes a significant amount of alias-free data at the lower radar frequencies ([215,400] MHz) where the beamwidth angle is well beyond 50 degrees. It turns out that most of the energy of a FOPEN radar that shows a significant return from a foliage area for concealed target detection is within this lower frequency band.

We should point out that for a planar aperture radar, or a parabolic radar with a feed, or a horn-type radar (which is used for collecting the P-3 data) where the beamwidth is a linear function of the wavelength, the slow-time Doppler bandwidth is approximately invariant in the radar frequency [S99]. For instance, if the P-3 radar beamwidth angle is approximately 50 degrees at 730 Mhz, then the target function spectral support is within the two solid lines which are shown in Figure 7. The resultant fold-over alising is shown by the dark shaded area in this figure.

In this report, we examine the P-3 reconstructions of a target area which is composed of a set of Corner Reflectors (CRs) on a clear land (Stakes 1-11), and CRs which are in a foliage region (Phenomenology Stakes 1-2). Figure 8a shows the location and the types of these targets; some of these targets (Stakes 1, 2, 10 and 11) are circled in the aerial optical image in Figure 8b; Phenomenology Stakes 1 and 2 are identified by two circles in the aerial optical image of Figure 8c.

Figures 9a-b show the SAR images of this target area which are formed by the MIT-LL processor. Figure 9a represents what is referred to as Level 3 reconstruction. Figure 9b corresponds to Level 4 reconstruction in which a calibration was performed to improve the results. (The calibration processing will be discussed.) The aimpoint which was used to form these images was at the center of the CRs on the clear land.

Figures 10a-b and 11a-b show the close-ups of Level 3 and Level 4 reconstructions of Stake 10 and Phenomenology Stake 2; the range and cross-range cuts for these targets are shown in Figures 10c and 11c. (These slices were obtained from the formed images which were clipped within a 45 dB power interval.)

Figures 12a-b show a portion of the reconstructed SAR image when the aimpoint is at a squint angle with respect to the CR area. (Level 4 reconstructions are shown.) For Figure 12a, the aimpoint is approximately 650 m from the center of the CR area; and the aimpoint is approximately -750 m from the center of the CR area in Figure 12b. Note that all the corner reflectors exhibit severe range domain side lobes in their images; this is due to the processing which was used for RFI suppression. In addition, there two other artifacts which can be seen in the MIT-LL processed P-3 images which are caused by processing Doppler-aliased data:

- i. The images contain fold-over aliasing from the targets which are outside the azimuth imaging area of $[-Y_0, Y_0] = [-819, 819]$ m. This form of Doppler aliasing appears as *background* (salt and pepper) noise that reduces the contrast of the image; these are clearly visible in Level 3 images of Figures 9a, 10a, and 11a. After Level 4 processing (i.e., calibration), the background noise becomes weaker. The calibration which was used for Level 4 processing was a one-dimensional gain (lowpass-type) and phase filter in the cross-range spatial frequency domain [Bes]. The fact that this filter was invariant in range can be observed by noting that the range domain side lobes are not changed after the calibration.
- ii. The images contain fold-over aliasing from the targets which are inside the azimuth imaging area of $[-Y_0, Y_0] = [-819, 819]$ m. This form of Doppler aliasing results in *ghost*-like structures which are the smeared and shifted replicas of the targets in the scene [S92], [S94], [S99]. These ghost artifacts get more severe as the target moves away in the cross-range domain from the aim point. This is due to the fact that the slow-time Doppler aliasing gets more severe as the target's squint angle with respect to the center of the synthetic aperture (or its azimuth location) increases [S92], [S94], [S99]. This phenomenon can be seen in Figures 12a-b where a squint aimpoint was used for image formation.

The authors of the MIT-LL processor had attributed the ghosting appearance of the squinted targets (that is, the targets away from the aimpoint) to the procedure which they used for GPS motion compensation [Bes]. It turns out that the MIT-LL processor performs GPS motion compensation with respect to the center of the target patch (aimpoint). This is analogous to using polar format processing (plane wave approximation) which introduces errors (smearing and shifting) for off-centered (squinted) targets. However, due to the large mean range swath of the P-3 radar and the relatively small range motion errors which were recorded by the GPS system of the P-3 SAR, the severe ghosts which appear in the MIT-LL images cannot be due to the GPS motion compensation.

Due to the effect of slow-time Doppler aliasing in the SAR image which is formed by the MIT-LL processor (which is most evident at the larger azimuth coordinates) as well as the incorrect processing of the wavefront reconstruction which resulted in the *circular convolution spatial aliasing*, the authors of the MIT-LL processor discarded about 75 percent of the azimuth data after image formation; that is, they kept the middle 4096 pixels in the azimuth, and discarded the other 12292 pixels at the azimuth boundaries (6146 on each side) of the image. The alias-free wavefront reconstruction processing which is described next also provides a tool (digital spotlighting) for forming an error-free image of an area without circular convolution spatial aliasing and discarding a large amount data.

B. Digital-Spotlight Alias-Free Wavefront Reconstruction

As we stated earlier, the echoed data in Figure 5b are aliased in the slow-time Doppler domain. We have developed a method to exploit certain redundancies in the SAR signal to circumvent this aliasing problem [S92], [S94], [S95], [S99]. The approach is similar to the processing which we used to retrieve the echoed data in the fast-time domain from the deramped data.

In this approach, the SAR data are first deramped in the slow-time domain; this is referred to as the slow-time compression. Then, a two-dimensional filtering method is used to extract the signature of the desired target area (that is, $|x - X_c| \leq X_0$ and $|y| \leq Y$) from the slow-time compressed SAR data; this is called *digital-spotlighting*. Then, based on the slow-time sampling constraint of

$$\Delta_u \leq \frac{\lambda_{\min}}{4 \sin \theta_{\max}},$$

the digitally-spotlighted slow-time compressed data are *upsampled* in the slow-time domain. Finally, the alias-free SAR signature of the desired target area is formed by *ramping* the upsampled data.

This scheme has been used for processing of the SRI data and Inverse SAR (ISAR) data of airborne targets [S94], [S99]. We use this procedure for imaging a *subpatch* target region with $X_0 = 150$ m and $Y_0 = 300$ m at the center of the CR area of Figure 8 using the P-3 data. For this, we only use 8192 pulses of the P-3 data, that is, one-half of the database which is used in the MIT-LL P-3 processor. There are two reasons for this. First, for alias-free processing of this target area, the slow-time sampling constraint becomes $\Delta_u \leq .19$ m; that is, we need to upsample the P-3 data by more than a factor of two that results in more than 32000 slow-time samples. Due to the memory limitation of the PC computer that we used, this was infeasible. (We should point out again that for alias-free reconstruction with the backprojection algorithm, the user has to process the same large upsampled database.)

The other reason for processing 8192 P-3 pulses is that the resultant alias-free reconstruction exhibits target images which are superior to the aliased MIT-LL reconstructions with 16384 pulses. (In fact, the alias-free reconstruction with 4096 P-3 pulses exhibit better *contrast* than the MIT-LL reconstructions with 16384 pulses.)

Based on the constraint $\Delta_u \leq \frac{\lambda_{\min}}{4 \sin \theta_{\max}}$, the 8192 P-3 pulses are upsampled (in addition to the digital spotlighting) to yield 11358 pulses with sample spacing of $\Delta_u = .2885$ m. Then, after slow-time Doppler processing (that is, Fourier transform from the synthetic aperture u domain to the k_u domain), the data could be down-sampled to satisfy the Nyquist constraint for imaging azimuth region $|y| \leq Y_0$; this yields about 3800 azimuth spatial frequency k_y samples.

After appropriate phase processing and interpolation, the two-dimensional target spectrum in Figure 13 is obtained. Figure 9c shows the resultant spatial domain reconstruction; this image is analogous to the uncalibrated Level 3 MIT-LL reconstruction in Figure 9a. Figure 9d shows the alias-free wavefront reconstruction after calibration. (The processing for calibration will be discussed.) Figures 10d-e, respectively, show the uncalibrated and calibrated alias-free reconstructions of Stake 10; the range and cross-range slices through these images are shown in Figure 10f. (MIT-LL reconstructions of Stake 10 are shown in Figures 10a-c.)

Figures 11d-f show similar alias-free wavefront reconstruction results for Phenomenology Stake 2. (MIT-LL reconstructions of Phenomenology Stake 2 are shown in Figures 11a-c.) Figures 14a-c, 15a-c, and 16a-c are the alias-free wavefront reconstructions for Stakes 2, 3, and 9. Note that all the image of the trihedrals also show a curved structure behind the CR. This signature is due to *multipath* effects inside a trihedral which results in multiple delayed echoes in addition to the primary echoed signal from the CR. Unlike the primary echo, the multipath echoes are not coherent in the slow-time (synthetic

aperture) domain to form a focused image. These echoes correspond to a (non-existent) moving target in the slow-time domain; the curved structure is the result of imaging such an imaginary moving target.

5. TWO-DIMENSIONAL SPECTRAL DOMAIN IN-SCENE TARGET CALIBRATION

SAR wavefront reconstruction theory shows that the SAR data has to be *matched-filtered* in the two-dimensional fast-time and slow-time spectral domains by appropriate phase functions of the radar radiation pattern and the radar signal spectrum [S99]. For this, the user should determine these phase functions in a noise-free environments; these are referred to as calibration measurements.

Calibration measurements for the P-3 radar were apparently made. The authors of MIT-LL processor used a scheme to exploit these measurements in conjunction with in-scene targets signatures to construct a *one-dimensional* calibration signal which varies with the azimuth (cross-range) spatial frequency, but was invariant in the range spatial frequency [Bes].

As we stated earlier, the calibrated MIT-LL P-3 image was called Level 4 reconstruction. The Level 3 (uncalibrated) and Level 4 MIT-LL reconstructions were shown in Figures 9a-b, and for two individual targets in Figures 10a-c and 11a-c. Note that these images show that the point spread function (PSF) improves in the azimuth domain for Level 4 reconstruction; the same is not true for the range domain PSF (the range side lobes) due to the invariance of the one-dimensional calibration in the range domain.

To determine the nature of the calibration phase, we examined the spectral phase of the CRs in the alias-free image of Figure 9c. These are shown in Figures 10g, 11g, 14d, 15d, and 16d for Stake 10, Phenomenology Stake 2, Stake 2, Stake 3, and Stake 9, respectively. (To exhibit the phase information, the magnitude of the real part of these two-dimensional spectra are shown in these figures; the multipath effects are not included to compute these spectra.) Note that these spectra all show a phase (fringe) pattern which varies with both the range spatial frequency k_x and the azimuth spatial frequency k_y . (Note that the calibration signal for the MIT-LL processor was a one-dimensional signal of the azimuth spatial frequency k_y .)

Moreover, this phase pattern is approximately radially symmetric in the spatial frequency domain. The latter observation indicates that the calibration phase is mainly due to the phase of the chirp radar spectrum; that is, an additional phase is modulating the chirp which has to be matched filtered (removed) [S99]. In our processing, we use the

two-dimensional spectral signature of Stake 10 for calibration. The results are shown in Figure 9d, and for the individual targets in Figures 10e, 11e, 14b, 15b, and 16b. (The range and cross-range slices for these targets appear as solid lines in Figures 10f, 11f, 14c, 15c, and 16c.)

6. SQUINT-MODE WAVEFRONT RECONSTRUCTION

In Figures 12a-b, we showed the reconstruction results for the MIT-LL processor when the aimpoint was moved to 650 m and -750 m from the center of the CR area; these are referred to as *squint* aimpoints. Due to the processing of aliased data by the MIT-LL P-3 processor, these squint-mode images exhibited severe ghost-like signatures of the CRs. The alias-free wavefront reconstruction of Section 4 may also be used for error-free mage formation in squint mode [S94], [S95], [S99].

To show this, consider a 200 m by 300 m target area ($X_0 = 100$ m and $Y_0 = 150$ m) centered at the range $X_c = 5600$ m. This area contains a CR (Stake 1) and some dense foliage. The *broadside* digitally-spotlighted reconstruction of this area with 512 P-3 pulses is shown in Figure 17a; for the broadside reconstruction, the 512 pulses are centered at the azimuth $y = 0$. Figure 17b is the spectrum of this reconstruction.

We should point out that after digital spotlighting, the Nyquist sampling constraint in the slow-time domain becomes $\Delta_u \leq 1.42$ m which is smaller than the slow-time sample spacing of $\Delta_u = .4$ m of the measured P-3 data. In this case, the digitally-spotlighted P-3 data is *down-sampled* accordingly by our processor. (Note that the azimuth spectral support in Figure 17b is *smaller* than the slow-time Doppler support of the original P-3 data in Figure 5b.) Figure 17c shows the spectral fringe pattern (real part) of Stake 1 signature.

Next, we consider 512 pulses of the P-3 data which are approximately 2500 m away from the center of the target area in the azimuth domain; this corresponds to an approximately 25 degrees of squint angle. Figure 18a shows the squint reconstruction of the target region; Figure 18b shows the reconstruction for the frequency band [230,515] MHz. It turns out that the P-3 data at this squint angle for the radar frequencies which are greater than 515 MHz are very weak. We believe that this is due to the 6:1 subsampling slow-time filter that was applied on the P-3 data [Bes], [Raw]. Figure 18c shows the spectrum of the reconstructed image; Figure 18d is the spectral fringe pattern of Stake 1.

7. REFERENCES

- [Bar] B. Barber, "Theory of digital imaging from orbital synthetic aperture radar," *International Journal of Remote Sensing*, vol. 6, no. 7, pp. 1009-1057, July 1985.
- [Bes] L. Bessette, M. Touns and B. Binder, "Documentation of the Lincoln Laboratory Processing for P-3 UWB imagery," S. Ayasli, Project Leader, *MIT Lincoln Laboratory Memorandum 47PM-STD-0010*, for Contract F19628-95-0002, August 21, 1996. (Unclassified property of the Government of the United States)
- [Caf] C. Cafforio, C. Prati, and F. Rocca, "SAR data focusing using seismic migration techniques," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 194-207, March 1991.
- [Car] W. Carrara, R. Goodman, and R. Majewski, *Spotlight Synthetic Aperture Radar*, Artech House, 1995.
- [Cur] J. Curlander and R. McDonough, *Synthetic Aperture Radar*, Wiley, 1991.
- [Dow] R. Downing, "FOPEN 1995 Camp Grayling, Michigan, Data Collection, Volume I: Ground Truth Report," *MIT Lincoln Laboratory Project Report STD-77*, for Contract F19628-95-0002, . (Unclassified property of the Government of the United States)
- [Gab] D. Gabor, "A new microscope principle," *Nature*, vol. 161, p. 777, 1948.
- [Goo] R. Goodman, S. Tummala and W. Carrara, "Issues in ultra-wideband, widebeam SAR image formation," *Proc. IEEE International Radar Conference*, pp. 479-485, 1995.
- [Kir] D. Kirk and P. Maloney, "A new autofocus technique for wide-band wide-angle synthetic aperture radar," *Proc. SPIE Aerosense Conf.*, Orlando, April 1998.
- [Raw] R. Rawson, R. Schneider, D. Sheen and G. Twigg, "P3/SAR Ultra Wideband upgrade (UWB), Final Report," *ERIM Report No. 241300-78-F*, volume 1, for Contract N62269-91-C-0526, March 3, 1996. (Unclassified property of the Government of the United States)
- [she] D. Sheen, C. Strawitch and T. Lewis, "UHF wideband SAR design and preliminary results," *Proc. of IEEE International Geoscience and Remote Sensing Symposium*, pp. 289-291, Pasadena (CA), August 1994.
- [S90] M. Soumekh, "Echo imaging using physical and synthesized arrays," *Optical Engineering*, vol. 29, no. 5, pp. 545-554, May 1990.
- [S91] M. Soumekh, "Bistatic synthetic aperture radar inversion with application in dynamic object imaging," *IEEE Transactions on Signal Processing*, vol. SP-39, no. 9, pp. 2044-2055, September 1991.

- [S92] M. Soumekh, "A system model and inversion for synthetic aperture radar imaging," *IEEE Transactions on Image Processing*, vol. IP-1, no. 1, pp. 64-76, January 1992.
- [S94] M. Soumekh, *Fourier Array Imaging*, Englewood Cliffs, NJ: Prentice Hall, 1994.
- [S95] M. Soumekh, "Reconnaissance with ultra wideband UHF synthetic aperture radar," *IEEE Signal Processing Magazine*, vol. 12, no. 4, pp. 21-40, July 1995.
- [S99] M. Soumekh, *Synthetic Aperture Radar Signal Processing with Matlab Algorithms*, New York: Wiley, to appear in 1999.

HIGH VELOCITY PENETRATION OF
LAYERED GROUT TARGETS

Joseph W. Tedesco
Professor
Department of Civil Engineering
Auburn University
Auburn, AL 36849

Final Report for:
Summer Faculty Research Program
Wright Laboratory Munitions Directorate

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

and

Wright Laboratory Munitions Directorate
Eglin AFB, FL

September 1998

HIGH VELOCITY PENETRATION OF LAYERED GROUT TARGETS

Joseph W. Tedesco

Introduction

In recent years the penetration of concrete and grout targets by small-scale projectiles has received considerable attention. The motivation for this directed effort is the trend in new designs for manned Air Force aircraft which employ internal weapon carriage. As a result of the mass and volume constraints imposed by this requirement, the development of smaller penetrators for air to surface munitions is being initiated by the Air Force.

A critical issue associated with the deployment of smaller penetrators (e.g. 250 lb class MMT) is that they meet or exceed the performance criteria of the currently deployed 2000 lb penetrator. To achieve this goal, the smaller penetrator must be capable of performing at impact velocities in excess of 4000 fps. Indeed, impact velocities in this range impose severe demands on the projectile which will experience significant nose erosion [1,2]. This has prompted the investigation of new, high-strength metal alloys for the penetrator. However, the performance of these new penetrator materials is beyond the scope of this report.

There currently exists a substantial database for high velocity penetration of small-scale projectiles into grout and concrete targets. The most comprehensive of these studies have been conducted at Sandia National Laboratories (SNL) by Forrestal et al [3-7]. However, the entire SNL database is for monolithic grout or concrete targets. This has raised several concerns among investigators with respect to the correlation of monolithic penetration data with layered target penetration.

Interest in the penetration of layered grout or concrete targets has been piqued for several reasons. One consideration is due to practical aspects associated with the construction of hardened facilities, specifically, the use of cold joint construction. A second consideration is more strategic in nature, that is the case of multiple burster layer slabs employed to influence penetrator trajectory. A third consideration is an idiosyncrasy of full scale testing, that is, the use of a multiple layer target configuration to represent an equivalent monolithic target. For these reasons, the present study was initiated.

Objectives and Scope

The primary objective of this study is to conduct scaled high velocity penetration experiments into layered grout targets and compare the results with monolithic target data in an attempt to

quantify penetration enhancement. A second objective is to identify characteristics of concrete penetration peculiar to layered target systems.

However, prior to implementation of the scaled experimental study, a test plan was based upon a comprehensive analytical study using the SAMPLL [8] computer program. The results of the analytical study were then used to establish the layered target configurations to be tested and the penetrator impact velocities.

Analytical Study

The analytical study was conducted to establish the layered target configurations to be investigated in the experimental study as well as the appropriate impact velocities for the small scale penetrators. To provide a meaningful comparison with the monolithic data, all layered target configurations had a total thickness of 36 in, since the monolithic targets were 30 in diameter cylinders of 36 in length. Furthermore, the layered configurations investigated were selected based upon the inventory of grout plate targets available. All plate targets were 32 in square, cast in four different thicknesses. The inventory of available grout targets is presented in Table 1. The physical characteristics of the small scale penetrator are illustrated in Figure 1 and summarized in Table 2.

Table 1. Inventory of Grout Plate Targets

Number	Thickness (in)	UNC (psi)	Density (lb/cu.ft.)
10	4	6409	140.4
10	8	8157	139.8
8	12	7856	139.1
5	16	7856	139.1

Table 2. Small Scale Projectile Data

Type	4340-1
Weight	0.15 lb
Length	3.5 in
Diameter	0.5 in
CRH	3.0
Yield Strength	250 ksi
Charpy Impact	18 ft-lb

The layered targets considered in the analytical study are summarized in Table 3. For each layered configuration (LC) a penetration curve was developed. Penetration curves for six different layered configurations and the corresponding monolithic target is presented in Figure 2. The percent increase in penetration depth exhibited by four of the layered target configurations relative to the monolithic target is presented in Figure 3.

Table 3. Layered Grout Targets Investigated in Analytical Study

Target Configuration	Layer Thickness (in)
LC1	4-16-16
LC2	12-12-12
LC3	4-8-8-16
LC4	4-4-8-8-12
LC5	8-16-12
LC6	8-12-16
LC7	8-8-8-12

From the results of the analytical study a test plan was devised. It was decided to investigate five different layered configurations at a penetrator velocity of 1000 m/s, and four of these same configurations at 1200 m/s. The final test plan is summarized in Table 4. Layered target configuration LC8 was tested at 1000 m/s only due to the availability of 4 in plate targets.

Table 4. Final Test Plan for Layered Grout Targets

Target Configuration	Layer Thicknesses (in)	Impact Velocity (m/s)
LC1	4-16-16	1000 and 1200
LC2	12-12-12	1000 and 1200
LC3	4-8-8-16	1000 and 1200
LC7	8-8-8-12	1000 and 1200
LC8	4-4-4-4-16	1000 only

Small Scale Experimental Study

The small scale projectile illustrated in Figure 1 was fired at the specified velocities into the layered grout target configurations described in Table 4. The delivery device used for the experiments was the 50 mm gun located on Range C-64 at Eglin Air Force Base, FL. A schematic of the gun and the high velocity concrete penetration experiments (HVCP) is presented in Figure 4. Between the front of the gun and the plate targets, two velocity screens (V_1 and V_2) and five x-ray cameras (T_1 through T_5) were situated. The pitch of the projectile, before impact, was ascertained from cameras T_1 , T_2 and T_3 , and the projectile yaw determined from cameras T_4 and T_5 . Moreover, the velocity determined from the velocity screens was verified from cameras T_2 and T_3 . A pre-test photograph of target configuration LC1 is presented in Figure 5. To insure the best possible contact between the individual plates, a system of angle irons and threaded rods was employed at both the top and bottom of the target configuration as illustrated in the photograph.

In an attempt to quantify the penetration enhancement exhibited by the layered grout targets, comparisons were made with the average penetration experienced by several monolithic targets at specified velocities. This was done for several reasons: 1) the projectiles used for the monolithic target tests were of a different material than those used for the layered target tests; and 2) the unconfined compressive strength (UNC) of the monolithic targets was the same as that for the 8 in plate targets (8157 psi), but different than that for the 4 in, 12 in and 16 in plates (refer to Table 1).

The HVCP depth of penetration results for the monolithic targets are presented in Table 5 for 1000 m/sec and in Table 6 for 1200 m/sec. A post-test photograph of monolithic test HVCP-41 (AF-1410 projectile at 1200 m/sec) is presented in Figure 6.

Table 5. Penetration into Monolithic Targets (1000 m/sec)

Test	Projectile	Velocity (m/sec)	Max. Pitch (deg)	Max Yaw (deg)	Penetration (in)
HVCP-58	ES-1	1073	6.5 up	9.0	12.0
HVCP-40	AF-1410	1000	0	2.40 r	13.25 12.63 (ave)

Table 6. Penetration into Monolithic Targets (1200 m/sec)

Test	Projectile	Velocity (m/sec)	Max. Pitch (deg)	Max Yaw (deg)	Penetration (in)
HVCP-53	AerMet-100	1220	1.0 up	3.0 r	19.69
HVCP-41	AF-1410	1218	1.70 dn	2.44 r	20.31 20.0 (ave)

The HVCP depth of penetration results for the layered grout targets are presented in Table 7 for 1000 m/sec and in Table 8 for 1200 m/sec. Post-test photographs of layered target test HVCP-69 are presented in Figures 7 through 10. Figure 7 shows the front face of the 4 in layer and Figure 8 shows the back face of the same layer. Figures 9 and 10 show the front face and back face, respectively of the first 16 in. layer. Post-test photographs of the front face and back face, respectively, of the first 12 in layer from test HVCP-64 are presented in Figures 11 and 12.

Table 7. Penetration into Layered Targets (4340-1 @ 1000 m/sec)

Test	Target	Velocity (m/sec)	Max Pitch (deg)	Max Yaw (deg)	Layers Penetrated	Penetration (in)	% Increase in Penetration
HVCP-64	LC2	982	4.5 dn	2 l	2	13.35	5.7
HVCP-65	LC1	1107	5.5 up	6 r	2	12.725	0.7
HVCP-66	LC3	1070	8.5 dn	7.5 r	3	15.31	21.2
HVCP-67	LC7	947	0	1.0 r	2	13.0	2.9
HVCP-72	LC8	970	17 dn	3.5 l	4	13.125	3.9 6.9 (ave)
HVCP-68	LC2	1227	1.5 dn	4.0 l	2	20.125	0.6
HVCP-69	LC1	1278	3.5 dn	0	3	22.345	11.9
HVCP-70	LC3	1236	1.5 dn	2.0 l	4	21.188	5.9
HVCP-71	LC7	1240	0	7.0 l	3	20.0	9.4 6.9 (ave)

Post-Test Observations

The results of the HVCP tests provided a number of interesting observations with respect to penetration, projectile tunneling through layers, front face and back face cratering in the individual layers, and the development of a compaction zone around the projectile tunnel. This is especially evident when the HVCP test results for the layered targets are compared with the HVCP monolithic test results.

With respect to penetration, the layered target configurations exhibited some enhancement, but not nearly as much as predicted by the analytical study. For example, for LC3, the predicted increase in penetration was almost 50% for 1000 m/sec, as illustrated in Figure 13. However, the actual penetration enhancement was 21.2% exhibited in test HVCP-66 as presented in Table 6. For this same configuration at 1200 m/sec, the predicted increase in penetration was approximately 45% as shown in Figure 14. The observed increase in penetration for layered test HVCP-70, however, was only 5.9%. In fact, the average increase in penetration for all layered configurations tested was approximately 7%, for both the 1000 m/sec and 1200 m/sec HVCP tests.

The tunneling phenomenon was readily observed through the individual layers. Well defined tunnels of approximately $\frac{1}{2}$ in diameter were clearly manifested throughout all layers penetrated. In direct contrast to monolithic targets, this same observation cannot be made without conducting laborious coring and slicing operations.

The cratering aspects of the penetration phenomenology presented some interesting observations in the layered target configurations. The cratering observed in the front faces of the first layer of the layered target configurations was generally consistent with the cratering observed in the faces of the monolithic targets, both in terms of crater depth and diameter. One distinct dissimilarity, however, was that the monolithic target faces generally exhibited a greater number of radial cracks than the first layer front faces of the layered target configurations. This observation is clearly evidenced in Figures 6 and 7.

Cratering on the target faces at layer interfaces exhibited two distinct, and quite different, appearances. In some instances, there was virtually no evidence of a crater forming. The tunnel was surrounded by a circular compaction zone having a radius of approximately 2 to 3 penetrator diameters. Where radial cracks developed, they were initiated outside the circumference of the compaction zone. This observation is clearly illustrated in Figures 8 and 9. However, in direct contrast to this behavior, in some instances cratering quite similar to that observed in the front face of the first layer in a layered target configuration, or on the face of a monolithic target, developed in the interior layer interfaces. This is clearly demonstrated in Figure 10.

At this early stage in the investigation of high velocity penetration into layered grout targets, it is quite difficult to ascertain the precise reason for these diverse observations at the layer interfaces. One possible explanation is the degree or completeness of contact between the individual layers.

Because many of the plate target faces were not level or bowed (either concave or convex), gaps were formed in many of the interfaces in the various layered target configurations tested. These gaps may have contributed to the cratering observed at the layer interfaces. However, at interfaces where good contact between layers was maintained, confinement was preserved and crater formation did not occur.

Conclusions and Recommendations

The results of the experimental study have uncovered several interesting aspects regarding the high velocity penetration into layered concrete or grout targets. There does appear to be some penetration enhancement in layered targets in comparison to corresponding monolithic targets. However, the increase in penetration does not appear to be as great as predicted by the preliminary analytical study. Moreover, there are several factors influencing penetration in layered targets which are not a consideration in penetration of monolithic targets. These factors include the number of layers penetrated, the relative thickness of the individual layers and the contact condition at layer interfaces.

The results of the experimental study generally indicate that penetration depth increases as the number of layers penetrated increases. The exception to this trend is HVCP-72 in which the penetration enhancement percentage is only 3.9%, well below the average for all targets of 6.9%. However, this particular test exhibited an extremely large amount of pitch (17 degrees) which undoubtedly reduced the potential penetration significantly.

Probably, the single most important factor influencing penetration of layered targets is the contact condition at layer interfaces. As discussed in the previous section, this condition affects the cratering phenomenon at layer interfaces. Interfaces experiencing good contact conditions and confinement exhibit less cratering around the tunnel than do interfaces with gaps. However, it cannot be ascertained whether these interface gaps enhance or mitigate penetration.

To address these uncertainties concerning penetration into layered target systems, it is recommended that a second phase of small scale experiments be conducted. In the Phase II experimental study it is recommended that penetration curves be developed for a limited number of layered target configurations. It is suggested that penetrator impact velocities range from 1000 m/s to 1600 m/s. Moreover, it is further recommended that two separate series of experiments be conducted for each target configuration. The first series should be conducted with full contact between all layer interfaces, and the second series be conducted on target systems having specified gaps between layers. These series experiments should lend some insight into quantifying the effect of the contact condition at layer interfaces on penetration. Some possible layered target configurations that might be investigated in the Phase II experiments are summarized in Table 8.

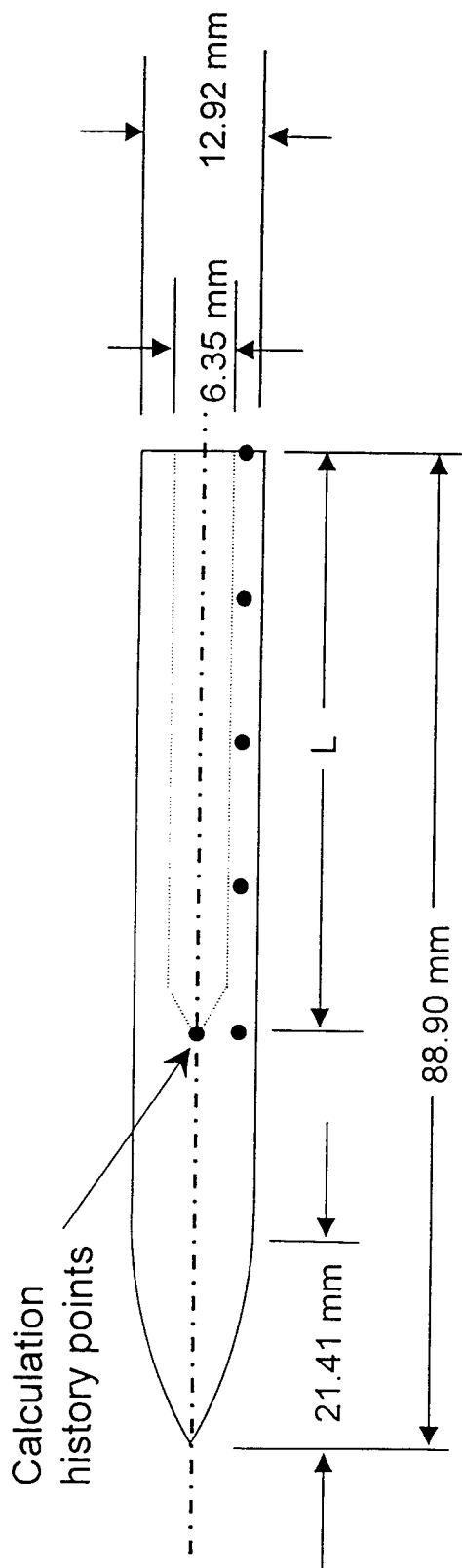
Table 8. Layered Target Configurations for Phase II Experiments

No. of Layers	Layer Thicknesses (in)
3	12-12-12
4	8-8-8-12
7	4-4-4-4-4-4-12

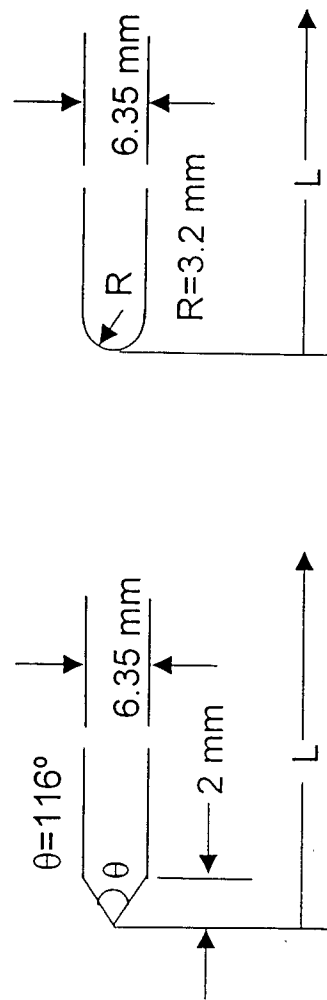
Finally, to further to investigate and understand the penetration phenomenon into layered target systems, it is recommended that a comprehensive parametric study be conducted using the Lagrangian finite element hydrocode EPIC [9]. The smooth particle hydrodynamics (SPH) algorithms in EPIC make it possible to analyze both severe distortions and structural responses with a single Lagrangian code.

References

1. Gold, V.M., Vradis, G.C. and Pearson, J.C., "Concrete Penetration by Eroding Projectiles: Experiment and Analysis", ASCE Journal of Engineering Mechanics, Vol. 122, No. 2, pp. 145-152, 1996.
2. Gold, V.M. and Vradis, G.C., "Analysis of Penetration Resistance of Concrete", ASCE Journal of Engineering Mechanics, Vol. 124, No. 3, pp. 328-338, 1998.
3. Forrestal, M.J., Frew, D.J., Hanchak, S.J. and Brar, N.S., "Penetration of Grout and Concrete Targets With Ogive-Nose Steel Projectiles", International Journal of Impact Engineering, submitted for publication.
4. Forrestal, M.J., Brar, N.S. and Hanchak, S.J., "Penetration of Grout Targets With Ogive-Nose Steel Projectiles", International Journal of Impact Engineering, submitted for publication.
5. Lak, V.K. and Forrestal, M.J., "Penetration Into Semi-Infinite Reinforced Concrete Targets With Spherical and Ogival Nose Projectiles", International Journal of Impact Engineering, Vol. 6, No. 4, pp. 291-301, 1987.
6. Forrestal, M.J., Altman, B.S., Cargile, J.D. and Hanchak, S.J., "An Empirical Equation For Penetration Depth of Ogive-Nose Projectiles Into Concrete Targets". International Journal of Impact Engineering, Vol. 15, No. 4, pp. 345-405, 1994.
7. Frew, D.J., Forrestal, M.J., Hanchak, S.J. and Green, M.L., "Penetration into Limestone Targets With Ogive-Nose Steel Projectiles", The 14th U.S. Army Symposium on Solid Mechanics, Oct. 16-18, 1996, Myrtle Beach, SC.
8. Young, C.W., Simplified Analytical Model of Penetration with Lateral Loading (SAMPLL), Sandia National Laboratories, Albuquerque, NM, Report SAND 88-0013, 1988.
9. Johnson, G.R., Stryk, R.A., Holmquist, T.J. and Beissel, S.R., Numerical Algorithms in a Lagrangian Hydrocode, Final Report, WL-TR-1977-7039, Wright Laboratories, Armament Directorate, Eglin Air Force Base, 1997.

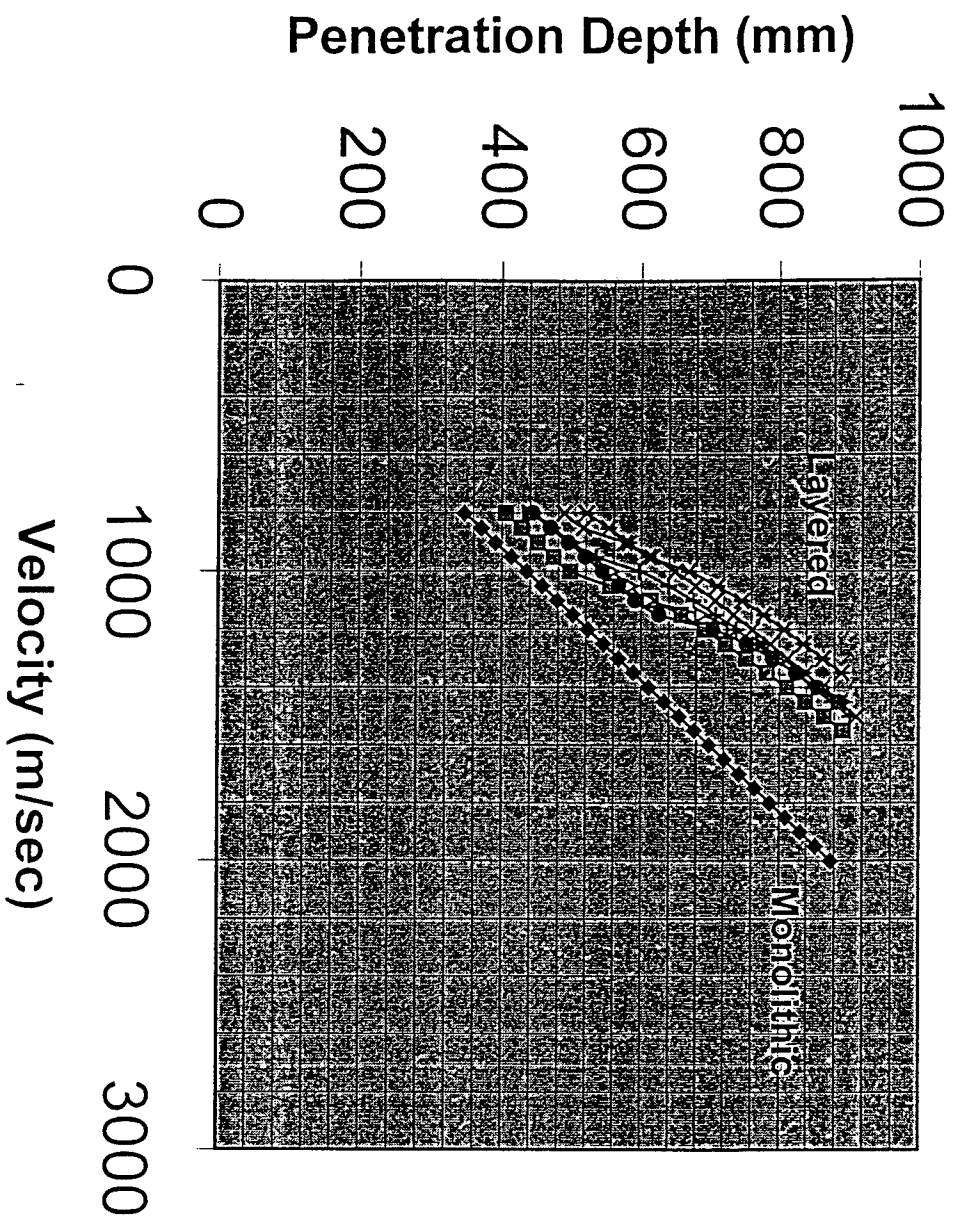


$L =$ approximately 50.8 mm (adjusted to obtain $m=65$ g)



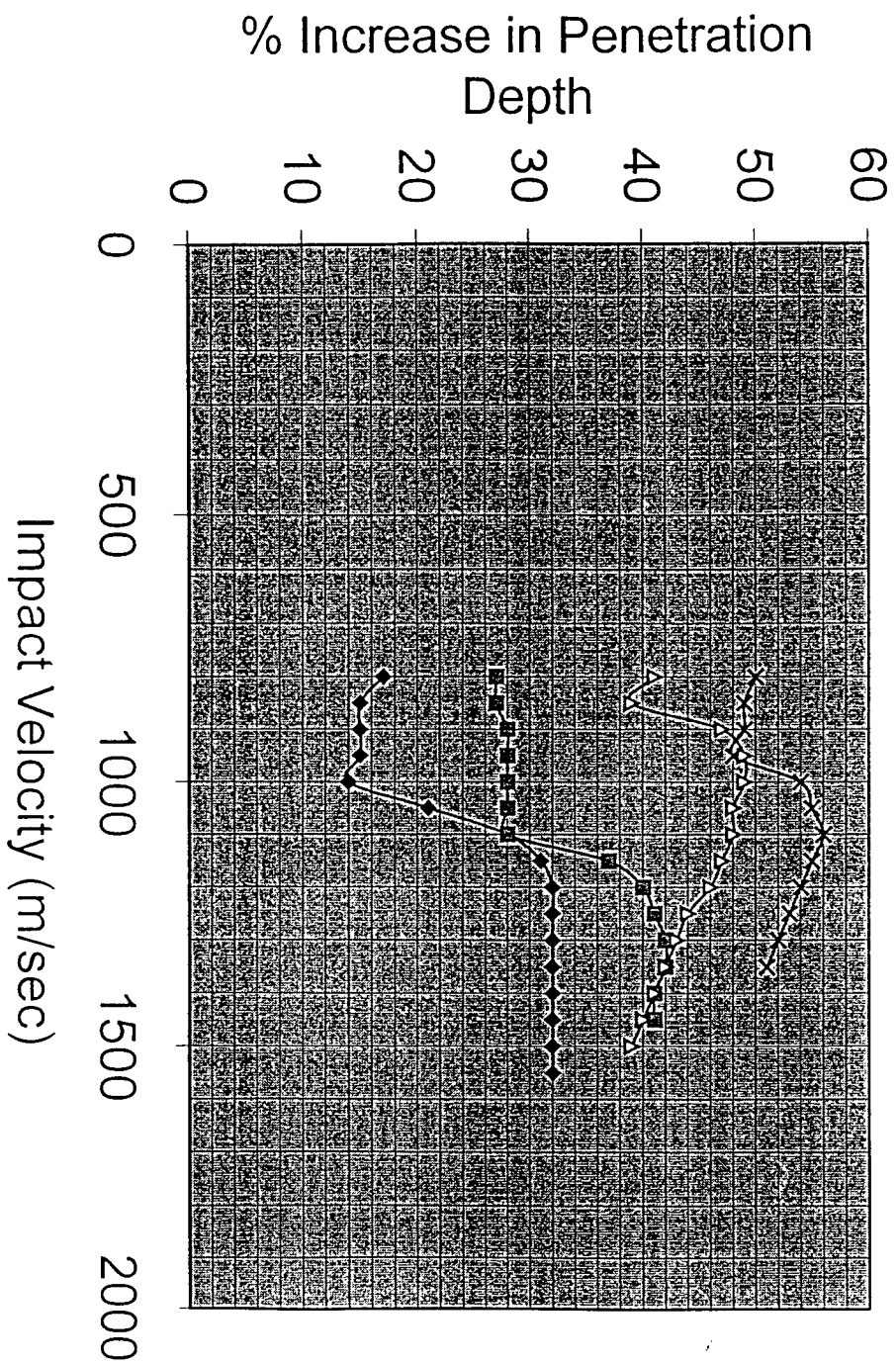
Internal cavity variations

Figure 1. Small scale projectile.



- ◆ Monolithic: 30 in dia. X 36 in lg.
- Layered: 4 in - 16 in - 16 in
- △ Layered: 12 in - 12 in - 12 in
- × Layered: 4 in - 8 in - 8 in - 16 in
- * Layered: 4 in - 4 in - 8 in - 8 in - 12 in
- Layered: 8 in - 16 in - 12 in
- + Layered 8 in - 12 in - 16 in

Figure 2. Penetration curves for layered grout targets.



- ◆ Layer Config. 1
- Layer Config. 2
- ▲ Layer Config. 3
- × Layer Config. 4

Figure 3. Increase in penetration depth in layered targets.

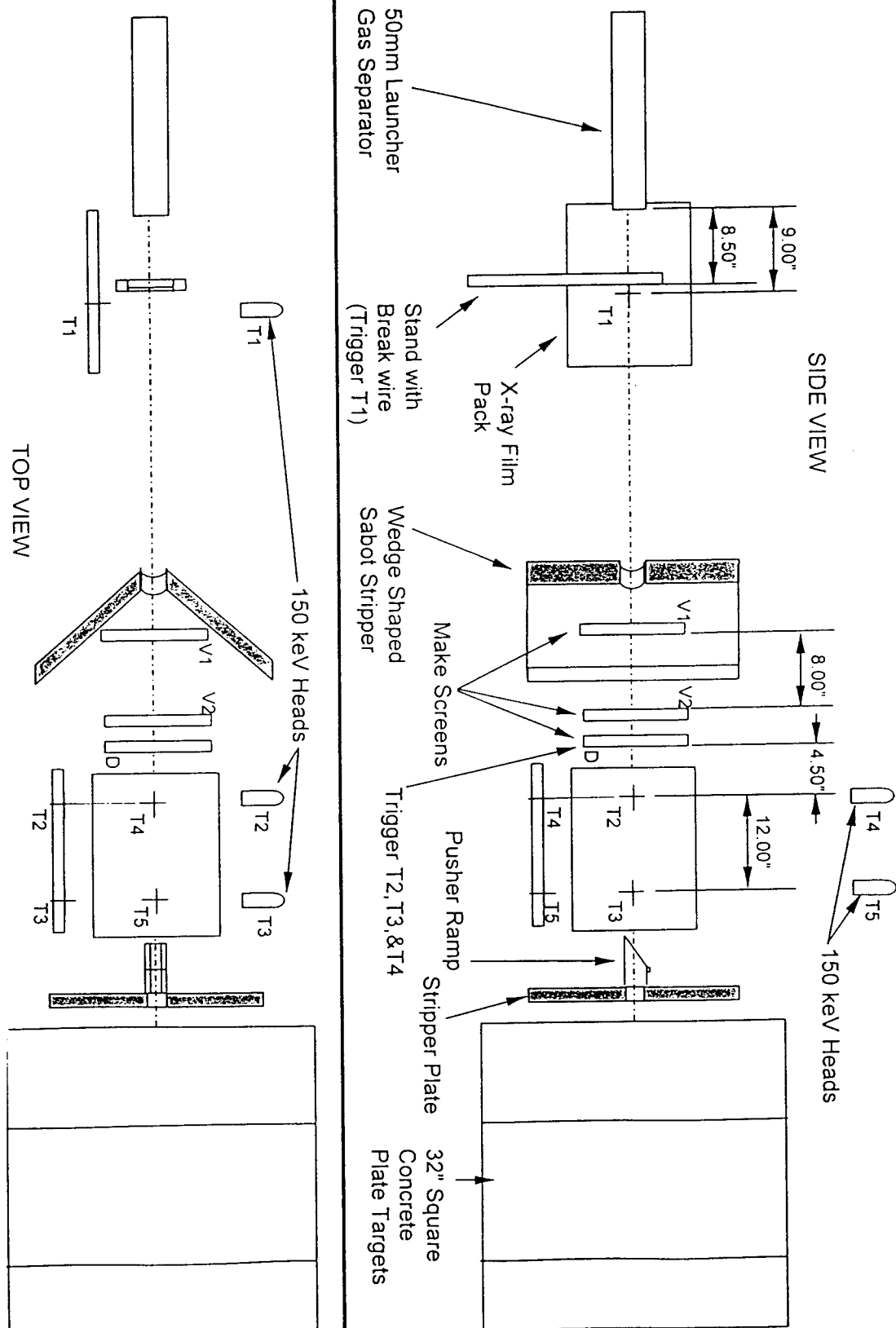


Figure 4. Schematic of 50 mm HVCP set-up.



Figure 5. Typical layered target configuration (LC1-4/16/16).



Figure 6. HVCP-41 (AF-1410 @ 1200 m/s).

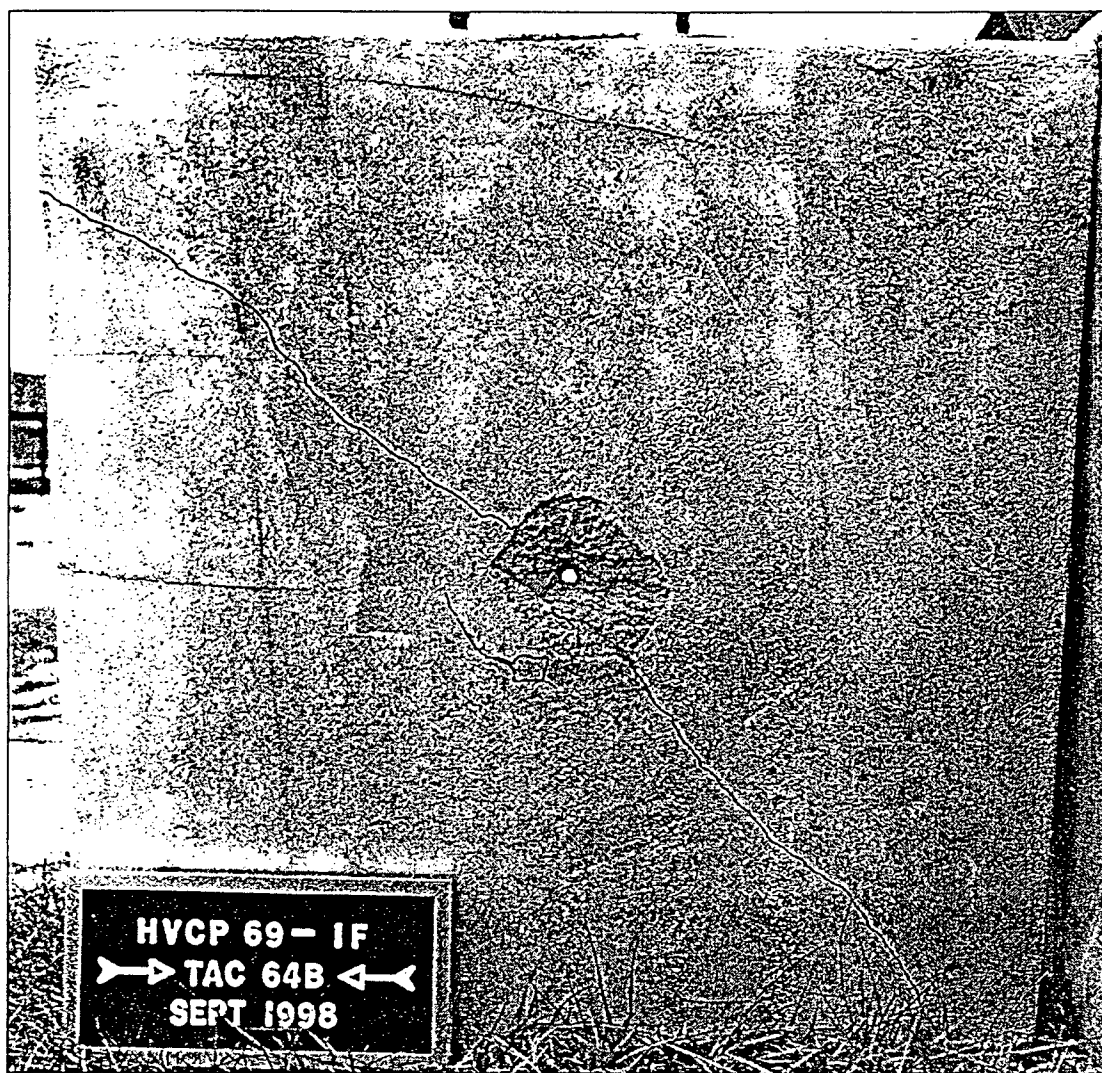


Figure 7. HVCP-69 1f (4340-1 @ 1200 m/s into 4/16/16).

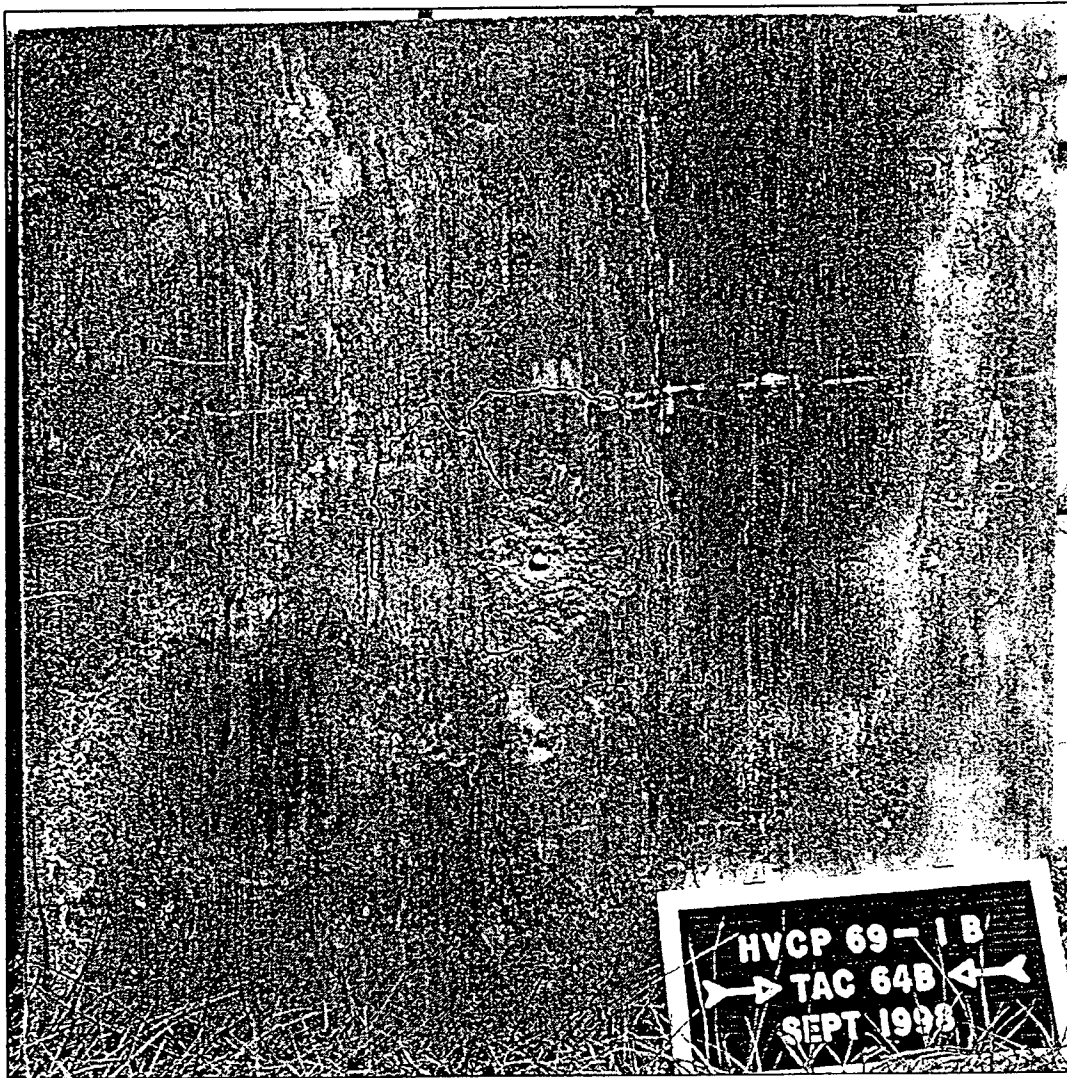


Figure 8. HVCP-69 1B.

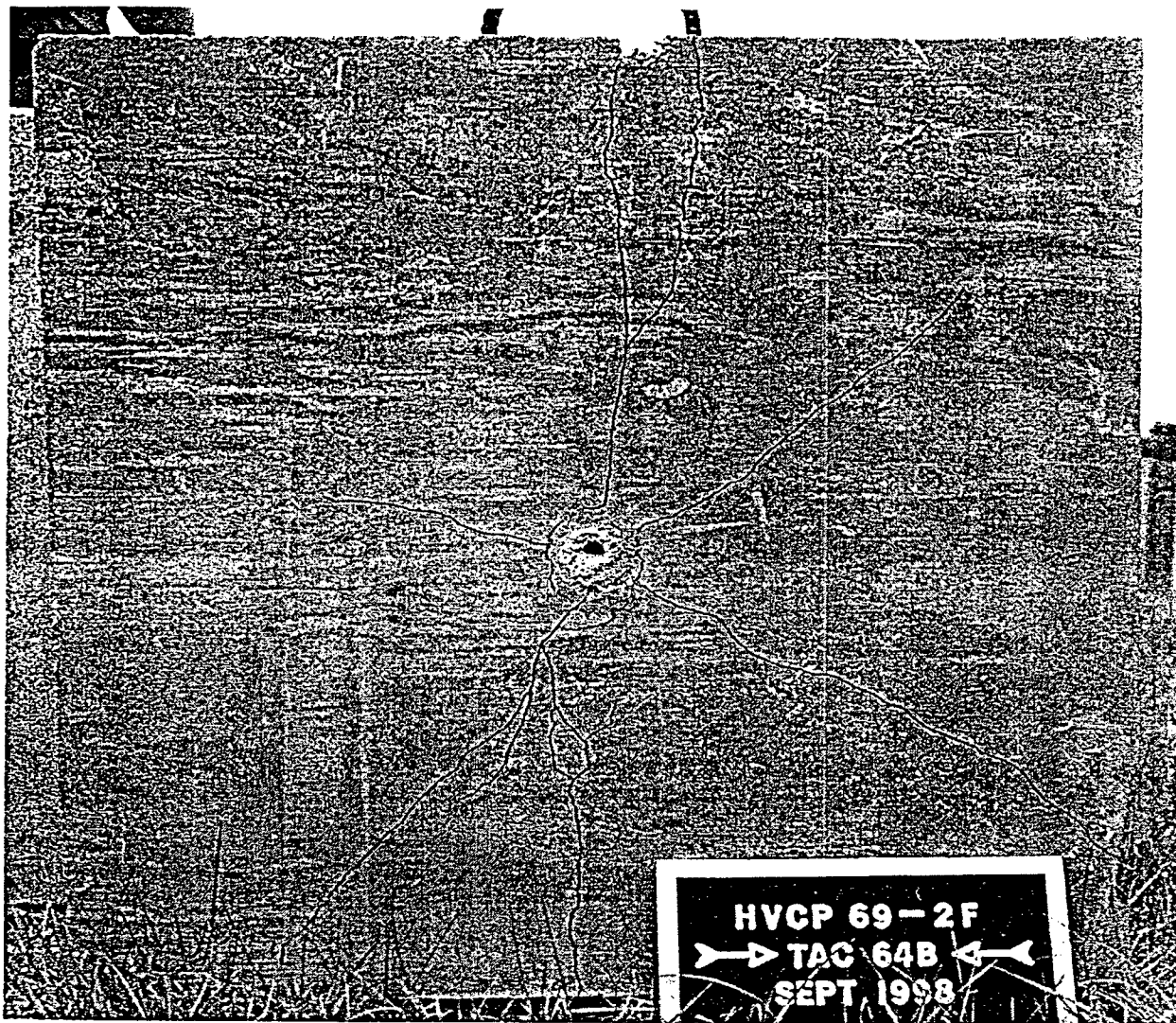


Figure 9. HVCP - 69 2F.

ENHANCEMENTS TO A DIRECT AEROELASTIC STABILITY COMPUTATIONAL MODEL

J. Mitch Wolff
Assistant Professor

Department of Mechanical and Materials Engineering

Wright State University
3640 Colonel Glenn Hwy.
Dayton, Ohio 45435-0001
Ph: (937) 775-5040
Fax: (937) 775-5009

Final Report for:
Graduate Student Research Program
Summer Faculty Research Program
Wright Laboratories

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

and

Wright Laboratories
Wright Patterson Air Force Base, OH

December 1998

ENHANCEMENTS TO A DIRECT AEROELASTIC STABILITY COMPUTATIONAL MODEL

J. Mitch Wolff
Assistant Professor

Department of Mechanical and Materials Engineering
Wright State University

Abstract

The ability to accurately and efficiently predict transonic flutter boundaries is investigated using an enhanced direct computational method. Steady characteristic and unsteady approximate non-reflecting characteristic farfield boundary conditions are utilized to more accurately model the aerodynamic flow physics in a direct method. In addition, the aerodynamic model is modified to lock the movement of the farfield grid points while allowing the airfoil surface points to move freely. This is accomplished by introducing a linear weighting function in the grid deformation model. The direct method is based on a discretization of the Euler equations and a coupled set of structural dynamics equations representative of a pitch-and-plunge airfoil with trailing edge flap. The coupled equations are expanded to specify a Hopf-bifurcation point, which defines an incipient flutter state. Since the enhanced farfield boundary conditions improve the modeling of the flow physics, the computational domains can be reduced in size over similar models with simple freestream conditions. A flapped NACA 64A006 airfoil, executing pitching and plunging motion, is utilized to demonstrate the ability of the enhanced direct method to accurately and efficiently calculate flutter boundaries for reduced domain sizes. Both zero and non-zero angle of attack results are shown to highlight the improved accuracy and efficiency of the enhanced model. Each model modification resulted in analysis improvements, with the steady characteristic model demonstrating significant improvements in the nonlinear flow regime. For a 1° static pre-twist analysis at a transonic freestream Mach number of 0.84, the enhanced model resulted in over a 75% decrease in the flutter speed error. Therefore, the capability to more accurately and efficiently model transonic flow conditions with strong shock interaction has been shown.

ENHANCEMENTS TO A DIRECT AEROELASTIC STABILITY COMPUTATIONAL MODEL

Tim J. Leger
J. Mitch Wolff

Introduction

In the design of flexible aircraft structures, it is important to determine accurately and efficiently the flutter boundaries within which aircraft can safely operate. Using conventional computational fluid dynamics (CFD) methods, these boundaries are estimated through time integration of the governing dynamical (fluid and structure) equations. In a design environment, where a large number of configurations are examined, time integration with CFD techniques can be unsystematic and prohibitively inefficient. However, efficient aerodynamic methods currently used within aeroelastic design procedures are typically based on linear aerodynamics and, thus, are not well suited to transonic flutter computations. There is a need for a robust method of flutter prediction of sufficiently high fidelity to capture properly the nonlinear aerodynamics characteristic of transonic flight, but with sufficient efficiency to bring nonlinear aerodynamics to the design environment.

Recently, direct, or non-time-integration, methods have been developed that are suitable for both flutter prediction and steady aeroelastic analysis. Morton and Beran (1995,1996A) developed, implemented and validated the first direct method for the calculation of flutter points in 2-D transonic flow. The implicit methodology provides the ability to compute the critical speed, as an unknown of an expanded system of equations, at which the equilibrium fluid-structure system loses stability to periodic motions. For a soft flutter condition, the critical speed is the speed at which limit-cycle oscillation (LCO) is first possible, and becomes critical at a supercritical Hopf-bifurcation point.

There are advantages and disadvantages of the use of direct methods for flutter prediction relative to that of time-integration methods. The advantages are primarily three-fold: (1) direct methods avoid the need for bracketing, i.e., the repetitious application of a time-integration method to obtain stable and unstable equilibrium solutions about a presumed critical state, (2) the computational effort of direct methods is not time-step sensitive, whereas the computational effort of time integration is inversely proportional to the time step, (3) direct methods obtain an explicit representation of the aeroelastic mode that is associated with the loss of stability in the system. The primary disadvantage of direct methods to date has been the fully implicit and computationally expensive solution strategies employed to compute simultaneously the aeroelastic solution, critical aeroelastic mode, flutter speed, and flutter frequency. Still, for grids typical of inviscid, 2-D,

transonic airfoil calculations (2000-3000 grid points), Beran (1998) observed significant improvements (a factor of 50) in computational performance of a direct approach over an explicit, total-variation diminishing, time-integration scheme.

Direct methods based on bifurcation analysis have been improved and applied in several recent studies. Beran and Morton (1997) added a solution-path tracing capability to the direct solver called BIFAE to allow flutter boundaries to be automatically computed, and made numerous fundamental improvements to their algorithmic procedure. Morton and Beran (1996B) studied the flutter characteristics of an airfoil with bi-linear structural coupling, including freeplay, and later (1997) implemented a deforming mesh strategy and investigated the effect of a trailing-edge flap on airfoil flutter speed. Beran (1998) developed a new direct methodology within BIFAE that allowed robust computation of flutter states throughout the transonic regime and at angle-of-attack, and adopted a block-relaxation strategy that reduced computational work and storage requirements, especially for large grids. Most recently, Lee, Jiang, and Wong (1998) applied a similar bifurcation analysis to a pitch-and-plunge airfoil with a cubic nonlinear restoring force. In this work, the aerodynamics was modeled (i.e., not computed with CFD), allowing a system of reduced order to be studied. They also applied asymptotic theory to the computation of LCO amplitude for the supercritical, hard-spring case.

Beran and Morton (1997) noted flutter-speed sensitivity to domain size in the transonic regime, which resulted from the following: growth of the lateral length scale in the transonic regime; the relatively small domains used in their analysis (6 to 12 airfoil chords), and the enforcement of far-field conditions based on freestream conditions and extrapolation. To minimize the sensitivity in the transonic regime, a corrective action is required. Grid insensitivity could be achieved through domain enlargement. However, owing to the current use of a fully implicit solution methodology, direct flutter analysis with computational grids of increased size incurs significant performance penalties (these penalties are reduced with a domain decomposition procedure, but not eliminated). A more reasonable approach to reducing grid sensitivity is to improve the analytical formulation of the farfield boundary conditions.

One of the most important factors determining the success or failure of a CFD code is the boundary conditions. The far field boundaries are usually set to freestream values. Although appropriate, these boundary conditions require domain enlargement as previously discussed. Theoretically, a more appropriate boundary condition for the far field boundary is the characteristic variable method (Janus (1984)), which is consistent with the concept of upwinding. Therefore, characteristic variable boundary conditions are implemented giving a more efficient smaller computational domain. In addition, moving airfoils are known to generate unsteady disturbances, which propagate from the airfoil. These propagating waves must not be reflected back into the solution domain by the farfield boundary conditions. This problem is avoided by utilizing

approximate non-reflecting boundary conditions for the unsteady aerodynamic analysis (Giles (1990)).

The objective of this research is to implement and verify improvements obtained by utilizing higher fidelity farfield boundary conditions. Characteristic variable and approximate non-reflecting farfield boundary conditions for steady and unsteady flows, respectively, will be implemented in the direct BIFAE aeroelastic model. A NACA 64A006 flapped airfoil executing pitching and plunging motion will be utilized for both zero and non-zero static pretwist configurations to demonstrate the accuracy and efficiency improvements. In particular, a nonlinear transonic flow configuration will be investigated to highlight improvements in the transonic analysis capability of the enhanced model.

Methodology

In this section, the mathematical model for a pitch-and-plunge airfoil in inviscid flow is briefly summarized, and the analytical conditions satisfied at flutter points provided. Further details are given in the references cited above.

The rigid, pitch-and-plunge airfoil with flap is constrained to plunge along a vertical line and is allowed to pitch about an elastic axis. See Figure 1. Linear and torsional springs, with specified stiffness and damping constants (K_α , K_h , K_β , D_α , D_h , D_β), mediate the motion of the airfoil according to three linear, non-homogeneous, second-order ordinary differential equations for pitch (represented by the angle of attack, α), plunge (represented by the vertical displacement, h), and flap deflection, β . These equations are expressed as six, first-order differential equations, and are a function of the reduced velocity, $u = v_\infty / (b\omega_\alpha)$, where v_∞ is the freestream velocity, b is the semi-chord, and ω_α is the pitch natural frequency.

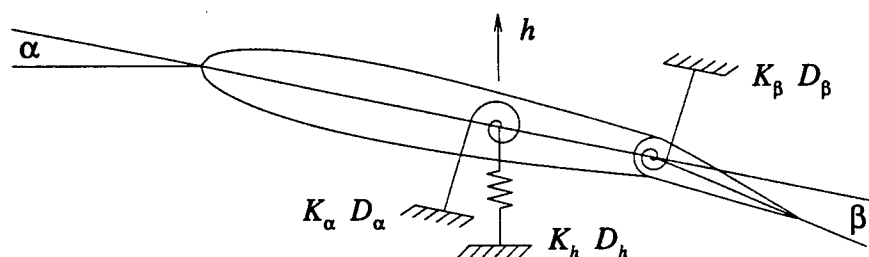


Figure 1: Schematic pitch-and-plunge airfoil with flap.

The governing aerodynamic equations are the Euler equations, cast in strong-conservative form for a general curvilinear coordinate system. These equations are placed in discrete form using a moving grid-point form of a standard upwind total-variation diminishing (TVD) scheme (Morton and Beran (1996A)). The systems of discretized fluid-dynamic equations and structural equations are combined into a single time-dependent system,

$$Y_t = G(y; u) \quad (1)$$

where Y is an array containing all the aerodynamic and structural dynamic variables, and G is an array of all the model equations. Equilibrium solutions of Equation 1 satisfy the nonlinear system $G(y; u) = 0$, and are computed with Newton's method. In this computation, the Jacobian G_Y is calculated numerically to second-order accuracy and fully stored. The structure of G_Y is that of a bordered, banded matrix.

The aeroelastic system governed by Equation 1 is also simulated through time integration following the approach detailed by Beran and Morton (1997). Using this approach, validation of stable equilibrium solutions obtained with Newton's method can be obtained, as well as validation of computed flutter points.

The equations representative of a Hopf-bifurcation, or flutter, point are developed by applying an assumed form of the solution Y . At a Hopf point, a steady-state solution transitions to an oscillatory solution with zero amplitude. The Hopf-point solution is assumed to be of the form, $Y(t; u) = Y_0(u) + \varepsilon P e^{\beta t}$, where ε is some small parameter and P is a coefficient vector. Substitution of this assumed form into Equation 1 yields the eigen-problem

$$G_Y P = \beta P \quad (2)$$

The Hopf-point condition is satisfied when the real part of β vanishes for some complex pair of eigenvalues:

$$\beta = i\Theta \quad (3)$$

where i is the imaginary number and Θ is the flutter frequency. With Eqn. 2 and Eqn. 3, the discrete equations $G = 0$ are expanded to yield a system of equations, approximately three times the size of the original discrete system, for the critical aeroelastic state, Y^* , the critical aeroelastic mode, P^* , the flutter velocity, u^* , and the flutter frequency, Θ^* :

$$F(x) = \begin{bmatrix} G \\ (G_Y - i\Theta)P \\ q^T P - I \end{bmatrix} = 0, \quad X = \begin{bmatrix} Y \\ P \\ u \\ \Theta \end{bmatrix} \quad (4)$$

with q defined as a user specified constant array. The solution of this expanded system within BIFAE is described in detail by Beran (1998).

Boundary Conditions

Theoretically speaking the characteristic variable boundary conditions should more accurately model nature due to the ability to pass information into and out of the computational domain. In order to correctly implement these characteristic variable boundary conditions however, the outer boundary of the computational domain must remain fixed in space. To accomplish this while allowing the airfoil to move, a weighting function grid deformation technique is employed.

Grid Deformation

In the deforming grid technique, the outer boundary remains fixed in space, with a deforming grid method used to locate the position of the airfoil. The grid lines connecting the inner and outer boundaries deform with the airfoil motion. Figure 2 shows a reduced computational grid with an exaggerated airfoil motion, which highlights the deforming grid method. The amount of deformation is a function of the distance from the airfoil surfaces. The grid deformation is defined as

$$\Delta x_{i,j} = W_{i,j}(\Delta x_{i,j}) \quad (5a)$$

$$\Delta y_{i,j} = W_{i,j}(\Delta y_{i,j}) \quad (5b)$$

where $\Delta x_{i,j}$ and $\Delta y_{i,j}$ are the spatial differences that would exist between successive time steps if the entire grid were to move as a rigid body. The weighting function, W , is defined as

$$W_{i,j} = W(\xi, \eta) = \left| \frac{s(\xi, \eta)}{s(\xi, \eta_{\max})} - 1 \right| \quad (6)$$

where s is the arc length of a grid line from the airfoil surface, $\eta = 1$, to some grid point along $\xi = \text{constant}$, and $\eta_{\max}$ is the outer boundary grid line as shown in Figure 2.

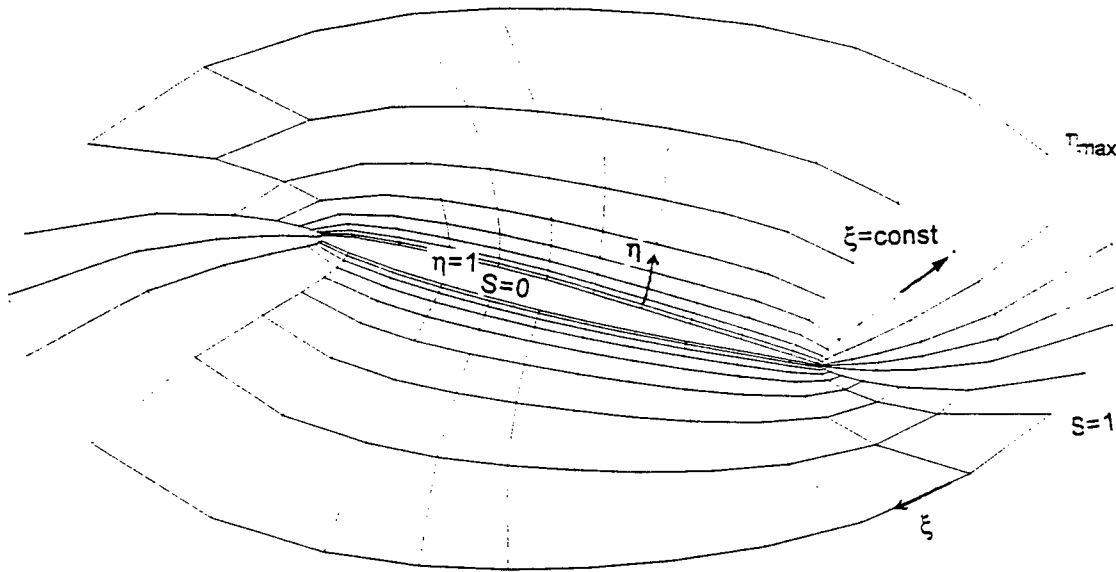


Figure 2: Deforming grid geometry.

From Equations 5 and 6, the nodes at the inner boundary, $s = 0$, give $W_{i,j} = 1$, which means the airfoil surface is deformed completely, as a rigid body motion. Conversely, the outer boundary nodes give $W_{i,j} = 0$, which signifies that these nodes remain fixed in space at their initial specified locations. The interior nodes shear in space relative to the initial grid as $W_{i,j}$ varies from 0 to 1. The grid speeds are then calculated by dividing the grid deformation by the time step used.

Steady State – Janus Characteristic Variable Method

The steady state characteristic variable boundary conditions are consistent with the concept of upwinding in which the signs of the characteristic velocities determine the appropriate propagation directions (Janus, 1984). A summary of their derivation is given in the following. The Euler equations written in their non-conservative form are

$$q_\tau + aq_\xi + bq_\eta = 0 \quad (7)$$

with the matrices a and b determined through an eigenvalue analysis.

The far field boundary conditions are obtained by multiplying Equation 7 by P_η^{-1} and neglecting the derivatives in the normal ξ direction to give

$$P_\eta^{-1} q_\tau + P_\eta^{-1} P_\eta \Lambda_\eta P_\eta^{-1} q_\eta = 0 \quad (8)$$

where Λ_η is a diagonal matrix containing the eigenvalues, λ_η , and P_η and P_η^{-1} are the left and right eigenvectors, respectively.

The characteristic vector is defined as

$$W_\eta = P_\eta^{-1} q \quad (9)$$

P_η is such that the elements of the characteristic vector become

$$w_\eta^1 = \frac{J}{|\nabla \eta|} \left[\eta, \left(\rho - \frac{P}{c_1^2} \right) \right] \quad (10a)$$

$$w_\eta^2 = \frac{J}{|\nabla \eta|} \left[\eta, \left(\rho - \frac{P}{c_1^2} \right) \right] \quad (10b)$$

$$w_\eta^3 = \frac{J}{\sqrt{2}|\nabla \eta|} \left[\frac{P \nabla \eta}{\rho, c_1} + (\eta, u + \eta, v) \right] \quad (10c)$$

$$w_\eta^4 = \frac{J}{\sqrt{2}|\nabla \eta|} \left[\frac{P \nabla \eta}{\rho, c_1} - (\eta, u + \eta, v) \right] \quad (10d)$$

The corresponding eigenvalues are

$$\lambda_\eta^1, \lambda_\eta^2 = \eta, u + \eta, v \quad (11a,b)$$

$$\lambda_\eta^3 = (\eta, u + \eta, v) + c|\nabla \eta| \quad (11c)$$

$$\lambda_\eta^4 = (\eta, u + \eta, v) - c|\nabla \eta| \quad (11d)$$

The implementation of the steady state far field characteristic variable boundary conditions is accomplished by computing the sign of λ to determine the directions of the characteristics. The subscript "a" denotes approaching the boundary, "b" refers to on the boundary, and "l" leaving the boundary. Then the characteristics approaching the boundary are set equal to the characteristics on the boundary using Equation 10.

$$\left[\eta_s \left(\rho - \frac{P}{c_1^2} \right) \right]_s = \left[\eta_s \left(\rho - \frac{P}{c_1^2} \right) \right]_s \quad (12a)$$

$$\left[\eta_s \left(\rho - \frac{P}{c_1^2} \right) \right]_s = \left[\eta_s \left(\rho - \frac{P}{c_1^2} \right) \right]_s \quad (12b)$$

$$\left[\frac{P \nabla \eta}{\rho_1 c_1} + \text{sign}(\lambda_s') (\eta_s u + \eta_s v) \right]_s = \left[\frac{P \nabla \eta}{\rho_1 c_1} + \text{sign}(\lambda_s') (\eta_s u + \eta_s v) \right]_s \quad (12c)$$

$$\left[\frac{P \nabla \eta}{\rho_1 c_1} + \text{sign}(\lambda_s') (\eta_s u + \eta_s v) \right]_s = \left[\frac{P \nabla \eta}{\rho_1 c_1} + \text{sign}(\lambda_s') (\eta_s u + \eta_s v) \right]_s \quad (12d)$$

The metrics at points "a" and "l" are assumed to be the same as those at point "b". Combining Equations 12(c,d) gives Equation 13a, with the remaining flow variables at the far field boundary found by simultaneous solution of Equations 12(a-d).

$$P_s = \frac{1}{2} \left(P_s + P_l - \rho_1 c_1 \left[\eta_s (u_s - u_l) + \eta_s (v_s - v_l) \right] \right) \quad (13a)$$

$$\rho_s = \rho_s + \frac{P_s - P_l}{c_1^2} \quad (13b)$$

$$u_b = u_a - \eta_x \frac{P_a - P_b}{\rho_1 c_1} \quad (13c)$$

$$v_b = v_a - \eta_y \frac{P_a - P_b}{\rho_1 c_1} \quad (13d)$$

Unsteady Flows – Giles Approximate Non-Reflecting Characteristic Method

Approximate non-reflecting unsteady farfield boundary conditions are developed by assuming that linear theory can be applied. Giles (1990) derived "non-reflecting" boundary conditions for a general turbomachinery Euler solver. First the steady flow is solved using the previously presented characteristic variable boundary conditions. The linearized Euler equations are then solved at the farfield boundary to determine the perturbation flow variables in terms of the characteristic variables. This allows time variations of static pressure at the farfield and reduces reflections from the boundaries. A brief outline of this method is given below.

The linearized, two-dimensional Euler equations are written in terms of primitive small-perturbation variables as

$$Q' + A Q' + B Q' = 0 \quad (14)$$

where

$$Q' = (\delta p, \delta u, \delta v, \delta p)^T \quad (15)$$

$$A = \begin{bmatrix} u & \rho & 0 & 0 \\ 0 & u & 0 & \frac{1}{P} \\ 0 & 0 & u & 0 \\ 0 & \gamma P & 0 & u \end{bmatrix}, \quad B = \begin{bmatrix} v & 0 & \rho & 0 \\ 0 & v & 0 & 0 \\ 0 & 0 & v & \frac{1}{P} \\ 0 & 0 & \gamma P & v \end{bmatrix} \quad (17)$$

The elements of the vector Q' represent perturbations from uniform flow conditions

$$\rho = \rho_0 + \delta\rho \quad (18a)$$

$$u = u_0 + \delta u \quad (18b)$$

$$v = v_0 + \delta v \quad (18c)$$

$$P = P_0 + \delta P \quad (18d)$$

where ρ_0 , u_0 , v_0 , and P_0 specify the steady solution.

The matrices A and B are evaluated using these same conditions. The analysis is greatly simplified if the unsteady perturbations and the steady variables in A and B are nondimensionalized using the steady density and speed of sound. With this choice of nondimensionalization, the final forms of the matrices A and B are

$$A = \begin{bmatrix} M_x & 1 & 0 & 0 \\ 0 & M_x & 0 & 1 \\ 0 & 0 & M_y & 0 \\ 0 & 1 & 0 & M_x \end{bmatrix}, \quad B = \begin{bmatrix} M_x & 0 & 1 & 0 \\ 0 & M_x & 0 & 0 \\ 0 & 0 & M_y & 1 \\ 0 & 0 & 1 & M_y \end{bmatrix} \quad (19)$$

where M_x and M_y are the Mach numbers in the x and y directions.

A set of eigenvectors representing an entropy wave, a vorticity wave and an upstream/downstream running pressure waves are determined by Fourier analysis of these equations. Using these eigenvectors and assuming locally one-dimensional flow at the boundary, the characteristic variables for unsteady flows are written in terms of the perturbation variables as

$$\begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \delta\rho \\ \delta u \\ \delta v \\ \delta P \end{bmatrix} \quad (20)$$

For a subsonic freestream flow, the amplitudes of the incoming unsteady characteristics (C_1 , C_2 , C_3) are set to zero and the outgoing characteristic (C_4) is computed using Equation 20. For subsonic downstream flow, $C_4=0$ and the remaining characteristics are computed. Once the characteristics are known, the perturbation variables are found using an inverse transform

$$\begin{bmatrix} \delta\rho \\ \delta u \\ \delta v \\ \delta P \end{bmatrix} = \begin{bmatrix} -1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix} \quad (21)$$

The primitive flow variables are found using Equation 18. With the non-conservative flow variables found, the conservative dependent variables are easily computed for the farfield boundaries.

Results

The direct computational procedure BIFAE has been successfully validated against other inviscid methods, including Euler and transonic small-disturbance methods (Morton and Beran (1997) and Beran and Morton (1997)). Validation of BIFAE was also the subject of work by Buxton and Beran (1997). In this report, both equilibrium and flutter point solutions are obtained for these

same configurations to validate the new farfield boundary condition modifications. The sensitivity of these new boundary conditions to domain size is examined through the use of flutter boundaries and Cp distributions. Results were obtained using a standardized set of structural and aerodynamic parameters for a baseline configuration of the BIFAE algorithm.

NACA 64A006 Flapped Airfoil Pitching and Plunging

The enhanced Hopf-bifurcation analysis is applied to a NACA 64A006 flapped airfoil with pitching and plunging motion for various computational grids and aerodynamic parameters. The structural parameters were not varied in this research, because the improvements to the analysis were related to the aerodynamic model. Therefore, only variations in the aerodynamic parameters were required to demonstrate the successful implementation of the improved flow physics model.

The structural parameters utilized for all the computational results are presented in Table 1.

Symbol	Description	Value
x_{cg}	center of gravity (from LE)	0.5
ζ_α, ζ_h	airfoil damping ratios	0.0
ζ_β	flap damping ratio	0.01
ω_h/ω_α	natural frequency ratio	0.3
$\omega_\beta/\omega_\alpha$	flap natural frequency ratio	1.5
x_α	elastic-axis location	-0.2
μ	airfoil-fluid mass ratio	50

Table 1: NACA 64A006 Structural Parameter Values

To demonstrate the increased solution accuracy obtained by modifying the farfield boundary conditions, three different computational domains were utilized. Table 2 gives specific details about each computational grid used to model the NACA 64A006 airfoil.

Grid	Δ_x	Δ_y	Δ_w	R_{max}	I	J
G-1	0.0010	0.0005	0.0006	5	84	31
G-3	0.0010	0.0005	0.0006	50	84	43
G-4	0.0010	0.0005	0.0006	150	84	49

Table 2: NACA 64A006 Computational Grids

As shown in Table 2, the three computational grids used are the same grid except for the location of the farfield boundary. The G-1 grid has a domain size of 5 airfoil chords with the G-3 and G-4 grids having 50 and 150 chords domain sizes, respectively. Utilizing the same grids except for in the J direction, ensures that any difference in the predictions is caused by modifications to the farfield boundary condition or boundary location.

Since the direct hopf-bifurcation model is optimized for calculating flutter boundaries, then flutter boundary studies will be utilized to demonstrate the enhancements to the model accuracy obtained with the new farfield boundary conditions. These results are typically obtained by

establishing all the aerodynamic parameters, then incrementally modifying only one parameter. Traditionally, the parameter which is incremented is the freestream Mach number. The initial investigation of the enhanced analysis was accomplished by varying the freestream Mach number from 0.7 to 0.9 with a static pre-twist angle of attack of 0° . The continuation feature designed into the analysis was utilized for these investigations. After the flutter point was calculated for the 0.7 Mach case, the Mach number was incremented by 0.01 with the previous flutter point solution used to begin the solution. The analysis was continued until the solution did not converge on a flutter point for a particular Mach number.

Initially, obtaining a flutter boundary that is independent of the computational domain size was accomplished. Figures 3 & 4 shows the flutter boundary analysis of the flutter speed and reduced frequency variation with Mach number for the 84x43 and 84x49 grids, which have domain sizes of 50 and 150 chords, respectively. The analysis did not converge above a Mach number of 0.88. The flutter boundaries for both domains are exactly the same. Therefore, the baseline grid independent solution for the specified structural and aerodynamic parameters is known. The flutter bucket for this configuration occurs between 0.83 and 0.85 Mach number. The flutter bucket location is of primary interest when flutter boundaries are analyzed.

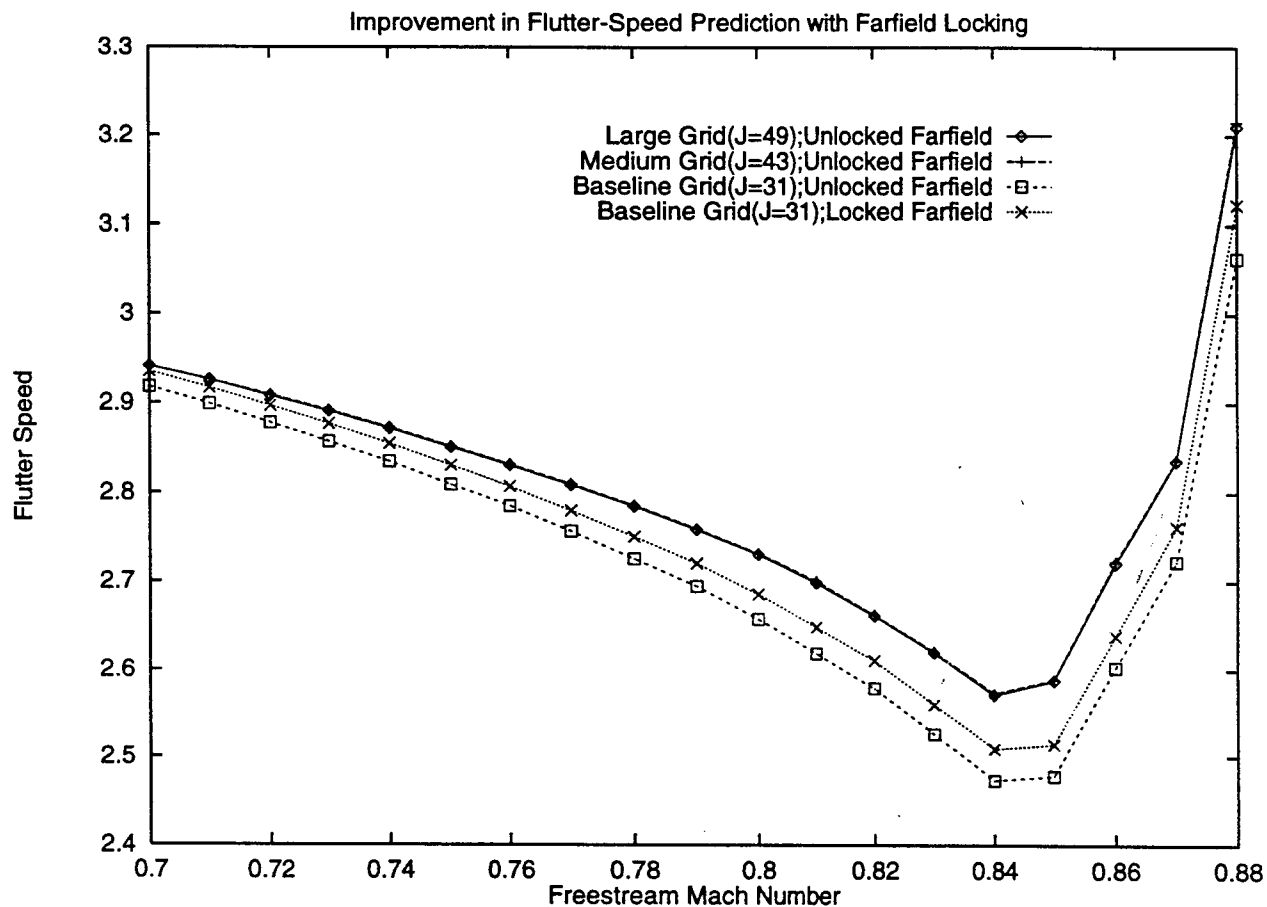


Figure 3: Reduced Velocity Flutter Boundary for a NACA 64A006 Airfoil with Flap

Included in Figures 3 & 4 is the unmodified BIFAE analysis using an 84x31 computational grid, which has a domain size of 5 chords. As shown in Figure 3, the 5 chord domain results under predict the flutter speed slightly at $M=0.7$ (0.8%), but the prediction is off by 3.8% at $M=0.84$, which is in the flutter bucket. With regards to the reduced frequency, shown in Figure 4, the unmodified model over predicts the correct frequency by 1.0 % for all Mach numbers. The difference in the model predictions just discussed are entirely attributed to the reduction in domain size from 50 chords to 5 chords. This means that the solution is dependent on the domain size of 5 chords for the farfield boundary.

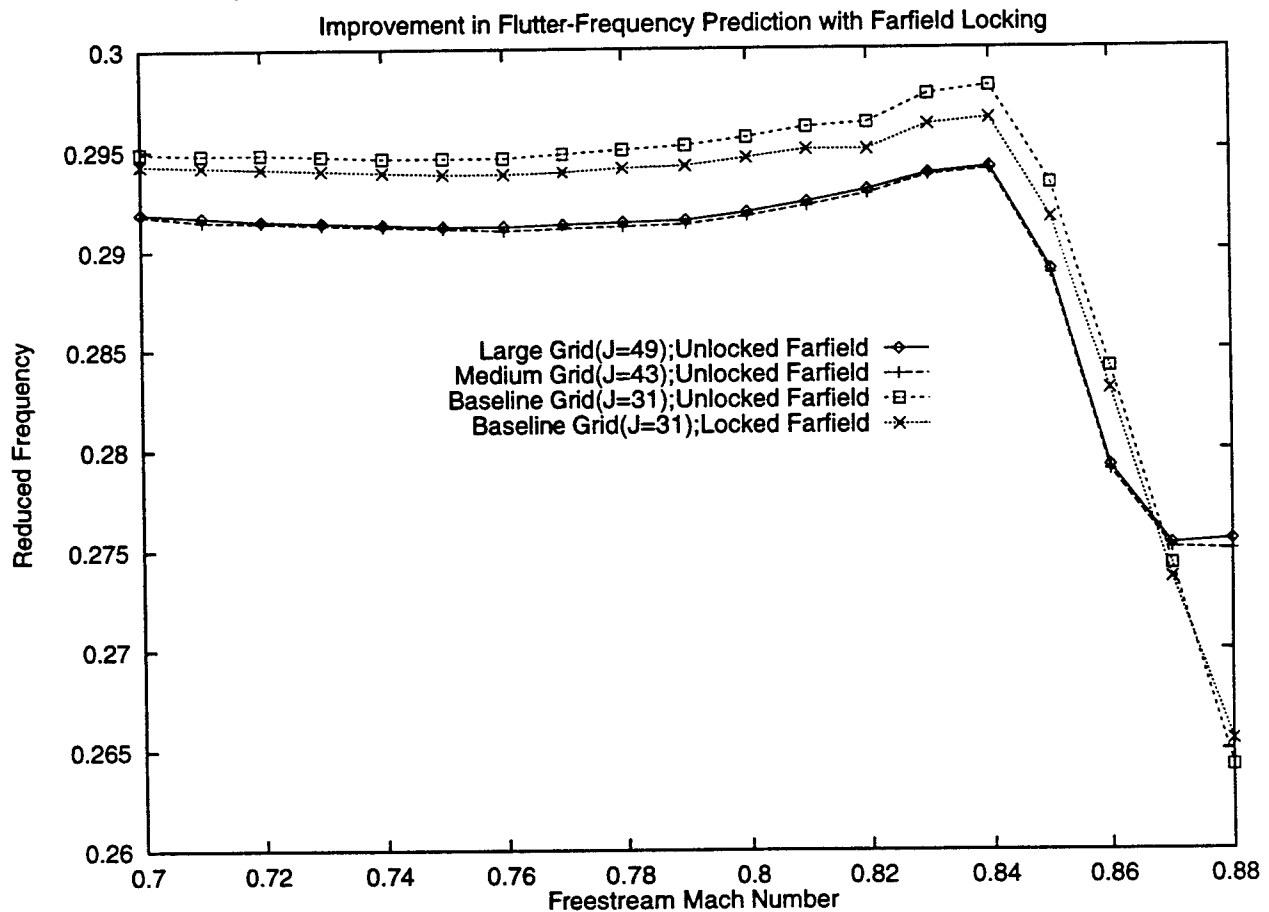


Figure 4: Reduced Frequency Flutter Boundary for a NACA 64A006 Airfoil with Flap

In addition, Figures 3 & 4 show the BIFAE analysis obtained with the grid deformation modification implemented. This modification locks the movement of the grid at the farfield boundary using a weighting function. As shown in Figure 3, the flutter speed prediction is improved at all Mach numbers. At $M=0.7$ the solution is off by 0.2% and at $M=0.84$ the reduced velocity predicted is off by 2.4%. Therefore, the percentage improvement increased with increasing Mach number. For the reduced frequency, Figure 4, the improvement was fairly uniform with Mach number, the predicted value was off by 0.8% from the grid independent analysis. These results show the improvement in the BIFAE model obtained by incorporating the grid deformation model.

The improvements in both the flutter speed and reduced frequency obtained by modifying the grid deformation model are more significant than expected.

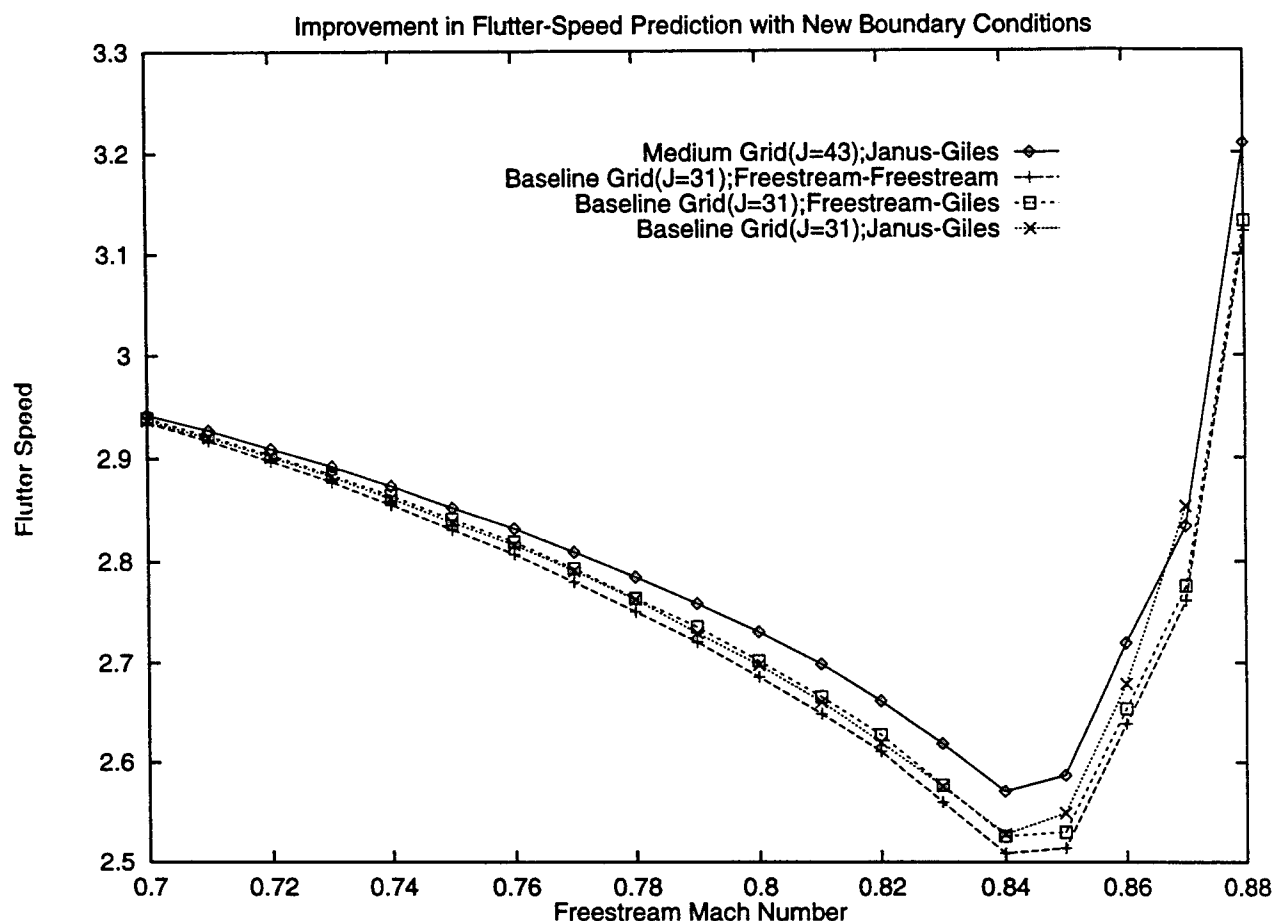


Figure 5: Reduced Velocity Flutter Boundary Effect of New Unsteady Boundary Conditions

To investigate the influence of improvements to the farfield boundary conditions, various combinations of the original freestream, the steady characteristic, and the unsteady approximate conditions are shown in Figures 5 and 6 for flutter speed and reduced frequency, respectively. All the results shown utilize the locked grid deformation model with the different farfield boundary condition. The grid independent result from a 50 chord domain analysis ($J=43$) is given as the reference solution for this configuration. Also, the locked original boundary condition solution shown in Figures 3 and 4 is given (i.e. 'Baseline Grid($J=31$);Freestream-Freestream'). At an inlet Mach number of 0.7 the flutter speed is only slightly off with the original boundary conditions, using both the steady characteristics of Janus and the unsteady approximate of Giles boundary conditions give a more accurate prediction, but the improvement is modest. As the Mach number is increased the improvement in the predictions of both flutter speed and reduced frequency from the original freestream to Janus-Giles conditions gets more significant. For the flutter speed prediction at $M=.85$, see Figure 5, utilizing the Giles unsteady conditions alone results in a 0.6% improvement while using both Janus's steady and Giles unsteady results in a 1.4% improvement in

flutter speed. While for the reduced frequency analysis, see Figure 6, the improvements were 0.2% and 0.8%, respectively. It is important to note, at a freestream Mach number of 0.84 the improvement from using Janus's steady boundary conditions is nearly zero in the flutter speed prediction (Figure 5), but is quite significant in terms of the reduced frequency prediction (Figure 6). These results warrant a more detailed investigation at this Mach number.

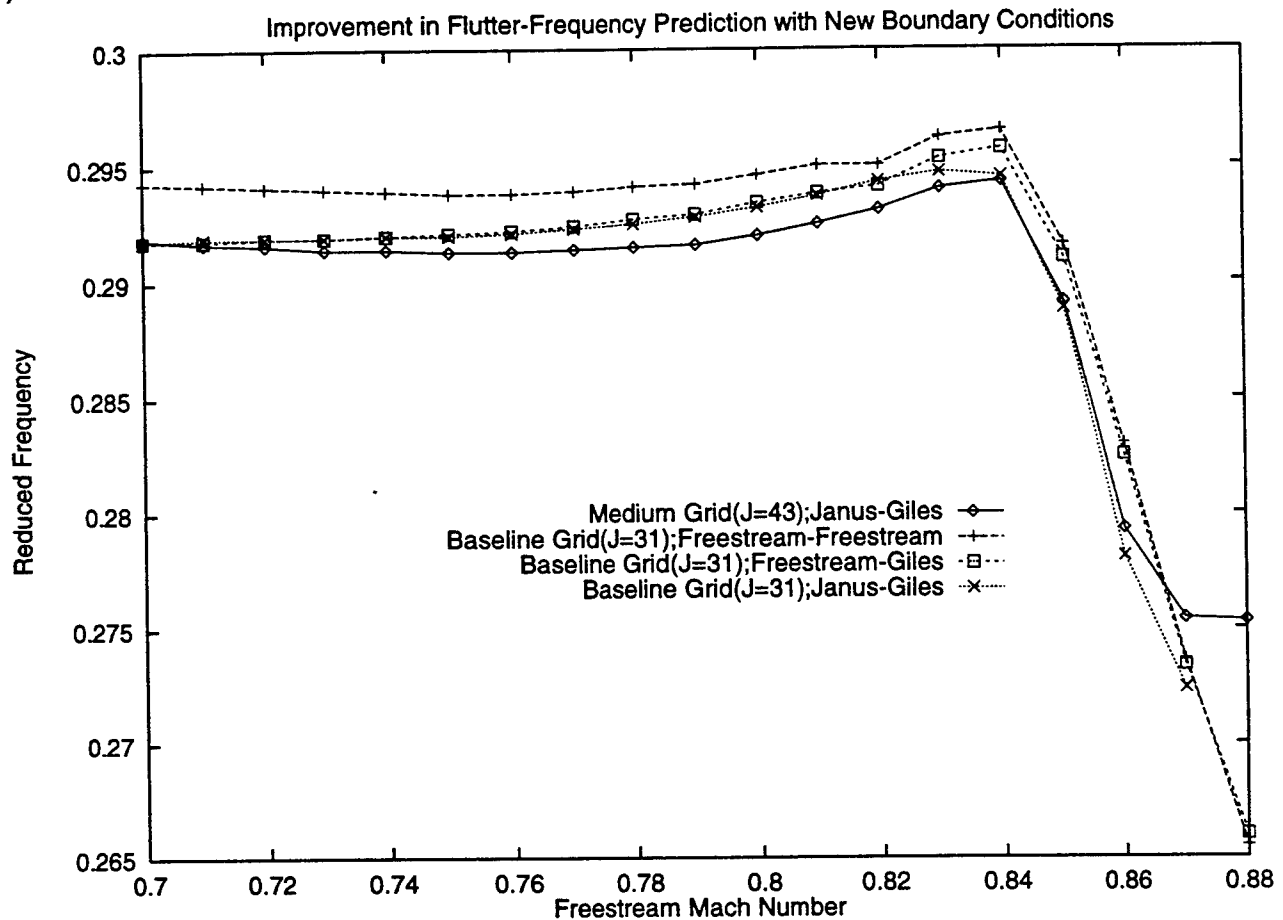


Figure 6: Reduced Frequency Flutter Boundary Effect of New Unsteady Boundary Conditions

Since the accurate prediction of the flutter bucket is of significant interest when doing flutter boundary studies, the significant change in solution accuracy from $M=0.84$ to $M=0.85$ was investigated in more detail. The only change between the results at $M=0.85$ and $M=0.84$ was in the steady farfield boundary conditions. At $M=0.85$ a significant effect was seen when Janus conditions were utilized but at $M=0.84$ no effect was shown. Therefore, the steady surface pressure distribution was investigated for differences as shown in Figure 7. It is quite apparent from Figure 7 that there is no effect on the steady pressure distribution no matter which farfield boundary conditions are used. Another interesting feature to note shown in Figure 7 is the lack of a shock wave. It is well known that the freestream boundary conditions have problems accurately modeling nonlinear flow phenomena. All the results given have been for zero static pre-twist. By increasing the airfoil angle of attack, a shock wave will form on the suction surface therefore, giving a better test of the

improved analysis model. Since the predictions at a freestream Mach number of 0.84 did not show a significant improvement in the flutter speed, the angle of attack is varied for this speed by increasing the static pre-twist angle in 0.1° increments until 1.0° is reached.

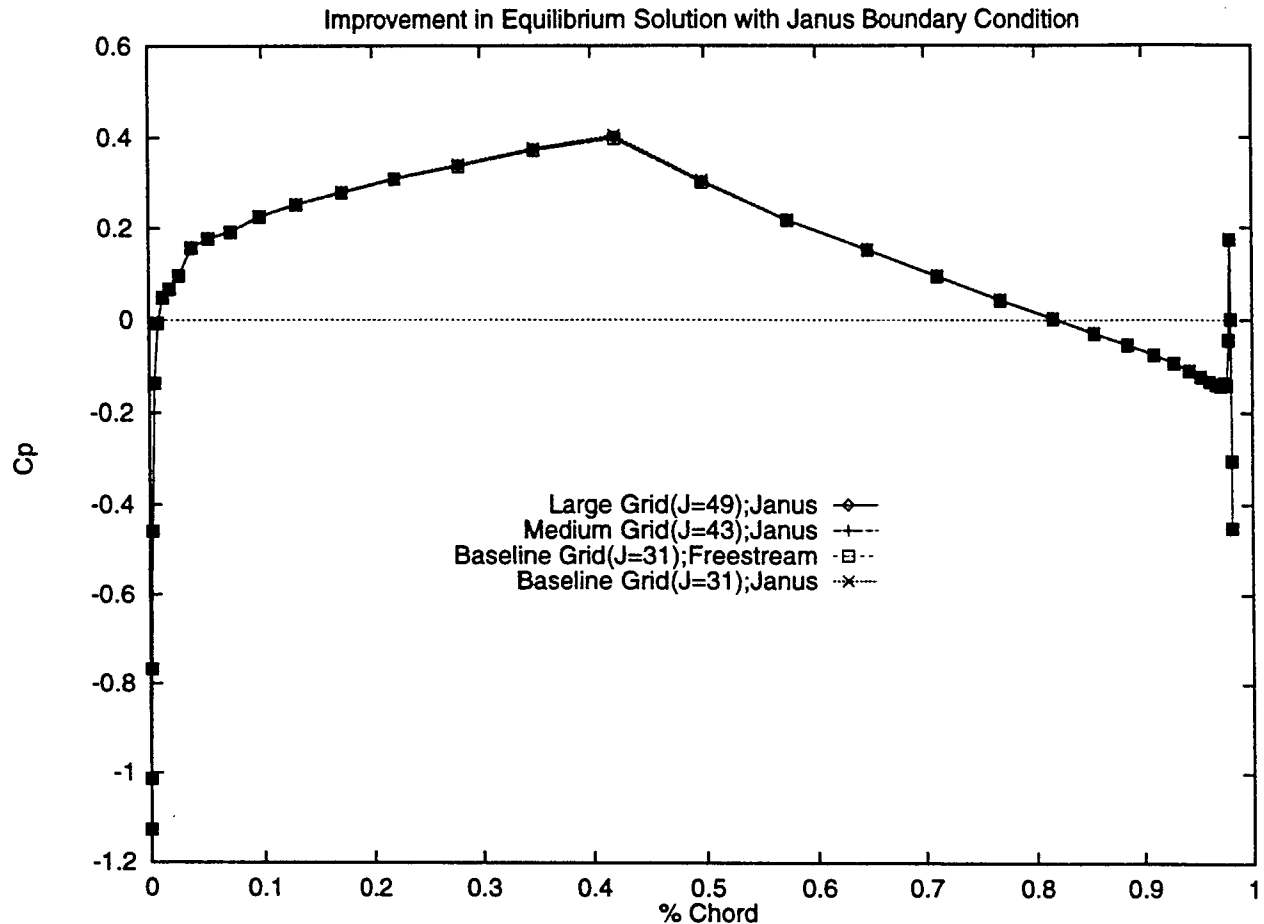


Figure 7: Steady C_p Distribution for $M=0.84$ and $\alpha_0 = 0^\circ$

Figure 8 shows the steady surface pressure distribution for the 1° static pre-twist analysis for 150, 50, and 5 chord domain sizes for both Janus and the original freestream farfield boundary conditions. At this angle of attack, the solution is grid independent at the 50 chord domain size as shown by comparing the 150 and 50 (i.e. $J=49$ and $J=43$, respectively) chord results. In addition, Figure 8 shows a normal shock on the suction surface at 75% of the airfoil chord. Comparing the predictions with Janus steady boundary conditions to the original freestream conditions shows a significant improvement in the C_p distribution at the shock location. The original freestream conditions predicted the shock starting at the 60% chord location, while the Janus conditions predicted the shock starting at 70% chord. In addition, the Janus calculations predicted the same ending location for the shock wave as the 50 chord solution, while the freestream conditions were 8% chord upstream.

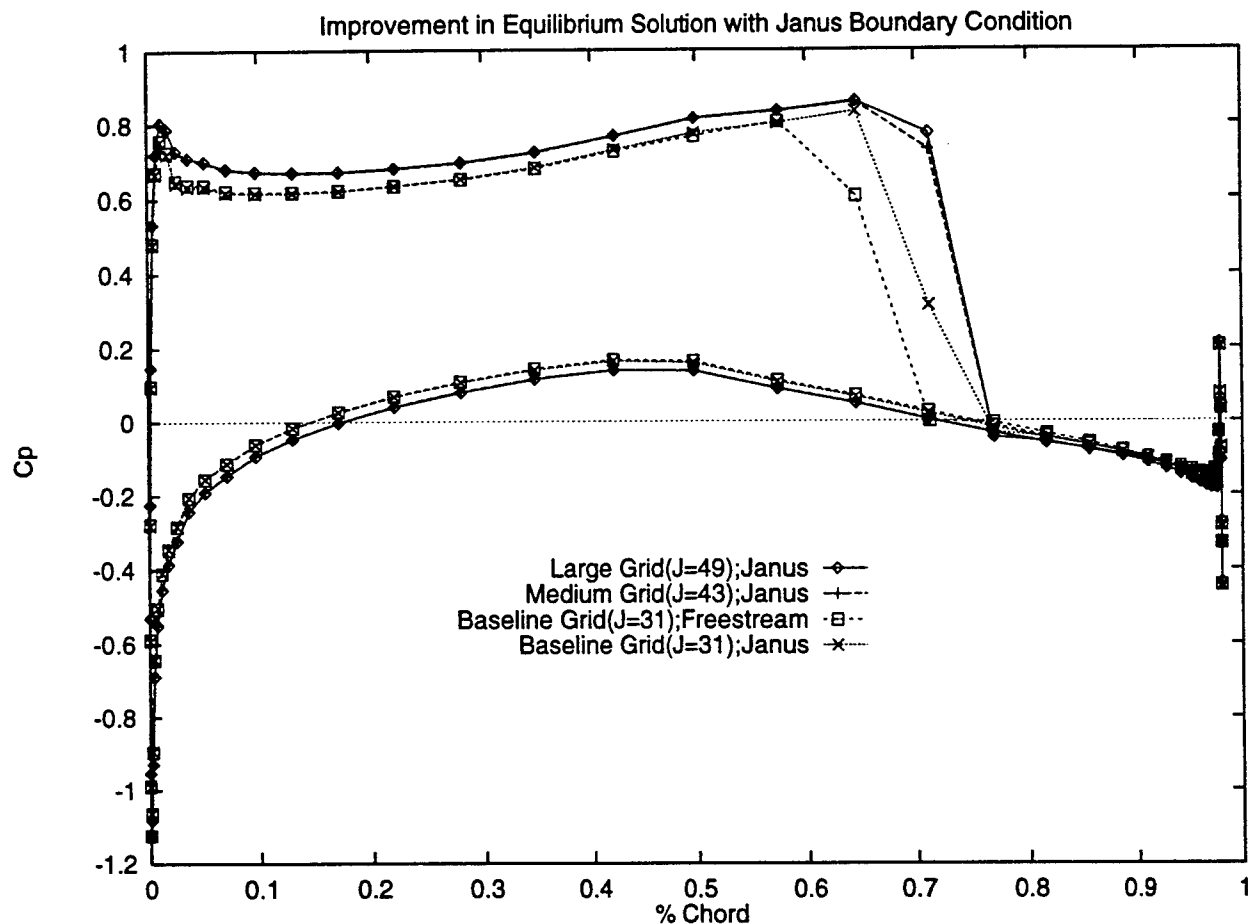


Figure 8: Steady C_p Distribution for $M=0.84$ and $\alpha_0 = 1^\circ$

Figure 9 shows the flutter speed variation with the different static pre-twist angles for a freestream Mach number of 0.84. Grid independence was achieved for the 150 and 50 chord domains solutions as shown in Figure 9. For the 5 chords domain ($J=31$) results, all the different model configurations are shown for the direct solver. The solution improvement from utilizing the locked grid deformation model is fairly consistent with changes in static pre-twist. Therefore, these improvements are not related to the nonlinear effects introduced by the strong shock wave. Comparing the approximate non-reflecting unsteady boundary conditions with the original freestream conditions resulted in another constant improvement in the flutter speed prediction for all static pre-twist values. Again, the conclusion is drawn that the approximate non-reflecting unsteady boundary conditions are not being significantly affected by the nonlinear nature of the flow physics.

When the steady characteristic boundary conditions are introduced with Giles unsteady conditions the improvement in flutter speed increases with increasing static pre-twist. At 0° static pre-twist, the improvement is nearly zero while at 1° the flutter speed prediction is 2.954 versus 2.812 without the new steady boundary conditions. The grid independent flutter speed is 3.005 at 1° . Therefore, the solution error is reduced nearly 60% by introducing the steady characteristic

boundary conditions, while each of the other modifications result in a 10% reduction, respectively. This is a strong indication that the steady characteristic boundary conditions do a significantly better job of modeling the nonlinear shock effects. Accurate shock predictions are required for transonic flight and beyond.

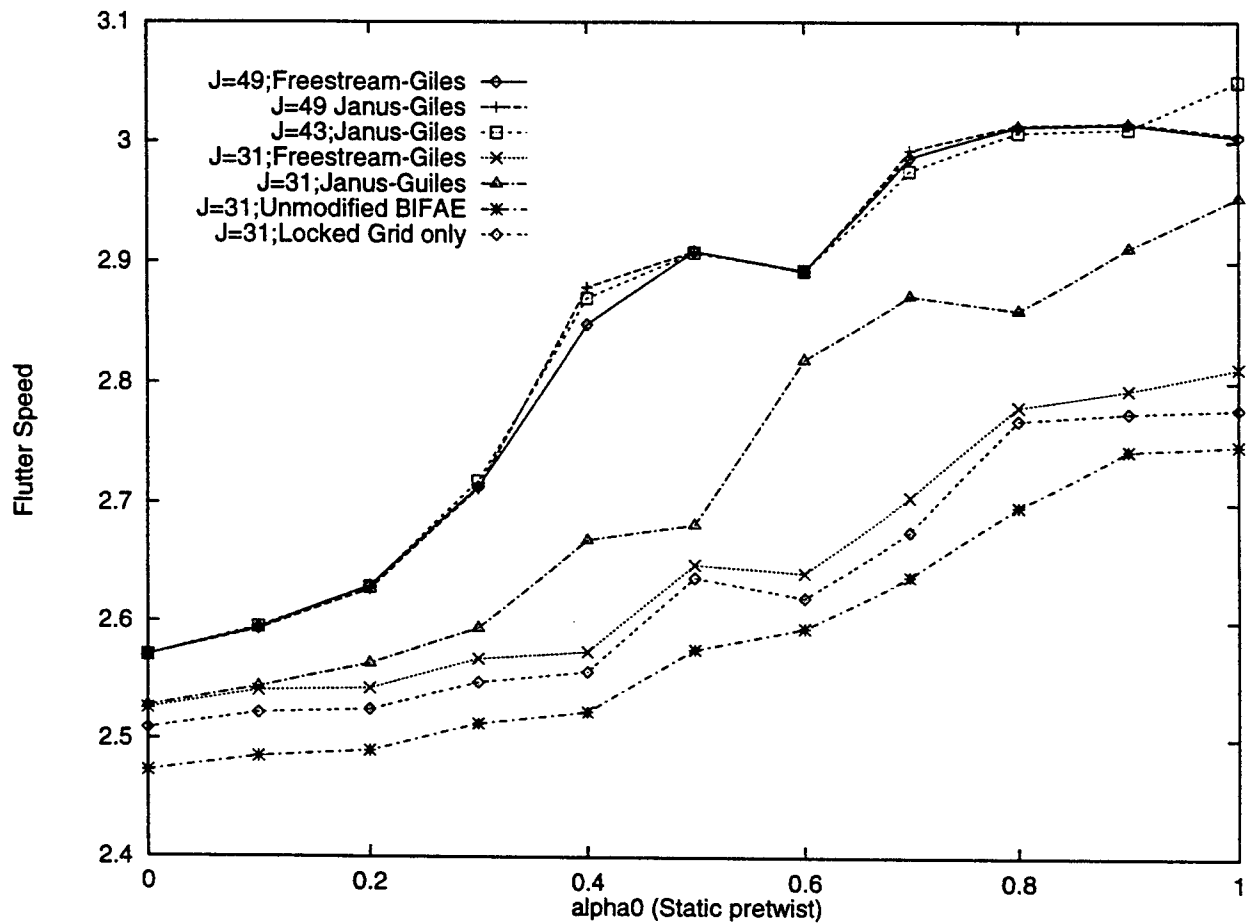


Figure 9: Flutter Speed Variations for $M=0.84$ with Static Pre-Twist

In addition, it is interesting to note the slight scatter in the results at 0.4° and 0.7° . For these conditions the analysis required reducing the CFL number, the under-relaxation parameter and increasing the number of iterations to get a converged flutter point solution. Typically, the flutter speed variation with static pre-twist is a smoother curve. Several possible reasons for the results shown in Figure 9 are inaccuracies in the trailing edge model and resolution of the shock location on the suction surface. As shown in Figure 8, the shock wave was smeared over three grid points. The shock position is extremely important in the calculation of the flutter speed, therefore any error in its location will have significant effects on the aeroelastic analysis.

Summary and Conclusion

The direct computational procedure BIFAE has been successfully enhanced by improved flow physics models of the farfield boundary conditions. In particular, a deforming grid model was

implemented which locks the farfield grid while allowing the grid near the airfoil surface to deform with the airfoil movement. In addition, steady characteristic and approximate non-reflecting unsteady farfield boundary conditions have been incorporated into the BIFAE model. A NACA 64A006 flapped airfoil, executing pitching and plunging motion, was utilized for verifying the successful implementation of these higher fidelity models. Both zero and non-zero angle of attack results were shown to highlight the improved efficiency of the enhanced model. For a 1° static pre-twist analysis at a freestream Mach number of 0.84, the new model resulted in over a 75% decrease in the flutter speed error. Therefore, the capability to more accurately model transonic flow conditions with strong shock interaction has been shown.

Potential enhancements to the direct BIFAE model as a consequence of the research described in this report are as follows: 1) implementation of the approximate non-reflecting unsteady 2-D farfield boundary conditions of Giles; 2) implement the Giles steady farfield boundary conditions; and 3) implement the geometric conservation law to properly account for the grid deformation. The results presented in this report could be improved by investigating the rounded trailing edge and chordwise grid resolution. In addition, it would be beneficial to identify the precise computational domain size required with the new farfield boundary conditions for a grid independent solution. These improvements were not completed in the current investigation because of a lack of time.

Acknowledgements

The authors would like to thank Dr. Phil Beran for providing this excellent educational opportunity and the Air Vehicles Directorate for the funding support required. In addition, we acknowledge all the support given us during this appointment by the other members of the Air Vehicles – Aeroelasticity Group. We have felt like one of the team and this has made the experience very rewarding. Finally, we appreciate the manner in which we have been able to continue working on this challenging and rewarding research after our summer appointment has ended.

References

- Beran, P.S., *A Domain-Decomposition Method for Airfoil Flutter Analysis*, 36th AIAA Aerospace Sciences Meeting, Reno, NV, Jan. 12-15, 1998, AIAA 98-0098.
- Beran, P.S. and Morton, S.A., *A Continuation Method for the Calculation of Airfoil Flutter Boundaries*, Journal of Guidance, Control and Dynamics, Vol. 20, No. 6, Nov.-Dec., 1997, pp. 1165-1171.
- Buxton, B.J. and Beran, P.S., *Validation of Two Shock-Capturing Methods for Calculation of Transonic Airfoil Flutter*, 35th AIAA Aerospace Sciences Meeting, Reno, NV, Jan. 6-10, 1997, AIAA 97-0834.

Giles, M.B., *Non-Reflecting Boundary Conditions for Euler Equation Calculations*, AIAA Journal, Vol. 28, No. 12, 1990, pp. 2050-2058.

Janus, J.M., *The Development of a Three-Dimensional Split Flux Vector Solver with Dynamic Grid Applications*, MS Thesis, Mississippi State University, August, 1984.

Lee, B.H.K., Yang, L.Y., and Wong, Y.S., *Flutter of an Airfoil with a Cubic Nonlinear Restoring Force*, Proceedings of the 39th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and Exhibit, Long Beach, CA, April 20-23, 1998, AIAA 98-1725-CP, pp. 237-257.

Morton, S.A. and Beran, P.S., *Hopf Bifurcation Analysis Applied to Deforming Airfoils at Transonic Speeds*, 13th AIAA Computational Fluid Dynamics Conference, Snowmass, CO, June 29-July 2, 1997, AIAA 97-1772.

Morton, S.A. and Beran, P.S., *Hopf-Bifurcation Analysis of Airfoil Flutter at Transonic Speeds*, 34th AIAA Aerospace Sciences Meeting and Exhibit, Reno NV, Jan. 15-18, 1996A, AIAA 96-0060.

Morton, S.A. and Beran, P.S., *Effects of Structural Nonlinearity in the Bifurcation Analysis of Transonic Airfoil Flutter*, 27th AIAA Fluid Dynamics Conference, New Orleans, LA, June 17-20, 1996B, AIAA 96-1975.

Morton, S.A. and Beran, P.S., *Nonlinear Analysis of Airfoil Flutter at Transonic Speeds*, Proceedings of the 13th AIAA Applied Aerodynamics Conference, San Diego, CA, June 19-22, 1995, AIAA 95-1905-CP, pp. 1169-1179.

A DETAILED STUDY OF THE NUMERICAL PROPERTIES
OF FDTD ALGORITHMS FOR DISPERSIVE MEDIA

Jeffrey L. Young
Associate Professor
Department of Electrical Engineering
and
Ronald O. Nelson
Graduate Research Assistant
Department of Electrical Engineering

University of Idaho
Moscow, ID 83844-1023

Final Report for:
Summer Faculty Research Program
Air Force Research Laboratory

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

and

Air Force Research Laboratory, Air Vehicles Directorate

August 1998

A DETAILED STUDY OF THE NUMERICAL PROPERTIES
OF FDTD ALGORITHMS FOR DISPERSIVE MEDIA

Jeffrey L. Young
Associate Professor
Department of Electrical Engineering
University of Idaho

Abstract

Numerous finite-difference, time-domain (FDTD) algorithms for dispersive media are presented and studied. To limit the scope of the discussion, only direct integration and recursive convolution techniques are considered in conjunction with the following media types: the isotropic cold plasma and the multi-pole Debye dielectric. The study primarily focuses on the numerical permittivity calculation that results from the discretization procedure. Where appropriate, this permittivity is used to quantify the temporal error mechanisms of the scheme. Discussions are provided that assess the overall strengths and weaknesses of each scheme.

A DETAILED STUDY OF THE NUMERICAL PROPERTIES OF FDTD ALGORITHMS FOR DISPERSIVE MEDIA

Jeffrey L. Young

1 Introduction

The finite-difference, time-domain method (FDTD) has emerged as a leading numerical tool for the solution of the temporally dependent Maxwell's equations [1, 2]. Originally the scheme was created to discretize Maxwell's equations, with the assumption that the medium that hosted the electromagnetic wave was isotropic, non-dispersive, linear and time-invariant. In many applications that arise in the study of electromagnetic waves, such assumptions are indeed valid.

As the method grew in popularity and proved its worth in the prediction of scattering, diffraction and propagation events, many researchers postulated and then devised ways that the algorithm could be applied to problems for which the medium was no longer simple in its composition. Such media types include the ionosphere, biological tissues, crystalline structures, ferrites, optical fibers, and radar absorbing materials, to name a few. Whatever the case may be, however, these media are subsets of what is generally termed as *complex* media – media that is described by one or many of the following descriptors: dispersive, anisotropic, nonlinear and time-variant [3].

This report examines the most recent and popular advances in FDTD algorithm development as applied to electromagnetic wave propagation in complex media. Several methodologies are presented and comparisons are made in order to demonstrate the strengths and weaknesses of the various approaches. Particularly, direct integration and recursive convolution schemes associated with wave propagation in a cold plasma and a Debye dielectric are considered.

Error analyses are also provided. Since the discretized equations are approximations of Maxwell's equations, the data obtained from the FDTD algorithm will be corrupted by errors of the dissipative or dispersive kind. These errors linearly accumulate after each time step and hence, the data will be attenuated and out of phase with the data obtained from the exact solution. To quantify these errors, numerical dispersion relationships are derived. These same dispersion relationships also provide information on the stability of the scheme; this information is couched in terms of a Courant–Friedrichs–Lewy (CFL) number [4]. Thus, for those schemes for which the stability properties are known, that information is also provided.

2 Maxwell's Equations and Complex Media

According to Maxwell, the electromagnetic field is comprised of an electric field $\mathbf{E}(\mathbf{x}, t)$ and a magnetic field $\mathbf{B}(\mathbf{x}, t)$ that satisfy, at a point, two curl equations and two divergence equations [5]:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} - \mathbf{M}, \quad (1)$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}, \quad (2)$$

$$\nabla \cdot \mathbf{D} = 0 \quad (3)$$

and

$$\nabla \cdot \mathbf{B} = 0. \quad (4)$$

Here $\mathbf{D}$ is the electric flux density and $\mathbf{H}$ is the magnetic intensity. The terms $\mathbf{J}$ and $\mathbf{M}$ are often regarded as source terms, and in that context they are the electric and magnetic current densities, respectively. For purposes of numerical computation, the two curl equations are discretized and advanced in time; the two divergence equations are additional constraints imposed upon the fields. To bring closure to this system of equations, a set of constitutive relationships are imposed that quantify the effects of a material medium on an applied electromagnetic field. In general, we write

$$\mathbf{D} = \mathbf{D}(\mathbf{E}, \mathbf{H}) \quad \mathbf{B} = \mathbf{B}(\mathbf{E}, \mathbf{H}). \quad (5)$$

In addition, the currents may also be coupled to the fields, in which case

$$\mathbf{J} = \mathbf{J}(\mathbf{E}, \mathbf{H}) \quad \mathbf{M} = \mathbf{M}(\mathbf{E}, \mathbf{H}). \quad (6)$$

It is in this latter context that we will invoke the notion of current. That is, rather than the current being typified as *impressed*, it will be considered *induced* – a current generated by a field. For example, in the study of wave propagation in a cold plasma, $\mathbf{J}$ is replaced with $\mathbf{J}_p$, the polarization current generated in the plasma by an applied field. By definition, the polarization current is viewed as the time rate of change of the polarization vector $\mathbf{P}$ (i.e. $\mathbf{J}_p = \partial \mathbf{P} / \partial t$). The inclusion of $\mathbf{P}$ into the field description proves useful when studying propagation in Debye-type materials.

If the material is *simple* and source-free, the previous relationships reduce to $\mathbf{D} = \epsilon \mathbf{E}$, $\mathbf{B} = \mu \mathbf{H}$, $\mathbf{J} = \sigma \mathbf{E}$ and $\mathbf{M} = 0$, where the constants ϵ , μ and σ are the electric permittivity, magnetic permeability and electric conductivity, respectively. Written as such, $\mathbf{J}$ is now interpreted as the induced conduction current. For *complex* materials, the constitutive relationships do not reduce to such simple forms. Instead, the relationships may be cast in terms of convolution integrals, matrices or derivatives. For these situations, the material may be characterized as dispersive, anisotropic, or non-linear, to name a few possibilities.

3 FDTD Method

Consider a medium that is simple and source-free, but lossy. The FDTD method is based upon the application of central differences for both temporal and spatial discretization [6] which, for the sake of future developments, will be considered separately. With respect to the former, the temporal discretization of Maxwell's curl equations yields

$$\mu \left(\frac{\mathbf{H}^{n+1/2} - \mathbf{H}^{n-1/2}}{\delta_t} \right) = -\nabla \times \mathbf{E}^n \quad (7)$$

and

$$\epsilon \left(\frac{\mathbf{E}^{n+1} - \mathbf{E}^n}{\delta_t} \right) = \nabla \times \mathbf{H}^{n+1/2} - \sigma \left(\frac{\mathbf{E}^{n+1} + \mathbf{E}^n}{2} \right), \quad (8)$$

where we have employed central averages for the loss term in Ampere's law. Here δ_t is the time increment and $\mathbf{E}^n$ denotes the value of $\mathbf{E}$ at time $n\delta_t$. Upon the rearrangement of terms,

$$\mu \mathbf{H}^{n+1/2} = \mu \mathbf{H}^{n-1/2} - \delta_t \nabla \times \mathbf{E}^n \quad (9)$$

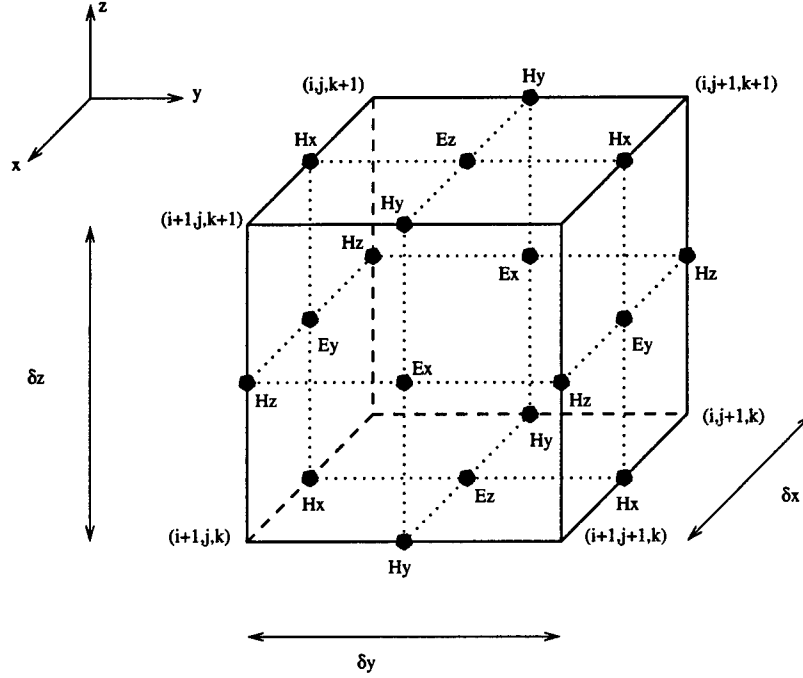


Figure 1: The Yee cell

and

$$\epsilon \mathbf{E}^{n+1} = \epsilon \left(\frac{1 - \sigma \delta_t/2}{1 + \sigma \delta_t/2} \right) \mathbf{E}^n + \left(\frac{\delta_t}{1 + \sigma \delta_t/2} \right) \nabla \times \mathbf{H}^{n+1/2}. \quad (10)$$

Whence, the update equation for the field constituents is readily observed. Given that $\mathbf{E}$ and $\mathbf{H}$ are synchronized, but syncopated by a half-time step, this type of integration scheme is commonly referred to as leap-frog integration. The previous equations constitute a second-order temporal method, as deduced from the truncation term of the central difference and average approximations.

To accomplish the spatial discretization, the Yee grid is invoked, as shown in Figure 1. The components of the Yee grid are arranged in accordance with the natural rotational structure of the curl operator acting upon $\mathbf{H}$, which is also spatially proportional to $\mathbf{E}$ (or vice-versa). For example, upon the spatial discretization of the z -component of Ampere's law, we obtain,

$$\mathbf{a}_z \cdot \nabla \times \mathbf{H} \Big|_{i+1/2, j+1/2, k} = \left[\frac{H_y|_{i+1} - H_y|_i}{\delta_x} \right]_{j+1/2, k} - \left[\frac{H_x|_{j+1} - H_x|_j}{\delta_y} \right]_{i+1/2, k}. \quad (11)$$

The indices (i, j, k) denote the spatial coordinates $(i\delta_x, j\delta_y, k\delta_z)$ with δ_x , δ_y and δ_z being the length of the cell in the x , y , z directions, respectively. The remaining five components of electromagnetic field are derived in a similar manner.

To quantify the FDTD method in terms of its dissipation and dispersion errors, consider a plane wave propagating in the $\mathbf{k}$ direction and in an homogeneous open domain. That is, let $\mathbf{E} = \mathbf{E}_0 e^{-j\mathbf{k} \cdot \mathbf{x}} e^{j\omega t}$ and $\mathbf{H} = \mathbf{H}_0 e^{-j\mathbf{k} \cdot \mathbf{x}} e^{j\omega t}$; $\mathbf{E}_0$ and $\mathbf{H}_0$ are constant vectors. If $\mathbf{E}$ and $\mathbf{H}$ are to be solutions of the discretized form of Maxwell's equations, then

$$-j\mathbf{K} \times \mathbf{E}_0 = -j\Omega\mu\mathbf{H}_0. \quad (12)$$

and

$$-j\mathbf{K} \times \mathbf{H}_0 = j\Omega\epsilon\mathbf{E}_0 + \Lambda\sigma\mathbf{E}_0; \quad (13)$$

moreover, $\mathbf{K} \cdot \mathbf{E}_0 = 0$ and $\mathbf{K} \cdot \mathbf{H}_0 = 0$. Here

$$\mathbf{K} = \left(\frac{2}{\delta_x}\right) \sin\left(\frac{k_x\delta_x}{2}\right) \mathbf{a}_x + \left(\frac{2}{\delta_y}\right) \sin\left(\frac{k_y\delta_y}{2}\right) \mathbf{a}_y + \left(\frac{2}{\delta_z}\right) \sin\left(\frac{k_z\delta_z}{2}\right) \mathbf{a}_z, \quad (14)$$

$$\Omega = \left(\frac{2}{\delta_t}\right) \sin\left(\frac{\omega\delta_t}{2}\right) \quad (15)$$

and

$$\Lambda = \cos\left(\frac{\omega\delta_t}{2}\right). \quad (16)$$

Obviously, $\mathbf{K} \rightarrow \mathbf{k}$, $\Omega \rightarrow \omega$ and $\Lambda \rightarrow 1$ as $\delta_x, \delta_y, \delta_z, \delta_t \rightarrow 0$, as required for a consistent scheme. Further manipulation of Eqns. (12) and (13) leads to the equation

$$\mathbf{K} \cdot \mathbf{K} = -j\Omega\mu(j\Omega\epsilon + \Lambda\sigma) = \Omega^2\mu\hat{\epsilon}, \quad (17)$$

which is the numerical dispersion equation and is the counterpart of the analytical dispersion equation $\mathbf{k} \cdot \mathbf{k} = -j\omega\mu(j\omega\epsilon + \sigma)$; here $\hat{\epsilon} = [\epsilon + \Lambda\sigma/(j\Omega)]$. In many of the subsequent sections this same equation will be referenced, but different permittivity relationships will be supplied. Whatever the case may be, Eqn. (17) provides the necessary information for understanding the scheme's dispersion, dissipation, anisotropy and stability properties [7, 8].

Stability (i.e., non-exponential growth in time) is insured if $\text{Im}\{\omega\delta_t\} > 0$. From Eqn. (17) and for $\sigma = 0$, this constraint is satisfied for all directions of propagation provided that [2]

$$\delta_t < \frac{1}{v_p} \left[\left(\frac{1}{\delta_x}\right)^2 + \left(\frac{1}{\delta_y}\right)^2 + \left(\frac{1}{\delta_z}\right)^2 \right]^{-1/2}. \quad (18)$$

Here $v_p = 1/\sqrt{\mu\epsilon}$, which is the phase velocity (if $\epsilon = \epsilon_0$ and $\mu = \mu_0$ then $v_p = c$, the speed of light). Note: If Eqn. (18) is satisfied, $\text{Im}\{\omega\delta_t\} = 0$, which implies a dissipationless numerical scheme. Finally, given that the FDTD method is both consistent and stable, Lax's theorem states that the data will converge to the exact solution as the cell and time-step become infinitesimally small [9].

In the ensuing sections, the FDTD method will be modified or extended to model wave propagation in complex media. However, in all cases considered, the Yee grid of Figure 1 and the discretization of the curl operators will be retained.

4 Cold Plasma

To demonstrate the wide variety of techniques available to model numerically wave propagation in linearly dispersive media, the case of wave propagation in an isotropic, lossy, cold plasma is considered next. For this scenario, Ampere's law and the momentum equation are [10]

$$\epsilon_0 \frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{H} - \mathbf{J}_p \quad (19)$$

and

$$\frac{\partial \mathbf{J}_p}{\partial t} = -\nu \mathbf{J}_p + \epsilon_0 \omega_p^2 \mathbf{E}, \quad (20)$$

where $\mathbf{J}_p$ is the polarization current, ν is the collision frequency and ω_p is the plasma frequency; specifically $\mathbf{J}_p = -n_0 q \mathbf{u}$ and $\omega_p = \sqrt{n_0 q^2 / (m \epsilon_0)}$. For time-harmonic signals of dependence $e^{j\omega t}$, it is a simple exercise to show that the momentum equation can be embedded into Ampere's law, from which a scalar frequency domain permittivity $\hat{\epsilon}$ can be defined. Suppressing that analysis, one finds that

$$\hat{\epsilon} = \epsilon_0 \left(1 + \frac{\omega_p^2}{\omega(j\nu - \omega)} \right) = \epsilon_0 (1 + \hat{\chi}(\omega)), \quad (21)$$

where $\hat{\chi}$ is the scalar frequency domain susceptibility. Since $\hat{\mathbf{D}} = \hat{\epsilon} \hat{\mathbf{E}}$, the corresponding time-domain constitutive relationship, cast in terms of a convolution integral, is deducible from Fourier theory [11]:

$$\mathbf{D}(\mathbf{x}, t) = \epsilon_0 \mathbf{E}(\mathbf{x}, t) + \epsilon_0 \int_{-\infty}^t \chi(\tau) \mathbf{E}(\mathbf{x}, t - \tau) d\tau, \quad (22)$$

where

$$\chi(\tau) = \frac{\omega_p^2}{\nu} [1 - e^{-\nu\tau}] U(\tau). \quad (23)$$

Here U is the unit step function.

Each equation in the previous development gives a different insight into the phenomenology of wave propagation in a cold plasma. And, for each of those insights, there exists a corresponding FDTD discretization methodology. In this section, we will explore several such methodologies and will classify those methodologies into two categories: Direct integration methods and recursive convolution methods. The direct integration methods are either based on the fundamental state equations of Eqns. (19) and (20) or on the constitutive relation $\hat{\mathbf{D}} = \hat{\epsilon} \hat{\mathbf{E}}$. The recursive methods incorporate the convolution integral of Eqn. (22) into the FDTD paradigm¹. Although each discretization procedure is consistent with the governing equations, their accuracy order and memory requirements are different. Such similarities and differences are documented in the last subsection.

In the following subsections, Eqn. (9) is the temporally discretized form of Faraday's law. Additionally, the components of $\mathbf{J}_p$, $\mathbf{D}$ and $\mathbf{E}$ are assumed to occupy the same locations in the Yee grid.

4.1 Direct Integration Method One: CP-DIM1

Consider first a state-space integration technique. For this situation, Eqns. (19) and (20) are discretized using central differences for the derivatives and central averages for the collision term. To maintain synchronization with the FDTD leap-frog integrator, the momentum equation and Faraday's law are advanced simultaneously and collectively leaped with Ampere's law. For source-free media, this procedure results in the following system of equations [15]:

$$\frac{\mathbf{J}_p^{n+1/2} - \mathbf{J}_p^{n-1/2}}{\delta_t} = -\nu \left(\frac{\mathbf{J}_p^{n+1/2} + \mathbf{J}_p^{n-1/2}}{2} \right) + \epsilon_0 \omega_p^2 \mathbf{E}^n \quad (24)$$

and

$$\epsilon_0 \left(\frac{\mathbf{E}^{n+1} - \mathbf{E}^n}{\delta_t} \right) = \nabla \times \mathbf{H}^{n+1/2} - \mathbf{J}_p^{n+1/2}. \quad (25)$$

Given the simplicity of the algebraic manipulations that solve for $\mathbf{J}_p^{n+1/2}$ in Eqn. (24), the final expression is not given. This will be the case for many of the other direct integration methods as well.

¹ Since the Z -transform method is a subset of the recursive method, it will not be documented herein. For more information on using Z -transforms in conjunction with the FDTD method consult [12]–[14].

4.2 Direct Integration Method Two: CP-DIM2

Instead of averaging the loss term, as in CP-DIM1, averaging the current term in Ampere's law is equally possible. For this case, the momentum equation and Ampere's law are synchronized and collectively leaped with Faraday's law [16]. That is,

$$\frac{\mathbf{J}_p^{n+1} - \mathbf{J}_p^{n-1}}{2\delta_t} = -\nu \mathbf{J}_p^n + \epsilon_0 \omega_p^2 \mathbf{E}^n \quad (26)$$

and

$$\epsilon_0 \left(\frac{\mathbf{E}^{n+1} - \mathbf{E}^n}{\delta_t} \right) = \nabla \times \mathbf{H}^{n+1/2} - \left(\frac{\mathbf{J}_p^{n+1} + \mathbf{J}_p^n}{2} \right). \quad (27)$$

4.3 Direct Integration Method Three: CP-DIM3

Next consider the case in which the polarization current in Ampere's law, the electric field term in the momentum equation and the dissipation term in the momentum equation are all time averaged. The synchronization of the momentum equation and Ampere's law is maintained [17]. The resulting scheme is

$$\frac{\mathbf{J}_p^{n+1} - \mathbf{J}_p^n}{\delta_t} = -\nu \frac{\mathbf{J}_p^{n+1} + \mathbf{J}_p^n}{2} + \epsilon_0 \omega_p^2 \left(\frac{\mathbf{E}^{n+1} + \mathbf{E}^n}{2} \right) \quad (28)$$

and

$$\epsilon_0 \left(\frac{\mathbf{E}^{n+1} - \mathbf{E}^n}{\delta_t} \right) = \nabla \times \mathbf{H}^{n+1/2} - \left(\frac{\mathbf{J}_p^{n+1} + \mathbf{J}_p^n}{2} \right). \quad (29)$$

4.4 Direct Integration Method Four: CP-DIM4

Observing that the momentum equation has solutions that decay exponentially like $e^{-\nu t}$, Cummer suggests that exponential fitting is necessary, particularly when the dissipation term is large [17]. To that end, the exponentially fitted equation is

$$\mathbf{J}_p^{n+1} = e^{-\nu \delta_t} \mathbf{J}_p^n + \frac{\epsilon_0 \omega_p^2}{\nu^2 \delta_t} (\nu \delta_t + e^{-\nu \delta_t} - 1) \mathbf{E}^{n+1} + \frac{\epsilon_0 \omega_p^2}{\nu^2 \delta_t} (1 - e^{-\nu \delta_t} - \nu \delta_t e^{-\nu \delta_t}) \mathbf{E}^n. \quad (30)$$

Ampere's law still requires the averaging of the polarization current:

$$\epsilon_0 \left(\frac{\mathbf{E}^{n+1} - \mathbf{E}^n}{\delta_t} \right) = \nabla \times \mathbf{H}^{n+1/2} - \left(\frac{\mathbf{J}_p^{n+1} + \mathbf{J}_p^n}{2} \right). \quad (31)$$

4.5 Direct Integration Method Five: CP-DIM5

The next direct integration technique considers the frequency domain constitutive relationship $\hat{\mathbf{D}} = \hat{\epsilon} \hat{\mathbf{E}}$ and Eqn. (21) as the basis of temporal discretization [18]. Combining these two equations, we obtain

$$(\omega^2 - j\omega\nu) \hat{\mathbf{D}} = \epsilon_0 (\omega^2 - j\omega\nu - \omega_p^2) \hat{\mathbf{E}}. \quad (32)$$

Since multiplications of $j\omega$ in the frequency domain are equivalent to time derivatives in the time-domain, Eqn. (32) is equivalent to

$$\frac{\partial^2 \mathbf{D}}{\partial t^2} + \nu \frac{\partial \mathbf{D}}{\partial t} = \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} + \epsilon_0 \nu \frac{\partial \mathbf{E}}{\partial t} + \epsilon_0 \omega_p^2 \mathbf{E}. \quad (33)$$

Discretization of Eqn. (33) with second-order central differences, along with Eqn. (2), yields the following system of semi-discrete numerical equations:

$$\mathbf{D}^{n+1} = \mathbf{D}^n + \delta_t \nabla \times \mathbf{H}^{n+1/2}, \quad (34)$$

and

$$\begin{aligned} \epsilon_0(2 + \nu\delta_t)\mathbf{E}^{n+1} &= (2 + \nu\delta_t)\mathbf{D}^{n+1} - 4\mathbf{D}^n + (2 - \nu\delta_t)\mathbf{D}^{n-1} \\ &- 2\epsilon_0(\omega_p^2\delta_t^2 - 2)\mathbf{E}^n - \epsilon_0(2 - \nu\delta_t)\mathbf{E}^{n-1}. \end{aligned} \quad (35)$$

4.6 Recursive Convolution Method One: CP-RCM1

The recursive convolution method is based upon the discretization of the convolution integral of Eqn. (22) [19]. At time $n\delta_t$,

$$\mathbf{D}(\mathbf{x}, n\delta_t) = \epsilon_0\mathbf{E}(\mathbf{x}, n\delta_t) + \epsilon_0 \int_0^{n\delta_t} \mathbf{E}(\mathbf{x}, n\delta_t - \tau)\chi(\tau) d\tau. \quad (36)$$

Under the assumption that $\mathbf{E}$ is piece-wise constant,

$$\mathbf{D}^n = \epsilon_0\mathbf{E}^n + \epsilon_0 \sum_{m=0}^{n-1} \mathbf{E}^{n-m} \int_{m\delta_t}^{(m+1)\delta_t} \chi(\tau) d\tau. \quad (37)$$

Likewise,

$$\mathbf{D}^{n+1} = \epsilon_0\mathbf{E}^{n+1} + \epsilon_0 \sum_{m=0}^n \mathbf{E}^{n-m+1} \int_{m\delta_t}^{(m+1)\delta_t} \chi(\tau) d\tau. \quad (38)$$

The discretization of Ampere's law, as given by Eqn. (2), requires us to take the difference of the two preceding equations. Doing so and replacing that difference into Ampere's law, we obtain

$$\mathbf{E}^{n+1} = \alpha\mathbf{E}^n + \alpha\Psi^n + \frac{\alpha\delta_t}{\epsilon_0} \nabla \times \mathbf{H}^{n+1/2}, \quad (39)$$

where

$$\Psi^n = \sum_{m=0}^{n-1} \mathbf{E}^{n-m} \Delta\chi^m \quad (40)$$

and

$$\alpha = \frac{1}{1 + \chi^0}. \quad (41)$$

Here

$$\Delta\chi^m = \chi^m - \chi^{m+1} \quad (42)$$

and

$$\chi^m = \int_{m\delta_t}^{(m+1)\delta_t} \chi(\tau) d\tau. \quad (43)$$

Performing the calculation for χ^0 and $\Delta\chi^m$ in conjunction with Eqn. (23), we obtain

$$\chi^0 = \frac{\omega_p^2\delta_t}{\nu} - \left(\frac{\omega_p}{\nu}\right)^2 (1 - e^{-\nu\delta_t}) \quad (44)$$

and

$$\Delta\chi^m = -\left(\frac{\omega_p}{\nu}\right)^2 e^{-m\nu\delta_t} (1 - e^{-\nu\delta_t})^2 = \Delta\chi^0 e^{-m\nu\delta_t}. \quad (45)$$

Although Maxwell's equations are temporally discretized, they are not in a form suitable for computation. For we note that to compute Eqn. (40), one would have to store the complete time history of $\mathbf{E}$. However, given the exponential form of $\Delta\chi^m$,

$$\Psi^n = \Delta\chi^0 \sum_{m=0}^{n-1} \mathbf{E}^{n-m} e^{-m\nu\delta_t}; \quad (46)$$

whence,

$$\Psi^n = \Delta\chi^0 \mathbf{E}^n + e^{-\nu\delta_t} \Psi^{n-1}. \quad (47)$$

Thus, the equation for Ψ reduces to a simple recursive relationship.

4.7 Recursive Convolution Method Two: CP-RCM2

Numerical accuracy of CP-RCM1 can be increased by assuming piecewise linear line segments in the integration of the convolution integral [20]. For this situation, one approximates $\mathbf{E}$ over the interval $[n\delta_t, (n+1)\delta_t]$ with

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}(\mathbf{x}, n\delta_t) + (t - n\delta_t) \left(\frac{\mathbf{E}(\mathbf{x}, (n+1)\delta_t) - \mathbf{E}(\mathbf{x}, n\delta_t)}{\delta_t} \right). \quad (48)$$

Substitution of this approximation into Eqn. (36) yields the following result:

$$\mathbf{D}^n = \epsilon_0 \mathbf{E}^n + \epsilon_0 \sum_{m=0}^{n-1} \mathbf{E}^{n-m} \chi^m + \epsilon_0 \sum_{m=0}^{n-1} (\mathbf{E}^{n-m-1} - \mathbf{E}^{n-m}) \xi^m, \quad (49)$$

where χ^m is given by Eqn. (43) and

$$\xi^m = \frac{1}{\delta_t} \int_{m\delta_t}^{(m+1)\delta_t} (\tau - m\delta_t) \chi(\tau) d\tau. \quad (50)$$

As before, the time derivative of $\mathbf{D}$ at time $(n+1/2)\delta_t$ is approximated to second-order by $(\mathbf{D}^{n+1} - \mathbf{D}^n)/\delta_t$. From Eqn. (2) and Eqn. (49),

$$\mathbf{E}^{n+1} = \alpha(1 - \xi^0) \mathbf{E}^n + \frac{\alpha\delta_t}{\epsilon_0} \nabla \times \mathbf{H}^{n+1/2} + \alpha \Psi^n. \quad (51)$$

Here

$$\Psi^n = \sum_{m=0}^{n-1} \mathbf{E}^{n-m} \Delta\chi^m + (\mathbf{E}^{n-m-1} - \mathbf{E}^{n-m}) \Delta\xi^m, \quad (52)$$

$$\alpha = \frac{1}{(1 + \chi^0 - \xi^0)} \quad (53)$$

and

$$\Delta\xi^m = \xi^m - \xi^{m+1}. \quad (54)$$

The expression for $\Delta\chi^m$ is given by Eqn. (45).

For the time-domain susceptibility function given by Eqn. (23), it is a simple matter to show that

$$\xi^m = \frac{\omega_p^2}{\delta_t \nu} \left[\frac{\delta_t^2}{2} + \frac{e^{-\nu m \delta_t}}{\nu} \left(\delta_t e^{-\nu \delta_t} + \frac{e^{-\nu \delta_t}}{\nu} - \frac{1}{\nu} \right) \right] \quad (55)$$

Method	Numerical Permittivity: $\hat{\epsilon}$
CP-DIM1	$\hat{\epsilon} = \epsilon_0 \left(1 + \frac{\omega_p^2}{\Omega(j\nu\Lambda - \Omega)} \right)$
CP-DIM2	$\hat{\epsilon} = \epsilon_0 \left(1 + \frac{\omega_p^2 \Lambda}{\Omega(j\nu - \Omega\Lambda)} \right)$
CP-DIM3	$\hat{\epsilon} = \epsilon_0 \left(1 + \frac{\omega_p^2 \Lambda^2}{\Omega(j\nu\Lambda - \Omega)} \right)$
CP-DIM4	$\hat{\epsilon} = \epsilon_0 \left[1 + \frac{\Lambda\omega_p^2 \{ \nu \sin[(\omega - j\nu)\delta_t/2] - \Omega \sinh(\nu\delta_t/2) \}}{j\Omega\nu^2 \sin[(\omega - j\nu)\delta_t/2]} \right]$
CP-DIM5	$\hat{\epsilon} = \epsilon_0 \left(1 + \frac{\omega_p^2}{\Omega(j\nu\Lambda - \Omega)} \right)$
CP-RCM1	$\hat{\epsilon} = \epsilon_0 \left(1 + \frac{\chi^0 e^{j\omega\delta_t/2}}{j\Omega\delta_t} + \frac{\Delta\chi^0 e^{\nu\delta_t/2}}{2\Omega\delta_t \sin[(\omega - j\nu)\delta_t/2]} \right)$
CP-RCM2	$\hat{\epsilon} = \epsilon_0 \left[1 - \xi^0 + \frac{\chi^0 e^{j\omega\delta_t/2}}{j\Omega\delta_t} + \left(\frac{\Delta\chi^0 - j\Omega\delta_t \Delta\xi^0 e^{-j\omega\delta_t/2}}{2\Omega\delta_t \sin[(\omega - j\nu)\delta_t/2]} \right) e^{\nu\delta_t/2} \right]$

Table 1: Numerical permittivities for various schemes associated with propagation in an isotropic cold plasma.

and

$$\Delta\xi^m = \frac{\omega_p^2}{\delta_t \nu^2} \left(\delta_t e^{-\nu\delta_t} + \frac{e^{-\nu\delta_t}}{\nu} - \frac{1}{\nu} \right) (1 - e^{-\nu\delta_t}) e^{-m\nu\delta_t} = \Delta\xi^0 e^{-m\nu\delta_t}. \quad (56)$$

Finally, the summation over all time in Eqn. (52) is replaced with a recursive factor – to wit,

$$\Psi^n = \Delta\chi^0 \mathbf{E}^n + (\mathbf{E}^{n-1} - \mathbf{E}^n) \Delta\xi^0 + e^{-\nu\delta_t} \Psi^{n-1}. \quad (57)$$

The expression for $\Delta\chi^0$ is given by Eqn. (45) when $m = 0$.

Although the previous equation appears to require an additional storage location for $\mathbf{E}^{n-1}$, such is not the case. As pointed out by Kelley *et al.* [20], a low storage implementation is possible by making judicious use of a temporary variable.

4.8 Comparative Analysis

The numerical errors of each of the aforementioned schemes are deduced in a similar way as in Section Three. Assuming again plane wave propagation and an unbounded domain, the numerical dispersion relationship is similar to that given by Eqn. (17). Of course, each scheme produces a different $\hat{\epsilon}$, as seen in Table 1. Note the equivalence between CP-DIM1 and CP-DIM5. Even though both of these procedures are derived from different equations, the final numerical permittivities are identical. However, as seen in Table 3, CP-DIM5 requires more memory. Next, as pointed out by Cummer [17], there exists an equivalence between exponential differencing (i.e., CP-DIM4) and the recursive convolution methods; this observation is confirmed in Table 1, where it is seen that the numerical permittivities bear some semblance to one another. Finally, if $\xi^0 = \Delta\xi^0 = 0$, CP-RCM1 and CP-RCM2 are one and the same, as expected.

To appreciate the subtle differences between each of the schemes, Table 2 is provided. In this table, each entry corresponds to the leading term of the Taylor series of the function $\Omega^2 \hat{\epsilon} / \epsilon_0$. As seen from Eqn. (17), the function $\Omega^2 \hat{\epsilon}$ contains the complete temporal information of the numerical scheme. Except for CP-RCM1, these numerical permittivity relationships reduce to the exact analytical permittivity relationship to the order of δ_t^2 ; CP-RCM1 is temporally accurate to first-order. Of course, for certain choices of ω , ω_p , ν , δ_t , the scheme with lowest truncation term is deemed temporally best, from an accuracy point of view.

Method	Truncation Term	Truncation Term: $\nu = 0$
CP-DIM1	$\frac{(\nu^2\omega^4 + 2j\nu\omega^5 - \omega^6 + j\nu\omega^3\omega_p^2)\delta_t^2}{12(\omega - j\nu)^2}$	$\frac{-\omega^4\delta_t^2}{12}$
CP-DIM2	$\frac{(\nu^2\omega^4 + 2j\nu\omega^5 - \omega^6 - 2j\nu\omega^3\omega_p^2)\delta_t^2}{12(\omega - j\nu)^2}$	$\frac{-\omega^4\delta_t^2}{12}$
CP-DIM3	$\left(-\omega^4 + \frac{2j\omega^3\omega_p^2}{\nu + j\omega} - \frac{\omega^4\omega_p^2}{(\nu + j\omega)^2}\right)\left(\frac{\delta_t^2}{12}\right)$	$\frac{\omega^2(-\omega^2 + 3\omega_p^2)\delta_t^2}{12}$
CP-DIM4	$\left(-\omega^4 + \frac{3j\omega^3\omega_p^2}{\nu + j\omega}\right)\left(\frac{\delta_t^2}{12}\right)$	$\frac{\omega^2(-\omega^2 + 3\omega_p^2)\delta_t^2}{12}$
CP-DIM5	$\frac{(\nu^2\omega^4 + 2j\nu\omega^5 - \omega^6 + j\nu\omega^3\omega_p^2)\delta_t^2}{12(\omega - j\nu)^2}$	$\frac{-\omega^4\delta_t^2}{12}$
CP-RCM1	$\frac{\omega_p^2\omega\delta_t}{2(\nu + j\omega)}$	$\frac{\omega_p^2\omega\delta_t}{2j}$
CP-RCM2	$\left(-\omega^4 + \frac{2j\omega^3\omega_p^2}{\nu + j\omega}\right)\left(\frac{\delta_t^2}{12}\right)$	$\frac{\omega^2(-\omega^2 + 2\omega_p^2)\delta_t^2}{12}$

Table 2: Truncation term from the Taylor analysis of $\Omega^2\hat{\epsilon}/\epsilon_0$: Cold plasma.

A special case of interest is when $\nu = 0$. From Table 2, it is readily seen that CP-DIM1, CP-DIM2 and CP-DIM5 all have the same leading truncation term, which is independent ω_p . CP-DIM3 and CP-DIM4 also have the same truncation term; the truncation term for CP-RCM2 is close in value to that of CP-DIM3 and CP-DIM4. Since the truncation term for CP-DIM3, CP-DIM4 and CP-RCM2 are dependent on ω_p^2 , the truncation terms are smaller than those associated with CP-DIM1, CP-DIM2 and CP-DIM5. Moreover, there exists a frequency for which the truncation terms of CP-DIM3, CP-DIM4 and CP-RCM2 vanish. Finally, if $\omega_p^2\delta_t^2 \ll 1$, then all of the second order schemes produce a truncation term of $-\omega^4\delta_t^2/12$, which is the truncation term of the standard FDTD algorithm.

Other properties of these algorithms, including memory and one-dimensional CFL equations are shown in Table 3 [17]. As expected from the previous discussion on the truncation terms, CP-DIM1, CP-DIM2 and CP-DIM5 have the same stability equations, which are dependent on the plasma frequency, when $\nu = 0$. When $\nu \neq 0$, CP-DIM2 is unconditionally unstable. CP-DIM3 and CP-DIM4 have a CFL of unity. This is to be expected since the leading truncation term of $\Omega^2\hat{\epsilon}/\epsilon_0$ for CP-DIM3 and CP-DIM4 contains the factor $(\omega_p\delta_t/2)^2$, which is also the same factor that appears in the CFL equation for CP-DIM1, CP-DIM2 and CP-DIM5. As for CP-RCM1 and CP-RCM2, the CFL is found experimentally to be near unity [17]. From a memory point of view, CP-DIM2 and CP-DIM5 require additional memory (compared to the other schemes) for the back storage of a vector.

5 Debye Dielectric

The Debye dielectric has received considerable attention in the literature due to the fast growing interest in electromagnetic wave interactions with biological and water-based substances [21]–[23]. Given the water content of these substances, they are well modeled by a combination of single-pole models, which account for permanent dipole moments established within the water [3]. Moreover, since the Debye model is consistent with the requirements associated with the Kramers–Kronig relationships, the model inherently satisfies causality requirements. For this reason, an M -pole model is often used to construct an analytical model

Method	3D Memory/Cell	1D Stability
CP-DIM1	9	$\delta_t < (\delta_x/c)\sqrt{1 - (\omega_p \delta_t/2)^2}; \nu = 0$
CP-DIM2	12	$\delta_t < (\delta_x/c)\sqrt{1 - (\omega_p \delta_t/2)^2}; \nu = 0$
CP-DIM3	9	$\delta_t < \delta_x/c$
CP-DIM4	9	$\delta_t < \delta_x/c$
CP-DIM5	12	$\delta_t < (\delta_x/c)\sqrt{1 - (\omega_p \delta_t/2)^2}; \nu = 0$
CP-RCM1	9	$\delta_t^{max} \approx \delta_x/c$
CP-RCM2	9	$\delta_t^{max} \approx \delta_x/c$

Table 3: Comparison table of various schemes in terms of memory per cell and stability equations: Cold plasma.

from empirical data. Since in the frequency-domain the permittivity of the Debye dielectric is given by

$$\hat{\epsilon} = \epsilon_0 \epsilon_\infty + \epsilon_0 \sum_{k=1}^M \frac{\epsilon_{sk} - \epsilon_\infty}{1 + j\omega\tau_k} = \epsilon_0 (1 + \hat{\chi}(\omega)), \quad (58)$$

the modeling task is to deduce values for τ_k , ϵ_∞ and ϵ_{sk} that best fit $\hat{\epsilon}$ to the data [24]. Here τ_k is the k th relaxation time, $\epsilon_0 \epsilon_\infty$ is the value of $\hat{\epsilon}$ at infinite frequency and $\epsilon_{sk} \epsilon_0$ is the zero frequency permittivity of the k th relaxation. Note: In the time-domain,

$$\chi(t) = \sum_{k=1}^M \left(\frac{\epsilon_{sk} - \epsilon_\infty}{\tau_k} \right) e^{-t/\tau_k} U(t). \quad (59)$$

The Debye dielectric may also be characterized in state variable form (excluding Faraday's law) in terms of the polarization vector $\mathbf{P}$ [25]:

$$\epsilon_0 \epsilon_\infty \frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{H} - \sum_{k=1}^M \frac{1}{\tau_k} [(\epsilon_{sk} - \epsilon_\infty) \epsilon_0 \mathbf{E} - \mathbf{P}_k] \quad (60)$$

and

$$\frac{\partial \mathbf{P}_k}{\partial t} = \frac{1}{\tau_k} [(\epsilon_{sk} - \epsilon_\infty) \epsilon_0 \mathbf{E} - \mathbf{P}_k], \quad (61)$$

where $k = 1, 2, \dots, M$ in the last equation (Note: The subscript k denotes the pole number, not the spatial index of z). These latter two equations form the basis for two of the direct integration schemes.

5.1 Direct Integration Method One: D-DIM1

For the first direct integration technique, consider initially a single-pole model. Then

$$\epsilon_0 \epsilon_\infty \frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{H} - \frac{1}{\tau} [(\epsilon_s - \epsilon_\infty) \epsilon_0 \mathbf{E} - \mathbf{P}] \quad (62)$$

and

$$\frac{\partial \mathbf{P}}{\partial t} = \frac{1}{\tau} [(\epsilon_s - \epsilon_\infty) \epsilon_0 \mathbf{E} - \mathbf{P}]. \quad (63)$$

The components of $\mathbf{P}$ and $\mathbf{E}$ are conjoined on the Yee grid and the discretization of the curl operators is accomplished in the usual manner. Regarding the temporal discretization, we invoke the notion of the central difference and average approximations and the leap-frog time advancement [26]. To this end,

$$\frac{\mathbf{P}^{n+1/2} - \mathbf{P}^{n-1/2}}{\delta_t} = \frac{1}{\tau} \left[(\epsilon_s - \epsilon_\infty) \epsilon_0 \mathbf{E}^n - \left(\frac{\mathbf{P}^{n+1/2} + \mathbf{P}^{n-1/2}}{2} \right) \right] \quad (64)$$

and

$$\epsilon_0 \epsilon_\infty \left(\frac{\mathbf{E}^{n+1} - \mathbf{E}^n}{\delta_t} \right) = \nabla \times \mathbf{H}^{n+1/2} - \frac{1}{\tau} \left[(\epsilon_s - \epsilon_\infty) \epsilon_0 \left(\frac{\mathbf{E}^{n+1} + \mathbf{E}^n}{2} \right) - \mathbf{P}^{n+1/2} \right]. \quad (65)$$

The proceeding equations can be algebraically arranged so that only $\mathbf{P}^{n+1/2}$ and $\mathbf{E}^{n+1}$ appear on the left-hand side of the equations; see [26]. Note that in this implementation, the polarization vector $\mathbf{P}$ and the magnetic field vector $\mathbf{H}$ are advanced simultaneously and leaped with $\mathbf{E}$.

The basic strategy for M -poles is essentially the same as before, with the exception that all of the polarization equations are advanced simultaneously with Faraday's law [26]. When temporal averaging is invoked, the discretized temporal form becomes

$$\frac{\mathbf{P}_k^{n+1/2} - \mathbf{P}_k^{n-1/2}}{\delta_t} = \frac{1}{\tau_k} \left[(\epsilon_{sk} - \epsilon_\infty) \epsilon_0 \mathbf{E}^n - \left(\frac{\mathbf{P}_k^{n+1/2} + \mathbf{P}_k^{n-1/2}}{2} \right) \right] \quad (66)$$

and

$$\epsilon_0 \epsilon_\infty \left(\frac{\mathbf{E}^{n+1} - \mathbf{E}^n}{\delta_t} \right) = \nabla \times \mathbf{H}^{n+1/2} - \sum_{k=1}^M \frac{1}{\tau_k} \left[(\epsilon_{sk} - \epsilon_\infty) \epsilon_0 \left(\frac{\mathbf{E}^{n+1} + \mathbf{E}^n}{2} \right) - \mathbf{P}_k^{n+1/2} \right]. \quad (67)$$

Further manipulation of Eqns. (66) and (67) leads to the following update equations for $\mathbf{P}$ and $\mathbf{E}$:

$$\mathbf{P}_k^{n+1/2} = \alpha_k \mathbf{P}_k^{n-1/2} + \beta_k \delta_t \tau_k^{-1} (\epsilon_{sk} - \epsilon_\infty) \epsilon_0 \mathbf{E}^n, \quad (68)$$

where

$$\alpha_k = \frac{2\tau_k - \delta_t}{2\tau_k + \delta_t}, \quad (69)$$

and

$$\beta_k = \frac{2\tau_k}{2\tau_k + \delta_t}; \quad (70)$$

$$\mathbf{E}^{n+1} = c_1 \mathbf{E}^n + c_2 \delta_t \nabla \times \mathbf{H}^{n+1/2} + c_2 \delta_t \sum_{k=1}^M \tau_k^{-1} \mathbf{P}_k^{n+1/2}, \quad (71)$$

where

$$c_1 = \left[\epsilon_\infty - \frac{\delta_t}{2} \sum_{k=1}^M \frac{\epsilon_{sk} - \epsilon_\infty}{\tau_k} \right] \left[\epsilon_\infty + \frac{\delta_t}{2} \sum_{k=1}^M \frac{\epsilon_{sk} - \epsilon_\infty}{\tau_k} \right]^{-1} \quad (72)$$

and

$$c_2 = \left[\epsilon_\infty \epsilon_0 + \frac{\delta_t}{2} \sum_{k=1}^M \frac{(\epsilon_{sk} - \epsilon_\infty) \epsilon_0}{\tau_k} \right]^{-1}. \quad (73)$$

5.2 Direct Integration Method Two: D-DIM2

To have greater control over the stability properties of the scheme, semi-implicit schemes, such as those proposed by Kashiwa [25] and Petropolous [27] must be considered. (The reader may also consult [28] for further information on the semi-implicit approach.) For example, for M th-order media,

$$\epsilon_0 \epsilon_\infty \left(\frac{\mathbf{E}^{n+1} - \mathbf{E}^n}{\delta_t} \right) + \sum_{k=1}^M \left(\frac{\mathbf{P}_k^{n+1} - \mathbf{P}_k^n}{\delta_t} \right) = \nabla \times \mathbf{H}^{n+1/2} \quad (74)$$

and

$$\left(\frac{\mathbf{P}_k^{n+1} - \mathbf{P}_k^n}{\delta_t} \right) = \frac{1}{\tau_k} \left[(\epsilon_{sk} - \epsilon_\infty) \epsilon_0 \left(\frac{\mathbf{E}^{n+1} + \mathbf{E}^n}{2} \right) - \left(\frac{\mathbf{P}_k^{n+1} + \mathbf{P}_k^n}{2} \right) \right]; \quad (75)$$

here $k = 1, 2, \dots, M$.

5.3 Direct Integration Method Three: D-DIM3

Before leaving this section on direct integration techniques, we wish to consider an integration technique based upon the frequency domain permittivity relationship: $\hat{\mathbf{D}} = \hat{\epsilon} \hat{\mathbf{E}}$ [18]. As an example, consider the technique for a one-pole Debye dielectric. For this situation, the constitutive relationship can be written as

$$\hat{\mathbf{D}} = \epsilon_0 \left[\epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{1 + j\omega\tau} \right] \hat{\mathbf{E}}. \quad (76)$$

Or, in the time-domain,

$$\mathbf{D} + \tau \frac{\partial \mathbf{D}}{\partial t} = \epsilon_s \epsilon_0 \mathbf{E} + \epsilon_\infty \epsilon_0 \tau \frac{\partial \mathbf{E}}{\partial t}. \quad (77)$$

As before, central differences and averages are applied to the previous ordinary differential equation, whereby an update equation for $\mathbf{E}$ is obtained. Using this update equation and the corresponding update equation for Ampere's law, we obtain

$$\mathbf{D}^{n+1} = \mathbf{D}^n + \delta_t \nabla \times \mathbf{H}^{n+1/2} \quad (78)$$

and

$$\mathbf{E}^{n+1} = \left(\frac{\delta_t + 2\tau}{\epsilon_0(2\tau\epsilon_\infty + \epsilon_s\delta_t)} \right) \mathbf{D}^{n+1} + \left(\frac{\delta_t - 2\tau}{\epsilon_0(2\tau\epsilon_\infty + \epsilon_s\delta_t)} \right) \mathbf{D}^n + \left(\frac{2\tau\epsilon_\infty - \epsilon_s\delta_t}{2\tau\epsilon_\infty + \epsilon_s\delta_t} \right) \mathbf{E}^n. \quad (79)$$

Note: The update equation for $\mathbf{E}^{n+1}$ must be rederived for each additional pole considered in the dielectric model.

5.4 Recursive Convolution Method One: D-RCM1

The formulation of the recursive convolution method follows one-for-one with that of CP-RCM1. The only difference in that formulation and the one presented below is the equation for the accumulator, since the time-domain susceptibility function is different. To account for ϵ_∞ , the first-order update equation is rewritten as[29]

$$\mathbf{E}^{n+1} = \epsilon_\infty \alpha \mathbf{E}^n + \alpha \Psi^n + \frac{\alpha \delta_t}{\epsilon_0} \nabla \times \mathbf{H}^{n+1/2}, \quad (80)$$

where

$$\alpha = \frac{1}{\epsilon_\infty + \chi^0}; \quad (81)$$

here

$$\chi^0 = (\epsilon_s - \epsilon_\infty) (1 - e^{-\delta_t/\tau}). \quad (82)$$

With respect to the accumulator,

$$\Psi^n = \Delta\chi^0 \mathbf{E}^n + e^{-\delta_t/\tau} \Psi^{n-1}, \quad (83)$$

where

$$\Delta\chi^0 = \chi^0 (1 - e^{-\delta_t/\tau}). \quad (84)$$

5.5 Recursive Convolution Method Two: D-RCM2

To improve the temporal accuracy, a piecewise linear model may be employed in the integrand of the convolution integral [20]. (See also CP-RCM2 for the cold plasma.) The final equations are

$$\mathbf{E}^{n+1} = \alpha(\epsilon_\infty - \xi^0) \mathbf{E}^n + \frac{\alpha\delta_t}{\epsilon_0} \nabla \times \mathbf{H}^{n+1/2} + \alpha \Psi^n, \quad (85)$$

where

$$\alpha = \frac{1}{(\epsilon_\infty + \chi^0 - \xi^0)} \quad (86)$$

and

$$\Psi^n = \Delta\chi^0 \mathbf{E}^n + (\mathbf{E}^{n-1} - \mathbf{E}^n) \Delta\xi^0 + e^{-\delta_t/\tau} \Psi^{n-1}. \quad (87)$$

Here χ^0 and $\Delta\chi^0$ are given by Eqns. (82) and (84), respectively. As for ξ^0 and $\Delta\xi^0$,

$$\xi^0 = -(\epsilon_s - \epsilon_\infty) \left(\frac{\tau}{\delta_t} \right) \left[\left(\frac{\delta_t}{\tau} + 1 \right) e^{-\delta_t/\tau} - 1 \right] \quad (88)$$

and

$$\Delta\xi^0 = \xi^0 (1 - e^{-\delta_t/\tau}). \quad (89)$$

5.6 Comparative Analysis

As with the cold plasma, Table 4 documents the numerical permittivities for the previously reported schemes. From this table, the equivalence between D-DIM2 and D-DIM3 is seen. However, since D-DIM3 is constructed from the frequency-domain constitutive equation, D-DIM3 requires more memory than D-DIM2. Finally, D-RCM1 is first-order accurate in time; all other schemes are temporally second-order accurate.

To appreciate many of the subtle features of each of the aforementioned schemes, Tables 5 and 6 are given. These tables show the leading truncation term of the function $\Omega^2 \hat{\epsilon}/\epsilon_0$. From a temporal accuracy point of view, the scheme with the lowest truncation term for a given ω , τ , ϵ_s , ϵ_∞ and δ_t is deemed best. Consider Table 6, which shows the leading truncation term for the two cases $\omega\tau \ll 1$ and $\omega\tau \gg 1$. Except for D-RCM1, all other schemes have identical high-frequency truncation terms. These terms are the same truncation terms that the standard FDTD algorithm would produce in a dielectric of permittivity $\epsilon_\infty \epsilon_0$. In the low-frequency limit only D-DIM2 and D-DIM3 predict the anticipated truncation term. Thus it can be argued that D-DIM2 and D-DIM3 are the most compatible with the standard FDTD algorithm.

A summary of the stability properties and memory requirements of the various methodologies is provided in Table 7. Here we note that D-DIM2 is the most memory intensive. With respect to the stability equations,

Method	Numerical Permittivity: $\hat{\epsilon}$	
D-DIM1	$\hat{\epsilon} = \epsilon_{\infty}\epsilon_0 + \epsilon_0(\epsilon_s - \epsilon_{\infty})$	$\frac{\Lambda + j(\delta_t/2)^2(\Omega/\tau)}{\Lambda + j\Omega\tau}$
D-DIM2	$\hat{\epsilon} = \epsilon_{\infty}\epsilon_0 + (\epsilon_s - \epsilon_{\infty})\epsilon_0$	$\frac{\Lambda}{\Lambda + j\Omega\tau}$
D-DIM3	$\hat{\epsilon} = \epsilon_{\infty}\epsilon_0 + (\epsilon_s - \epsilon_{\infty})\epsilon_0$	$\frac{\Lambda}{\Lambda + j\Omega\tau}$
D-RCM1	$\hat{\epsilon} = \epsilon_0$	$\epsilon_{\infty} + \frac{\chi^0 e^{j\omega\delta_t/2}}{j\Omega\delta_t} - \frac{\Delta\chi^0 e^{\delta_t/(2\tau)}}{2j\Omega\delta_t \sinh[(\delta_t/\tau)(1+j\omega\tau)/2]}$
D-RCM2	$\hat{\epsilon} = \epsilon_0$	$\epsilon_{\infty} - \xi^0 + \frac{\chi^0 e^{j\omega\delta_t/2}}{j\Omega\delta_t} - \left(\frac{\Delta\chi^0 - j\Omega\delta_t \Delta\epsilon^0 e^{-j\omega\delta_t/2}}{2j\Omega\delta_t \sinh[(\delta_t/\tau)(1+j\omega\tau)/2]} \right) e^{\delta_t/(2\tau)}$

Table 4: Numerical permittivities for various schemes associated with propagation in single-pole Debye dielectric.

Method	Truncation Term	
D-DIM1	$\frac{\omega^4 \delta_t^2}{12}$	$(\epsilon_s - \epsilon_{\infty}) \left(\frac{3j}{\omega\tau} + \frac{j}{j-\omega\tau} - \frac{1}{(j-\omega\tau)^2} \right) - \epsilon_{\infty}$
D-DIM2	$\frac{\omega^4 \delta_t^2}{12}$	$(\epsilon_{\infty} - \epsilon_s) \left(\frac{2j}{j-\omega\tau} + \frac{1}{(j-\omega\tau)^2} \right) - \epsilon_{\infty}$
D-DIM3	$\frac{\omega^4 \delta_t^2}{12}$	$(\epsilon_{\infty} - \epsilon_s) \left(\frac{2j}{j-\omega\tau} + \frac{1}{(j-\omega\tau)^2} \right) - \epsilon_{\infty}$
D-RCM1	$\frac{j\omega^3 \delta_t}{2}$	$\frac{\epsilon_s - \epsilon_{\infty}}{1+j\omega\tau}$
D-RCM2	$\frac{\omega^4 \delta_t^2}{12}$	$\frac{j(\epsilon_{\infty} - 2\epsilon_s) + \epsilon_{\infty}\omega\tau}{j-\omega\tau}$

Table 5: Truncation term from the Taylor analysis of $\Omega^2 \hat{\epsilon}/\epsilon_0$: Debye dielectric.

Method	Truncation Term: $\omega\tau \ll 1$	Truncation Term: $\omega\tau \gg 1$
D-DIM1	$\left(\frac{\omega^4 \delta_t^2}{12} \right) [(\epsilon_s - \epsilon_{\infty}) (2 + \frac{3j}{\omega\tau}) - \epsilon_{\infty}]$	$\frac{-\omega^4 \delta_t^2 \epsilon_{\infty}}{12}$
D-DIM2	$\frac{-\omega^4 \delta_t^2 \epsilon_s}{12}$	$\frac{-\omega^4 \delta_t^2 \epsilon_{\infty}}{12}$
D-DIM3	$\frac{-\omega^4 \delta_t^2 \epsilon_s}{12}$	$\frac{-\omega^4 \delta_t^2 \epsilon_{\infty}}{12}$
D-RCM1	$\left(\frac{j\omega^3 \delta_t}{2} \right) (\epsilon_s - \epsilon_{\infty})$	$\left(\frac{\omega^2 \delta_t}{2\tau} \right) (\epsilon_s - \epsilon_{\infty})$
D-RCM2	$\left(\frac{\omega^4 \delta_t^2}{12} \right) (\epsilon_{\infty} - 2\epsilon_s)$	$\frac{-\omega^4 \delta_t^2 \epsilon_{\infty}}{12}$

Table 6: Truncation term from the Taylor analysis of $\Omega^2 \hat{\epsilon}/\epsilon_0$ for two special cases: Debye dielectric.

Method	3D Memory/Cell	1D Stability
D-DIM1	9	$\delta_t < \delta_x / v_p^{max}$
D-DIM2	9	$\delta_t < \sqrt{\epsilon_{\infty}} \delta_x / c$
D-DIM3	12	$\delta_t < \sqrt{\epsilon_{\infty}} \delta_x / c$
D-RCM1	9	$\delta_t < \delta_x / v_p^{max}$
D-RCM2	9	$\delta_t < \delta_x / v_p^{max}$

Table 7: Comparison table of various schemes in terms of memory per cell and stability equations: Single-pole Debye dielectric.

the entries for D-DIM1, D-RCM1 and D-RCM2 have not been rigorously established. The entry for D-DIM3 follows directly from the semi-implicit nature of the scheme [30]. Finally, since D-DIM2 and D-DIM3 are equivalent to one another, they also have the same stability equation.

To close this section, a comment on multi-pole Debye dielectrics is in order. Each of the schemes reviewed herein are capable of modeling the multi-pole dielectric. Some schemes like D-DIM1, D-RCM1 and D-RCM2 can be directly extended without any special treatment. Other schemes like D-DIM3 require a new derivation for each new pole added to the model. As for D-DIM2, the semi-implicit nature requires the loading and the inverting of a new matrix (each time a pole is added) to find the necessary coefficients of the algorithm.

5.7 Parameter Selection

To set the parameters δ_t and δ_x for the simulation, consider the example in which muscle is the medium. For a two-pole model, $\epsilon_\infty = 40$, $\epsilon_{s1} = 3948$, $\epsilon_{s2} = 59.09$, $\tau_1 = 46.25$ ns and $\tau_2 = .0907$ ns [23]. One possible way to deduce a value for δ_t is to make it a fraction of the smallest relaxation; for this example, let $\delta_t = \tau_2/10 = 9.07$ ps. To minimize dispersion errors, the CFL number should be maximized. Thus, from the definition of CFL let $\delta_x = \delta_t v_p^{max} = .430$ mm, where, for this example, $v_p^{max} = 4.74 \times 10^7$ m/s. A second way is to set δ_x in accordance with the minimum wavelength of interest. Let $\delta_x = \lambda^{min}/20$, where λ^{min} is the minimum wavelength at some angular frequency; for this example let $\omega = 10/\tau_2$, in which case $\lambda^{min} = 2.69$ mm and $\delta_x = .135$ mm. Finally, from the stability relationship, the maximum time step is $\delta_t = \delta_x/v_p^{max} = 2.84$ ps.

Based upon the previous example, the more conservative choice for the values of δ_x and δ_t is based upon the minimum wavelength argument. Unfortunately, this argument produces a time-step that over-resolves the time-domain driving signal. Moreover, by doing so, we incur the cost of increasing the simulation time by about a factor of 3.2 and the computational memory by a factor of 32.8 (three-dimensional domain), as compared with the other method. Had we chosen the first method, the high-frequency information of the medium would not be captured and hence, the early-time information of the output data would be in error. One possible way to resolve these contrary specifications is to increase the spatial accuracy order of the scheme [27]. For example, if a fourth-order in space and a second-order in time scheme is employed, the sampling requirements on the minimum wavelength can be relaxed and the resulting time-step will be commensurate with the smallest relaxation.

6 Summary

A survey of some of the more popular schemes associated with FDTD analysis of electromagnetic wave propagation in complex media has been given. These schemes have been characterized in terms of their numerical permittivities in conjunction with a numerical dispersion relationship. In addition, the CFL number of each scheme is given or postulated when such information has been provided in the literature. In some cases, discussions on memory requirements have also been provided.

It is not the intention of this report to rank such schemes in terms of *best* or *worst*. Such terms are subjective, since best or worst designators require one to quantify both the tangible (e.g., accuracy) and intangible (e.g., simplicity) features of a scheme into a single number. Instead, each scheme should be analyzed with a particular application in mind and, from that analysis, an informed selection can be made.

References

- [1] K. S. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Trans. Ant. Propagat.*, vol. 14, no. 3, pp. 302-307, 1966.
- [2] A. Taflov and M. E. Brodwin, "Numerical solution of steady-state electromagnetic scattering problems using the time-dependent Maxwell's equations," *IEEE Trans. Microwave Theory Tech.*, vol. 23, no. 8, pp. 623-630, 1975.
- [3] A. Ishimaru, *Electromagnetic Wave Propagation, Radiation and Scattering*, Prentice-Hall, Englewood Cliffs, NJ, 1991.
- [4] R. Richtmyer and K. Morton, *Difference Methods for Initial-Value Problems*, Wiley, New York, NY, 1967.
- [5] R. F. Harrington, *Time-Harmonic Electromagnetic Fields*, McGraw-Hill, New York, NY, 1961.
- [6] K. S. Kunz and R. J. Luebbers, *The Finite Difference Time Domain Method for Electromagnetics*, CRC Press, Boca Raton, FL, 1993.
- [7] Y. Liu, "Fourier analysis of numerical algorithms for the Maxwell equations," *31st Aerospace Sciences Meeting & Exhibit*, AIAA-93-0368, Reno, NV, 1993.
- [8] A. Taflov and K. R. Umashankar, "The finite-difference time-domain method for numerical modeling of electromagnetic wave interactions," *Electromagnetics*, vol. 10, no. 1/2, pp. 105-126, 1990.
- [9] D. A. Anderson, J. C. Tannehill and R. H. Pletcher, *Computational Fluid and Mechanics and Heat Transfer*, Taylor & Francis, Bristol, PA, 1984.
- [10] J. R. Wait, *Electromagnetic Waves in Stratified Media*, Pergamon Press, Oxford, UK, 1970.
- [11] Jackson, J. D., *Classical Electrodynamics*, Wiley, New York, NY, 1975.
- [12] D. M. Sullivan, "Frequency-dependent FDTD methods using Z transforms," *IEEE Trans. Ant. Propagat.*, vol. 40, no. 10, pp. 1223-1230, 1992.
- [13] D. M. Sullivan, "Z-transform theory and the FDTD method," *IEEE Trans. Ant. Propagat.*, vol. 44, no. 1, pp. 28-34, 1996.
- [14] W. H. Weedon and C. M. Rappaport, "A general method for FDTD modeling of wave propagation in arbitrary frequency-dependent media," *IEEE Trans. Ant. Propagat.*, vol. 45, no. 3, pp. 401-410, 1997.
- [15] J. L. Young, "A full finite difference time domain implementation for radio wave propagation in a plasma," *Radio Sci.*, vol. 29, no. 6, pp. 1513-1522, 1994.
- [16] L. J. Nickisch and P. M. Franke, "Finite-difference time-domain solution of Maxwell's equations for the dispersive ionosphere," *IEEE Ant. Propagat. Mag.*, vol. 34, no. 5, pp. 33-39, 1992.
- [17] S. A. Cummer, "An analysis of new and existing FDTD methods for isotropic cold plasma and a method for improving their accuracy," *IEEE Trans. Ant. Propagat.*, vol. 45, no. 3, pp. 392-400, 1997.
- [18] R. M. Joseph, S. C. Hagness and A. Taflov, "Direct time integration of Maxwell's equations in linear dispersive media with absorption for scattering and propagation of femtosecond electromagnetic pulses," *Opt. Lett.*, vol. 16, no. 18, pp. 1412-1414, 1991.
- [19] R. J. Luebbers, F. Hunsberger, and K. S. Kunz, "A frequency-dependent finite-difference time-domain formulation for transient propagation in plasma," *IEEE Trans. Ant. Propagat.*, vol. 39, no. 1, pp. 29-34, 1991.
- [20] D. F. Kelley and R. J. Luebbers, "Piecewise linear recursive convolution for dispersive media using FDTD," *IEEE Trans. Ant. Propagat.*, vol. 44, no. 6, pp. 792-797, 1996.
- [21] D. M. Sullivan, "A frequency-dependent FDTD method for biological applications," *IEEE Trans. Microwave Theory Tech.*, vol. 40, no. 3, pp. 532-539, 1992.
- [22] O. P. Gandhi and C. M. Furse, "Currents induced in the human body for exposure to ultrawideband electromagnetic pulses," *IEEE Trans. Electromagn. Compat.*, vol. 39, no. 2, pp. 174-180, 1997.

- [23] C. M. Furse, J.-Y. Chen and O. P. Ghandi, "The use of frequency-dependent finite-difference time-domain method for induced current and SAR calculations for a heterogeneous model of the human body," *IEEE Trans. Electromagn. Compat.*, vol. 36, no. 2, pp. 128-133, 1994.
- [24] W. D. Hurt, "Multiterm Debye dispersion relations for permittivity of muscle," *IEEE Trans. Biomed. Engr.*, vol. 32, no. 1, pp. 60-64, 1985.
- [25] T. Kashiwa, N. Yoshida and I. Fukai, "A treatment by the finite-difference time-domain method of the dispersive characteristics associated with orientation polarization," *IEICE Transactions*, vol. E73, no. 8, pp. 1326-1328, 1990.
- [26] J. L. Young, "Propagation in linear dispersive media: Finite difference time-domain methodologies," *IEEE Trans. Ant. Propagat.*, vol. 43, no. 4, pp. 422-426, 1995.
- [27] P. G. Petropoulos, "The computation of linear dispersive electromagnetic waves," *Applied Computational Electromagnetics Society Journal*, vol. 11, no. 1, pp. 8-16, 1996.
- [28] M. Okoniewski, M. Mrozowski and M.A. Stuchly, "Simple treatment of multi-term dispersion in FDTD," *IEEE Microwave Guided Wave Lett.*, vol. 7, no. 5, pp. 121-123, 1997.
- [29] R. Luebbers, F. P. Hunsberger, K. S. Kunz, R. B. Standler and M. Schneider, "A frequency-dependent finite-difference time-domain formulation for dispersive materials," *IEEE Trans. Electromagn. Compat.*, vol. 32, no. 3, pp. 222-227, 1990.
- [30] P. G. Petropoulos, "Stability and phase error analysis of FD-TD in dispersive dielectrics," *IEEE Trans. Ant. Propagat.*, vol. 42, no. 1, pp. 62-69, 1994.