UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1997

SUMMER FACULTY RESEARCH PROGRAM FINAL REPORTS

VOLUME 5A

WRIGHT LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Major Linda Steel-Goodwin

Program Manager, RDL
Scott Licoscos

Program Administrator, RDL
Johnetta Thompson

Program Administrator, RDL
Rebecca Kelly

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1997

20010319 033

AQM01-06-1194

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruc
the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, includ
Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budg

AFRL-SR-BL-TR-00-
0757

ing and reviewing
s for Information

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. F
	December, 1997	

4. TITLE AND SUBTITLE
1997 Summer Research Program (SRP), Summer Faculty Research Program (SFRP), Final Reports, Volume 5A, 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	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	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.
NASA -	Leave blank.
NTIS -	Leave blank.

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-xviii

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

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 three parts, 5A, 5B and 5C. Volume 5A contains #1-24. Volume 5B contains reports #25-48 and 5C contains #49-70. The Table of Contents for Volume 5 is included in all parts.

This document is one of a set of 16 volumes describing the 1997 AFOSR Summer Research Program. The following volumes comprise the set:

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

SRP Final Report Table of Contents

Author	University/Institution Report Title	Armstrong Laboratory Directorate	Vol-Page
DR Jean M Andino	University of Florida , Gainesville , FL Atmospheric Reactions of Volatille Paint Components a Modeling Approach	AL/EQL	2- 1
DR Anthony R Andrews	Ohio University , Athens , OH Novel Electrochemiluminescence Reactions and Instrumentation	AL/EQL	2- 2
DR Stephan B Bach	Univ of Texas at San Antonio , San Antonio , TX Investigation of Sampling Interfaces for Portable Mass Spectormetry and a survery of field Portable	AL/OEA	2- 3
DR Marilyn Barger	Florida A&M-FSU College of Engineering , Tallahassee , FL Analysis for The Anaerobic Metabolites of Toulene at Fire Training Area 23 Tyndall AFB, Florida	AL/EQL	2- 4
DR Dulal K Bhaumik	University of South Alabama , Mobile , AL The Net Effect of a Covariate in Analysis of Covariance	AL/AOEP	2- 5
DR Marc L Carter, PhD, PA	Hofstra University , Hempstead , NY Assessment of the Reliability of Ground Based Observeres for the Detecton of Aircraft	AL/OEO	2- 6
DR Huseyin M Cekirge	Florida State University , Tallahassee , FL Developing a Relational Database for Natural Attenuation Field Data	AL/EQL	2- 7
DR Cheng Cheng	Johns Hopkins University , Baltimore , MD Investigation of Two Statistical Issues in Building a Classification System	AL/HRM	2- 8
DR Gerald P Chubb	Ohio State University , Columbus , OH Use of Air Synthetic Forces For GCI Training Exercises	AL/HR1	2- 9
DR Sneed B Collard, Jr.	University of West Florida , Pensacola , FL Suitability of Ascidians as Trace Metal Biosenosrs-Biomonitors In Marine Environments An Assessment	AL/EQL	2- 10
DR Catherine A Cornwell	Syracuse University , Syracuse , NY Rat Ultrasoud Vocalization Development and Neurochemistry in Stress-Sensitive Brain Regions	AL/OER	2- 11

Author	University/Institution Report Title	Armstrong Laboratory Directorate	Vol-Pag
DR Baolin Deng	New Mexico Tech , Socorro , NM Effect of Iron Corrosion Inhibitors on Reductive Degradation of Chlorinated Solvents	AL/EQL	2- 1
DR Micheal P Dooley	Iowa State University , Ames , IA Copulatory Response Fertilizing Potential, and Sex Ratio of Offsprings Sired by male rats Ecposed in	AL/OER	2- 1
DR Itiel E Dror	Miami University , Oxford , OH The Effect of Visual Similarity and Reference Frame Alignment on the Recognition of Military Aircraf	AL/HRT	2- 1
DR Brent D Foy	Wright State University , Dayton , OH Advances in Biologivcally-Based Kinetic Modeling for Toxicological Applications	AFRL/HES	2- 1
DR Irwin S Goldberg	St. Mary's Univ , San Antonio , TX Mixing and Streaming of a Fluid Near the Entrance of a Tube During Oscillatory Flow	AL/OES	2- 1
DR Ramesh C Gupta	University of Maine at Orono , Orono , ME A Dynamical system approach in Biomedical Research	ALOES	2- 1
DR John R Herbold	Univ of Texas at San Antonio , San Antonio , TX A Protocol for Development of Amplicons for a Rapid and Efficient Methoiid of Genotyping Hepatitis C	AL/AOEL	2- 1
DR Andrew E Jackson	Arizona State University , Mesa , AZ Development fo a Conceptual Design for an Information Systems Infrastructure To Support the Squadron	AL/HRA	2- 1
DR Charles E Lance	Univ of Georgia Res Foundation , Athens , GA Replication and Extension of the Schmidt, Hunter, and Outerbridge (1986) Model of Job Performance R	AL/HRT	2- 2
DR David A Ludwig	Univ of N.C. at Greensboro , Greensboro , NC Mediating effect of onset rate on the relationship between+ Gz and LBNP Tolerance	AL/AOCY	2- 2
DR Robert P Mahan	University of Georgia , Athens , GA The Effects of Task Structure on Cognitive Organizing Principles Implaicatins for Complex Display	AL/CFTO	2- 2

Author	University/Institution Report Title	Armstrong Laboratory Directorate	Vol-Page
DR Phillip H Marshall	Texas Tech University , Lubbock , TX Preliminary report on the effects of varieties of feedback training on single target time-to-contac	AL/HRM	2- 23
DR Bruce V Mutter	Bluefield State College , Bluefield , WV	AL/EQP	2- 24
DR Allen L Nagy	Wright State University , Dayton , OH The Detection of Color Breakup In Field Sequential Color Displays	AL/CFHV	2- 25
DR Brent L Nielsen	Auburn University , Auburn , AL Rapid PCR Detection of Vancomycin Resistance of Enteroccus Species in infected Urine and Blood	AL/AOEL	2- 26
DR Thomas E Nygren	Ohio State University , Columbus , OH Group Differences in perceived importance of swat workload dimensions: Effects on judgment and perf	AL/CFHP	2- 27
DR Edward H Piepmeier	Oregon State University , Corvallis , OR	AL/AOHR	2- 28
DR Judy L Ratliff	Murray State Univ , Murray , KY Accumulation of Storntium and Calcium by Didemnum Conchyliatum	AL/EQL	2- 29
DR Joan R Rentsch	Wright State University , Dayton , OH the Effects of Individual Differences and Team Processed on Team Member Schema Similarity and task P	AL/CFHI	2- 30
DR Paul D Retzlaff	Univ of Northern Colorado , Greeley , CO The Armstrong Laboratory Aviation Personality Survey (ALAPS) Norming and Cross - Validation	AL/AOCN	2- 31
DR David B Reynolds	Wright State University , Dayton , OH Modeling Heat Flux Through Fabrics Exposed to a Radiant Souurce and Analysis of Hot Air Burns	AL/CFBE	2- 32
DR Barth F Smets	University of Connecticut , Storrs , CT Desorption and Biodegradation of Dinitrotoluenes in aged soils	AL/EQL	2- 33

Author	University/Institution Report Title	Phillips Laboratory Directorate	Vol-Page
DR Graham R Allan	National Avenue , Las Vegas , NM Temporal and Spatial Characterisation of a Synchronously-Pumped Periodically-Poled Lithium Niobate O	PL/LIDD	3- 1
DR Mark J Balas	Univ of Colorado at Boulder , Boulder , CO Nonlinear Tracking Control for a Precision Deployable Structure Using a Partitioned Filter Approach	PL/SX	3- 2
DR Mikhail S Belen'kii	Georgia Inst of Technology , Atlanta , GA Multiple Aperture Averaging Technique for Measurment Full Aperture Tilt with a Laser Guide Star and	PL/LIG	3- 3
DR Gajanan S Bhat	Univ of Tennessee , Knoxville , TN Spinning Hollow Fibers From High Performance Polymers	PL/RK	3- 4
DR David B Choate	Transylvania Univ , Lexington , KY Blackhole Analysis	PL/VTMR	3- 5
DR Neb Duric	University of New Mexico , Albuquerque , NM Image Recovery Using Phase Diversity	AFRL/DEB	3- 6
DR Arthur B Edwards	9201 University City Blvd. , Charlotte , NC Theory of Protons in Buried Oxides	PL/VTMR	3- 7
DR Gary M Erickson	Boston University , Boston , MA Modeling The Magnetospheric Magnetic Field	PL/GPSG	3- 8
DR Hany A Ghoneim	Rochester Inst of Technol , Rochester , NY Focal Point Accuracy Assessement of an Off-Axis Solar Caoncentrator	PL/RKES	3- 9
DR Subir Ghosh	Univ of Calif, Riverside , Riverside , CA Designing Propulsion Reliability of Space Launch Vehicles	PL/RKBA	3- 10
DR George W Hanson	Univ of Wisconsin - Milwaukee , Milwaukee , WI Asymptotic analysis of the Natural system modes of coupled bodies in the large separatin, Low-Freque	AFRL/DEH	3- 11

SRP Final Report Table of Contents

Author	University/Institution Report Title	Phillips Laboratory Directorate	Vol-Page
DR Brian D Jeffs	Brigham Young University , Provo , UT Blind Bayyesian Restoration of Adaptive Optics Images Using Generalized Gaussian Markov Random Field	AFRL/DES	3- 12
DR Christopher H Jenkins	S Dakota School of Mines/Tech , Rapid City , SD Mechnics of Surface Precosion for Membrane Reflectors	PL/VTVS	3- 13
DR Dikshitulu K Kalluri	University of Lowell , Lowell , MA Mode Conversion in a Time-Varying Magnetoplasma Medium	PL/GPID	3- 14
DR Aravinda Kar	University of Central Florida , Orlando , FL Measurement of the Cutting Performance of a High Beam Quality Chemical Oxygen-Iodine Laser on Aerosp	AFRL/DEO	3- 15
DR Bernard Kirtman	Univ of Calif, Santa Barbara , Santa Barbara , CA Quantum Chemical Characterization of the elkectronic Structure and Reactions of Silicon Dangling Bon	PL/VTMR	3- 16
DR Spencer P Kuo	Polytechnic University , Farmingdale , NY Excitation of Oscillating Two Stream Instability by Upper Hybrid Pump Waves in Ionospheric Heating	PL.GPI	3- 17
DR Henry A Kurtz	Memphis State University , Memphis , TN H2 Reactions at Dangling Bonds in SIO2	PL/VTMR	3- 18
DR Min-Chang Lee	Massachusetts Inst of Technology , Cambridge , MA Laboratory Studies of Ionospheric Plasma Effects Produced by Lightning-induced Whistler Waves	PL/GPSG	3- 19
DR Donald J Leo	University of Toledo , Toledo , OH Microcontroller-Based Implementation of Adaptive Structural Control	AFRL/VSD	3- 20
DR Hua Li	University of New Mexico , Albuquerque , NM	PL/LIDD	3- 21
DR Hanli Liu	Univ of Texas at Arlington , Arlington , TX Experimental Validation of Three-Dimensional Reconstruction of Inhomogenety Images in Turbid Media	AFRL/DEB	3- 22

Author	University/Institution Report Title	Phillips Laboratory Directorate	Vol-Pag
DR M. Arfin K Lodhi	Texas Tech University , Lubbock , TX Thermoelectric Energy Conversion with solid Electolytes	PL/VTRP	3- 2
DR Tim C Newell	University of New Mexico , Albuquerque , NM Study of Nonlineaar Dynamics in a Diode Pumped Nd:YAG laser	PL/LIGR	3- 2
DR Michael J Pangia	Georgia College & State University , Milledgeville , GA Preparatory Work Towards a Computer Simulation of Electron beam Operations on TSS 1	PL/GPSG	3- 2
DR Vladimir O Papitashvili	Univ of Michigan , Ann Arbor , MI Modeling of Ionospheric Convectin from the IMF and Solar Wind Data	PL/GPSG	3- 2
DR Jaime Ramirez-Angulo	New Mexico State University , Las Cruces , NM	PL/VTMR	3- 2
DR Louis F Rossi	University of Lowell , Lowell , MA Analysis of Turbulent Mixing in the Stratosphere & Troposphere	PL/GPOL	3- 2
DR David P Stapleton	University of Central Oklahoma , Edmond , OK Atmospheric Effects Upon Sub-Orbital Boost glide Spaceplane Trajectories	PL/RKBA	3- 2
DR Jenn-Ming Yang	Univ of Calif, Los Angeles , Los Angeles , CA Thermodynamic Stability and Oxidation Behavior of Refractory (Hf, Ta, Zr) Carbide/boride Composites	PL/RKS	3- 3

SRP Final Report Table of Contents

Author	University/Institution Report Title	Rome Laboratory Directorate	Vol-Page
DR A. F Anwar	University of Connecticut , Storrs , CT Properties of Quantum Wells Formed In AlGaN/GaN Heterostructures	RL/ERAC	4- 1
DR Milica Barjaktarovic	Wilkes University , Wilkes Barre , PA Assured Software Design: Privacy Enhanced Mail (PEM) and X.509 Certificate Specification	AFRL/IFG	4- 2
DR Stella N Batalama	SUNY Buffalo , Buffalo , NY Adaptive Robust Spread-Spectrum Receivers	AFRL/IFG	4- 3
DR Adam W Bojanczyk	Cornell Univesity , Ithaca , NY Lowering the Computational Complexity of Stap Radar Systems	RL/OCSS	4- 4
DR Nazeih M Botros	So. Illinois Univ-Carbondale , Carbondale , IL A PC-Based Speech Synthesizing Using Sinusoidal Transform Coding (STC)	RL/ERC-1	4- 5
DR Nikolaos G Bourbakis	SUNY Binghamton , Binghamton , NY Eikones-An Object-Oriented Language Forimage Analysis & Process	AFRL/IF	4- 6
DR Peter P Chen	Louisiana State University , Baton Rouge , LA Reconstructing the information Warfare Attack Scenario Guessing what Actually Had Happened Based on	RL/CA-II	4- 7
DR Everett E Crisman	Brown University , Providence , RI A Three-Dimensional, Dielectric Antenna Array Re-Configurable By Optical Wavelength Multiplexing	RL/ERAC	4- 8
DR Digendra K Das	SUNYIT , Utica , NY A Study of the Emerging Dianostic Techniques in Avionics	RL/ERSR	4- 9
DR Venugopala R Dasigi	Southern Polytechnic State Univ , Marietta , GA Information Fusion for text Classification-an Expjerimental Comparison	AFRL/IFT	4- 10
DR Richard R Eckert	SUNY Binghamton , Binghamton , NY Enhancing the rome Lab ADII virtual environment system	AFRL/IFSA	4- 11

Author	University/Institution Report Title	Rome Laboratory Directorate	Vol-Page
DR Micheal A Fiddy	University of Lowell , Lowell , MA Target Identification from Limited Backscattered Field Data	RL/ERCS	4- 12
DR Lili He	Nothern Illinois University , Dekalb , IL the Study of Caaractreistics of CdS Passivation on InP	RL/EROC	4- 13
DR Edem Ibragimov	Michigan Tech University , Houghton , MI Effects of Surface Scattering in 3-D Optical Mass Storage	RL/IRAP	4- 14
DR Phillip G Kornreich	Syracuse University , Syracuse , NY Analysis of Optically Active Material Layer Fibers	RL/OCPA	4- 15
DR Kuo-Chi Lin	University of Central Florida , Orlando , FL A Study on The Crowded Airspace Self Organized Criticality	AFRL/IFSB	4- 16
Dr. Beth L Losiewicz	Colorado College , Colorado Spring , CO The Miami Corpus Latin American Dialect Database continued Research and Documentation	RL/IRAA	4- 17
DR John D Norgard	Univ of Colorado at Colorado Springs , Colorado Spring , CO Microwave Holography using Infrared Thermograms of Electromagnetic Fields	RL/ERST	4- 18
DR Jeffrey B Norman	Vassar College , Poughkeepsie , NY Gain Spectra of Beam-Coupling In Photorefractive Semiconductors	RL/OCPA	4- 19
DR Dimitrios N Pados	State Univ. of New York Buffalo , Buffalo , NY Joint Domain Space-Time Adaptive Processing w/Small Training Data Sets	AFRL/SNR	4- 21
DR Brajendra N Panda	University of North Dakota , Grand Forks , ND A Model to Attain Data Integrity After System Invasion	AFRL/IFG	4- 22
DR Michael A Pittarelli	SUNY OF Tech Utica , Utica , NY Phase Transitions in probability Estimation and Constraint Satisfaction Problems	AFRL/IFT	4- 23

SRP Final Report Table of Contents

Author	University/Institution Report Title	Rome Laboratory Directorate	Vol-Page
DR Salahuddin Qazi	SUNY OF Tech Utica , Utica , NY Low Data rate Multimedia Communication Using Wireless Links	RL/IWT	4- 24
DR Arindam Saha	Mississippi State University , Mississippi State , MS An Implementationa of the message passing Interface on Rtems	RL/OCSS	4- 25
DR Ravi Sankar	University of South Florida , Tampa , FL A Study ofIntegrated and Intelligent Network Management	RL/C3BC	4- 26
DR Mark S Schmalz	University of Florida , Gainesville , FL Errors inherent in Reconstruction of Targets From multi-Look Imagery	AFRL/IF	4- 27
DR John L Stensby	Univ of Alabama at Huntsville , Huntsville , AL Simple Real-time Tracking Indicator for a Frequrncy Feedback Demodulator	RL/IRAP	4- 28
DR Micheal C Stinson	Central Michigan University , Mt. Pleasant , MI Destructive Objects	RL/CAII	4- 29
DR Donald R Ucci	Illinois Inst of Technology , Chicago , IL Simulation of a Robust Locally Optimum Receiver in correlated Noise Using Autoregressive Modeling	RL/C3BB	4- 30
DR Nong Ye	Arizona State University , Tempe , AZ A Process Engineering Approach to Continuous Command and Control on Security-Aware Computer Networks	AFRL/IFSA	4- 31

SRP Final Report Table of Contents

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Pag
DR William A Baeslack	Ohio State University , Columbus , OH	WL/MLLM	5-
DR Bhavik R Bakshi	Ohio State University , Columbus , OH Modeling of Materials Manufacturing Processes by NonlinearContimuum Regression	WL/MLIM	5-
DR Brian P Beecken	Bethel College , St. Paul , MN Contribution of a Scene Projector's Non-Uniformity to a Test Article's Output Image Non-Uniformity	AFRL/MN	5-
DR John H Beggs	Mississippi State University , Mississippi State , MS The Finite Element Method in Electromagnetics For Multidisciplinary Design	AFRL/VA	5-
DR Kevin D Belfield	University of Detroit Mercy , Detroit , MI Synthesis of Novel Organic Compounds and Polymers for two Photon Asorption, NLO, and Photorefractive	WL/MLBP	5-
DR Raj K Bhatnagar	University of Cincinnati , Cincinnati , OH A Study of Intra-Class Variability in ATR Systems	AFRL/SN	5-
DR Victor M Birman	Univ of Missouri - St. Louis , St Louis , MO Theoretical Foundations for Detection of Post-Processing Cracks in Ceramic Matrix Composites Based o	WL/FIBT	5-
DR Gregory A Blaisdell	Purdue University , West Lafayette , IN A Review of Benchmark Flows for Large EddySimulation	AFRL/VA	5-
DR Octavia I Camps	Pennsylvania State University , University Park , PA MDL Texture Segmentation Compressed Images	WL/MNGA	5-
DR Yiding Cao	Florida International Univ , Miami , FL A Feasibility Study of Turbine Disk Cooling by Employing Radially Rotating Heat Pipes	WL/POTT	5- 10
DR Reaz A Chaudhuri	University of Utah , Salt Lake City , UT A Novel Compatibility/Equilibrium Based Interative Post-Processing Approach For Axisymmetric brittle	WL/MLBM	5- 11

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Page
DR Mohamed F Chouikha	Howard University , Washington , DC Detection Techniques Use in Forward-Looking Radar Signal Procesing a Literature Review	WL/AAMR	5- 12
DR Milton L Cone	Embry-Riddle Aeronautical University , Prescott , AZ Scheduling in the Dynamic System Simulation Testbed	WL/AACF	5- 13
DR Robert C Creese	West Virginia University , Morgantown , WV Feature Based Cost Modeling	WL/MTI	5- 14
DR William Crossley	Purdue University , West Lafayette , IN Objects and Methods for Aircraft Conceptual Design and Optimization in a Knowledge-Based Environment	WL/FIBD	5- 15
DR Gene A Crowder	Tulane University , New Orleans , LA Vibrational Analysis of some High-Energy Compounds	WL/MNM	5- 16
DR Richard W Darling	University of South Florida , Tampa , FL Geometrically Invariant NonLinear recursive Filters, with Applicaation to Target Tracking	WL/MNAG	5- 17
DR Robert J DeAngelis	Univ of Nebraska - Lincoln , Lincoln , NE Quantitative Description of Wire Tecxtures In Cubic Metals	WL/MNM	5- 18
DR Bill M Diong	Pan American University , Edinburg , TX Analysis and Control Design for a Novel Resonant DC-DC Converter	WL/POOC	5- 19
DR John K Douglass	University of Arizona , Tucson , AZ Guiding Missiles "On The Fly:" Applications of Neurobiologica Princioles to Machine Vision For Arma	AFRL/MN	5- 20
DR Mark E Eberhart	Colorado School of Mines , Golden , CO Modeling The Charge Redistribution Associated with Deformation and Fracture	WL/MLLM	5- 21
DR Gregory S Elliott	Rutgers:State Univ of New Jersey , Piscataway , NJ On the Development of Planar Doppler Velocimetry	WL/POPT	5- 22

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Pag
DR Elizabeth A Ervin	University of Dayton , Dayton , OH Eval of the Pointwise K-2 Turbulence Model to Predict Transition & Separtion in a Low Pressure	WL/POTT	5- 2
DR Altan M Ferendeci	University of Cincinnati , Cincinnati , OH Vertically Interconnected 3D MMICs with Active Interlayer Elements	WL/AADI	5- 2
DR Dennis R Flentge	Cedarville College , Cedarville , OH Kinetic Study of the Thermal Decomposition of t-Butylphenyl Phosphate Using the System for Thermal D	WL/POSL	5- 2
DR George N Frantziskonis	University of Arizona , Tuson , AZ Multiscale Material Characterization and Applications	WL/MLLP	5- 2
DR Zewdu Gebeyehu	Tuskegee University , Tuskegee , AL Synthesis and Characterization of Metal-Xanthic Acid and -Amino Acid Com[lexes Useful Ad Nonlinear	WL/MLPO	5- 2
DR Richard D Gould	North Carolina State U-Raleigh , Raleigh , NC Reduction and Analysis of LDV and Analog Raw Data	WL/POPT	5- 2
DR Michael S Grace	University of Virginia , Charlottesville , VA Structure and Function of an Extremely Sensitive Biological Infrared Detector	WL/MLPJ	5- 2
DR Gary M Graham	Ohio University , Athens , OH Indicial Response Model for Roll Rate Effects on A 65-Degree Delta wing	WL/FIGC	5- 3
DR Allen G Greenwood	Mississippi State University , Mississippi Sta , MS An Object-Based approach for Integrating Cost Assessment into Product/Process Design	WL/MTI	5- 3
DR Rita A Gregory	Georgia Inst of Technology , Atlanta , GA Range Estimating for Research and Development Alternatives	WL/FIVC	5- 3
DR Mark T Hanson	University of Kentucky , Lexington , KY Anisotropy in Epic 96&97: Implementation and Effects	WL/MNM	5- 3

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Page
DR Majeed M Hayat	University of Dayton , Dayton , OH A Model for Turbulence and Photodetection Noise in Imaging	WL/AAJT	5- 34
DR Larry S Helmick	Cedarville College , Cedarville , OH NMA Study of the Decomposition Reaction Path of Demnum fluid under Tribological Conditions	WL/MLBT	5- 35
DR William F Hosford	Univ of Michigan , Ann Arbor , MI INTENSITY OF [111]AND [100] TEXTURAL COMPONENTS IN COMPRESSION-FORGED TANTALUM	AFRL/MN	5- 36
DR David E Hudak	Ohio Northern University , Ada , OH A Study fo a Data-Parallel Imlementation of An Implicit Solution fo the 3D Navier-Stokes Equations	WL/FIMC	5- 37
DR David P Johnson	Mississippi State University , Mississippi , MS An Innovative Segmented Tugsten Penetrating Munition	WL/MNAZ	5- 38
DR Ismail I Jouny	Lafayette College , Easton , PA	WL/AACT	5- 39
DR Edward T Knobbe	Oklahoma State University , Stillwater , OK Organically Modified silicate Films as Corrosion Resistant Treatments for 2024-T3 Alumium Alloy	WL/MLBT	5- 40
DR Seungug Koh	University of Dayton , Dayton , OH Numerically Efficinet Direct Ray Tracing Algorithms for Automatic Target Recognition using FPGAs	WL/AAST	5- 41
DR Ravi Kothari	University of Cincinnati , Cincinnati , OH A Function Approximation Approach for Region of Interest Selection in synthetic Aperture Radar Image	WL/AACA	5- 42
DR Douglas A Lawrence	Ohio University , Athens , OH On the Analysis and Design of Gain scheduled missile Autopilots	WL/MNAG	5- 43
DR Robert Lee	Ohio State University , Columbus , OH Boundary Conditions applied to the Finite Vlume Time Domain Method for the Solution of Maxwell's Equ	WL/FIM	5- 44

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Pag
DR Junghsen Lieh	Wright State University , Dayton , OH Develop an Explosive Simulated Testing Apparatus for Impact Physics Research at Wright Laboratory	WL/FIV	5- 45
DR James S Marsh	University of West Florida , Pensacola , FL Distortion Compensation and Elimination in Holographic Reocnstruction	WL/MNSI	5- 46
DR Mark D McClain	Cedarville College , Cedarville , OH A Molecular Orbital Theory Analysis of Oligomers of 2,2'-Bithiazole and Partially Reduced 3,3'-Dimet	WL/MLBP	5- 47
DR William S McCormick	Wright State University , Dayton , OH Some Observations of Target Recognition Using High Range Resolution Radar	WL/AACR	5- 48
DR Richard O Mines	University of South Florida , Tampa , FL Testing Protocol for the Demilitarization System at the Eglin AFB Herd Facility	WLMN/M	5- 49
DR Dakshina V Murty	University of Portland , Portland , OR A Useful Benchmarking Method in Computational Mechanics, CFD, adn Heat Tansfer	WL/FIBT	5- 50
DR Krishna Naishadham	Wright State University , Dayton , OH	WL/MLPO	5- 51
DR Serguei Ostapenko	University of South Florida , Tampa , FL	WL/MLPO	5- 52
DR Yi Pan	University of Dayton , Dayton , OH Improvement of Cache Utilization and Parallel Efficiency of a Time-Dependnet Maxwell Equation Solver	AFRL/VA	5- 53
DR Rolfe G Petschek	Case Western Reserve Univ , Cleveland , OH AB INITIO AUANTUM CHEMICAL STUDIES OF NICKEL DITHIOLENE COMPLEX	WL/MLPJ	5- 54
DR Kishore V Pochiraju	Stevens Inst of Technology , Hoboken , NJ Refined Reissner's Variational Solution in the Vicinity of Stress Singularities	AFRL/ML	5- 55

SRP Final Report Table of Contents

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Page
DR Muhammad M Rahman	University of South Florida , Tampa , FL Computation of Free Surface Flows with Applications in Capillary Pumped Loops. Heat Pipes, and Jet I	WL/POOB	5- 56
DR Mateen M Rizki	Wright State University , Dayton , OH Classification of High Range Resolution Radar Signatures Using Evolutionary Computation	WL/AACA	5- 57
DR Shankar M Sastry	Washington University , St Louis , MO	WL/MLLM	5- 58
DR Martin Schwartz	University of North Texas , Denton , TX Computational Studies of Hydrogen Abstraction From Haloalkanes by the Hydroxyl Radical	WL/MLBT	5- 59
DR Rathinam P Selvam	Univ of Arkansas , Fayetteville , AR Computation of Nonlenear Viscous Panel Flutter Using a Full-Implicit Aeroelastic Solver	WL/FIMC	5- 60
DR Yuri B Shtessel	Univ of Alabama at Huntsville , Huntsville , AL Smoothed Sliding Mode control Approach For Addressing Actuator Deflection and Deflection Rate Satura	AFRL/VA	5- 61
DR Mario Sznaier	Pennsylvania State University , University Park , PA Suboptimal Control of Nonlenear Systems via Receding Horizon State Dependent Riccati Equations	WL/MNAG	5- 62
DR Barney E Taylor	Miami Univ. - Hamilton , Hamilton , OH Photoconductivity Studies of the Polymer 6FPBO	WLMLBP	5- 63
DR Joseph W Tedesco	Auburn University , Auburn , AL high Velocity Penetration of Layered Concrete Targets with Small Scale Ogive-nose Steel projectiles	WL/MNSA	5- 64
DR Krishnaprasad Thirunarayan	Wright State University , Dayton , OH A VHDL MODEL SYNTHESIS APPLET IN TCL/TK	WL/AAST	5- 65

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Pag
DR Karen A Tomko	Wright State University , Dayton , OH Grid Level Parallelization of an Implicit Solution of the 3D Navier-Stokes Equations	WL/FIMC	5- 6(
DR Max B Trueblood	University of Missouri-Rolla , Rolla , MO AStudy of the Particulate Emissions of a Well-Stirred Reactor	WL/POSC	5- 6'
DR Chi-Tay Tsai	Florida Atlantic University , Boca Raton , FL Dislocation Dynamics in Heterojunction Bipolar Transistor Under Current Induced Thermal St	WL/AA	5- 6(
DR John L Valasek	Texas A&M University , College Station , TX Two Axis Pneumatic Vortex Control at High Speed and Low Angle-of-Attack	WL/FIMT	5- 6(
DR Mitch J Wolff	Wright State University , Dayton , OH An Experimaental and Computational Analysis of the Unsteady Blade Row Potential Interaction in a Tr	WL/POTF	5- 7(
DR Rama K Yedavalli	Ohio State University , Columbus , OH Improved Aircraft Roll Maneuver Performance Using Smart Deformable Wings	WL/FIBD	5- 7'

Author	University/Institution Report Title	Arnold Engineering Development Center Directorate	Vol-Page
DR Csaba A Biegl	Vanderbilt University , Nashville , TN Parallel processing for Tubine Engine Modeling and Test Data validation	AEDC/SVT	6- 1
DR Frank G Collins	Tennessee Univ Space Institute , Tullahoma , TN Design of a Mass Spectrometer Sampling Probe for The AEDC Impulse Facility	AEDC	6- 2
DR Kenneth M Jones	N Carolina A&T State Univ , Greensboro , NC	AEDC/SVT	6- 3
DR Kevin M Lyons	North Carolina State U-Raleigh , Raleigh , NC Velocity Field Measurements Using Filtered-Rayleigh Scattering	AEDC/SVT	6- 4
DR Gerald J Micklow	Univ of Alabama at Tuscaloosa , Tucasloosa , AL	AEDC/SVT	6- 5
DR Michael S Moore	Vanderbilt University , Nashville , TN Extension and Installation of the Model-Integrated Real-Time Imaging System (Mirtis)	AEDC/SVT	6- 6
DR Robert L Roach	Tennessee Univ Space Institute , Tullahoma , TN Investigation of Fluid Mechanical Phenomena Relating to Air Injection Between the Segments of an Arc	AEDC	6- 7
DR Nicholas S Winowich	University of Tennessee , Knoxville , TN	AEDC	6- 8
DR Daniel M Knauss	Colorado School of Mines , Golden , CO Synthesis of salts With Delocalized Anions For Use as Third Order Nonlinear Optical Materials	USAFA/DF	6- 9
DR Jeffrey M Bigelow	Oklahoma Christian Univ of Science & Art , Oklahoma City , OK Raster-To-Vector Conversion of Circuit Diagrams: Software Requirements	OCALC/TI	6- 10

Author	University/Institution Report Title	Arnold Engineering Development Center Directorate	Vol-Pag
DR Paul W Whaley	Oklahoma Christian Univ of Science & Art , Oklahoma City , OK A Probabilistic framework for the Analysis of corrosion Damage in Aging Aircraft	OCALC/L	6- 1
DR Bjong W Yeigh	Oklahoma State University , Stillwater , OK Logistics Asset Management : Models and Simulations	OCALC/TI	6- 1
DR Michael J McFarland	Utah State University , Logan , UT Delisting of Hill Air Force Base's Industrial Wastewater Treatment Plant Sludge	OC-ALC/E	6- 1
DR William E Sanford	Colorado State University , Fort Collins , CO Nuerical Modeling of Physical Constraints on in-SItu Cosolvent Flushing as a Groundwater Remedial Op	OO-ALC/E	6- 1
DR Sophia Hassiotis	University of South Florida , Tampa , FL Fracture Analysis of the F-5, 15%-Spar Bolt	SAALC/TI	6- 15
DR Devendra Kumar	CUNY-City College , New York , NY A Simple, Multiversion Concurrency Control Protocol For Internet Databases	SAALC/LD	6- 16
DR Ernest L McDuffie	Florida State University , Tallahassee , FL A Propossed Exjpert System for ATS Capability Analysis	SAALC/TI	6- 17
DR Prabhaker Mateti	Wright State University , Dayton , OH How to Provide and Evaluate Computer Network Security	SMALC/TI	6- 18
DR Mansur Rastani	N Carolina A&T State Univ , Greensboro , NC Optimal Structural Design of Modular Composite bare base Shelters	SMALC/L	6- 19
DR Joe G Chow	Florida International Univ , Miami , FL Re-engineer and Re-Manufacture Aircraft Sstructural Components Using Laser Scanning	WRALC/TI	6- 20

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

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 1997 are shown in this table.

1997 Applicants and Participants			
PARTICIPANT CATEGORY	TOTAL APPLICANTS	SELECTEES	DECLINING SELECTEES
SFRP	490	188	32
(HBCU/MI)	(0)	(0)	(0)
GSRP	202	98	9
(HBCU/MI)	(0)	(0)	(0)
HSAP	433	140	14
TOTAL	1125	426	55

4. SITE VISITS

During June and July of 1997, 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/MIs)

Before 1993, an RDL program representative visited from seven to ten different HBCU/MIs annually to promote interest in the SRP among the faculty and graduate students. These efforts were marginally effective, yielding a doubling of HBCI/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 the 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. This year, 5 HBCU/MI SFRPs declined after they were selected (and there was no one qualified to replace them with). The following table records HBCU/MI participation in this program.

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 1997 SRP were the AFOSR-provided slots for the basic contract and laboratory funds. Funding sources by category for the 1997 SRP selected participants are shown here.

1997 SRP FUNDING CATEGORY	SFRP	GSRP	HSAP
AFOSR Basic Allocation Funds	141	89	123
USAF Laboratory Funds	48	9	17
HBCU/MI By AFOSR (Using Procured Addn'l Funds)	0	0	N/A
TOTAL	9	98	140

SFRP - 188 were selected, but thirty two canceled too late to be replaced.
GSRP - 98 were selected, but nine canceled too late to be replaced.
HSAP - 140 were selected, but fourteen canceled too late to be replaced.

7. COMPENSATION FOR PARTICIPANTS

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

1997 SRP Associate Compensation

PARTICIPANT CATEGORY	1991	1992	1993	1994	1995	1996	1997
Faculty Members	$690	$718	$740	$740	$740	$770	$770
Graduate Student (Master's Degree)	$425	$442	$455	$455	$455	$470	$470
Graduate Student (Bachelor's Degree)	$365	$380	$391	$391	$391	$400	$400
High School Student (First Year)	$200	$200	$200	$200	$200	$200	$200
High School Student (Subsequent Years)	$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 $50/day for faculty and $40/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 % (194 out of 286) claimed travel reimbursements at an average round-trip cost of $776.

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 $388. By contrast, in 1993, 58 % of SFRP associates took

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. These 1993 numbers are included to show the fluctuation which can occur in these numbers for planning purposes.

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 second 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, only 7% of the associates' information needed clarification in order for direct deposit to properly function as opposed to 10% 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 1997 REPORT

The complete set of reports for the 1997 SRP includes this program management report (Volume 1) augmented by fifteen volumes of final research reports by the 1997 associates, as indicated below:

1997 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, WHMC	6	11	16

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	189
GSRP	97
HSAP	140
TOTAL	426

Degrees Represented			
	SFRP	GSRP	TOTAL
Doctoral	184	0	184
Master's	2	41	43
Bachelor's	0	56	56
TOTAL	186	97	298

SFRP Academic Titles	
Assistant Professor	64
Associate Professor	70
Professor	40
Instructor	0
Chairman	1
Visiting Professor	1
Visiting Assoc. Prof.	1
Research Associate	9
TOTAL	186

Source of Learning About the SRP		
Category	Applicants	Selectees
Applied/participated in prior years	28%	34%
Colleague familiar with SRP	19%	16%
Brochure mailed to institution	23%	17%
Contact with Air Force laboratory	17%	23%
IEEE Spectrum	2%	1%
BIIHE	1%	1%
Other source	10%	8%
TOTAL	100%	100%

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	275
HSAPs	113
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. 1997 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)

		How Many Associates Would You Prefer To Get ? (% Response)											
		SFRP				GSRP (w/Univ Professor)				GSRP (w/o Univ Professor)			
Lab	Evals Recv'd	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
USAFA	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 Reponses to above questions

	AEDC	*AL*	*USAFA*	*PL*	*RL*	*WHMC*	*WL*
# Evals Recv'd	0	7	1	14	5	0	46
Question #							
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. 1997 SFRP & GSRP EVALUATION RESPONSES

The summarized results listed below are from the 257 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 1997 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. 1997 USAF LABORATORY HSAP MENTOR EVALUATION RESPONSES

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

5. 1997 HSAP EVALUATION RESPONSES

The summarized results listed below are from the 113 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%

Associate did not participate in the program.

MODELING OF MATERIALS MANUFACTURING PROCESSES BY NONLINEAR CONTINUUM REGRESSION

Bhavik R. Bakshi
Assistant Professor
Department of Chemical Engineering

The Ohio State University
Columbus, OH 43210

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory, Materials Directorate

September 1997

MODELING OF MATERIALS MANUFACTURING PROCESSES BY NONLINEAR CONTINUUM REGRESSION

Bhavik R. Bakshi
Assistant Professor
Department of Chemical Engineering
The Ohio State University

Abstract

Since processes for the manufacturing of microelectronic materials are not very well understood, their modeling has to rely on methods that extract models from measured data. A wide variety of these empirical modeling methods is available such as various neural and statistical methods. This report describes the application of a new empirical modeling method called nonlinear continuum regression (NLCR) to various problems relevant to materials science and manufacturing. NLCR is a method that unifies neural and statistical modeling methods that combine inputs by linear projection before transformation by the activation function. These methods include ordinary least squares regression, principal component regression, partial least squares, backpropagation networks, projection pursuit regression, nonlinear principal component regression, and nonlinear partial least squares. Since these methods lie on a continuum the NLCR methodology automatically specializes to the method on this continuum that provides the smallest error of approximation. The NLCR method is applied to data generated by a cellular automata-based molecular model of molecular beam epitaxy. The objective of this case study is to determine if NLCR can provide accurate and compact models for large-scale on-line simulation. The second case study is that of modeling the relationship between the structure and properties of various materials. These case studies demonstrate the ability of NLCR to provide accurate and compact models, and encourage further research.

MODELING OF MATERIALS MANUFACTURING PROCESSES BY NONLINEAR CONTINUUM REGRESSION

Bhavik R. Bakshi

1. Introduction

Processing of semiconductor materials is a multi-billion dollar industry, and is essential for a variety of applications such as the manufacture of computer chips and integrated circuits, and several techniques have been developed for their manufacture. Efficient manufacturing with consistent product quality requires process monitoring and feedback control, which in turn require a detailed understanding of the underlying physico-chemical processes, and models for their simulation. Unfortunately, most microelectronic manufacturing processes are not understood well enough to develop accurate fundamental models for process control and monitoring. Consequently, it is common to derive models for such processes based on experimental data using various empirical modeling methods.

Artificial neural networks are among the most popular class of methods for empirical modeling due to their ability to capture arbitrary nonlinear relationships from multivariate data. Among the wide variety of neural networks, the most popular for empirical modeling are the backpropagation network (BPN) and radial basis function network (RBFN). In addition to artificial neural networks (ANN), several statistical methods are also available for nonlinear empirical modeling. These methods include, projection pursuit regression (PPR), nonlinear principal component regression (NLPCR), nonlinear partial least squares regression (NLPLS), classification and regression trees (CART), and multivariate adaptive regression splines (MARS). These neural and statistical empirical modeling methods differ in their modeling approach, causing some methods to perform better for certain types of modeling problems. For example, backpropagation networks (BPN) often require a large amount of training data to obtain an acceptable model for a given number of input variables, whereas, statistical methods such as, NLPCR and NLPLS can perform equally well with a smaller ratio of training data to input variables. ANN usually provide black box models, whereas the model obtained by CART or MARS may be represented in terms of simple rules. Statistical methods with adaptive basis functions such as PPR, usually require less basis functions for comparable performance than neural techniques such as, BPN.

Given this broad variety of empirical modeling methods, it is important to select the best method for a given modeling task. Proper selection requires a deep understanding of all the modeling methods and a systematic approach to model selection. In response to this need, Bakshi and Utojo

(1997a) developed a common framework for comparing empirical modeling methods and enabling greater understanding of their similarities and differences. This framework is based on representing the model developed by any empirical modeling method as a weighted sum of basis functions, and showing how various methods can be derived depending on decisions about the nature of the input transformation, the type of activation or basis functions, and the optimization criteria for determining the adjustable parameters.

The insight provided by this comparison framework is then used to unify linear and nonlinear empirical modeling methods that combine the inputs as a linear weighted sum before operation of the basis function (Bakshi and Utojo, 1997b). These methods based on linear projection include linear methods such as, ordinary least squares regression (OLS), partial least squares regression (PLS), principal components regression (PCR) and ridge regression (RR), and nonlinear methods such as, backpropagation networks with a single hidden layer, projection pursuit regression, nonlinear partial least squares regression, and nonlinear principal component regression. The comparison framework shows that all methods based on linear projection are special cases along a continuum of methods. The result of their unification is a new method called NonLinear Continuum Regression (NLCR) that can specialize to any existing method or to a method along the continuum between existing methods, with the help of an additional tuning parameter. An efficient hierarchical training methodology is developed for NLCR modeling that trains one node at a time to reduce the residual error of approximation. Since NLCR subsumes all methods based on linear projection, the resulting models are at least as good, if not better, than those obtained by existing methods based on linear projection.

The primary objective of the research in the summer faculty research program at Wright Laboratories was to gain greater insight into the challenges in modeling of semiconductor materials manufacturing processes. The ability of NLCR to model various materials processes was to be explored, and its performance compared with existing empirical modeling methods. Areas of future collaboration and common interest were also to be identified.

These objectives were accomplished by studying a molecular beam epitaxy process for making GaAs thin films. The NLCR method was used to model this process, with data generated from a fundamental model developed by Jackson et al. (1997). The NLCR method was also used for prediction of materials properties from a structure-property data base. Such a model may be useful for predicting the properties of new materials.

The rest of this report is organized as follows. A description of the common framework for empirical modeling methods is provided in Section 2. This is followed by a description of the

NLCR training methodology in Section 3. The two case studies are described in Section 4 and 5, followed by the conclusions and discussion in Section 6.

2. A Common Comparison Framework for Empirical Modeling Methods

The model determined by any empirical modeling method may be represented as a weighted sum of basis functions,

$$\hat{y}_k = \sum_{m=1}^{M} \beta_{mk} \theta_m(\phi_m(\alpha; x_1, x_2, ..., x_J)) \quad (1)$$

where, $\hat{y}_k$ is the k-th predicted output or response variable, θ_m is the m-th basis or activation function, β_{mk} is the output weight or regression coefficient relating the m-th basis function to the k-th output, α is the matrix of basis function parameters, ϕ_m represents the input transformation, and $x_1, ..., x_J$ are the inputs or predictor variables. The variable obtained by transforming the inputs, $\mathbf{z}_m = \phi_m(\alpha; \mathbf{x})$, is often referred to as the latent variable or projected input. Specific empirical modeling methods may be derived from Equation (1) depending on decisions about the nature of input transformation, type of activation or basis functions, and optimization criteria. These decisions form the basis of the common framework developed in this paper for comparing all empirical modeling methods, and are described in the rest of this section.

Nature of Input Transformation. Reducing the dimensionality of the input space is essential for improving the complexity of the modeling task, and the quality of the extracted model. Empirical modeling techniques fight this "curse of dimensionality" by transforming the inputs to latent variables that capture the relation between the inputs with less latent variables than the number of inputs. Such dimensionality reduction is usually accomplished by exploiting the relationship among inputs, or distribution of training data in the input space, or relevance of input variables for predicting the output. Thus, empirical modeling methods may be divided into the following three categories depending on the nature of input transformation.

- *Methods based on linear projection* exploit the linear relationship among inputs by projecting them on a linear hyperplane, as shown in Figure 1a, before applying the basis function. Thus, the inputs are transformed by combination as a linear weighted sum to form the latent variables.
- *Methods based on nonlinear projection* exploit the nonlinear relationship between the inputs by projecting them on a nonlinear hypersurface resulting in latent variables that are nonlinear functions of the inputs, as shown in Figures 1b and c. If the inputs are projected on a localized hypersurface then the basis functions are local, as depicted in Figure 1c. Otherwise, the basis functions are non-local in nature.

- *Partition-based methods* fight the curse of dimensionality by selecting input variables that are most relevant to efficient empirical modeling. The input space is partitioned by hyperplanes that are perpendicular to at least one of the input axes, as depicted in Figure 1d.

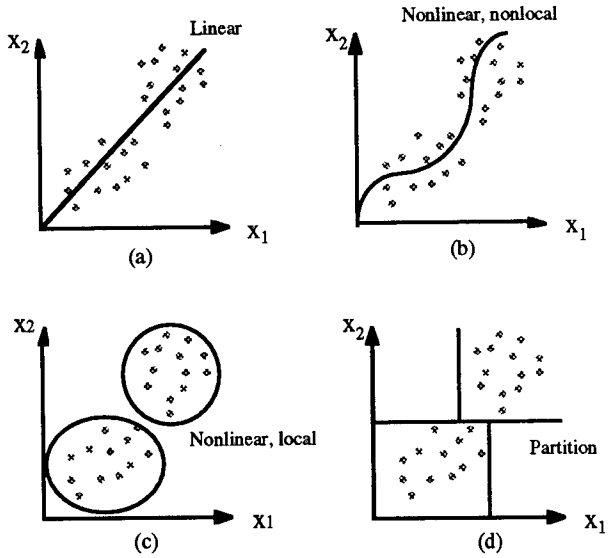


Figure 1. Input transformation in (a) methods based on linear projection, (b) and (c) methods based on nonlinear projection, non-local and local transformation respectively, and (d) partition-based methods.

Type of Activation or Basis Functions. The wide variety of activation or basis functions used in empirical modeling methods may be broadly divided into the following two categories:

- *Fixed-shape basis functions.* The basis functions in several empirical modeling methods are of a fixed shape such as, linear, sigmoid, Gaussian, wavelet, or sinusoid. Adjusting the basis function parameters changes their location, size, and orientation, but their shape is decided a priori, and remains fixed.
- *Adaptive-shape basis functions.* Some empirical modeling methods relax the fixed-shape requirement and allow the basis functions to adapt their shape, in addition to their location, size, and orientation, to the training and testing data.

Optimization Criteria. The aim of any empirical modeling method is to extract the underlying input-output relationship and/or input transformation from the available data. The input transformation is determined by the function, ϕ, and parameters, α, whereas the model relating the transformed inputs to the output is determined by the parameters, β, and basis functions, θ. Empirical modeling methods often use different objective functions for determining the input transformation, and the model relating the transformed inputs to the output. This separation of the empirical modeling optimization criteria provides explicit control over the dimensionality reduction by input transformation, and often results in more general empirical models. Most empirical modeling methods minimize the mean square error of approximation to determine the basis function, θ and regression coefficients, β. The criterion used to determine the input transformation parameters, ϕ and α differ for each method depending on the emphasis on transforming the inputs versus minimizing the output error of approximation. For example, PCR and NLPCR focus entirely on obtaining an optimum transformation of the inputs by maximizing the variance captured by the latent variables, whereas, OLS, BPN, and PPR transform the inputs to minimize the output

prediction error, and PLS and NLPLS maximize the covariance between the projected inputs and output.

The nature of the input transformation, type of basis functions, and optimization criteria discussed in this section provide a common framework for comparing the wide variety of techniques for input transformation and input-output modeling, as depicted in Table 1. This comparison framework is useful for understanding the similarities and differences between various methods, and may be used for unifying methods based on linear projection, as described in the next section.

3. Unification of Methods Based on Linear Projection

The latent variable for methods based on linear projection is a weighted sum of the inputs. The resulting model may be represented by specializing Equation (1) to,

$$\hat{y}_k = \sum_{m=1}^{M} \beta_{mk} \theta_m \left(\sum_{j=1}^{J} \alpha_{jm} x_j \right) \tag{2}$$

The comparison framework described in the previous section indicates that unification of methods based on linear projection requires common methods for determining the different shapes of basis functions, a common objective function and a common training methodology. Such a unified method is developed in this section for modeling with multiple inputs and a single output.

Techniques for Determining Basis Functions. Each basis function for methods based on linear projection maps the linearly projected input, z_m to the output, y. Unification of the variety of basis functions used in these methods requires a general approach that can provide any linear or nonlinear relationship between the latent variable and output, depending on the nature of the training data. Such basis functions may be obtained by using univariate smoothing techniques for approximating the training data in the projected input-output space. A variety of such smoothing techniques are available including, variable span smoothers (Friedman, 1984), Hermite functions (Hwang et al., 1994), automatic smoothing splines (Roosen and Hastie, 1994), and backpropagation networks. The NLCR method developed in this paper can use any of these smoothing techniques to determine the appropriate basis functions.

General Optimization Criterion for Projection Directions. Unification of methods based on linear projection requires a general optimization criterion that consists of information from both the inputs and output, and can specialize to the criterion used by existing methods based on linear projection. Thus, the optimization criterion should span the continuum between different methods based on linear projection. The techniques of PCR and NLPCR lie at one extreme of this continuum, since their optimization criterion is unaffected by the nature of the outputs or basis

functions. Both methods focus on transforming only the input space by maximizing the variance captured by the projected inputs as,

$$\max_{\alpha_m} \{ \text{var}(\mathbf{X}\alpha_m) \} \tag{3}$$

At the other extreme of the continuum of methods based on linear projection, are the techniques of OLS, PPR and BPN, since their optimization criterion focuses entirely on minimizing the output prediction error. This optimization criterion is equivalent to maximizing the square of the correlation between the actual and approximated outputs (Bakshi and Utojo, 1997) and may be written as,

$$\max_{\alpha_m} \{ \text{corr}^2(\mathbf{y}, \theta_m(\mathbf{X}\alpha_m)) \} \tag{4}$$

The optimization criteria at two extremes of the continuum of methods given by Equations (3) and (4) may be combined as,

$$\max_{\alpha_m} \{ \text{corr}^2(\mathbf{y}, \theta_m(\mathbf{X}\alpha_m)) \, \text{var}(\mathbf{X}\alpha_m) \} \tag{5}$$

and should result in a method between PPR and NLPCR. Indeed, Equation (5) has been used as the optimization criterion for NLPLS modeling by Wold et al. (1989) for quadratic PLS, Wold (1992) for spline PLS, and Holcomb and Morari (1992) for neural net/PLS.

Equations (3), (4) and (5) may be combined to obtain a general optimization criterion that subsumes all methods based on linear projection as,

$$\max_{\alpha_m} \{ [\text{corr}^2(\mathbf{y}, \theta_m(\mathbf{X}\alpha_m))] \, [\text{var}(\mathbf{X}\alpha_m)]^{\gamma} \} \tag{6}$$

where values of γ equal to 0, 1, and ∞ result in BPN, PPR or OLS; NLPLS or PLS; and NLPCR or PCR, respectively. Equation (6) is a nonlinear version of the optimization criterion suggested by Stone and Brooks (1990) to unify OLS, PLS, and PCR. The exponents in Equation (6) may be modified to,

$$\max_{\alpha_m} \{ [\text{corr}^2(\mathbf{y}, \theta_m(\mathbf{X}\alpha_m))]^{1+\gamma-2\gamma^2} \, [\text{var}(\mathbf{X}\alpha_m)]^{3\gamma-2\gamma^2} \} \tag{7}$$

This objective function reduces to various existing methods, as summarized in Table 2. The effect of the adjustable parameter, γ on the generality of the empirical model may be understood in terms of the bias-variance trade-off. As γ increases from 0 to 1, the model bias increases, while the variance decreases, causing the mean-squares error of approximation to go through a minimum.

The NLCR training methodology aims to find this value of γ that optimizes the bias-variance trade-off as described in the next section.

The remaining adjustable parameters, namely the regression coefficients, β_m and basis functions, θ_m are determined by minimizing the mean-squares error of approximation,

$$\min_{\beta_m,\, \theta_m} \frac{1}{I} \sum_{i=1}^{I} (y_i - \hat{y}_i)^2 \tag{8}$$

Equations (7) and (8) constitute the general objective function that unifies all methods based on linear projection.

Hierarchical Training Methodology. The final challenge for the unification of empirical modeling methods based on linear projection is the development of a common training methodology that uses the general basis functions, and the common optimization criterion, to determine the empirical model in an efficient manner. Training methodologies for empirical model building may determine the model parameters simultaneously for all the basis functions, or hierarchically for one basis function at a time. Examples of the simultaneous approach include eigenvalue decomposition for computing the projection directions in PCR and PLS, and the error backpropagation algorithm for BPN (Rumelhart and McCleland, 1986). Examples of the hierarchical approach include the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm (Martens and Naes, 1989) for PCR and PLS, cascade correlation for BPN (Fahlman and Lebiere, 1990), and the PPR algorithm (Friedman and Stuetzle, 1984). Hierarchical modeling methods are usually more efficient than their simultaneous modeling counterparts since an existing model may be easily adapted by adding new nodes to capture the residual error of approximation as necessary.

The steps comprising the hierarchical, node-by-node NLCR training methodology, are shown below.

1) For $\gamma \leftarrow 1$ to 0,
2) Add new node and optimize,
3) Projection directions, α_m
4) Basis functions, θ_m
5) Regression coefficients, β_m
6) Update model
7) Update output residual
8) Update input residuals or backfit previously added nodes
9) If prediction error is acceptable, go to 10, else go to 2
10) End

The projection directions are computed by optimizing the general objective function for the selected value of γ, for the basis function and regression coefficient determined in the previous iteration. If orthonormal projection directions are desired, as in PCR and PLS, then both the input and output residuals need to be updated, otherwise, the input residual is left unchanged. The modeling ability of each node may be improved in Step (8) by accounting for the nature of previously added basis functions by adjusting their parameters by backfitting or backward pruning (Friedman, 1985).

The NLCR training methodology can specialize to hierarchical algorithms for existing methods based on linear projection. For example, the NIPALS algorithm for PLS may be obtained by restricting the basis functions to be linear, selecting γ=0.5, and determining the input and output residuals after training each node. Backfitting is not needed since the projection directions are fixed by the orthogonality requirement. Specializing the general method to PPR, requires determining the projection directions, basis functions, and regression coefficients by maximizing the objective function for γ=0, and computing the output residual only.

Efficient techniques for finding the best value of γ are essential for the application of NLCR modeling to practical problems. The optimum value of γ may be found by from models developed for several values, and selecting the γ and number of basis functions that result in the smallest error of approximation for testing data. Unfortunately, the nonlinear nature of the model can make this trial-and-error approach computationally expensive for large problems. Furthermore, the modeling with several different initial values of the parameters may be necessary to avoid local minima. These practical and computational issues may be addressed by exploiting the following properties of NLCR models.

- Unique values of the projection directions for γ=1 may be determined by maximizing the variance captured by the projected inputs. If orthogonal projection directions are not required, then the projection directions for all nodes will be equal to the first principal component of the input data matrix, which is the eigenvector of the input covariance matrix.
- Decreasing the value of γ causes the projection directions to gradually rotate away from those capturing the relationship between the inputs to those minimizing the output prediction error.

Thus, the NLCR model may be first determined for γ=1, and the resulting parameters used as initial values of the parameters for modeling at smaller values of γ.

4. Molecular Beam Epitaxy

Molecular Beam Epitaxy (MBE) is a method for growing a thin-film semiconductor by depositing atoms on a surface. A predictive model of film growth would be useful for determining the appropriate process conditions for obtaining the desired product quality in an efficient manner. A

fundamental model for MBE has been developed by Jackson et al. (1997) based on cellular automaton methods. This model predicts the state of a site, that is the type of occupation, and the index of the new position to be occupied based on information about the state and type of atoms on adjacent sites on a cube. The cellular automata utilize several if-then rules for prediction the state of the central site. Unfortunately, the rule structure is too complex for real-time simulation of the MBE process. The speed of the simulation may be improved by capturing the cellular automata model in the form of an empirical model.

Ideally, the input to the empirical model should consist of information about all the 26 positions surrounding the central position as well as information about the number of occupied neighbors, probability of like-like and like-unlike bonding, and current state of the central position. The very large number of inputs makes the modeling process extremely slow. Consequently, it was decided to decrease the number of inputs by eliminating information about the 26 adjacent positions. The final training data consisted of four inputs and one output, and 1500 exemplars.

Of the available data, 1000 were used for training and 500 for testing. The best model was determined by crossvalidation with training data. Since the ratio of training data to number of inputs was very large, the NLCR model with γ=0 resulted in the best model, as expected. The best mean-squares error of approximation on testing data was found to be 0.00997 for 25 hidden nodes. Only 4 nodes were enough for reducing this error to 0.0100. The error of approximation was comparable with that obtained by the orthogonal functional basis neural network (OFBNN) of Chen et al. (1997), but the NLCR model was much more compact than the OFBNN which required more than 100 hidden nodes.

5. Material Structure-Property Prediction

This case study models the relationship between various material properties with the objective of predicting the properties of new materials. The training and testing data for all the case studies are identical to those used by Chen et al. (1997) for modeling by the OFBNN. Additional details of the data set are described by Pao and Meng(1997).

Tables 3, 4 and 5 show the data used in this case study. In each of the three tables, the last column (atomic weight) contains the dependent variable. The first five columns contain the input variables. The last row in each table, contains the testing set, while the remaining rows make up the training set. The data in the three tables have been broken down into subgroups from a larger data base according to clusters determined by Chen et al (1997).

For the data in Table 3, the best NLCR model was obtained for $\gamma = 0$, with eight basis functions determined by the supersmoother (Friedman, 1985). The training MSE is 0.00085 based on

original data. The desired feature value is 181.8360 and the predicted value is 181.8357. The loading directions, α, were initialized using results from principal component regression with $\gamma = 1$. Models for smaller values of γ were developed with the initial parameters determined by the previous larger value of γ. Such an initialization of the model parameters, instead of a random initialization, is likely to decrease the chances of getting caught in local minima.

For the data in Table 4, the modeling approach was similar to that for the previous example. The best result was obtained for $\gamma = 1.0$, with ten hidden nodes determined by the supersmoother. Training MSE is 215.832 based on scaled data . The desired feature value is 100.69 and the predicted value is 97.3181.

For the data in Table 5, the best result was again obtained for $\gamma = 1.0$, with loading directions being initialized by linear PCR. The training MSE based on the scaled input data is 0.0168. The desired feature value is 199.9 and the predicted value is 217.4902. The results of NLCR modeling are compared with those obtained by OFBNN in Table 6. This indicates that the results of NLCR modeling are comparable to those of OFBNN.

6. Conclusions and Discussion

The research conducted as a summer faculty associate has met the objective of evaluating the modeling ability of nonlinear continuum regression. The various case studies indicate that NLCR results in models that are better than conventional statistical and neural network methods based on linear projection, and comparable to the orthogonal functional basis neural network of Chen et al. (1997). An important difference between the two methods is that the basis functions in NLCR adapt to the data. This usually results in less number of hidden nodes than OFBNN and other methods with basis functions of a previously determined fixed shape. Furthermore, since the nonlinear model may get stuck in suboptimal local minima, it is impossible to guarantee that the results reported are the best ones possible for the selected method. Further modeling with different initial conditions may result in better models.

Since the OFBNN method is the closest competitor to NLCR, the similarities and differences between these methods are discussed in this paragraph. The primary difference between the two approaches is that NLCR is a method based on linear projection, whereas OFBNN is a method based on nonlinear local projection. This implies that in NLCR, the inputs are projected as shown in Figure 1a, whereas OFBNN projects the inputs as shown in Figure 1c. The training methodology for NLCR and OFBNN is also different. The NLCR modeling approach is hierarchical, and trains one node at a time to minimize the residual error of approximation. The OFBNN training selects the structure of the network and determines the values of all the

parameters simultaneously. The hierarchical approach provides greater flexibility and efficiency in modeling since the results of previous modeling can be used to e model. In contrast, simultaneous modeling methods rely on trial-and-error and develop several models with different number of nodes, and select the best one.

7. References

Bakshi, B. R., Utojo, U., Neural and Statistical Methods for Empirical Modeling: A Common Framework and Overview, *Chemometrics and Intelligent Laboratory Systems*, submitted, (1997a)

Bakshi, B. R., Utojo, U., Unification of Neural and Statistical Modeling Methods that Combine Inputs by Linear Projection, *Computers and Chemical Engineering*, accepted, (1997b)

Chen, C. L. P., Cao, Y., LeClair, S. R., Material Structure-Property Prediction Using an Orthogonal Functional Basis Neural Network, *Australasia-Pacific Forum on Intelligent Processing and Manufacturing of Materials, IPMM'97*, Gold Coast, Australia, July 14-17, (1997)

Fahlman S. E. and C. Lebiere, The Cascaded-Correlation Learning Architecture, *Advances in Neural Information Processing Systems*, 2, 524-532, Morgan Kaufmann (1990).

Friedman J. H., Classification and Multiple Regression Through Projection Pursuit, *Technical Report* No. 12, Dept. of Statistics, Stanford University, Stanford, CA (1985).

Friedman, J.H. and W. Stuetzle, Projection pursuit regression. *J. Amer. Stat. Assoc.*, 76, 817-823 (1981).

Friedman, J.H., Multivariate adaptive regression splines. *Annals of Statistics*, 19, 1-141 (1991).

Holcomb T. R. and M. Morari, PLS/Neural Networks, *Computers and Chem. Engg.*, 16, 4, 393-411 (1992).

Hwang, J. N., Lay, M., R. D. Martin, and J. Schimert, Regression modeling in back-propagation and projection pursuit learning. *IEEE Trans. Neural Networks*, 5 (1994).

Jackson, A. G., Benedict, M. D., LeClair, S. R., Cellular Automaton-Based Models of Thin Film Growth, *Australasia-Pacific Forum on Intelligent Processing and Manufacturing of Materials, IPMM'97*, Gold Coast, Australia, July 14-17, (1997)

Martens H. and T. Naes, *Multivariate Calibration*, Wiley, New York (1989).

Roosen C. B. and T. J. Hastie, Automatic Smoothing Spline Projection Pursuit, *J. Comput. Graph. Stat.*, 3, 3, 235-248 (1994).

Rumelhart, D. E., J. L. McClelland, et al., *Parallel Distributed Processing, Vol. 1*, The MIT Press, Cambridge, MA (1986).

Stone, M. and R. J. Brooks, Continuum regression cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression, *J. Royal Stat. Soc.*, Ser. B., Vol. 52, pp. 237-269 (1990).

Wold S., Nonlinear Partial Least Squares Modeling II. Spline Inner Relation, *Chemometrics and Intelligent Laboratory Systems*, 14, 71-84 (1992).

Wold, S., N. Kettaneh-Wold and B. Skagerberg, Nonlinear PLS modeling. *Chemometrics and Intelligent Laboratory Systems*, 7, 53-65 (1989).

Table 1. Comparison matrix for empirical modeling methods

Method	Input Transformation	Basis Function	Optimization Criteria
OLS	Linear projection	Fixed shape, linear	α - max. squared correlation between projected inputs and output β - min. output prediction error
PLS	Linear projection	Fixed shape, linear	α - max. covariance between projected inputs and output β - min. output prediction error
PCR	Linear projection	Fixed shape, linear	α - max. variance of projected inputs β - min. output prediction error
BPN single	Linear projection	Fixed shape, sigmoid	[α, β] - min. output prediction error
PPR	Linear projection	Adaptive shape, supersmoother	[α, β, θ] - min. output prediction error
BPN mult.	Nonlinear proj., nonlocal	Fixed shape, sigmoid	[α, β] - min. output prediction error
NLPCA	Nonlinear proj., nonlocal	Adaptive shape	[α, ϕ] - min. input prediction error
RBFN	Nonlinear projection, local	Fixed shape, radial	[σ, t] - min. distance between inputs and cluster center β - min. output prediction error
CART	Input partition	Adaptive shape, piecewise constant	[β, t] - min. output prediction error
MARS	Input partition	Adaptive shape, spline	[β, t] - min. output prediction error

Table 2. Specialization of objective function for projection directions to existing methods based on linear projection.

γ	Linear basis functions	Nonlinear basis functions
0	OLS	PPR/BPN
1/2	PLS	NLPLS
1	PCR	NLPCR

Table 3. Data from Table 1,cluster 2 of Chen et al.(1997) as applied to NLCR.

gap	a	c	Radiuslon	Density	Atomic-wt
2.1	4.61	4.61	29	0.0001	84.6
2.26	4.359	4.359	29	3.191	40.09
3.3	3.251	5.209	22	5.651	81.369
3.9	3.823	6.261	53	3.536	97.434
4	5.481	5.171	22	4.502	181.836
5.9	5.58	4.69	22	0.0001	60.069
6	4.359	4.359	29	3.191	40.09
6.2	3.11	4.98	25	3.255	40.99
7	7.45	6.97	59	0.0001	136.086
8.4	4.9134	5.4052	22	2.65	60.078
4	5.481	5.171	22	4.502	181.836

Table 4. Data from Table 2, cluster 1 of Chen et al.(1997) as applied to NLCR.

gap	a	c	Radiuslon	Density	Atomic-wt
0.57	5.943	11.217	71	5.6	334.97
1.2	6.099	11.691	66	5.808	286.798
1.53	5.489	11.101	53	4.73	242.468
1.7	5.606	11.006	66	4.73	242.468
1.74	5.606	10.88	71	4.7	243.43
1.8	5.981	10.865	66	5.759	335.51
2.1	4.145	9.496	53	7.101	232.654
2.43	5.351	10.47	53	4.332	197.388
2.638	5.751	10.238	53	4.66	241.718
2.91	5.74	10.776	59	4.549	246.93
3.05	5.568	10.04	53	3.97	380.096
2.05	5.463	10.731	59	4.105	199.9

Table 5. Data from Table 2, cluster 3 of Chen et al.(1997) as applied to NLCR.

gap	a	c	Radiuslon	Density	Atomic-wt
0.23	6.479	6.479	89	5.777	236.55
0.33	4.457	5.939	82	6.25	236.55
0.36	6.268	6.479	71	5.72	189.79
0.72	6.095	6.095	89	5.615	191.47
1.35	5.868	5.868	59	4.798	145.77
1.4	5.653	5.653	71	5.316	144.71
1.7	4.361	4.954	66	4.819	78.96
2.3	6.101	6.101	82	5.924	192.97
2.7	5.667	5.668	66	5.318	144.33
2.8	6.473	6.473	126	6.0	234.77
2.91	5.69	5.69	82	4.72	143.449
2.95	6.042	6.042	96	5.667	190.44
3.05	5.568	5.568	53	3.97	80.096
3.17	5.405	5.405	77	4.137	98.993
2.3	5.45	5.45	59	4.135	100.69

Table 6. Comparison of % prediction error of NLCR and OFBNN

Data set	NLCR % prediction error	OFBNN % prediction error
Table - 1	0	0
Table - 2	3.35	13.26
Table - 3	8.80	4.17

CONTRIBUTION OF A SCENE PROJECTOR'S NON-UNIFORMITY TO A TEST ARTICLE'S OUTPUT IMAGE NON-UNIFORMITY

Brian P. Beecken, Professor
T. James Belich, Graduated Student
Department of Physics

Bethel College
3900 Bethel Drive
St. Paul, MN 55112

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory
Eglin Air Force Base, FL

August 1997

CONTRIBUTION OF A SCENE PROJECTOR'S NON-UNIFORMITY TO A TEST ARTICLE'S OUTPUT IMAGE NON-UNIFORMITY

Brian P. Beecken, Professor
T. James Belich, Graduated Student
Department of Physics
Bethel College

ABSTRACT

A mathematical model of the contribution of the non-uniformity of a projector array to the non-uniformity of a test article's output image was developed. Using this model the maximum theoretical limit for the output image non-uniformity was determined. The realistic situations likely to be encountered during simulation testing were all found to be significantly below the theoretical maximum. The output image non-uniformity is dependent upon the non-uniformity of the projector array, as well as a weighting factor which results from the contribution of the different emitters upon individual detector elements. It is through this weighting factor that parameters such as the sampling ratio, the fill factor of the detector array, the optical blur of the emitters, and the alignment of the emitters with respect to the detectors influence the non-uniformity. A computer program has been written to numerically approximate the weighting factor for a user defined set of parameters.

CONTRIBUTION OF A SCENE PROJECTOR'S NON-UNIFORMITY TO A TEST ARTICLE'S OUTPUT IMAGE NON-UNIFORMITY

Brian P. Beecken
T. James Belich

Introduction

The mission of the KHILS (Kinetic Kill Vehicle Hardware-In-the-Loop Simulator) facility at Wright Laboratory's Armament Directorate is to test infrared imaging sensors in the lab by projecting simulated IR scenes that subtend the sensor's field of view.[1] The IR projector, an integral component of the simulation testing, is produced under Wright Lab's WISP (Wideband Infrared Scene Projector) program.

The WISP projector consists of an array of emitters at least as large in number as the array of detectors on the sensor's focal plane.[2] Unfortunately, when the emitter array is set to produce a uniform IR scene, there is significant non-uniformity in the array's output.[3] Although this non-uniformity can be corrected to a large extent, it can never be completely eliminated. Thus, the non-uniformity of the projector will exist as an artifact in the test article's output image. This artifact is solely the result of the simulation, and represents a degradation of the actual scene that would be encountered during the sensor's mission.

The goal of this paper is to predict analytically how the non-uniformity of the projector impacts the uniformity of the test article's output image. Such information will be valuable for determining both the realism of the simulation and the non-uniformity correction requirements. The prediction should be as general as possible, without any reference to a particular emitter array or sensor. These components will be changed as program requirements change, but the need to know the projector's impact on the test article's output image uniformity will always be present.

In order to make our prediction as general as possible, a number of parameters will be

required. These parameters are the ratio of the projector's non-uniformity to the detector array's non-uniformity, the linear ratio of the emitters to detectors (i.e., the sampling ratio), the relative size of the optical blur on the detector array, the alignment of the emitters relative to the detectors, and the fill factor of the detector array. These are believed to be the test parameters of primary importance and will be incorporated into the following analysis.

1 Theory

In this section, we will derive the primary statistical equation for predicting the contribution of projector non-uniformity to the output image non-uniformity. First, we will consider the simplest case for which there is so little blur that a one-to-one mapping of emitters to detectors results. Next, this case will be generalized by increasing the ratio of emitters to detectors and allowing one detector to receive the radiation from multiple emitters. There will still be no blurring of the incident radiation, however, so each emitter will still be detected by only one detector. These two cases will be examined first because they set the pattern for the most general case.

The most general situation will be the last case examined in this section. This case will include parameters such as the relative optical blur size (the size of the radiation from one emitter on the detector array measured in detector pitch), the alignment of the emitters relative to the detectors, and the fill factor of the detector array.

1.1 "No Blur" with 1 to 1 Mapping of Emitter to Detector

The output signal v of a detector is often approximated by

$$v = g\eta\Phi + b \tag{1}$$

where g is the detector's conversion gain, η is the detector's quantum efficiency, Φ the average number of incident photons per integration time, and b is the signal offset that

exists even for zero incident photon flux.[1] In this paper, we will only be considering those situations in which non-uniformity noise dominates over the other noise sources that are more apparent at low flux levels. Consequently, the incident flux will be assumed to be large enough that any variation in the offset b can be ignored.[2]

Each detector on the array produces a signal that differs from the other detectors if either the conversion gain or the quantum efficiency of the detector varies, or if the incident photon flux is different. We are not interested in whether a variation in signal is due to a detector's conversion gain or quantum efficiency. Therefore, these terms will be combined together to represent the detector's sensitivity $g\eta$. The non-uniformity of the detector array can then be conveniently represented as the standard deviation of all the detector sensitivities divided by the average sensitivity, $\frac{\sigma_{g\eta}}{\overline{g\eta}}$. Similarly, the non-uniformity of the emitter array can be represented by $\frac{\sigma_\phi}{\overline{\phi}}$, where $\overline{\phi}$ is the average emittance across the emitter array and σ_ϕ is the standard deviation. Because we are discussing the one-to-one mapping case, the non-uniformity of the emittances will be identical to the non-uniformity of the incident flux:

$$\frac{\sigma_\Phi}{\overline{\Phi}} = \frac{\sigma_\phi}{\overline{\phi}}. \tag{2}$$

Eq. (1) is a product of two terms (detector sensitivity and incident flux) that each have an uncertainty. The standard error propagation equation in this situation is

$$\left(\frac{\sigma_v}{\overline{v}}\right)^2 = \left(\frac{\sigma_{g\eta}}{\overline{g\eta}}\right)^2 + \left(\frac{\sigma_\Phi}{\overline{\Phi}}\right)^2. \tag{3}$$

Simple algebra puts the equation in a more useful form,

$$\frac{\frac{\sigma_v}{\overline{v}}}{\frac{\sigma_{g\eta}}{\overline{g\eta}}} = \sqrt{1 + \left(\frac{\frac{\sigma_\Phi}{\overline{\Phi}}}{\frac{\sigma_{g\eta}}{\overline{g\eta}}}\right)^2}. \tag{4}$$

This result gives the ratio of the output image non-uniformity to the detector's non-uniformity as a function of the ratio of the projector's non-uniformity to the detector's non-uniformity. It is easily seen, for example, that if the projector has a non-uniformity

[1]The output signal v can only be represented by a linear function over a limited region. It is this region, however, for which two point calibration (offset and gain correction) is performed.

[2]Further work should enable us to remove this limitation so that the non-uniformity's contribution to the total noise could be predicted even when other noise sources are comparable in size.

equal to that of the detector's non-uniformity, then the output image non-uniformity will be increased by a factor of $\sqrt{2}$ over the detector's non-uniformity. As expected, if either the projector or the detector array has significantly larger non-uniformity, it will dominate the output image.

This equation applies in the simplest case where each emitter has an output of ϕ detected by only one detector, and each detector receives the radiation emitted by only one emitter. This simple case will serve as the baseline for further calculations.

1.2 "No Blur" with a Sampling Ratio of n to 1

The *sampling ratio* n is defined as the ratio of emitters to detectors when counted in one dimension. The previous case was restricted to a sampling ratio of one-to-one. Now we will allow for any integer sampling ratio, but we will again restrict ourselves to virtually "no blur." All radiation emitted by an emitter is still received by one and only one detector (Fig. 1).

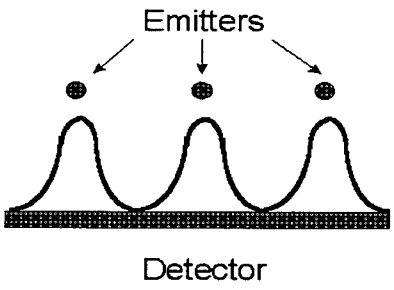


Figure 1: Illustration of a sampling ratio of 3 to 1 with essentially "no blur."

Now there are several emitters contributing to the signal of one detector. Thus, the Φ in Eq. (1) represents the *total* radiation landing on the detector from all the emitters in its field of view:

$$\Phi = \phi_1 + \phi_2 + \phi_3 + \ldots + \phi_{n^2}. \tag{5}$$

Each ϕ_i refers to both the radiation received from emitter i as well as the flux produced by emitter i (since all of each emitter's flux lands on one detector). If one assumes that the

sampling ratio is the same in both directions, then there will be n^2 emitters per detector. Each individual emitter's flux ϕ_i can be approximated by the average flux $\overline{\phi}$ of all the emitters. Therefore,

$$\Phi \approx n^2 \overline{\phi}. \tag{6}$$

The goodness of the approximation of each ϕ_i by $\overline{\phi}$ is given by the standard deviation of the emittances in the emitter array σ_ϕ. Thus,

$$\phi_i \approx \overline{\phi} \pm \sigma_\phi \tag{7}$$

The uncertainty of a sum of variables is simply the square root of the sum of the squares of each variable's uncertainty. Since each emittance ϕ_i is approximated in the same way, we have

$$\sigma_\Phi^2 = n^2 \sigma_\phi^2. \tag{8}$$

A combination of Eqs. (6) and (8) can be used to represent the non-uniformity of the flux incident on the detector array,

$$\frac{\sigma_\Phi}{\overline{\Phi}} = \frac{n\sigma_\phi}{n^2\overline{\phi}} = \frac{\sigma_\phi}{n\overline{\phi}}. \tag{9}$$

Thus, Eq. (4) becomes

$$\frac{\frac{\sigma_u}{\overline{v}}}{\frac{\sigma_{gn}}{\overline{g\eta}}} = \sqrt{1 + \left(\frac{\frac{\sigma_\phi}{\overline{\phi}}}{\frac{\sigma_{gn}}{\overline{g\eta}}}\right)^2 \frac{1}{n^2}}. \tag{10}$$

This result is plotted in Fig. 2 for three different sampling ratios. It will turn out that each line in Fig. 2 represents the greatest non-uniformity possible for a particular sampling ratio. Thus, we will refer to these lines as the *no blur limits.*

1.3 The General Case

Up until now, we have assumed that each emitter illuminates only one detector—what we have referred to as the "no blur" case. Now we will allow the image of the emitters on the focal plane array to be blurred out so that each emitter is seen by multiple detectors. Under these circumstances, it is still true that there will be n^2 emitters per detector. Thus, as before,

$$\Phi \approx n^2 \overline{\phi}. \tag{11}$$

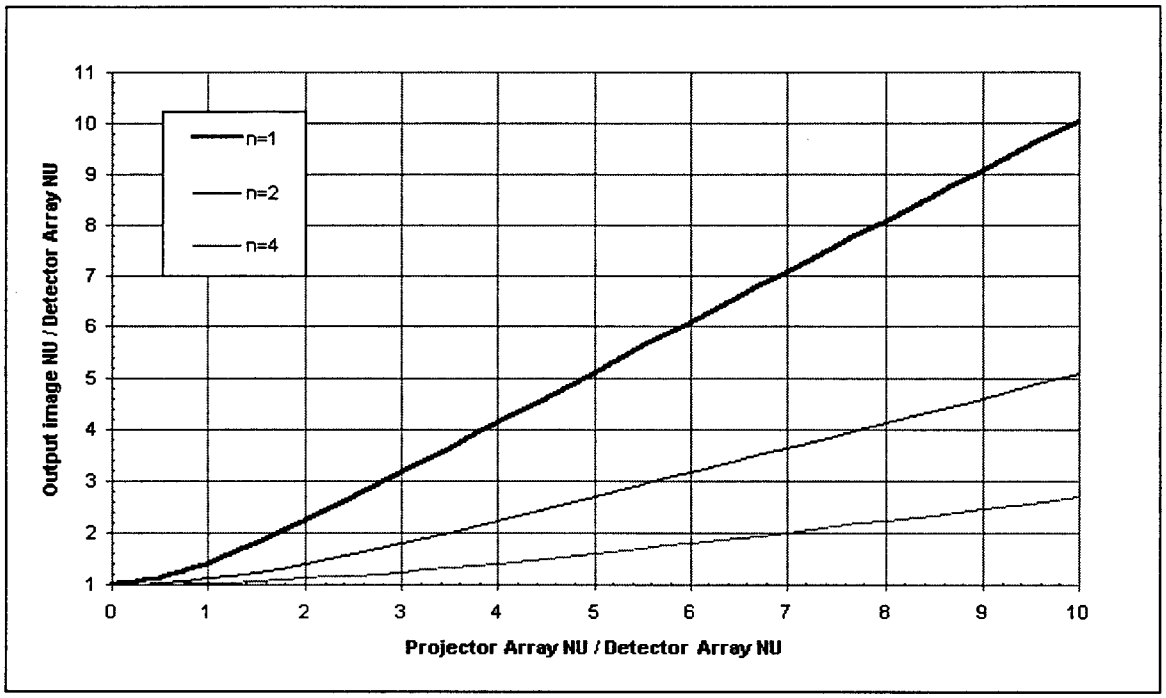


Figure 2: A plot of Eq. (10) for three different sampling ratios. Each line serves as the upper limit of non-uniformity, regardless of blur size.

But Eq. (5) now becomes

$$\Phi = y_1\phi_1 + y_2\phi_2 + y_3\phi_3 + \ldots + y_M\phi_M, \tag{12}$$

where y_i is some weighting function that accounts for how much a particular emitter contributes based on its position relative to the detector. Here M is used to represent the total number of emitters seen by one detector. If the blur is large, then M can become quite large. Because of Eq. (11),

$$\sum_{i=1}^{M} y_i = n^2. \tag{13}$$

It is convenient to normalize this weighting function by defining x_i as $\frac{y_i}{n^2}$. Then, due to normalization,

$$\sum_{i=1}^{M} x_i = 1. \tag{14}$$

The normalized *weighting function* x_i is simply the geometry dependent fraction of each emitter i's contribution to the total flux received by the detector. It will be discussed more fully in the next section.

The uncertainties add as the sum of the squares:

$$\begin{aligned}\sigma_\Phi^2 &= y_1^2\sigma_{\phi_1}^2 + y_2^2\sigma_{\phi_2}^2 + y_3^2\sigma_{\phi_3}^2 + \ldots + y_M^2\sigma_{\phi_M}^2 \\ &= n^4\left[x_1^2\sigma_{\phi_1}^2 + x_2^2\sigma_{\phi_2}^2 + x_3^2\sigma_{\phi_3}^2 + \ldots + x_M^2\sigma_{\phi_M}^2\right].\end{aligned}$$

But all the uncertainties are the same, because they are all represented by the same standard deviation of the emittances across the emitter array. Thus, $\sigma_{\phi_i} = \sigma_\phi$, so

$$\sigma_\Phi^2 = n^4\sigma_\phi^2\sum_{i=1}^{M} x_i^2. \tag{15}$$

Substituting Eq. (11) and Eq. (15) into Eq. (4) we get

$$\frac{\frac{\sigma_v}{\bar{v}}}{\frac{\sigma_{g\eta}}{\overline{g\eta}}} = \sqrt{1 + \left(\frac{\frac{\sigma_\phi}{\bar{\phi}}}{\frac{\sigma_{g\eta}}{\overline{g\eta}}}\right)^2 \sum_{i=1}^{M} x_i^2}. \tag{16}$$

This is our *primary equation* which will be used to calculate the non-uniformity of the output image. Parameters such as fill factor, alignment (or registration), and relative optical blur size affect the output's uniformity only through the *weighting factor* $\sum_{i=1}^{M} x_i^2$. As discussed in the next section, the calculation of this weighting factor is done through numerical approximations.

Before moving on, it is instructive to ensure that our primary equation reduces to Eq. (10) for cases of small blur. When the blur is sufficiently small, $M \longrightarrow n^2$ because the detector sees only those emitters whose unblurred images fall directly on top of it. Since each emitter's radiation is received only by one detector, all $x_i \longrightarrow \frac{1}{n^2}$. Thus, the weighting factor becomes

$$\sum_{i=1}^{n^2} x_i^2 = n^2\left(\frac{1}{n^2}\right)^2 = \frac{1}{n^2}, \tag{17}$$

which makes Eq. (16) reduce to Eq. (10) as it should.

2 Determination of the Weighting Function

In the previous section, we derived the primary equation, Eq. (16), that predicts the non-uniformity of the detector's output image. In order to use this equation, however, it is necessary to first calculate the weighting factor, $\sum_{i=1}^{M} x_i^2$. Unfortunately, the weighting

function x_i cannot be solved analytically. In this section, we will first describe our numerical approach to finding the weighting function. Next we will discuss the calculation of the *relative optical blur size* K, a crucial parameter that must be used in the calculation of the weighting function.

2.1 Numerical Calculation of the Weighting Function

Determining the contribution of each emitter to the total flux received by a given detector is a difficult task because the optical blur is governed by Fraunhofer diffraction. We chose to approximate the distribution of the flux incident on the FPA from a single emitter with a two dimensional Gaussian function. The portion of the emitter's flux which lands on a detector is determined by integrating the two dimensional Gaussian over the detector's area. The fractional contribution of each individual emitter to the total flux incident on a detector is what we have defined as the weighting function x_i.

Because the emitter array is two dimensional, it is necessary to relabel our weighting function as x_{ij} in order to obtain a realistic calculation. Each x_{ij} is then the energy from emitter ij divided by the total flux falling upon the detector. Because the weighting function is normalized, $\sum_i \sum_j x_{ij} = 1$. Although the idea behind this calculation is relatively simple, it becomes difficult in practice for two reasons. First, since we cannot analytically solve the integral of the two dimensional Gaussian over square areas, we approximated it numerically by doing Riemann-like sums. Second, the equation for x_{ij} becomes rather complex once parameters such as fill factor and alignment are taken into account. Our approximation for the weighting function is:

$$x_{ij} \approx \frac{1}{(nA)^2} \sum_{p=-S}^{S} \sum_{q=-S}^{S} \frac{\exp\left\{-\frac{1}{2}\left[\left(\frac{i-T_x-\frac{p}{A}F_x}{0.2616Kn}\right)^2 + \left(\frac{j-T_y-\frac{q}{A}F_y}{0.2616Kn}\right)^2\right]\right\}}{\sum_{a=-L}^{L}\sum_{b=-L}^{L} \exp\left\{-\frac{1}{2}\left[\left(\frac{a-T_x-\frac{p}{A}F_x}{0.2616Kn}\right)^2 + \left(\frac{b-T_y-\frac{q}{A}F_y}{0.2616Kn}\right)^2\right]\right\}} \tag{18}$$

where

$$S = \tfrac{1}{2}(nA-1) \quad \text{and} \quad L = 0.7848Kn + \tfrac{n}{2}.$$

The integration of the two dimensional Gaussian is approximated by using a number of "detection points" distributed across the active detector area. The total number of detection points is given by $(nA)^2$, where A^2 is the number of detection points per emitter. The value of the Gaussian at each detection point is multiplied by the fraction $\frac{1}{(nA)^2}$ of the detector's area being represented by the detection point. The sum of these values yields a good approximation of the integral, assuming that a large enough value of A has been chosen. The computer program we developed to calculate x_{ij} will automatically choose an appropriate value for A when given values for the sampling ratio n and the relative optical blur size K. (Determination of K is discussed in the next subsection.) Clearly, a larger A will provide a better approximation, but computer processing time increases rapidly with A.

In order for x_{ij} to be normalized, the denominator of the larger fraction in Eq. (18) must be the sum of the total energy. Therefore, the sum limits $\pm L$ are chosen so that the contribution of the emitters within at least three σ of optical blur from the detector will be counted. Emitters outside this range are assumed to be far enough away from the detector that their impact upon it is negligible. The value of L also determines the number M of weighting functions x_{ij} that must be summed to find the weighting factor in Eq. (16). In terms relevant to this numerical calculation, the weighting factor is now

$$\sum_{i=-L}^{L} \sum_{j=-L}^{L} x_{ij}^2. \tag{19}$$

These implications are all incorporated into our computer program which calculates the weighting factor.

In Eq. (18), T_x and T_y account for the alignment of the emitter array relative to the detector array. Each variable represents a translation of the emitter centered over the detector in the x and y directions (measured as a fraction of the emitter spacing).[3] Their values may range from 0 to 0.5. The detector's fill factor is accounted for by F_x and F_y,

[3]Zero translation is defined as one emitter being centered over each detector element. In our computer program, in an effort to be as user friendly as possible, zero translation is defined as the emitters being symmetrically distributed over the detector. This difference in definitions only occurs when the sampling ratio n is even.

which are the linear fill factors in the x and y directions. Each must be in the range of 0 to 1, where 1 is a 100% linear fill factor.

2.2 Determination of the Relative Optical Blur Size K

Determination of K, the size of the optical blur relative to the detectors on the FPA, is crucial to determining the weighting factor in Eq. (16). In our weighting function, Eq. (18), we have implicitly defined:

$$K \equiv \text{number of detector pitches in the diameter of a circular blur containing 83.9\% of the radiation.}$$

Now we will derive a means of determining K from the characteristics of the optical system. For a distant point source viewed through a circular aperture, it is well known that the angular diameter (in radians) of the central maximum is given by

$$\theta = \frac{2.44\lambda}{D} \tag{20}$$

where D is the aperture diameter and λ is the wavelength of the incident radiation. When the radiation is focused on the FPA, this equation becomes

$$d = \frac{2.44\lambda}{D} f, \tag{21}$$

where f is the focal length and d is the diameter of the central maximum, which is called the Airy disk. The Airy disk contains 83.9% of the incident radiation.[4] The value for K can be calculated by dividing Eq. (21) by the detector pitch w (the center-to-center detector spacing):

$$K = 2.44\frac{\lambda f}{Dw}. \tag{22}$$

In Eq. (18) we used a two dimensional Gaussian blur to approximate the actual radiation distribution, and we can now examine how it relates to the actual diffraction pattern. The number of σ that correlates to the energy in the Airy disk can be found by integrating the two dimensional Gaussian. Surprisingly, the integral in two dimensions is not difficult

because the blur is radially symmetric. First, we choose a dimensionless variable r equal to $(x-\mu)/\sigma$. Next, we set the mean μ equal to zero so that r is simply the radius in σ. Then we solve for the value for R, the radius which contains 83.9% of the two dimensional Gaussian:

$$\int_0^R re^{-\frac{1}{2}r^2}dr = 0.839\int_0^\infty re^{-\frac{1}{2}r^2}dr$$

becomes

$$R = 1.911.$$

Consequently, a Gaussian blur will include the same energy as the Airy disk when its diameter is 3.822σ. The σ of the Gaussian is given by

$$\sigma = \frac{K}{3.822} = 0.2616K. \tag{23}$$

This form is found in our weighting function, Eq. (18).

3 Results

Several parameters are required to determine the weighting factor in Eq. (16). Therefore, one simple plot of the primary equation cannot be made. In this section, we will display several plots that are chosen to demonstrate general trends over regions of interest.

Figures (3) and (4) illustrate the output image non-uniformity as a function of the projector non-uniformity for sampling ratios of 1:1 and 4:1. As in Fig. (2), the comparison is done with the projector non-uniformity varying from perfect uniformity to a non-uniformity that is an order of magnitude greater than the detector array's non-uniformity. In these two figures, however, each line represents a different optical blur, unlike Fig. (2), which was for virtually "no blur." Recall that the relative optical blur size K represents the detector pitch per diameter of 83.9% of the blurred radiation. Thus when $K = 1$, one detector gets 83.9% of the radiation from one emitter. Note that the "no blur" limit coincides with the appropriate sampling ratio line in Fig. (2). If the blur is infinite, then each detector sees every emitter and the non-uniformity of the projector array cannot be observed. Consequently, large K values cause the lines to move toward

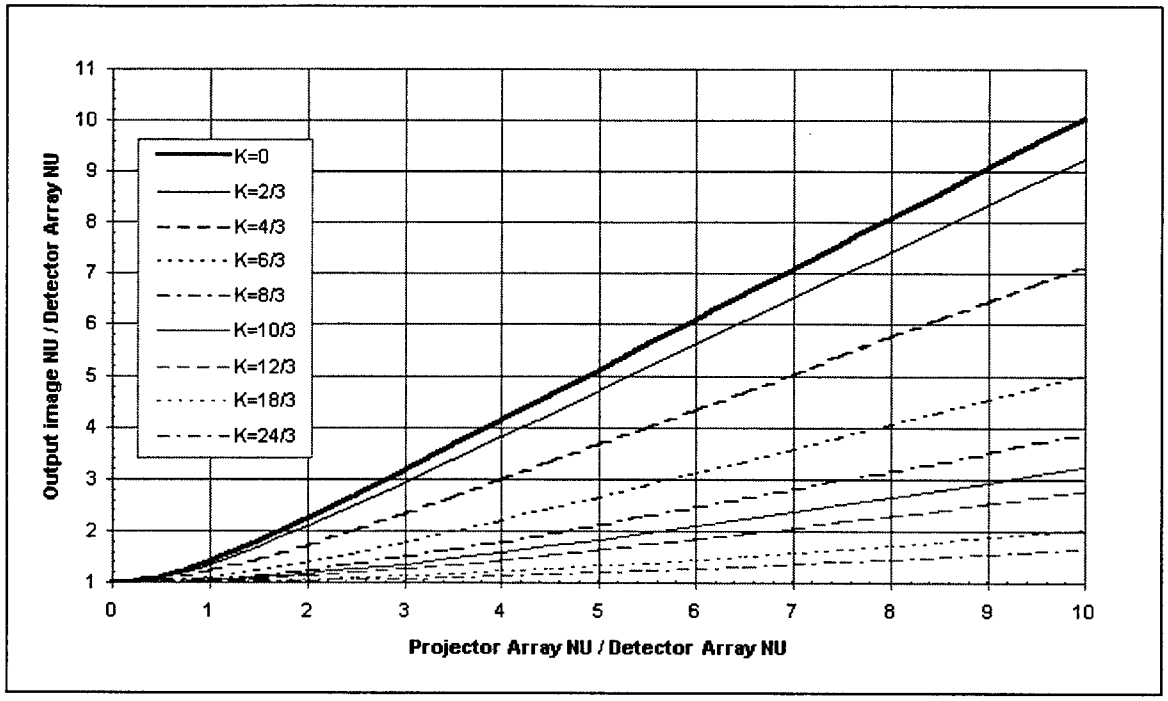


Figure 3: Output image NU vs. projector NU for different relative optical blur sizes K. Sampling ratio is 1:1, fill factor is 100%, and each emitter is centered over one detector.

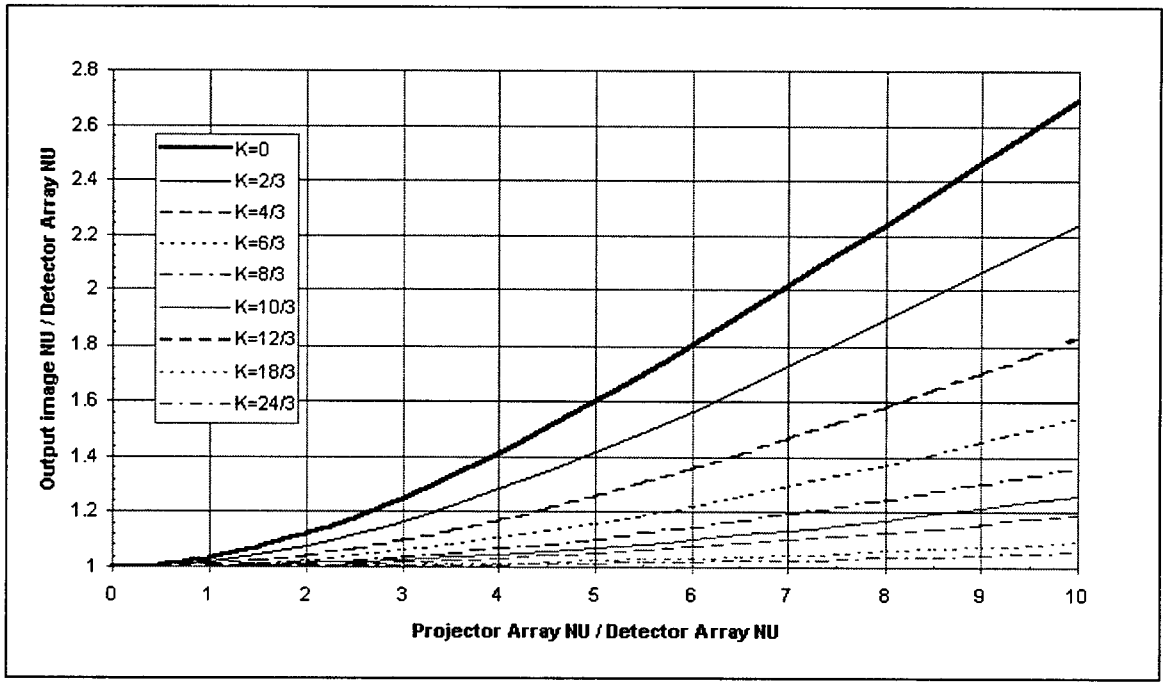


Figure 4: Output NU vs. projector NU for different relative optical blur sizes K. Sampling ratio is 4:1, fill factor is 100%, and emitters are symmetrically centered over detectors.

the horizontal axis where the output image non-uniformity is entirely due to the detector array's non-uniformity.

In order to display the effect of the detector array's fill factor, we chose the particular case of a projector non-uniformity five times greater than the detector array's non-uniformity. Then we plotted the non-uniformity of the output image resulting from different fill factors (in terms of detector area) as a function of the relative optical blur size K. The results are shown in Fig. (5) for a sampling ratio of 1:1 and in Fig. (6) for a sampling ratio of 4:1. Clearly, fill factor makes a difference if it changes significantly, but the more important parameter is the blur size.

As a side point, note that in Fig. (6) the no blur limit of Fig. (2) appears to be violated. The output image non-uniformity in Fig. (6) is as high as 170% above the detector array non-uniformity, whereas the no blur limit for this case should be a 60% increase according to Fig. (2). The apparent discrepancy occurs only when there is virtually no blur and a sampling ratio greater than 1:1. Very small fill factors cause the emittance of some emitters to effectively drop off the edge of the detector's active area. The result is essentially a smaller sampling ratio and a higher no blur limit.

Two more plots, Figs. (7) and (8), were also made with the output image non-uniformity as a function of optical blur, only this time the alignment of the arrays was varied by as much as half an emitter spacing. (Further translations would be the mirror images of lesser translations.) Apparently, alignment makes relatively little difference when there is 100% fill factor, especially if there is a large sampling ratio. To check this, Fig. (9) was done with only 64% fill factor at different alignments. As might be expected, the lower fill factors made a greater difference than the effects of alignment.

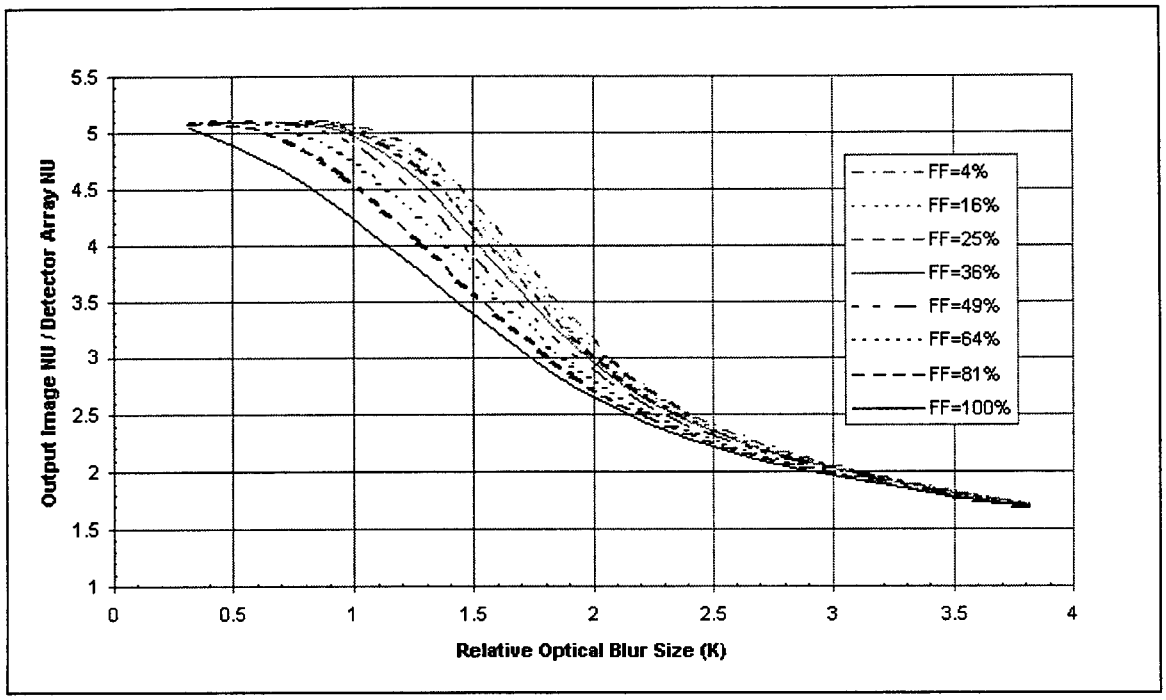


Figure 5: Output image NU vs. relative optical blur size for different fill factors. Sampling ratio is 1:1, projector NU/detector NU is 5, and each emitter is centered over one detector.

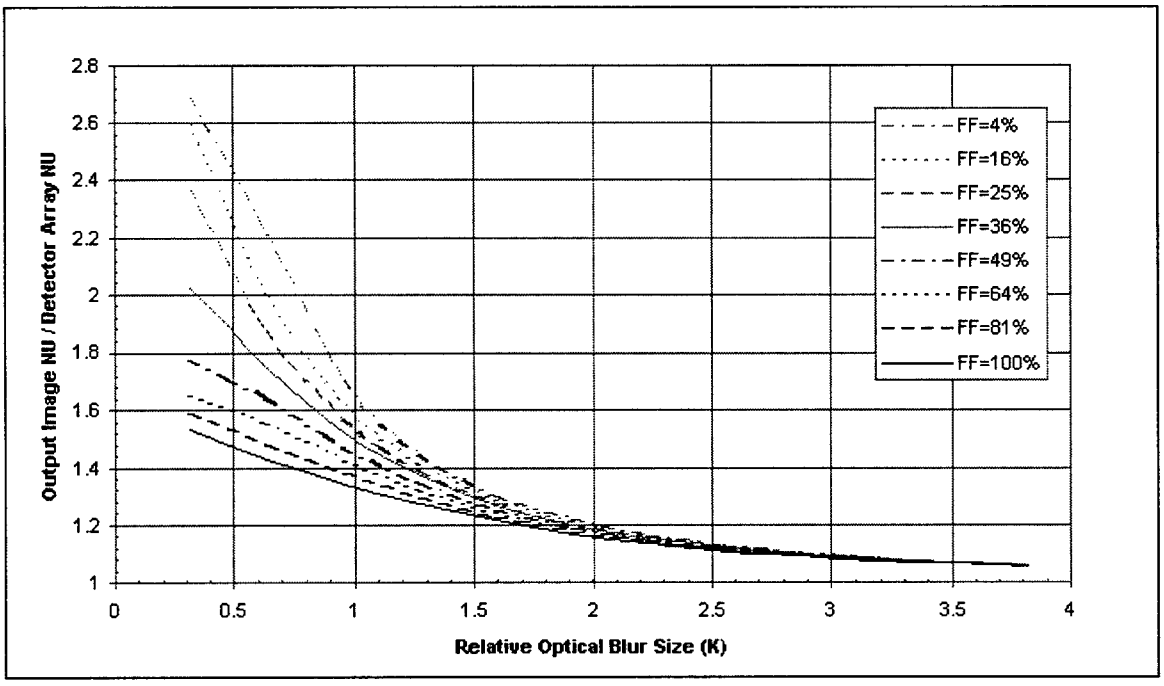


Figure 6: Output NU vs. relative optical blur size for different fill factors. Sampling ratio 4:1, projector NU/detector NU is 5, and emitters symmetrically centered over detectors.

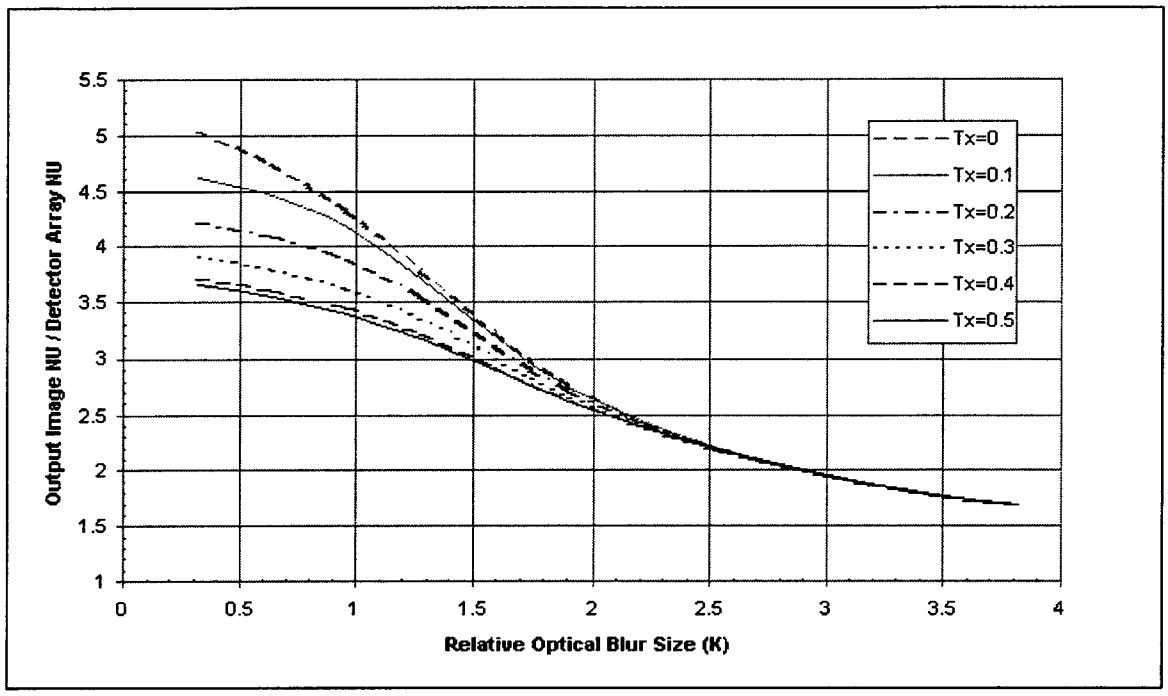


Figure 7: Output image NU vs. relative optical blur size for different alignments. Sampling ratio is 1:1, projector NU/detector NU is 5, and fill factor is 100%.

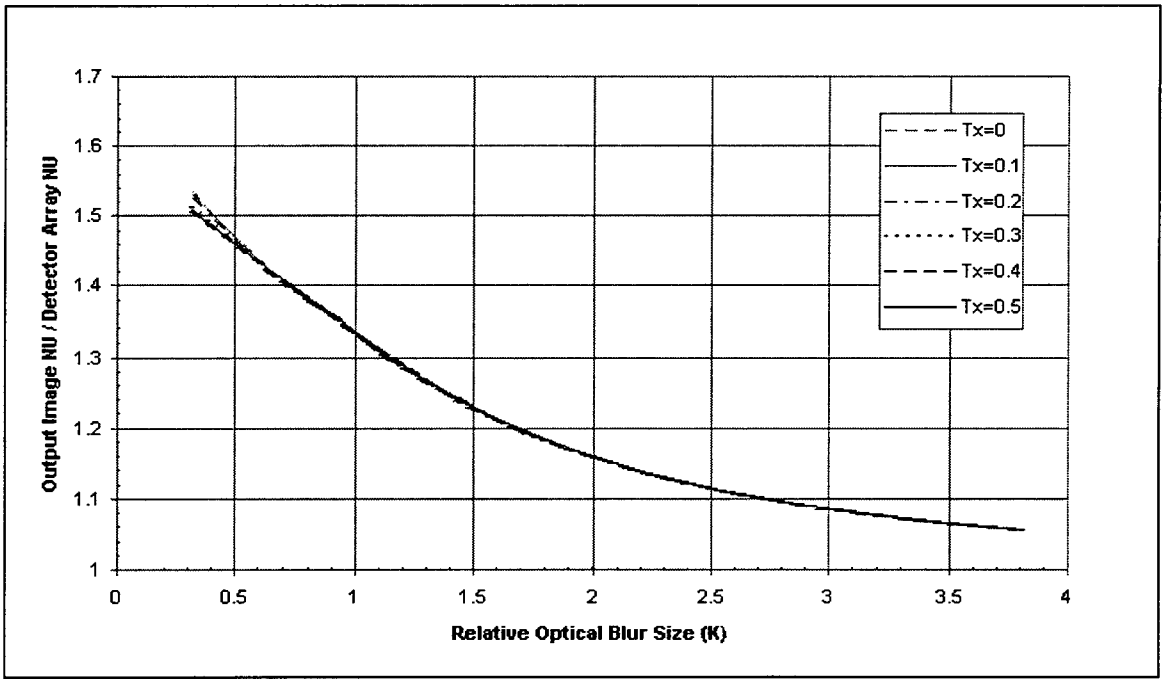


Figure 8: Output image NU vs. relative optical blur size for different alignments. Sampling ratio is 4:1, projector NU/detector NU is 5, and fill factor is 100%.

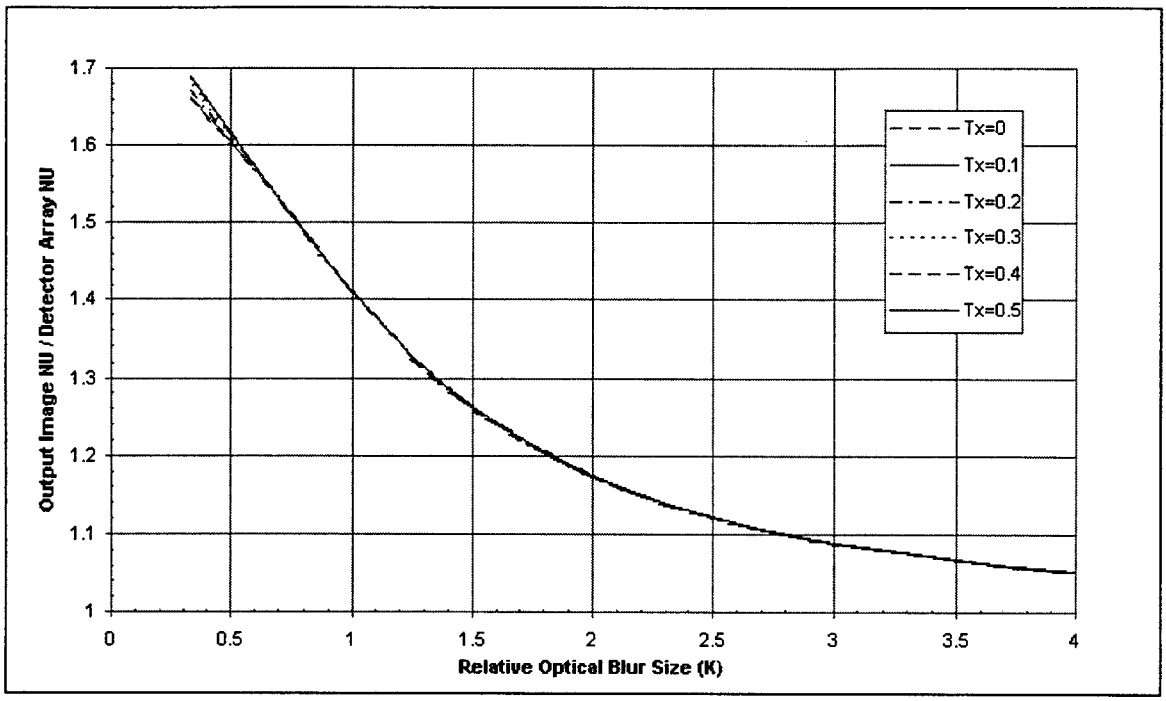


Figure 9: Output image NU vs. relative optical blur size for different alignments. Sampling ratio is 4:1, projector NU/detector NU is 5, and fill factor is 64%.

Conclusion

We have derived an equation (16) that predicts analytically how the non-uniformity of the projector will impact the uniformity of the output image. In order to be applicable to the widest range of possible systems, the equation is written in terms of the ratio of the projector's non-uniformity to the detector array's non-uniformity. The most significant parameter in determining the output image non-uniformity is the sampling ratio (linear number of emitters per detector), which sets an upper limit on the output non-uniformity, as given in Eq. (10). This upper limit decreases dramatically with an increase in sampling ratio n. Other parameters, such as relative optical blur size, fill factor of the detector array, and alignment of the emitters relative to the detectors (also called registration), have a lesser impact on the output image uniformity. In order to account for these parameters, it was necessary to develop a weighting function that determines the contribution of each emitter to the flux incident on a particular detector. The necessary integrals could not

be done analytically, so a computer program was written to obtain approximate values for the necessary weighting function.

The most important of the lesser parameters is the relative optical blur size. As intuition would suggest, the larger the optical blur, the smaller the output image non-uniformity. The reason is simply that as the blur increases each detector "sees" more emitters and so the differences between the emitters tend to balance out. The effect is quite dramatic for large blur, as illustrated in Figs. (3–4), but large blur also has a negative impact on things such as resolution. Fill factor and alignment have relatively little impact, especially as the blur is increased, as shown in Figs. (5–9). As can be seen from these figures, if the blur is large enough so that more than one detector is needed to collect 83.9% of one emitter's radiation ($K > 1$), then fill factor and alignment make relatively small differences in the output image uniformity.

This effort has provided a statistical model which predicts the impact of a projector's non-uniformity on output image uniformity. The contribution of the sampling ratio, the relative optical blur size, the alignment of emitters with the detectors, and the fill factor of the detector array can all be quantitatively predicted. It is clear from this work that the best way to reduce output non-uniformity, other than improving the uniformity of the projector, is to increase the sampling ratio.

Despite the progress that has been made, there are several more factors that should be considered. Spatial droop (also called busbar robbing) occurs when bright images are being projected by the emitter array.[5] Another issue is the accuracy of the measurement of the projector array's non-uniformity. The accuracy can be compromised both by measurement limitations and by temporal changes.

One important restriction to the model developed in this paper is the requirement that the sampling ratio n always be an integer. It is easy to show that a non-integer sampling ratio will cause a regular pattern of non-uniformity even for a perfectly uniform emitter array and a detector array with 100% fill factor. Intuitively, the amplitude of this pattern will decrease as the relative optical blur is increased. This non-uniformity is different in nature from the non-uniformity discussed in this paper and should be explored in depth.

4 Acknowledgments

The authors are grateful for many productive conversations with Lawrence Jones, Walt Krawczyk, and Eric Olson of Science Applications International Corp. and Dave Flynn and Steve Marlow of SeeTec. We also appreciate the support of WISP program manager Robert Stockbridge and KHILS program manager Lee Murrer, both of WL/MNGI.

This work was performed at the Armament Directorate, Wright Laboratory, Eglin AFB, and was funded by summer research fellowships provided by the Air Force Office of Scientific Research, Bolling AFB.

References

1. D.L. Garbo, E.M. Olson, C.F. Coker, and D.R. Crow, "Real-time Three Dimensional Infrared Scene Generation Utilizing Commercially Available Hardware," The 10th Annual International Aerospace Symposium, April 1996.

2. L.E. Jones, R.G. Stockbridge, A.R. Andrew, W.L. Herald, and A.W. Guertin, "Characterization Measurements of the Wideband Infrared Scene Projector Resistor Array (Part II)," presented at SPIE, Orlando, Florida, April 1997.

3. L.E. Jones, E.M. Olson, R.L. Murrer, and A.R. Andrew, "A Simplified Method for the Implementation of Nonuniformity Correction on a Resistor Array Infrared Scene Projector," presented at SPIE, Orlando, Florida, May 1997.

4. W.L. Wolfe and G.J. Zissis, Eds., *The Infrared Handbook*, prepared by the Infrared Information Analysis Center, Ann Arbor, MI, pg. 8-28. (1989)

5. A.P. Pritchard, M.A. Venables, and D.W. Gough, "Output accuracy and resolution limitations in resistor array infra-red scene projection systems," *Proceedings of SPIE*, vol. 2742, p. 15.

THE FINITE ELEMENT METHOD IN ELECTROMAGNETICS FOR MULTIDISCIPLINARY DESIGH OPTIMIZATION

John H. Beggs
Assistant Professor
Department of Electrical and Computer Engineering

Mississippi State University
Box 9571
Mississippi State, MS 39762

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

And

Wright Laboratory

August 1997

THE FINITE ELEMENT METHOD IN ELECTROMAGNETICS FOR MULTIDISCIPLINARY DESIGN OPTIMIZATION

John H. Beggs
Assistant Professor
Department of Electrical and Computer Engineering
Mississippi State University

Abstract

Several methods in computational electromagnetics were surveyed to determine the best approach for radar signature prediction/reduction calculations as part of a multidisciplinary design optimization (MDO) program. The finite-element method was chosen as the most suitable approach because of its versatility in simulating complex geometries and because of its similarity and interfacing with structural analysis finite-element grid generation routines. A specific finite-element electromagnetics code was then chosen, studied and tested on canonical and realistic problems. A brief description of the finite-element method for electromagnetics is given, including a description of the code and the computations. A simple example of Radar Cross Section reduction using a finite element code was also provided as a demonstration.

THE FINITE ELEMENT METHOD IN ELECTROMAGNETICS FOR MULTIDISCIPLINARY DESIGN OPTIMIZATION

John H. Beggs

1 Introduction

An ongoing program in multidisciplinary design optimization (MDO) exists within the Design and Analysis Branch of the Structures Division, which is part of the Flight Dynamics Directorate at the Air Force's Wright Laboratory. This program is ultimately designed to simultaneously examine constraints from aerodynamics, structures and electromagnetics to provide an optimized vehicle design which meets certain criteria in these areas. This is a formidable task since changing materials, parameters and/or vehicle shape to optimize the design in one particular area can have significant impact on the desired constraints in the remaining two areas.

The current state of the MDO program is that a vehicle design can be optimized for structural integrity, loadbearing and vibrations and at the same time accounting for certain aerodynamic factors. However, a fully coupled and three-dimensional aerodynamic and structural analysis is not yet tractable with current computational resources. However, the current MDO program can provide workable and reliable vehicle designs.

The main MDO program goal for this Summer Faculty research project was to begin incorporating electromagnetics as the third area of optimization. In the electromagnetics area, the vehicle will be optimized for minimal radar signature. Therefore, the specific objective of this Summer Faculty research project was to examine several methods in Computational Electromagnetics (CEM) for suitability in performing radar signature prediction and analysis for the MDO program and to select one of these methods for further study and analysis. A simple demonstration of the code capabilities on canonical and realistic problems was also desired. Another objective was to develop a strategic plan to further the EM optimization effort based upon the outcome of this Summer Faculty project. This final report outlines the work performed for this Summer Faculty project. Section 2 provides a statement of the problem, Section 3 outlines the methodology and description of the finite element method, Section 4 provides some theory behind the finite element code that was chosen, Section 5 discusses the results of the code demonstrations and Section 6 provides some concluding remarks.

2 Problem Statement

In the discipline of electromagnetics, there are three primary areas of interest in optimizing the performance of a vehicle: minimizing radar signature, electromagnetic compatibility (EMC) (minimizing electromagnetic interference) and electronic countermeasures (ECM). If a vehicle can operate in a combat environment undetected, that eases the burden on EMC management and on the ECM systems. Therefore, RCS reduction was chosen as the primary focus of an initial electromagnetic optimization initiative within the MDO program. Radar Cross Section (RCS) reduction is accomplished mainly through the use of appropriate vehicle shaping and material coatings on the vehicle skin. Vehicle skins are typically metal because they provide the most immunity from electromagnetic interference (EMI), but they can also be constructed of composite materials. However, composite materials make the vehicle more susceptible to electronic countermeasures, or electromagnetic interference (intentional or not). Therefore, additional constraints must be considered when using composite materials. Composite materials also have a definite effect on structural integrity and aerodynamics. A material that may be well suited for electromagnetic optimization may have undesirable aerodynamic or structural properties. Therefore, the advantage of an MDO program is clear as it provides the opportunity to do quick, efficient and cost effective studies on vehicle designs without the costly construction, testing and maintenance of prototype vehicles.

To incorporate electromagnetics into the existing MDO program, this Summer Faculty project must examine various CEM techniques to determine which approach would be the best match with the current MDO program and with future MDO goals and objectives. Once a specific method is chosen, a suitable code is to be selected, examined and executed on canonical problems. A simple demonstration of RCS reduction is to be provided along with a plan of action for future efforts in this area.

In order to provide a comprehensive RCS optimization as part of the overall MDO program, a separate optimization project should be undertaken to optimize a material layer using material parameters and thickness to provide the lowest RCS. This would not only be of benefit to the MDO program but to others within the Department of Defense. Perhaps such a project could be jointly funded by several DoD agencies. This type of project is beyond the scope of the current effort and will not be undertaken at this time. However, a related topic will be the subject of the Summer Research Extension Program (SREP) proposal.

3 Methodology

The methodology behind the current Summer Faculty project is to evaluate various CEM methods to determine which technique would be the best match with the MDO program. The CEM methods are evaluated based upon accuracy and efficiency, modeling of arbitrary and complex geometries and treatment of material media. A good tutorial on many of the more popular methods can be found in [1]. For the sake of brevity, a simple description of various CEM methods could not be included in this report, but the reader

should refer to reference [1]. However, the Method of Moments, Finite-Difference Time-Domain method, Transmission Line Method and characteristic-based methods were all reviewed as part of this project.

The method chosen for the electromagnetics portion of the MDO program was the Finite Element Method (FEM) [2]. There were several reasons for this choice. First, the FEM can easily treat very complicated objects with arbitrary material composition using a variety of element types such as surface quadrilaterals and volumetric hexahedrals. This can be a problem with other techniques such as Finite-Difference Time-Domain or Transmission Line Method. Second, the extension to higher-order element types and algorithms is relatively straightforward. Third, the FEM results in a sparse system of equations which can be solved more rapidly than dense matrices (such as from the Method of Moments) using specialized linear algebra techniques. For further efficiency and better accuracy, these matrices can also be symmetric by the use of a Galerkin approach with a self-adjoint operator. Finally, and perhaps most importantly, the FEM for electromagnetics can use the same finite element grid generation tools as a structural analysis code with little or no modifications. This is a significant advantage because a separate grid for the EM code does not need to be created which provides increased efficiency and productivity in the optimization procedure. Various software packages have been developed for providing finite element grids for structural analysis programs and these can be used directly for the finite element EM codes.

After the FEM was chosen as the preferred technique for RCS prediction in the MDO program, several computer codes were examined. Commercial FEM codes do exist; however, the general high cost of these codes was prohibitive for an initial optimization research and development effort. After looking at several non-commercial FEM computer codes, a code called SWITCH was chosen for implementation in the MDO program. SWITCH was chosen over other FEM codes for several reasons. First, it uses a generalized coordinate system incorporating different types of surface and volume elements to model complex geometries. Second, it is relatively straightforward to use higher-order elements for increased accuracy and resolution. Third, it interfaces with MacNeal Schwendler Corp.'s PATRAN code which is used in the WL Structures Division for computer modeling of three-dimensional geometries and finite element mesh generation. Fourth, the SWITCH code runs on both serial and massively parallel platforms which provides for a large variety in the number and type of problems that can be analyzed. Finally, SWITCH is designed to automatically calculate RCS for each run. This makes it ideal for the optimization program as the RCS is the desired output constraint for the electromagnetics code. The next section provides a brief description of the SWITCH code, its capabilities and some RCS results for an ogive and the NASA almond.

4 SWITCH Code Theory and Formulation

The SWITCH code [3]–[7] is produced by Northrop Grumman Corporation and is a frequency-domain, hybrid, finite element/integral equation code. It is specifically designed to compute RCS from geometrically complex bodies with arbitrary material composition. The FEM is used to solve for the field unknowns inside

and on the surface of the body and an integral equation (IE) method is used to enforce the Sommerfeld radiation condition on the object surface. This avoids the need to discretize a large portion of free space surrounding the object, which would lead to a large increase in the number of unknowns in 3D for a pure finite element approach. However, the integral equation method does create a dense matrix that needs to be inverted for the complete solution. This hybrid formulation is highly flexible and can be used for many different geometries. SWITCH also uses generalized coordinates for excellent modeling accuracy and further extending the flexibility of the code. This curvilinear formulation uses the covariant projection edge-based vector field expansion on brick elements in 3D. The next part of this section provides a brief description of the theory behind the SWITCH code and for a full exposition, the reader is referred to [3]. The intent of this section is to provide only a very basic introduction to the theory of the code and its input/output structure.

SWITCH solves the time harmonic vector wave equation given by

$$\vec{\nabla} \times \left[\bar{\mu}_r^{-1} \cdot \vec{\nabla} \times \vec{E}\right] - k_0^2 \bar{\epsilon}_r \cdot \vec{E} = -jk_0 Z_0 \vec{J}^i \tag{1}$$

where $\bar{\epsilon}_r$ and $\bar{\mu}_r$ are the 3×3 relative permittivity and permeability tensors describing the material characteristics, $\vec{E}$ is the electric field intensity, k_0 is the free space wave number, Z_0 is the free space wave impedance, $j = \sqrt{-1}$ and $\vec{J}^i$ are impressed current sources. For perfectly conducting boundaries, the Dirichlet boundary condition is used

$$\hat{n} \times \vec{E} = 0 \tag{2}$$

or equivalently, the Neumann boundary condition can be used

$$\hat{n} \times \vec{\nabla} \times \vec{H} = 0. \tag{3}$$

To solve (1) using a finite element method, SWITCH uses a symmetric scalar product given by

$$\left\langle \vec{W}, \vec{E} \right\rangle = \iiint \vec{W} \cdot \vec{E} dv \tag{4}$$

where $\vec{W}$ is the weighting function. A residual $\vec{R}$ is formed by putting the approximated electric field $\vec{E}_a$ into the original vector wave equation

$$\vec{R} = \mathcal{L}\left(\vec{E}_a\right) = \vec{\nabla} \times \left[\bar{\mu}_r^{-1} \cdot \vec{\nabla} \times \vec{E}_a\right] - k_0^2 \bar{\epsilon}_r \cdot \vec{E}_a + jk_0 Z_0 \vec{J}_i \tag{5}$$

This residual is minimized in a weighted sense by

$$\left\langle \vec{W}, \vec{R} \right\rangle = 0 \tag{6}$$

The inner product has the form

$$\begin{aligned} \left\langle \vec{W}, \mathcal{L}\left(\vec{E}_a\right) \right\rangle \;=\; & \iiint_V \vec{W} \cdot \left\{ \vec{\nabla} \times \left[\bar{\mu}_r^{-1} \cdot \vec{\nabla} \times \vec{E}_a\right] - k_0^2 \vec{\epsilon}_r \cdot \vec{E}_a \right\} dv + \\ & jk_0 Z_0 \iiint_{V_{s,j}} \vec{W} \cdot \vec{J}^i dv = 0 \end{aligned} \tag{7}$$

Applying a vector identity along with the divergence theorem (see Appendix A of [3]) gives the final finite element equation to be solved of

$$\iiint_V \left\{ \left[\bar{\mu}_r^{-1} \cdot \vec{\nabla} \times \vec{E}_a \right] \cdot \vec{\nabla} \times \vec{W} - k_0^2 \left(\bar{\epsilon}_r \cdot \vec{E}_a \right) \cdot \vec{W} \right\} dv \quad - \quad jk_0 Z_0 \iint_{Sa} \vec{W} \cdot \left(\hat{n} \times \vec{H} \right) ds$$
$$= -jk_0 Z_0 \iiint_{V_{sj}} \vec{W} \cdot \vec{J}^i dv \qquad (8)$$

where V is the volume of the scatterer, $\vec{E}_a$ is the approximated electric field, S_a is the aperture surface connecting the finite element volume to free space and V_{sj} is the volume that encloses the impressed (source) currents. The aperture S_a couples the finite element portion of the solution to the integral equation solution for the Sommerfeld radiation condition.

The integral equation portion is formulated for the free space region exterior to the finite element region. It can involve either the electric field integral equation (EFIE) or magnetic field integral equation (MFIE). For open metal geometries, the EFIE must be used because the MFIE results in an unsolvable system of equations. The best choice for solving the integral equation portion is through the combined field integral equation (CFIE) which is a linear interpolation between the EFIE and MFIE given by

$$(1-\alpha)\, MFIE + \frac{\alpha}{Z_0} EFIE \qquad (9)$$

where α is a mixing parameter and Z_0 is the free space wave impedance given by $Z_0 = \sqrt{\mu_0/\epsilon_0}$. A standard Green's theorem derivation to handle the radiation boundary condition exactly results in the MFIE given by

$$\hat{n} \times \vec{H}^i(\vec{R}) \quad = \quad \frac{\vec{J}(\vec{R})}{2} - \iint_{S_a} \hat{n} \times \left[\vec{\nabla} \times \vec{G}_1(\vec{R}, \vec{R}') \cdot \vec{J}(\vec{R}') \right] ds' +$$
$$jk_0 Y_0 \hat{n} \times \iint_{S_a} \vec{G}_2(\vec{R}, \vec{R}') \cdot \vec{M}(\vec{R}') ds' \qquad (10)$$

where $\vec{J} = \hat{n} \times \vec{H}$, $\vec{M} = \vec{E} \times \hat{n}$ and $Y_0 = 1/Z_0$ is the free space admittance. The Green's function $\vec{G}_2$ represents the magnetic half-space dyadic Green's function given by

$$\vec{G}_2\left(\vec{R}, \vec{R}'\right) = \left(\bar{I} - \frac{\vec{\nabla}\vec{\nabla}'}{k_0^2} \right) \left[\vec{g}(\vec{R}, \vec{R}') + \vec{g}(\vec{R}, \vec{R}_i') \right] - 2\hat{z}\hat{z}\vec{g}(\vec{R}, \vec{R}_i') \qquad (11)$$

where

$$\vec{\nabla} \times \vec{G}_1(\vec{R}, \vec{R}') = \vec{\nabla}\vec{g}(\vec{R}, \vec{R}') \times \bar{I} - \vec{\nabla}\vec{g}(\vec{R}, \vec{R}_i') \times \bar{I}_i \qquad (12)$$

and

$$\vec{g}\left(\vec{R}, \vec{R}'\right) = \frac{e^{-jk_0 \left| \vec{R} - \vec{R}' \right|}}{4\pi \left| \vec{R} - \vec{R}' \right|} \qquad (13)$$

with

$$\vec{R}' = x'\,\hat{x} + y'\,\hat{y} + z'\,\hat{z}, \qquad \vec{R}_i' = x'\,\hat{x} + y'\,\hat{y} - z'\,\hat{z}$$
$$\bar{I} = \hat{x}\hat{x} + \hat{y}\hat{y} + \hat{z}\hat{z} \qquad \bar{I}_i = \hat{x}\hat{x} + \hat{y}\hat{y} - \hat{z}\hat{z} \qquad (14)$$

Through a similar analysis, the EFIE is given by

$$\hat{n} \times \vec{E}^i(\vec{R}) = -\frac{\vec{M}(\vec{R})}{2} + \iint_{S_a} \hat{n} \times \left[\vec{\nabla} \times \vec{G}_2(\vec{R}, \vec{R}') \cdot \vec{M}(\vec{R}')\right] ds' + jk_0 Z_0 \hat{n} \times \iint_{S_a} \vec{G}_1(\vec{R}, \vec{R}') \cdot \vec{J}(\vec{R}') ds' \tag{15}$$

with

$$\vec{G}_1\left(\vec{R}, \vec{R}'\right) = \left(\bar{I} - \frac{\vec{\nabla}\vec{\nabla}'}{k_0^2}\right)\left[\vec{g}(\vec{R}, \vec{R}') - \vec{g}(\vec{R}, \vec{R}_i')\right] + 2\hat{z}\hat{z}\vec{g}(\vec{R}, \vec{R}_i') \tag{16}$$

and

$$\vec{\nabla} \times \vec{G}_2(\vec{R}, \vec{R}') = \vec{\nabla}\vec{g}(\vec{R}, \vec{R}') \times \bar{I} + \vec{\nabla}\vec{g}(\vec{R}, \vec{R}_i') \times \bar{I}_i \tag{17}$$

The solution of the integral equation portion follows the same Galerkin form of the method of weighted residuals for the finite element method and then the MFIE of (10) becomes

$$\iint_{S_a} \frac{\vec{W}(\vec{R}) \cdot \vec{J}(\vec{R})}{2} ds - \iint_{S_a} \vec{W}(\vec{R}) \cdot \iint \hat{n} \times \left[\vec{\nabla} \times \vec{G}_1(\vec{R}, \vec{R}') \cdot \vec{J}(\vec{R}')\right] ds' ds + jk_0 Y_0 \iint_{S_a} \vec{W}(\vec{R}) \cdot \hat{n} \times \iint_{S_a} \vec{G}_2(\vec{R}, \vec{R}') \cdot \vec{M}(\vec{R}') ds' ds = \iint_{S_a} \vec{W}(\vec{R}) \cdot \hat{n} \times \vec{H}^i(\vec{R}) ds \tag{18}$$

The hybrid finite element/integral equation method solves equations (8) and (18). The coupling between methods occurs by maintaing continuity of the $\vec{E}$ and $\vec{H}$ fields across the aperture surface S_a. However, there are other theoretical and numerical issues such as enforcing this field continuity, numerical discretization, higher order basis functions, etc. that will not be covered in this report for the sake of brevity. Instead, the reader must be referred to [3] for more details on these issues.

For element types, SWITCH currently uses 9 node curvilinear quadrilaterals for perfectly conducting (metal) surfaces and either 8 or 27 node curvilinear hexahedral elements for volumetric scatterers. SWITCH requires four input files: a setup file, a frequency file, an element file and a node file. The setup file is user defined and it contains file names (including directory paths) of all input and output files. The frequency file is also user-defined and it contains information on the start/stop frequencies and angles of interest for both the scattered and incident fields. The element file is not generally defined by the user, but it is user generated from a translation program. The translation program converts a PATRAN (version 3) neutral file for an object into the element and node files in the required grid format for SWITCH. The element file contains information on all elements in the grid such as the node numbers and the unknown numbers. The node file contains the x, y and z coordinates of each node in the finite element grid. More detailed information on these files can be found in [4].

SWITCH outputs a diagnostics file with various messages produced during code execution and it also outputs a plot file containing the RCS data. This RCS data is suitable for plotting with any one of several

public domain or commercial plotting programs. The examples to follow will show RCS versus angle obtained using SWITCH.

The next section describes some of the example problems that were analyzed using the SWITCH code, including difficulties or limitations encountered.

5 Results

The version of SWITCH that was first obtained from Northrop Grumman was a parallel version designed to run on an Intel IPSC/860 Hypercube, which no longer exists at Wright-Patterson AFB. Instead, a serial version of the code was desired to run simple test cases for the initial part of the electromagnetics MDO initiative. An updated serial version of SWITCH was then obtained and compiled on a Silicon Graphics PowerChallenge L server with two R10000 processors and 256 MB of total RAM. Several example problems came with the code distribution, and these problems were run using SWITCH to verify this code was functioning properly. These example problems were benchmark problems designed, outlined and maintained by the ElectroMagnetics Code Consortium (EMCC) to test and validate computational electromagnetics codes. The benchmark solution for these problems is measured RCS data taken in a highly controlled anechoic chamber environment. This measured data can be obtained from the EMCC.

One example of these benchmark problems is the finite element grid for the ogive geometry shown in Figure 1. The finite element SWITCH analysis for this geometry corresponded to 4640 surface quad

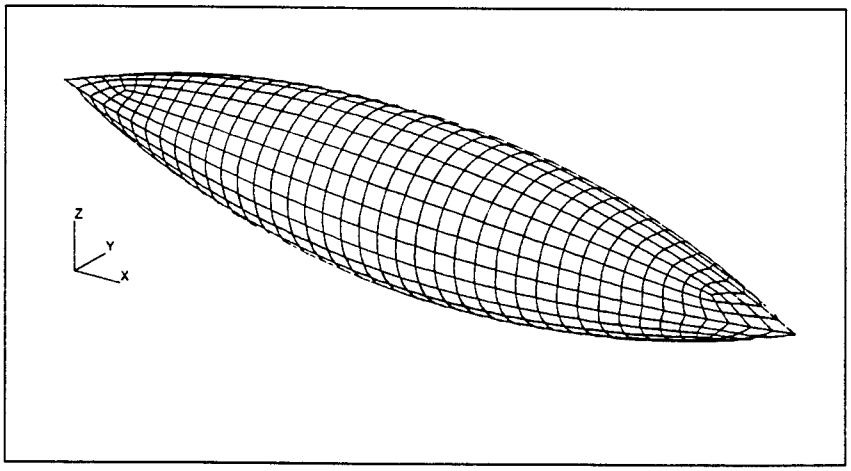


Figure 1: Finite element grid for ogive geometry.

patches, 9280 unknowns and the frequency was 9 GHz. The RCS data was calculated on an azimuth sweep ($\theta = 90°$) from $\phi = 0°$ to $\phi = 180°$, where ϕ is the angle measured from the positive x axis in the $x-y$ plane in Figure 1. Both the horizontal (in the $x-y$ plane) and vertical (perpendicular to the $x-y$ plane)

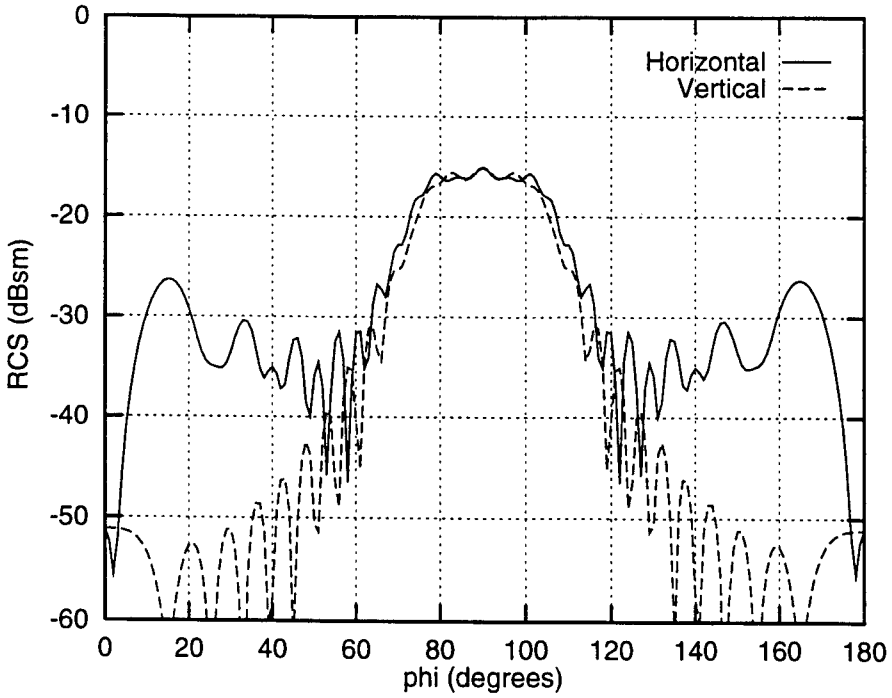


Figure 2: Horizontal and vertical polarization Radar Cross Section results for ogive geometry for $\phi = 0°$ to $\phi = 180°$ with $\theta = 90°$.

polarizations were considered. The RCS results for this geometry are shown in Figure 2. Note that the largest RCS occurs for $\phi = 90°$ at broadside incidence and the smallest RCS occurs occurs at $\phi = 0°$. This is to be expected since the largest physical cross section is visible for $\phi = 90°$ and the smallest physical cross section is visible for $\phi = 0°$. The remaining peaks and nulls in the RCS pattern come from phase additions and cancellations of the electromagnetic fields that are radiated from various parts of the object back toward the radar source and receiver. The results presented in Figure 2 agree with the ogive results presented in [6]. All other example problems provided in the code distribution were run, and the results obtained agreed with the published results in [6]. Therefore, initially the code was functioning as expected.

After running the example problems, the next problem to be analyzed was the Northrop VFY218 notional aircraft. A computer-aided design representation for this geometry is shown in Figure 3. The finite

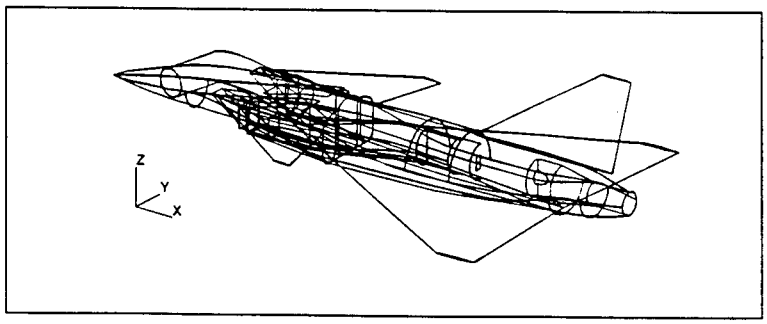


Figure 3: Computer-aided design representation of Northrop VFY218 notional aircraft.

element grid for this geometry consisted of 12,754 surface quadrilateral elements with 25,508 unknowns. When this problem was first attempted on the Silicon Graphics PowerChallenge L server, the code execution was abnormally terminated due to a code error in reading formatted input from the input element file. After contacting Northrop Grumman and after a substantial delay, this problem was resolved. However, at that time, Northrop disclosed that this notional aircraft geometry would take 10 GB of storage for the matrix and that 10 GB of disk space would be required in addition to a fast out-of-core solver. The original goal of this project was to perform a simple demonstration of the SWITCH code RCS calculations on a realistic problem, but the sheer size of this problem prohibited its completion.

As an alternative problem, a simple RCS reduction problem was chosen to illustrate that the SWITCH code could be used in an optimization environment. With the SWITCH code distribution, a NASA almond geometry was provided in both an uncoated and coated form. The uncoated geometry is shown in Figure 4 and is another one of the benchmark problems defined by the EMCC. This almond is a perfectly conducting

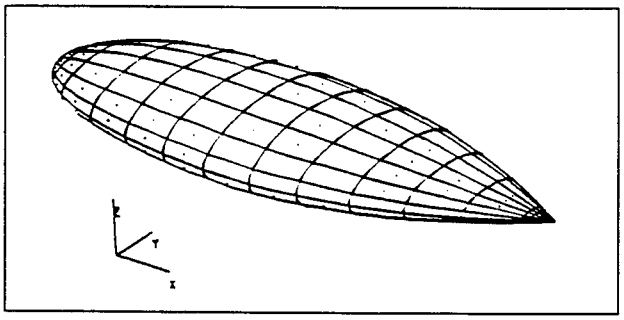


Figure 4: Finite element grid for perfectly conducting NASA almond geometry.

geometry with 248 surface quadrilateral elements having 496 unknowns. The frequency of interest was 3 GHz, the angular sweep was for $\phi = 0°$ to $\phi = 180°$ and $\theta = 90°$ (i.e. the $x - y$ plane) and both the horizontal and vertical polarizations were considered. The RCS results for this geometry are shown in Figure 5. Note that the RCS for the vertical polarization is about 5-10 dB *lower* on average than the horizontal polarization. This is to be expected because the horizontally polarized incident field sees a larger "electrical" area of the almond, hence the larger RCS. A version of this NASA almond with a material coating was also provided as part of the SWITCH code distribution. Since the VFY218 problem was too large, a RCS reduction problem was then chosen as an alternative demonstration. The idea of this problem was to use the coated almond geometry with certain material coating properties to see if the RCS could be reduced. This would demonstrate the viability of using the SWITCH code in the MDO program. Before the actual problem is outlined, a *very* brief review of RCS reduction techniques is in order.

There are fundamentally two methods for reducing RCS: working with the vehicle size and/or shape and material coatings. The simplest (and most naive) way to reduce the RCS for a vehicle is to reduce its size. However, to provide a workable aerodynamic design and a useful vehicle, this is not an option.

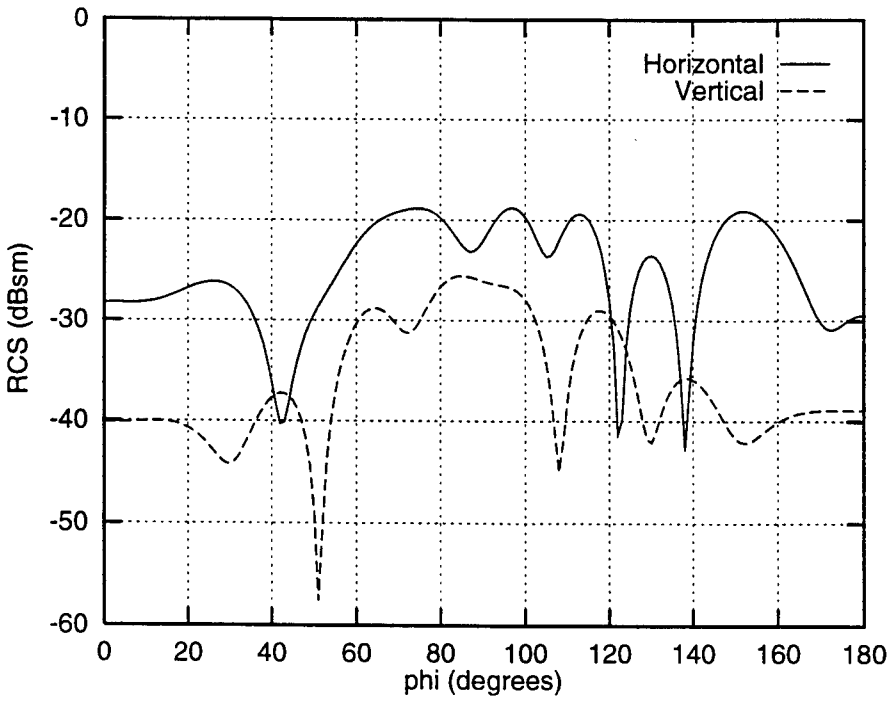


Figure 5: Horizontal and vertical polarization Radar Cross Section results for NASA almond geometry for $\phi = 0°$ to $\phi = 180°$ with $\theta = 90°$.

Therefore, the shape of the vehicle becomes a method for RCS reduction. The idea of shaping the vehicle is to provide a curved or flat-faceted surface to the vehicle so that incident electromagnetic energy will be scattered off in certain directions. For example, if the surface normals for a flat-faceted vehicle point in a certain direction, then phase cancellations could occur to give a null in the RCS signature in that direction. These phase cancellations are frequency sensitive and could even result in phase additions (i.e. a larger RCS). But the overall idea with vehicle shaping is to reduce the amount of EM energy that is reflected directly back to the radar, but instead is scattered off in some other direction.

The other method of RCS reduction is to coat the skin of the vehicle with Radar Absorbing Material (RAM). There are many different methods of coating a vehicle with a material to reduce RCS, but most of these are very frequency sensitive. The current trend in RCS reduction is to look for materials that can support RCS reduction over a large bandwidth to encompass many different threat radars. When using material coatings, there are four main parameters of interest: the relative permittivity, ϵ_r; the relative permeability, μ_r, the relative wave impedance, $Z_r = \sqrt{\mu_r/\epsilon_r}$ and the coating thickness. The basic idea with the coatings is to absorb and dissipate the incident EM energy inside the coating. It is best if $Z_r = 1$ or $\mu_r = \epsilon_r$, because this will result in no reflections of EM energy when the incident wave encounters the

surface of the coating. The permeability and permittivity are, in general, complex and are given by

$$\mu_r = \mu' - j\mu'' \tag{19}$$

$$\epsilon_r = \epsilon' - j\epsilon'' \tag{20}$$

The imaginary part of the permittivity and permeability correspond to the loss (or attenuation) in the material coating. Thus, materials are desired that have $\mu_r = \epsilon_r$ and large values of μ'' and/or ϵ''. With this situation, once the EM energy is inside the coating, it will be attenuated and the rate of attenuation is directly proportional to μ'' and ϵ''. However, the difficulty in many of these RAM coatings (especially magnetic coatings) is fabrication and weight.

To demonstrate that SWITCH can be used in RCS reduction analysis, the finite element grid for the coated NASA almond shown in Figure 6 was used. The geometry consisted of 280 hexahedral volume

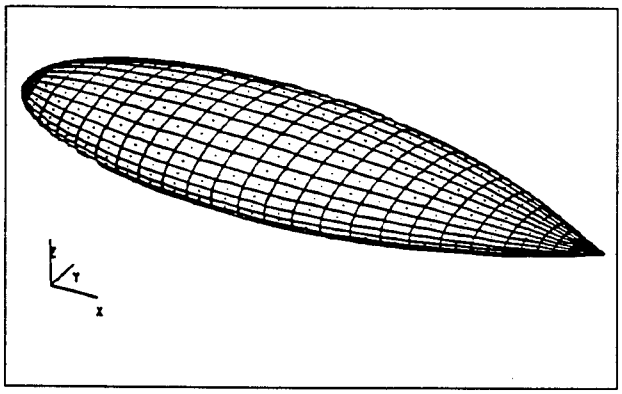


Figure 6: Finite element grid for coated NASA almond geometry.

elements with 842 electric field unknowns and 560 magnetic field unknowns. The frequency of interest was 3 GHz and the material parameters were arbitrarily chosen as $\mu_r = \epsilon_r = 1 - j5000$, with both the horizontal and vertical polarizations being considered. The RCS over three different angle sweeps was computed for $\phi = 0°$, $-90° \leq \theta \leq 90°$; $\phi = 90°$, $-90° \leq \theta \leq 90°$ and $\theta = 90°$, $0° \leq \phi \leq 180°$. These will be referred to as sweeps 1, 2 and 3, respectively. The coating was 5 mils (1 mil = 0.001 in) thick on the top and bottom of the almond and was tapered out to 35 mils thick at the edges and the point. The results for sweep 1 are shown in Figures 7 and 8 for the horizontal and vertical polarizations, respectively. Note that the RCS is reduced by about 5 dB for both polarizations over about a $50°$ angular range of θ. This angular region is approaching the point of the almond (at $\theta = 90°$ and it most likely corresponds to the thicker portions of the material coating. At broadside incidence ($\theta = 0°$), the RCS is not substantially reduced due to the lower thickness of the layer. A thicker layer would substantially reduce this broadside RCS. Figures 9 and 10 show the RCS results for sweep 2. Note that a 2-3 dB reduction in RCS is present over the entire angular range for the horizontal polarization and for about a $50°$ angular range for the vertical polarization. Although 2-3 dB may seem like a small reduction in RCS, it is still a significant amount. Figures 11 and 12 show the RCS results for sweep 3. Note again for the horizontal polarization about a 5 dB RCS reduction

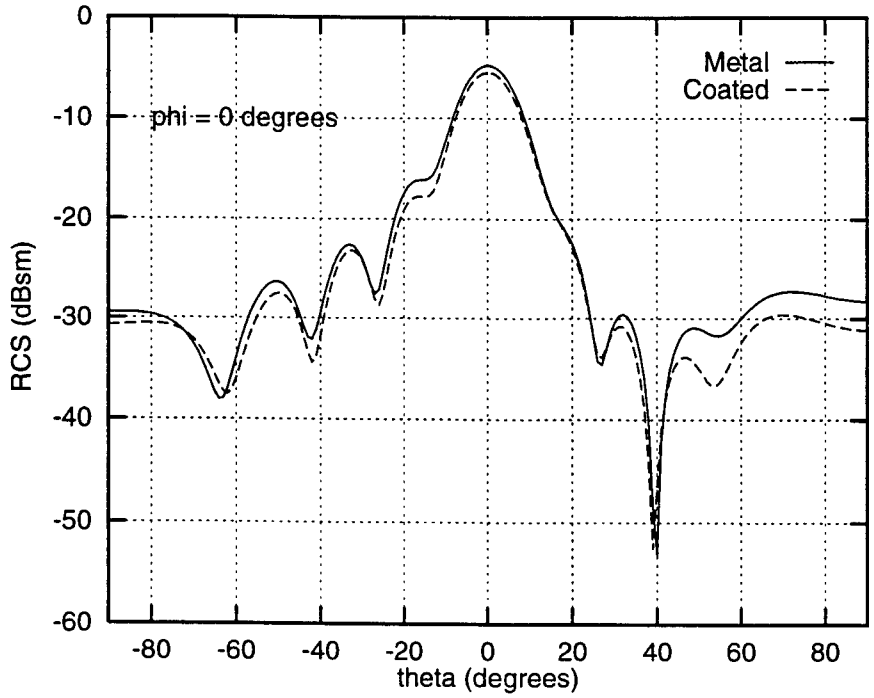


Figure 7: Horizontal polarization Radar Cross Section results for coated NASA almond geometry for $\phi = 0°$.

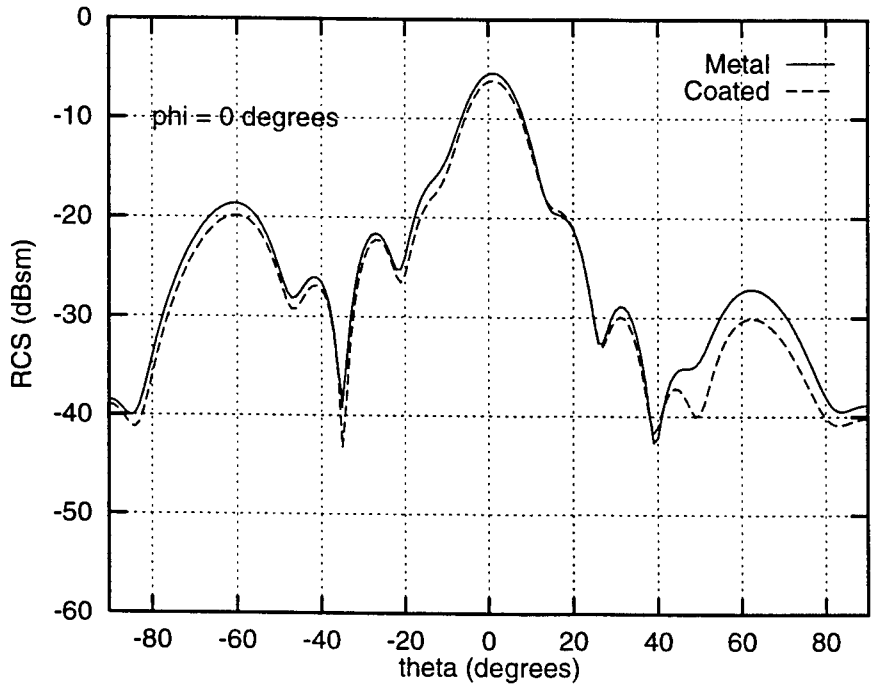


Figure 8: Vertical polarization Radar Cross Section results for coated NASA almond geometry for $\phi = 0°$.

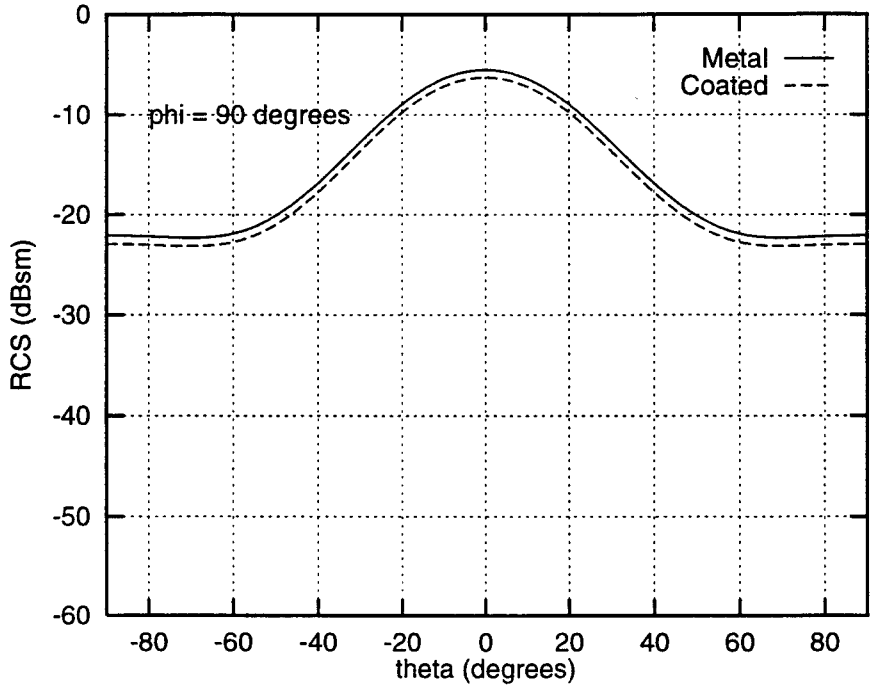


Figure 9: Horizontal polarization Radar Cross Section results for coated NASA almond geometry for $\phi = 90°$.

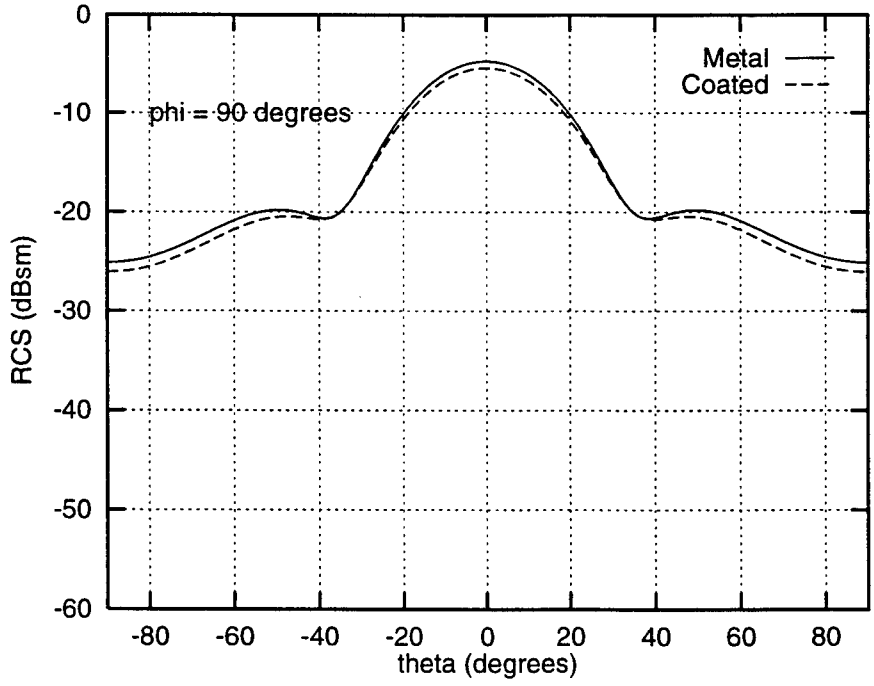


Figure 10: Vertical polarization Radar Cross Section results for coated NASA almond geometry for $\phi = 90°$.

in approximately a 30° angular range starting at $\phi = 0°$ and about a 2-3 dB reduction otherwise. Even the vertical polarization exhibits a 2-3 dB reduction in RCS over much of the angular sweep. Therefore, by applying a material coating and with a certain selection of material parameters, the SWITCH code has successfully predicted a reduction in Radar Cross Section of 2-3 dB on average at all angles in space for a canonical object. This problem could be optimized even further by applying a separate optimization routine to the material layer itself to determine layer thickness and material parameters to reduce reflections within a certain angular region to a minimum.

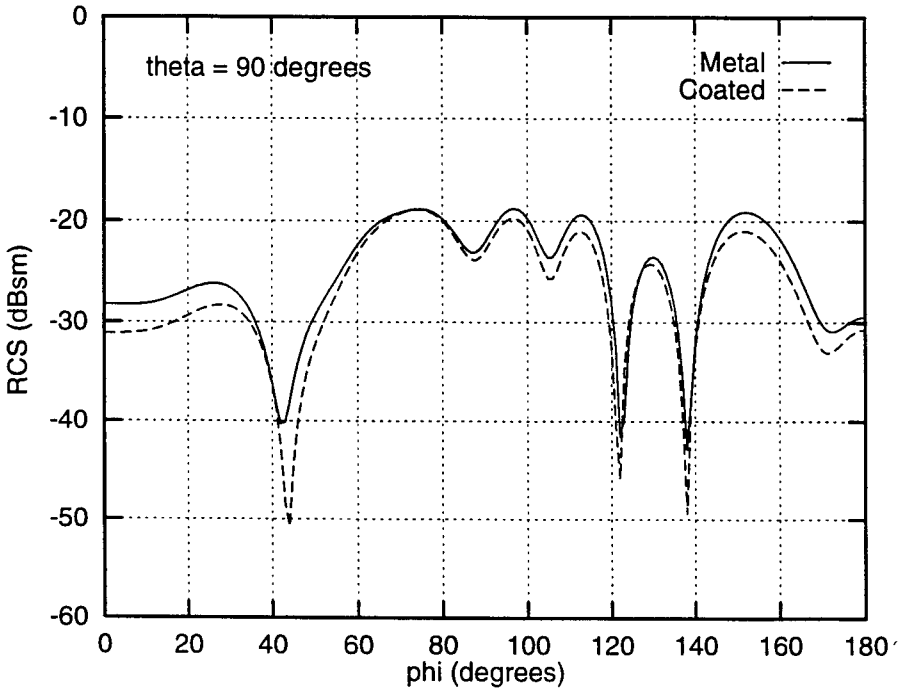


Figure 11: Horizontal polarization Radar Cross Section results for coated NASA almond geometry for $\theta = 90°$.

Several observations were made during this project about the SWITCH code and the hybrid finite-element/integral equation approach. First, the SWITCH code turns into a Method of Moments code when the object is only perfectly conducting. This is a serious disadvantage because of the dense matrix structure of the MoM and because electrically large objects consume enormous computer resources as demonstrated by the attempt to analyze the VFY218 notional aircraft. In optimizing a vehicle for RCS, the untreated metal vehicle must be analyzed to provide the "benchmark" or worst-case RCS. If the computational method renders this problem nearly intractable, that is a significant disadvantage. Perhaps a pure finite element method may be more appropriate with a sparse matrix. The disadvantage with a pure FEM is that a portion of the space surrounding the object must also be included in the grid and the outer grid boundaries must be appropriately terminated. However, the Perfectly Matched Layer (PML) [8] has shown to be an effective outer boundary treatment for the FEM. Second, the code should probably be written in C++ to take advantage of the object-oriented (OO) structure of C++ in creating data structures

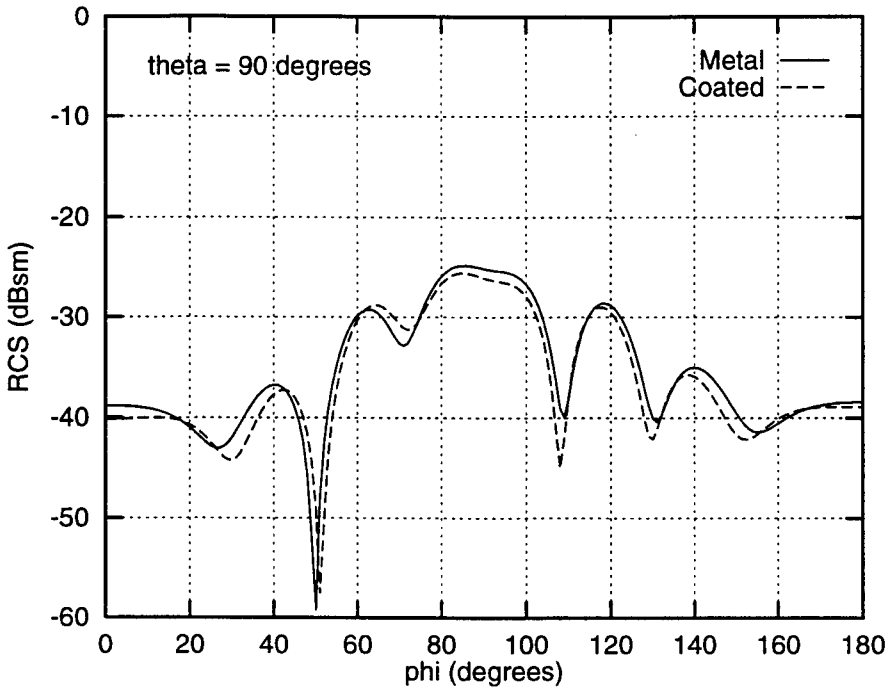


Figure 12: Vertical polarization Radar Cross Section results for coated NASA almond geometry for $\theta = 90°$.

and classes. This will lead to a more compact and efficient code in addition to being more maintainable, more portable, more extensible and easier to interface with parallel programming libraries. Third, there are no formal guidelines as to when the SWITCH code can break down or what its limitations are. Therefore, to continue using the SWITCH code, a study must be performed to determine the performance limitations of this code.

6 Conclusion

A hybrid finite element/integral equation method has been reviewed and chosen as a potential candidate for the electromagnetics portion of the MDO program. The FEM has the advantages of easily treating complicated objects with curvilinear elements, straightforward extension to higher order elements, a sparse system of equations, and easily interfacing with finite element grid generation tools for structural analysis. The integral equation method avoids the requirement of discretizing space surrounding the object and terminating that with an outer boundary radiation condition. However, the integral equation portion does have some disadvantages. It reverts to a MoM for perfectly conducting objects which is a serious problem in obtaining the benchmark RCS for the entire vehicle. This problem can be mitigated by the use of the geometrical components approach (described shortly). A code called SWITCH was obtained and tested on canonical problems and was also used to demonstrate a simple RCS reduction problem.

Based upon the work completed in this project, the following strategic plan is proposed for continuation of this work:

1. To continue using the SWITCH code, determine the performance limitations of the SWITCH code using various resolutions of vehicle grids.
2. Determine if a suitable pure FEM code is available that uses curvilinear or unstructured elements.
3. Obtain any codes in C++ or translate to C++.
4. Begin a material coating optimization project to optimize a material layer backed by a perfect conductor for minimal reflection.
5. Use the geometrical components approach to the electromagnetics MDO problem.
6. Obtain a graduate student and/or post-doctoral fellow with experience in using the finite element method for electromagnetics.

Pursuing each of these initiatives should result in a highly efficient, flexible, maintainable and accurate electromagnetics optimization. Having full time researchers available with direct experience in the finite element method for electromagnetics will allow this effort to get off the ground much more rapidly. Also, Dr. Gary Thiele of the University of Dayton has proposed using a geometrical components approach to the electromagnetics MDO problem. This type of approach is used in practice by current RCS designers. The idea is to align as many surface normals as possible in a given direction. Then various features that need RCS reduction are identified and a code (or codes) is applied to certain portions of the vehicle independently and the RCS is minimized for that portion only. This approach is also physically accurate for higher frequencies because different portions of the aircraft will be separated by several wavelengths and will not interact with each other. This geometrical components approach avoids the need to model the entire vehicle and makes the problem much more tractable both from an electromagnetics point of view and from the overall MDO program point of view. Perhaps the geometrical components approach could be applied to the structures and aerodynamics portion to maximize the flexibility and benefit to the MDO program. In the PI's opinion, this approach would probably be the most accurate method considering both the vehicle shape and material coatings. Another alternative would be high frequency methods such as Physical Optics (PO) or Physical Theory of Diffraction (PTD), but the geometrical components approach should be more accurate than these techniques.

7 Acknowledgements

The PI would like to thank and acknowledge the Wright Laboratory, Structures Division, Design and Analysis Branch for their financial and technical support and assistance throughout this Summer Faculty Program. The support and encouragement of Dr. Vipperla ("Van") Venkayya was especially helpful.

References

[1] M. N. O. Sadiku, *Numerical Techniques in Electromagnetics*, CRC Press, Boca Raton, FL:, 1992.

[2] Jianming Jin, *The Finite Element Method in Electromagnetics*, John Wiley & Sons, New York, 1993.

[3] G. E. Antilla, Y. C. Ma, M. I. Sancer, R. L. McClary, P. W. Van Alstine and A. D. Varvastis, "Development and implementation of computational electromagnetic techniques on massively parallel computing architectures, Volume 1: SWITCH code theory manual", Final Report, WL-TR-94-6009, Wright Laboratory, Wright Patterson AFB, OH, July 1994.

[4] G. E. Antilla, "Development and implementation of computational electromagnetic techniques on massively parallel computing architectures, Volume 2: SWITCH code user's manual", Final Report, WL-TR-94-6010, Wright Laboratory, Wright Patterson AFB, OH, July 1994.

[5] B. M. Sherrill and G. E. Antilla, "Development and implementation of computational electromagnetic techniques on massively parallel computing architectures, Volume 3: SWITCH code installation and porting guide", Final Report, WL-TR-94-6011, Wright Laboratory, Wright Patterson AFB, OH, July 1994.

[6] G. E. Antilla, "Development and implementation of computational electromagnetic techniques on massively parallel computing architectures, Volume 4: SWITCH code test case manual", Final Report, WL-TR-94-6021, Wright Laboratory, Wright Patterson AFB, OH, July 1994.

[7] M. I. Sancer, G. E. Antilla, B. M. Sherrill, Y. C. Ma, and D. E. Cass, "Development and implementation of computational electromagnetic techniques on massively parallel computing architectures, Volume 5: SWITCH code final report summary", Final Report, WL-TR-94-6020, Wright Laboratory, Wright Patterson AFB, OH, July 1994.

[8] J.-P. Berenger, "A perfectly matched layer for the absorption of electromagnetics waves", *J. Comput. Phys.*, vol. 114, no. 1, pp. 185–200, 1994.

SYNTHESIS OF NOVEL ORGANIC COMPOUNDS AND POLYMERS FOR TWO PHOTON ABSORPTION, NLO, AND PHOTOREFRACTIVE PHOTONICS APPLICATIONS, AND VISIBLE DYE-SENSITIZED PHOTOPOLYMERIZATION

Kevin D. Belfield
Associate Professor
Department of Chemistry

University of Detroit Mercy
4001 West McNichols, P.O. Box 19900
Detroit, MI 48219-0900

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

July 1997

SYNTHESIS OF NOVEL ORGANIC COMPOUNDS AND POLYMERS FOR TWO PHOTON ABSORPTION, NLO, AND PHOTOREFRACTIVE PHOTONICS APPLICATIONS, AND VISIBLE DYE-SENSITIZED PHOTOPOLYMERIZATION

Kevin D. Belfield
Associate Professor
Department of Chemistry
University of Detroit Mercy

Abstract

The synthesis of novel low molar mass organic chromophores, bearing arylamine electron-donating and phosphonate, nitro, or benzothiazole electron-withdrawing functionalities was undertaken. Two new fluorene-derived molecules were synthesized via amination and Heck coupling reactions. 2-(4-Iodophenyl)benzothiazole, the penultimate precursor for 2-(4-vinylphenyl)benzothiazole, was prepared. Regiospecific bromination methodology was investigated to prepare specifically brominated aromatic amines, including N-4-bromophenyl N,N-diphenylamine, bis(4-bromophenyl)amine, poly(3-bromo-9-vinylcarbazole), and poly(3,6-dibromo-9-vinylcarbazole). Attempted dibromination of triphenylamine, however, lead to a mixture of mono-, di-, and tribrominated products. Regiospecific O-allylation of N-(3-hydroxyphenyl)-N-phenylamine was attempted under two sets of reaction conditions, affording, in each case, a mixture of predominately O-allylation accompanied by a lesser amount of N-allylation. Heck coupling of diethyl 4-vinylbenzene phosphonate and N-4-bromophenyl N,N-diphenylamine was conducted to form 4'-N,N-diphenylamino-4-diethylphosphonostilbene. In a similar manner, Heck reaction of poly(3-bromo-9-vinylcarbazole) and 4-nitrostyrene was carried out to produce poly(3-(4'-nitrostilbene)carbazole).

Preliminary visible dye photoinitiated polymerizations were conducted to assess the possibility of developing a polymerization initiator system that can utilize infrared two-photon pumped up-conversion fluorescene as a means to generate spatially resolved visible photons. Photopolymerization of an acrylate/epoxy functionalized monomer was accomplished with a commercial dye/coinitiator system and one based on one of the aforementioned new fluorene-derived compounds.

SYNTHESIS OF NOVEL ORGANIC COMPOUNDS AND POLYMERS FOR TWO PHOTON ABSORPTION, NLO, AND PHOTOREFRACTIVE PHOTONICS APPLICATIONS, AND VISIBLE DYE-SENSITIZED PHOTOPOLYMERIZATION

Kevin D. Belfield

Introduction

Multiphoton absorption

Multiphoton absorption can be defined as a simultaneous absorption of two or more photons via virtual states in a medium. The process requires high peak power which is available from pulsed lasers. Even though multiphoton processes have been known for some time, materials that exhibit a multiphoton absorption have yet to find widespread applications. The reason for this is that most materials have a relatively low multiphoton absorption cross sections, σ. The discovery of multifunctional organic materials with large multiphoton absorption cross sections has opened up a new area of research in the photonic and biophotonic fields.[1-5] Two-photon pumped up-conversion lasing, multiphoton absorption-induced optical power limiting, multiphoton laser scanning microscopy, two-photon three dimensional optical data storage, and two-photon photodynamic therapy are some of the promising applications.[6,7]

A material can interact with optical fields in two ways: through a dissipative process, or through a parametric process. In a dissipative process, exchange of energy between molecules and the optical field takes place through absorption and emission. In the parametric process, on the other hand, energy is exchanged between different modes of the optical field but no energy is exchanged between the optical field and the molecules of the system.[9] The two photon absorption process is a nonlinear dissipative process. The energy exchanged between the light beam and the medium, per unit time and unit volume, is:

$$\frac{dW}{dt} = [\, E.\, P \,] \qquad (1)$$

Where E and P are the electric field and the polarization vector respectively, and the brackets indicate a time average over several cycles of the field. Two photon absorption is often described in terms of the two photon absorption cross section σ_2:

$$\frac{dn_p}{dt} = \sigma_2 N F^2 \qquad (2)$$

Where N is the number of absorbing molecules per unit volume, F = I / hν, the photon flux. The two photon absorption cross section is usually determined experimentally based on the change in intensity of an incident laser beam.

In many multiphoton absorption systems, a fluorescence emission at a wavelength shorter than that of the exciting laser light is observed, this is referred to as "up-converted fluorescence emission". Detection of such fluorescence takes advantage of the dependence of the fluorescence intensity on the excitation intensity. For example, a quadratic dependence corresponds to two photon absorption while a cubic dependence implies a three photon absorption process.

Two-photon absorption and optical limiting

Optical limiters rely on liquid or solid nonlinear media. The optical limiter is transparent at low light intensity and has a threshold above which the transmitted intensity remains constant. Liquid limiters are desirable because of their resilience. The heat from high intensities is usually dispersed in the solvent, affording protection for the sensor. Solid hosts on the other hand suffer from thermal heating and hence the composite may irreversibly get damaged. Most research efforts to design optical limiting materials have focused on carbon black suspensions, organometallics, fullerenes, semiconductors, liquid crystals, and polymer composites. However, it is worth mentioning that the most important and amazing sensor is, unquestionably, the human eye.

New applications have recently come to the surface in the area of optical power limiting, such as eye and sensor protection against intense light. Optical limiters can be divided into two categories, active and passive limiters. Active limiting is too slow to be used for practical applications. On the other hand, passive limiting is a promising area of research. It uses the inherent nonlinear optical properties of the material to sense the incident intensity and alter the transmittance. Passive limiters are referred to as smart materials. In the following overview, the development of the most widely studied optical limiters are briefly discussed along with their limitations.

Carbon Black Suspensions (CBS) Carbon is a uniform absorber in the visible region. Three mechanisms have been proposed for the limiting action of CBS: nonlinear absorption, nonlinear scattering, and nonlinear refraction. It was found that a limiting action occurs when CBS is placed in a cell in an intermediate focal plane and the intensity of the incident light is raised.[11] When CBS is subjected to visible radiation, small particles will absorb the radiation, and the energy is transferred to break the particles apart and into heating the system. This leads to thermal lensing, which can be used for optical limiting. When the incident light is of sufficient intensity, the suspension gives off a white flash. The lifetime of this flash is in the order of 30 ns. A major disadvantage of CBS systems is that under fast repetitive pulse limiting, the material suffers a limiting reduction because of degradation and particle breakage.

Organometallics Due to their widely varying optical properties and the potential for molecular engineering to allow tailoring of a molecule for a specific application, organometallic compounds have attracted the attention of researchers for optical limiting. The two classes of organometallics which have been extensively investigated are metal macrocycles and metal cluster compounds. Blau has reported reversed saturable absorption properties in free tetraphenyl porphyrins (H2TPP) and metallated complexes (ZnTPP and CoTPP).[12] Perry has recently reported silicon naphthalocyanine for broadband limiting.[13] The same material was previously reported to have a high third-order susceptibility at 1907 nm.[14] On the other hand, metal clusters were reported to have shown limiting actions. In this regard, a number of iron-tricobalt cluster compounds have shown limiting behavior.[15,16] The limiting action depends on the nature of the ligands. A new class of materials known as the "King complex" have recently become very important as organometallics for photonic applications. The King complex was discovered by R. B. King in 1966.[17] Numerous derivatives of the King complex have been studied ever since.[18,19] The King complex has been doped into a MMA polymer and the optical limiting response was measured. The presence of a transition metal adds a number of optical transitions that do not occur in organic compounds. The stronger absorption found was due to d-d electronic transitions.

Fullerenes In 1985, a new class of materials, known as Buckminster fullerenes was found in laser ablation products.[20] Optical properties of fullerenes have been investigated by a number of researchers. Early studies revealed that C_{60}, a fullerene, had a higher excited state absorption cross section than the ground state absorption cross section, over the visible spectrum.[21] This information implies that fullerenes may have application in optical limiting for sensor protection. To understand the optical limiting mechanism of fullerenes, C_{60} was embedded in a solid host of PMMA.[22] The transmitted intensity was measured with respect to incident intensity and the threshold was found to be relatively high. The results were explained by the nonlinear absorption. In solution, the nonlinear absorption is supplemented by nonlinear scattering and refraction. Sun *et al.* have studied the optical limiting properties of methano[60]fullerene and pyrrolidino[60]fullerene derivatives.[23] Figure 1 shows the optical limiting results of methano-C_{60} benzoyl derivative in toluene solutions. The functionalized C_{60} derivatives studied have shown optical limiting efficiencies similar to that of the parent C_{60}. However, the linear absorption and emission properties were found to be very different from those of the parent C_{60}. There has been some research done on mixtures of $C_{60/70}$, which also exhibited promising limiting behavior.[24]

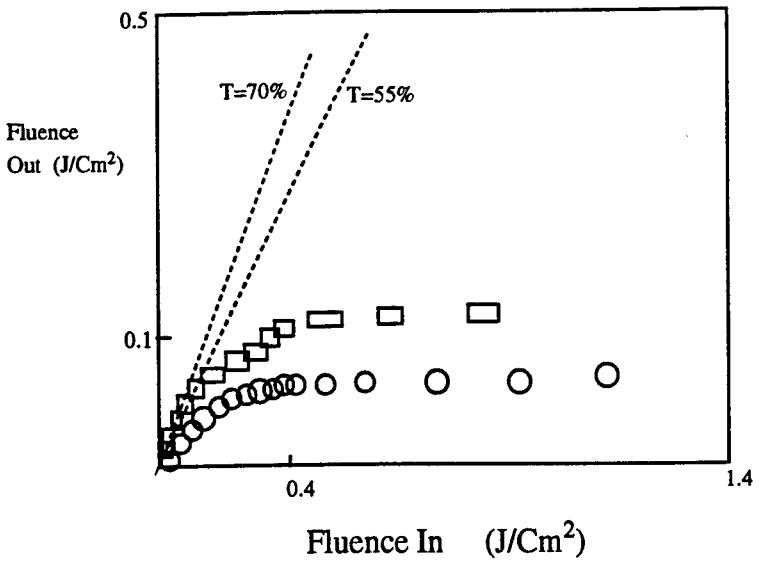


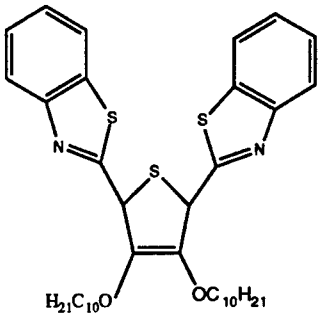
Figure 1. Optical limiting of methano-C60 benzoyl derivative in toluene solutions with linear transmittances of 55% (O) and 70% (□).

Semiconductors The broad range of diverse nonlinearities semiconductors exhibit can be applied to passive optical limiting. In fact, optical limiting based on nonlinear absorption in semiconductors has been extensively demonstrated.[25-27] In 1984, the concept of nonlinear absorption was applied in a semiconductor to construct an optical limiter. The Si and GaAs optical limiters are considered models for most of the semiconductor optical limiters. Semiconductors are attractive as elements in nonlinear optical devices because of their large and fast optical nonlinearity. Despite the fact that semiconductors exhibit varied optical linearity, these materials have yet to meet all requirements needed for an optical limiter. Short pulses for activation are often required for materials exhibiting broad band operation. Materials that operate as optical limiters at eye-safe levels have been reported.[28]

Liquid Crystals Liquid crystals have been used in active devices such as watches, calculators, and other electronic devices. Despite the fact that liquid crystals can and have been used for protection of sensors, they suffer a problem of speed. Liquid crystals can be used for passive optical limiting.[29] This class of optical limiters has recently received much interest in industry for their application to imaging devices. The possibility of orentationally ordering liquid crystals in response to light gives rise to many of their interesting properties. Research has been conducted to study liquid crystals of various mesophases.[30-34] Khoo *et al.* have studied the response of nematic liquid crystals.[35] They observed that certain liquid crystals have two features in the

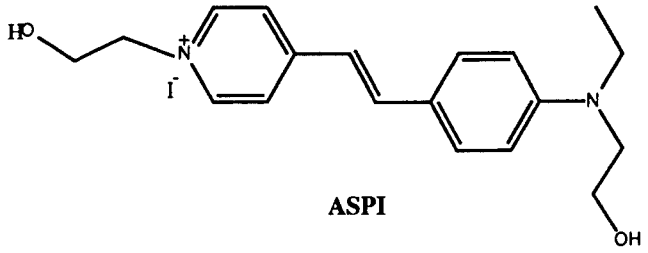
nonlinear refractive index: a fast component, 10,000 ns, and a slow component, 100 ms. Further improvement in liquid crystals should yield further improvement in optical limiting performance.

Polymer Composites The last class of materials for optical limiting that will be discussed is polymer composites. The optical limiting behavior of polymer composites has lately become of interest for researchers in the photonic and biophotonic fields due to the many advantages polymers afford, such as minimal loss due to volatilization during processing and, ideally, no phase separation.[1,2,4,36,37] Silica gel-PMMA and PMMA composites have been reported as hosts for organic optical limiting materials. Organic fluorophores (dyes that emit light) have been doped in optically transparent polymers. Prasad, Reinhardt, and coworkers have doped 2,5-bisbenzothiazole 3,4-didecyloxy thiophene (**BBTDOT**) in PMMA and studied the two-photon absorption of the system.[5]

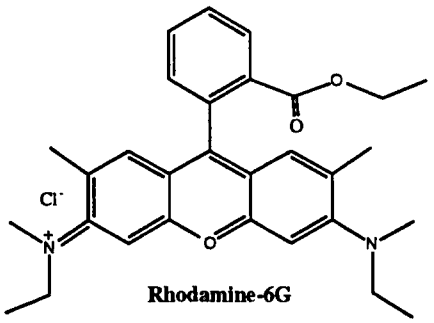


BBTDOT

Free radical polymerization of MMA was carried out using AIBN as initiator. Under illumination of intense visible laser radiation, the composite emits upconverted fluorescence frequencies. This result implies that a strong two-photon absorption process occurred. In 1995, Prasad *et al.* have doped two lasing dyes, Rhodamine-6G and trans-4-[p-(N-ethyl-N-(hydroxyethylamino)phenylstyryl]-N-(hydroxyethyl)pyridinium iodide (**ASPI**), in PMMA.[38] In this solid state laser system, energy transfer between Rhodamine-6G and **ASPI** was studied.



ASPI



Rhodamine-6G

The optical response of an ideal optical power limiter is shown in Figure 2. The ideal optical limiter is completely transparent at low light intensities until a certain intensity level is reached. Above this threshold, the transmitted intensity remains at a constant value.

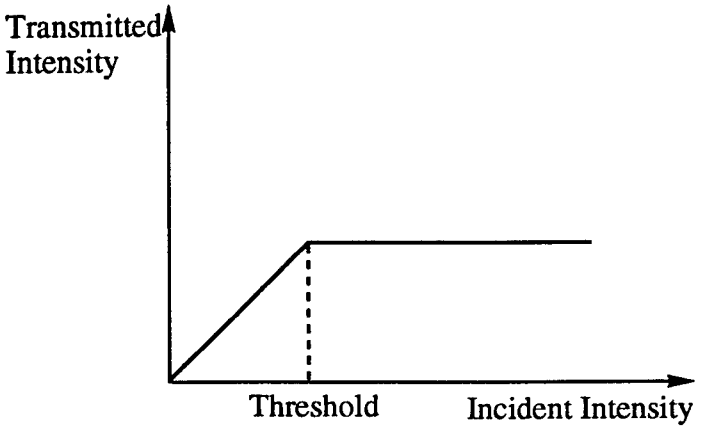


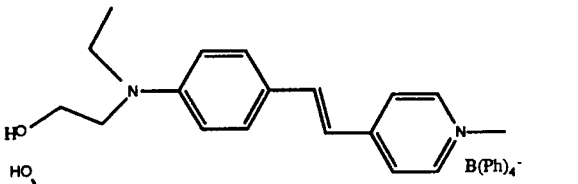
Figure 2. The optical response of an ideal optical power limiter.

Solubility limitations of two-photon chromophores in polymeric composites could be eliminated through covalent attachment of the fluorophores to polymers. The limitation of very low dye concentration can be overcome by incorporating the dye molecules onto every repeat unit. Another advantage of using polymers is that, in principle, they could easily be processed, making them potentially important in a number of different multiphotonic applications. Efforts directed toward this goal will be presented in this report.

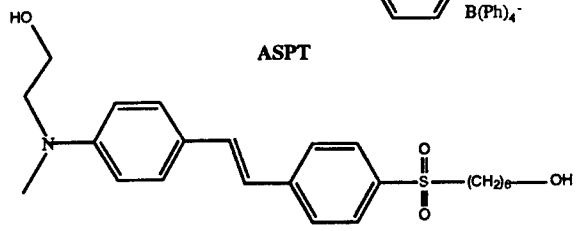
Two-Photon Pumped Lasing in Novel Dyes

Two-photon pumped lasing in organic dyes is an area of extensive ongoing research. It involves a direct absorption of two photons through virtual states. Commercial dyes such as Rhodamine 6G have been extensively studied. Recently, several materials were reported and the

field has been steadily growing.[37,39] The advantage of two-photon pumped lasing in dyes, as compared to conventional single photon pumped lasing, is that the pump wavelength can be shifted to longer wavelengths where dyes are relatively photostable. Mukherjee reported two-photon pumped upconverted lasing in a waveguide.[37] The material consisted of laser dye, 4-dicyanomethylene-2-methyl-6-*p*-dimethylaminostyryl-4H-pyrane (commercially known as DCM) doped at a 5 mmol concentration, into the polymer PMMA. Prasad and coworkers have reported two materials: trans-4-[N-ethyl-N-hydroxylethyl-amino)styryl]-N-methylpyridinium tetraphenylborate (**ASPT**) and 4-[N-(2-hydroxyethyl)-N-(methyl) amino phenyl]-4'-(6-hydroxysulfonyl)stilbene (**APSS**).[3,40]



ASPT



APSS

It was found that these materials have a significantly larger two-photon absorption cross section than that of Rhodamine 6G. The dyes were found to exhibit two-photon pumped lasing at very low pump energies. **ASPT** shows a strong two-photon absorption with upconverted yellow-red fluorescence and lasing at 600 nm. **APSS** exhibits two-photon pumped lasing at 555 nm. The dye was doped in a methyl methacrylate polymer at a concentration of 8×10^{-3} mole.

Solid-state dye lasers were reported to have advantages over liquid dyes as to compactness, absence of toxic solutions, and being environmentally more benign. However, in single-photon pumping, solid dyes suffer degradation, making it difficult to achieve long lasing lifetimes in solid state. This problem is not as severe in the case of two-photon pumping where the lifetime becomes longer as the pump operates at a longer wavelength. **ASPT** and **APSS** have such high solubilities in common organic solvents that it was possible to dope them in sol-gel processed glass and obtain two-photon pumped lasing.[40] Sol-gel processing has become an increasingly more important class of materials called multiphasic nanostructured composites. In this technique, it is possible to control the pore size and density. Usually, a methacrylate monomer is used to fill the pores. It is then polymerized *in situ*. Sol-gel processing has improved the quality of

composites because of the low phase separation associated with the process under lasing conditions. It was found that the lasing lifetime of such solid state dye lasers is limited as a result of damage in the host matrix. Improving the dye system and the matrices will remain a big challenge to enhance the lifetime even furthe

Multiphoton Confocal Microscopy for Material Science

In 1990, Denk *et al.* coupled two-photon induced fluorescence with laser scanning microscopy to probe surfaces.[41] The major advantage of this method over single-photon scanning is that the fluorescence intensity of a two-photon process is quadratically dependent on the illumination intensity. This makes the fluorescence emission limited to vicinity of the focal point and hence, it is possible to achieve depth discrimination. Multiphoton confocal laser scanning microscopy could be a useful tool to study surfaces, interface and fractures in polymer or glass specimens. Bhawalkar *et al.* reported images of fractures in polymer samples. The images were of a methacrylate polymer matrix containing organic fluorophores.[39] Two-photon multichannel confocal microscopy was demonstrated to be useful to probe and construct images of multilayered coatings.

Two-photon pumped-upconversion fluorescence has potential for three dimensional or spatially resolved photointiated polymerization,particularly in composites, adhesives, and sealant applications. For example, if an IR absorbing two-photon dye is used, IR radiation could be used to induce two-photon absorption of the dye, followed by subsequent emission of a visible photon. This visible photon could be absorbed by a visible absorbing dye capable of sensitizing a photoinitiator. The deep penetration of IR radiation can be exploited to achieve deep curing of monomers, for example, in cracks and seals. Utilization of a confocal configuration, described below, true stereolithography should be possible.

Photorefractive Materials

The photorefractive effect, found in materials that are both photoconductive and have nonlinear optical properties, is avidly being pursued for optical processing applications. Photorefractivity holds great potential in holographic optical data storage,[42] optical computing and switching, integrated optics, as well as frequency doubling of laser light. Photorefractive materials can, in principle, execute such integrated optoelectronic operations as switching, modulation, threshholding, and parallel processing for image processing and display. Until 1990, only inorganic materials were found to be photorefractive. Since then, organic crystals and, more recently, polymers have been synthesized that display photorefractivity.[42]

Photorefractive polymeric materials promise many of the traditional advantages associated with polymers, such as good thermal stability, low dielectric constant, geometric flexibility, and ease of

processing. In order to manifest the photorefractive effect, it is thought that these polymers must contain photocharge generating (CG) and transporting (CT) functionality, charge trapping sites, and nonlinear optical (NLO) chromophores. Among the major issues to be elucidated in the field of organic photorefractive materials are: increase stability from phase separation, increase temporal stability, increase thermal stability, increase diffraction efficiency, improve charge transport efficiency, develop wavelength sensitivity, improve electro-optic coefficients, create efficient synthetic methods, and develop reproducible processing methods.

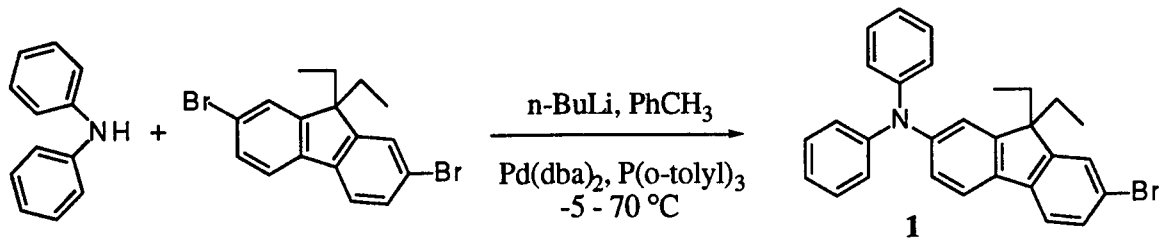
Objectives

Described herein are efforts to prepare polymers in which each repeat unit serves both CT and NLO functions. We have successfully demonstrated this approach[43] in the preparation of photorefractive polysiloxanes. Commercially available poly(9-vinylcarbazole) served as the starting substrate for the synthesis of highly functionalized polymers, described in the following section. Structural motifs of two-photon absorbing dyes, nonlinear optical chromophores, and photorefractive materials are intimately interrelated. The research described herein is directed at the synthesis of low molar mass organic compounds and polymers designed to function as two-photon absorbing dyes, nonlinear optical chromophores, and photorefractive materials. In addition, dye-sensitized visible photoinitiated polymerizations experiments were conducted to gather preliminary data for two-photon up-converted emission-initiated polymerization.

Results and Discussion

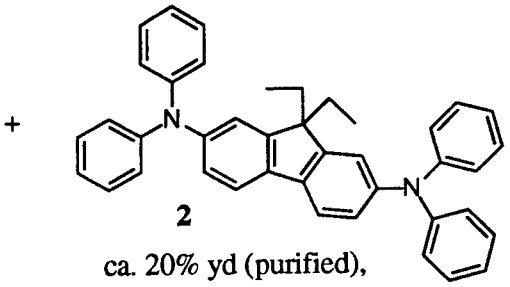
Two new fluorene-derived molecules were synthesized via amination and Heck coupling reactions, as illustrated in Schemes 1 and 2. Pd-catalyzed amination afforded the diarylaminefluorene derivative **1** as colorless cubic crystals in low yield. Meanwhile, disubstitution product **2** was also obtained as orange platelets. Quantitative Pd-catalyzed Heck coupling[44] of diarylfluoreneyl bromide **1** with either diethyl 4-vinylbenzene phosphonate or 4-nitrostyrene afforded novel fluorene dyes **3** and **4**, respectively. Fluorene dye **3** was fluorescent yellow with two λ_{max}, one at 308 nm and the other at 383 nm. The visible absorption of **3** extended out to about 480 nm. Fluorene dye **4** was fluorescent orange-red also with two λ_{max}, one at 309 nm and the other at 414 nm. The visible absorption of **4** extended out to about 550 nm.

Scheme 1



1

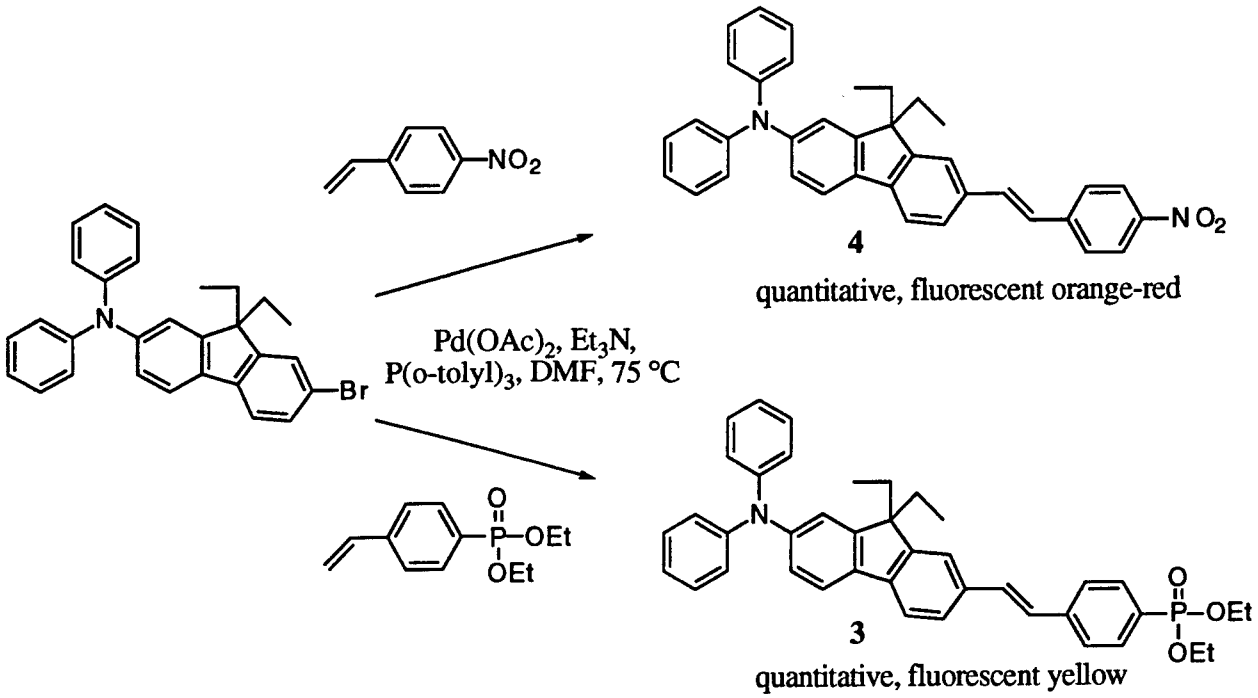
18% yd (purified), colorless cubic crystals
mp = 141.5-142.5 °C, MS (EI) 467, 469 (M^+),
C theor: 74.36 found: 74.37
H theor: 5.59 found: 5.90
N theor: 2.99 found: 3.30
Br theor: 17.06 found: 17.10



2

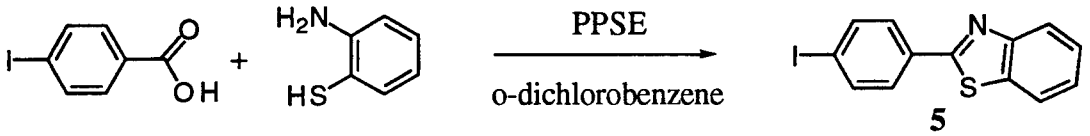
ca. 20% yd (purified),
orange platelets,
MS (EI) 556 (M^+)

Scheme 2



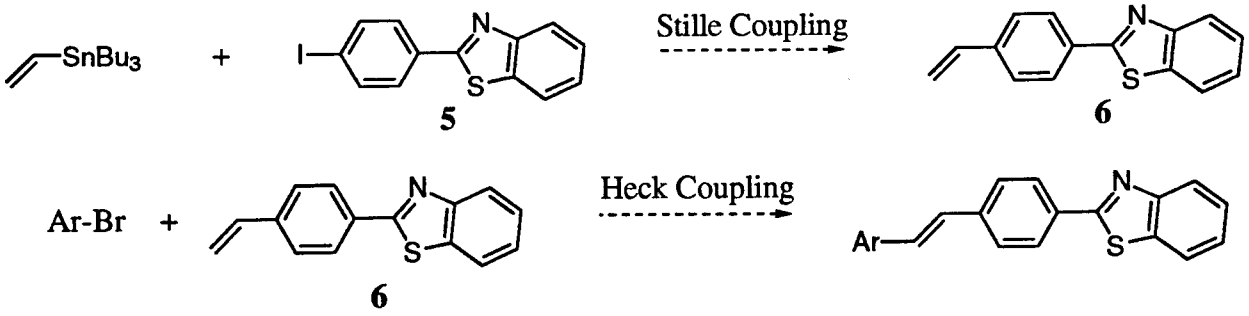
2-(4-Iodophenyl)benzothiazole (**5**), the penultimate precursor for 2-(4-vinylphenyl)benzothiazole (**6**), was prepared by PPSE-catalyzed condensation of 4-iodobenzoic acid and 2-aminothiophenol[45] (Scheme 3). White needles of **5** were obtained in ca. 60% yield. This iodo derivative was made to facilitate the preparation of 2-(4-vinylphenyl)benzothiazole (**6**) via Stille coupling with tri-*n*-butylvinyltin, the subject of future research efforts (Scheme 4). The benzothiazolestyrene derivative (**6**) will then be employed in Heck coupling reactions (Scheme 4) to create, e.g., a novel analog of **3**.

Scheme 3



61 & 57% yds. (recryst.)

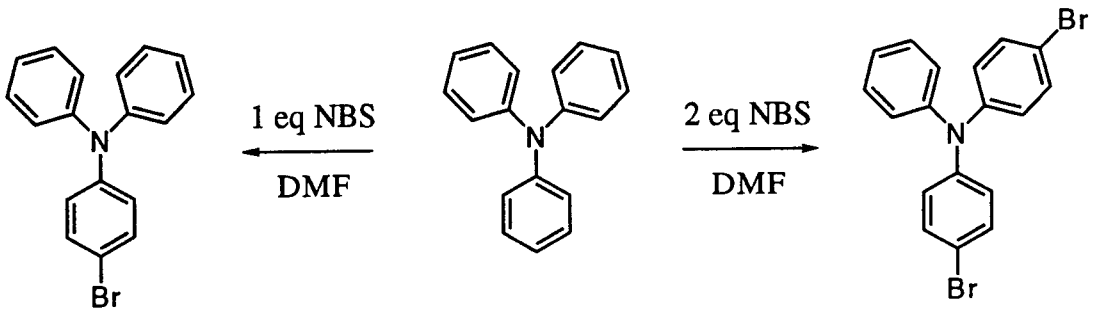
Scheme 4



Regiospecific bromination methodology, using NBS,[46] was investigated to prepare specifically brominated aromatic amines, including N-4-bromophenyl N,N-diphenylamine (**7**), bis(4-bromophenyl)amine (**8**), poly(3-bromo-9-vinylcarbazole) (**9**), and poly(3,6-dibromo-9-vinylcarbazole) (**10**).

When triphenylamine was treated with one equivalent of NBS in DMF at room temperature, N-4-bromophenyl N,N-diphenylamine (**7**) was obtained in rather high purity after three recrystallizations from EtOH (Scheme 5). The yield, however, was only about 42% after the recrystallizations. Considering the purification and yield, an Ullman reaction between diphenylamine and 4-iodobromobenzene is likely as good a procedure for the preparation of **7**. Attempted dibromination of triphenylamine, however, lead to a mixture of mono-, di-, and tribrominated products.

Scheme 5



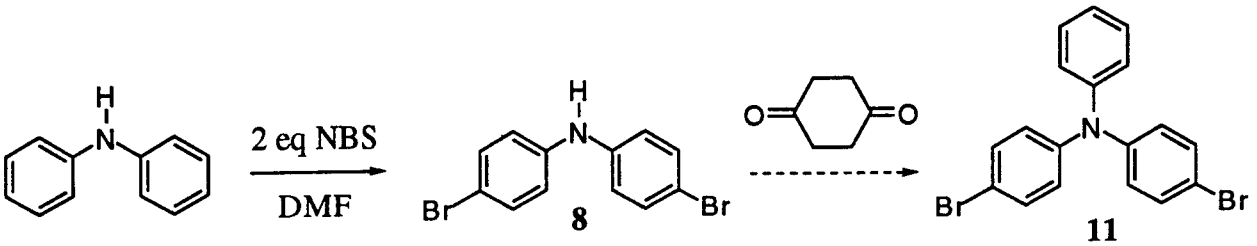
7

after 3 recrystallizations
mp = 110.5-111.5 °C (lit. 111-112, 115 °C)
C theor: 66.68 found: 67.24
H theor: 4.35 found: 4.43
N theor: 4.32 found: 3.82
Br theor: 24.65 found: 25.17

mixture of mono-, di-, and tri-bromination

Regiospecific *para*-dibromination of diphenylamine was achieved with 2 equivalents of NBS in DMF at room temperature, affording bis(4-bromophenyl)amine (**8**) in 90% yield (Scheme 6). **8** will be reacted with 1,4-cyclohexanedione to prepare dibromotriphenylamine (**11**), a valuable substrate for Heck and Stille coupling reactions.[47]

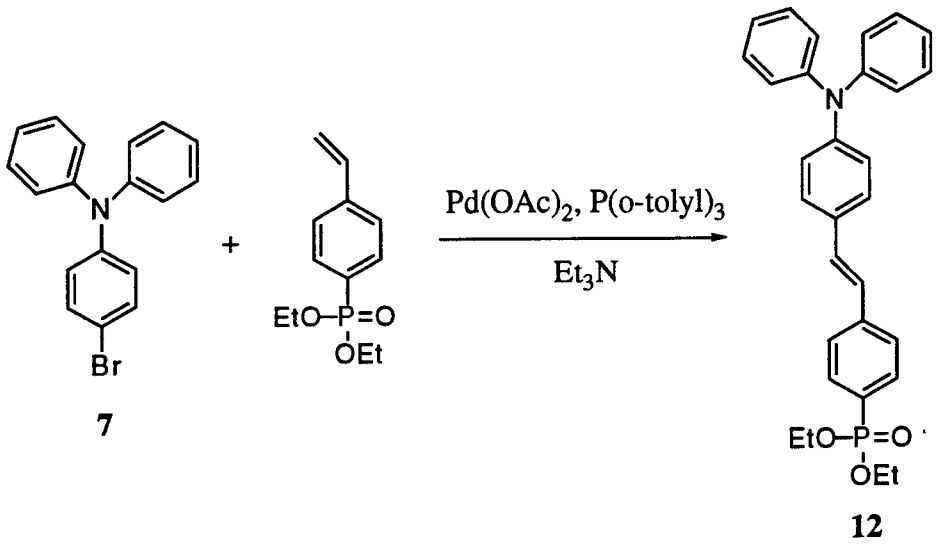
Scheme 6



mp = 103-104 °C (lit. 103-105 °C), 90% yield

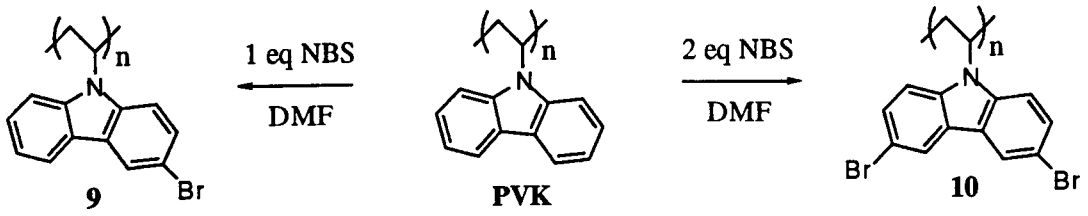
Heck coupling of N-4-bromophenyl N,N-diphenylamine (**7**) and diethyl 4-vinylbenzene phosphonate afforded potential NLO chromophore, charge (hole) transporter, and photorefractive component 4'-N,N-diphenylamino-4-diethylphosphonostilbene **12** (Scheme 7), as a fluorescent yellow material. Final purification is needed for final characterization, but tlc results strongly support formation of the desired product.

Scheme 7



With the commercial availability of poly(9-vinylcarbazole) (PVK) and its useful electronic properties, it is desirable to develop well-defined, efficient derivatization methodologies. PVK was treated with one equivalent of of NBS at room temperature (Scheme 8), yielding poly(3-bromo-9-vinylcarbazole) (**9**) with 64% bromination, determined by CHNBr elemental analysis, (slightly more than the 50% or one bromine per repeat unit that was expected). Poly(3,6-dibromo-9-vinylcarbazole) (**10**) was obtained in an analogous manner with two equivalents of NBS (96% bromination by CHNBr elemental analysis), as illustrated in Scheme 8. Thus, regiospecific bromination was achieved with the degree of bromination dictated by the stoichiometry of the brominating reagent and polymer.

Scheme 8



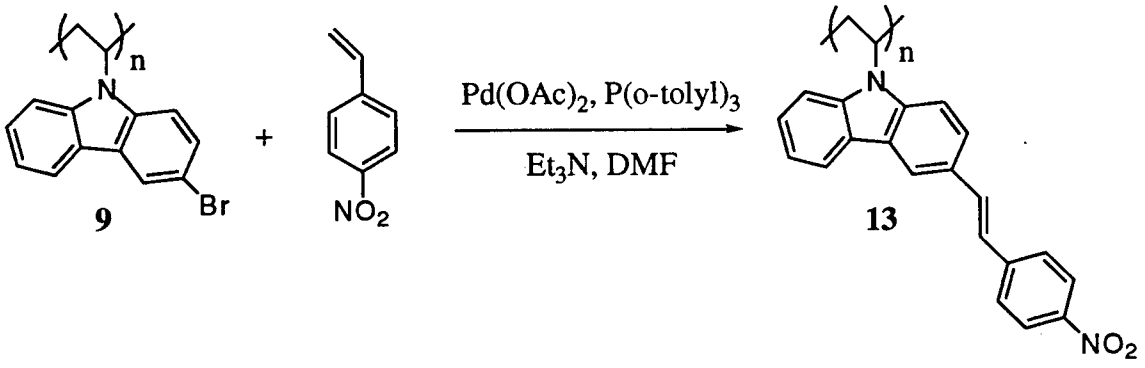
64% bromination
C theor: 61.79 found: 61.36
H theor: 3.70 found: 3.50
N theor: 5.15 found: 4.80
Br theor: 29.36 found: 29.33
Tg = 235 °C

96% bromination
C theor: 47.90 found: 48.75
H theor: 2.58 found: 2.58
N theor: 3.99 found: 3.44
Br theor: 45.53 found: 43.82

compare to bromination w/ $BTMACIBr_2$, only 82% bromination, both regiospecific (NMR)

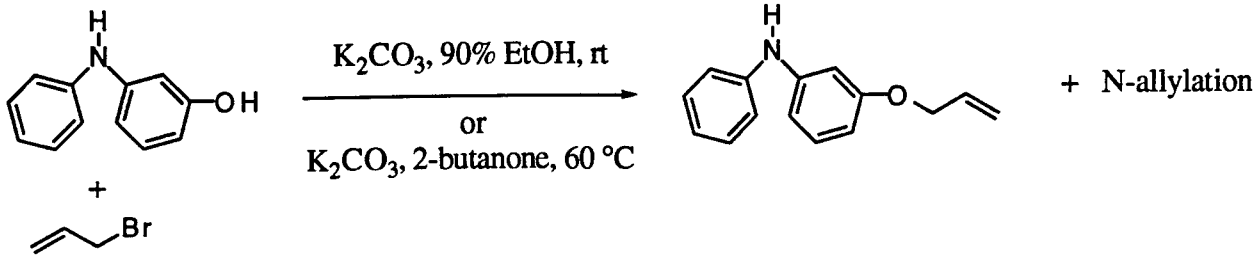
Monobromo-PVK **9** was subjected to Heck coupling conditions with 4-nitrostyrene, affording the dark orange-brown polymer poly(3-(4'-nitrostilbene)carbazole) (**13**), as shown in Scheme 9. Characterization of this polymer will be conducted, followed by electro-optic, photoconductive, and photorefractive characterization.

Scheme 9

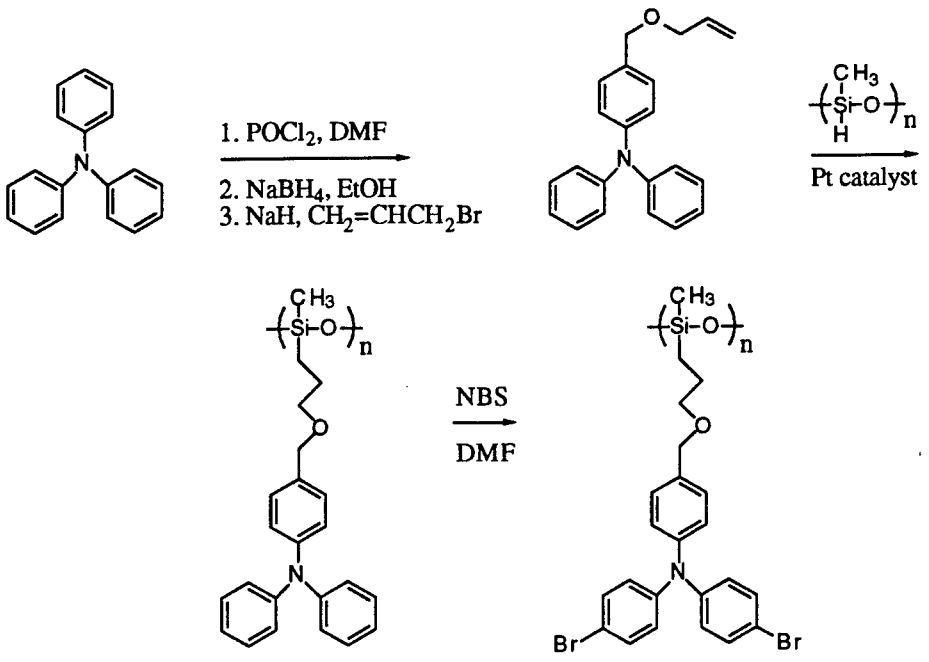


In order to prepare an arylamine substrate for hydrosilylation with poly(hydrogen methylsiloxane), the regiospecific O-allylation of N-(3-hydroxyphenyl)-N-phenylamine was attempted under two sets of reaction conditions (Scheme 10). In each case, a mixture of predominately O-allylation accompanied by a lesser amount of N-allylation. Unfortunately, the two isomers couldn't be separated, even after repeated column chromatography. Meanwhile, an alternate synthetic strategy (Scheme 11) was recently accomplished in K. Belfield's laboratory[48] which can be used to make polymer-bound two-photon dyes and photorefractive polymers.

Scheme 10



Scheme 11



Preliminary visible dye photoinitiated polymerizations were conducted to assess the possibility of developing a polymerization initiator system that can utilize infrared two-photon pumped up-conversion fluorescence as a means to generate spatially resolved visible photons. Photopolymerization of an acrylate/epoxy functionalized monomer was accomplished with a commercial dye/coinitiator system and one based on one of the aforementioned new fluorene-derived compound **4**. The commercial dye system, supplied by Spectra Group Ltd., Inc.[49] included the photooxidizing dye visible dye H-Nu 470 (5,7-diiodo-3-butoxy-6-fluorone), iodonium coinitiator CD1012 (4-(1-tetradecyl)phenyl)phenyliodonium hexafluoroantimonate), and DIDMA (N,N-dimethyl-2,6-diisopropylaniline). These were combined with the polymer in a 0.05 wt% H-Nu 470,).12 wt% iodonium salt, and 4 wt% DIDMA. Two drops of the monomer/dye/coinitiator solution was placed between two microscope slides and irradiated with an overhead projector at the high intensity setting. Photobleaching was observed after about 20 s. Polymerization appeared to occur as the slides could not be separated. About a 1 mm thick layer was placed in a N_2-purged screw cap vial and irradiated for 10 min. The once mobile mass was immobilized, indicative of polymerization. Irradiation of a thicker sample, ca. 3 mm, resulted in polymerization but incomplete curing appeared to occur due to a persistence of the orange color.

Similar conditions were used except fluorene **4** was used in place of the dye (H-Nu 470) in the same ratios as above. Irradiation did not produce photobleaching, though polymerization occurred as evidenced by the inability to separate the slides. Polymerization seemed to take somewhat longer with **4**. Control experiments were conducted: the monomer itself was irradiated for over 30 min and no polymerization occurred. Irradiation of a mixture of the monomer, iodonium salt, and DIDMA resulted in only partial cure after 10 min irradiation, thus demonstrating that fluorene **4** sensitized the photopolymerization.

In addition to the activities described above, AM1 Time-dependent Hartree-Fock calculations were carried out in collaboration with Dr. Doug Dudis (MLBP) to calculate NLO beta values for the two donor-acceptor substituted phosphorylated stilbene derivatives described in ref. 44. Dr. Richard Vaia (MLBP) performed preliminary electro-optic characterization (r_{33}) of one of these compounds in PMMA.

Acknowledgments

Special thanks are due Bruce Reinhardt, the lab focal point, and Robert Evers, MLBP Branch Chief, for generous assistance, and facilitating a productive, and truly collaborative summer research experience.

References

1. He, G. S.; Xu, G. C.; Prasad, P. N.; Reinhardt, B. A.; Bhatt, J. C.; Dillard, A. G. *Opt. Lett.* **1994**, *20*, 435.
2. He, G. S.; Bhawalkar, J. D.; Zhao, C. F.; Prasad, P. N. *Appl. Phys. Lett.* **1995**, *67*, 2433.
3. He, G. S.; Zhao, C. F.; Bhawalkar, J. D.; Prasad. P. N. *Appl. Phys. Lett.* **1995**, *67*, 3703.
4. Bhawalkar, J. D.; He, G. S.; Park, C.; Zhao, C. F.; Ruland, G.; Prasad, P. N. *Opt. Comm.* **1995**, *124*, 33.
5. He, G. S.; Gvishi. R.; Prasad, P. N.; Reinhardt, B. A. *Opt. Comm.* **1994**, *117*, 133.
6. Bhawalkar, J. D.; He, G. S.; Prasad, P. N. *Rep. Prog. Phys.* **1996**, *59*, 1041.
7. Tutt, L, W.; Boggess, T. F. *Prog. Quant. Electr.* **1993**, *17*, 299.
8. Goppert-Mayer, M.; *Ann. Phys.* **1931**, *9*, 237.
9. Dick, B.; Hochstrasser, R. M. *Resonant Molecular Optics Nonlinear Optical Properties of Organic Molecules and Crystals*; Vol 2.; Academic; New York, 1987.
10. Zhao, M. T.; Gui, Y.; Samoc, M.; Prasad, P. N. *J. Chem. Phys.* **1995**, 3991.
11. Nashold K. M.; Brown, R. A.; Hony, R. C.; Walter, D. P. *Liquid Cell Power Limiters*; NRL: Washington D.C. 1991.
12. Blau, W.; Byrne, H.; Dennis, W. M.; Kelly, J. M. *Opt. Comm.* **1985**, *56*, 25.
13. Perry, J. W.; Khundar, L. R.; Coulter, D.; Alvarez, Jr.; Marder, S. R.; Wei, T. H.; Sence, M. J.; Van Stryland, E. W.; Hagan, D. J.; Messier, J. *Nato ASI Series E.* **1991**, *194*, 369.
14. Wang. N. Q.; Cai, Y. M.; Heflin, J. R.; Garito, A. F. *Mol. Cryst. Liq. Cryst.***1990**, *189*, 39.
15. Tutt, L. W.; McCahon, S. W. *Opt. Lett.* **1990**, *15*, 700.
16. Tutt, L. W.; McCahon, S. W.; Klein, J. M. Opt. *Proc. SPIE*, **1990**, *1307*, 327.
17. King. R. B. *J. Inorg. Chem.* **1966**, *5*, 2227.
18. Allan, G. R.; Labergerie, D.; Rychnovsky, S. J.; Boggess, T. F.; Smirl, A. L.; Tutt, L. W.; McCahon, S. W.; Klein, M. B. *Tech. Dig. Ser.* **1991**, *10*, 172.
19. Allan, G. R.; Labergerie, D.; Rychnovsky, S. J.; Boggess, T. F.; Smirl, A. L.; Tutt, L. W.; McCahon, S. W.; Klein, M. B. *J. Phys. Chem.* **1992**, *96*, 6313.
20. Kroto, H. W.; Heath, J. R.; O'Brien, S. C.; Curl, R. F.; Smalley, R. D. *Nature* **1985**, *318*, 162.
21. Arbogast, J. W.; Darmayan, A. P.; Foote, C. S.; Rubian, Y.; Diederich, F. N.; Alverez, M. M.; Anz, S. J.; Whetten, R. L. *J. Phys. Chem.* **1991**, *95*, 6075.
22. Kost, A.; Tutt, L. W.; Klein, M. B.; Dougherty, T. K.; Elias, W. E. *Opt. Lett.* **1993**, *18*, 334.
23. Sun, Y.; Riggs, J. E.; Liu, B. *Chem. Mater.* **1997**, *9*, 1272.

24. Brandelink, D.; Mclean, D.; Schmitt. M.; Epling, B.; Colclasure, C.; Tondiglia, V.;Pachter, R.; Obermier, K., Crane, R. *Proc. Mat. Res. Soc.* **1991**, *247*, 361.
25. Geusic, J. E.; Singh, S.; Tipping, D. W.; Rich, T. C. *Phys. Rev. Lett.* **1969**, *19*, 1126.
26. Ralston, J. M.; Chang, K. R. *Appl. Phys. Lett.* **1969**, *15*, 164.
27. Arsenev, V. V.; Dneprovskii, V. S.; Klyshke, D. N.; Penin, A. N. *Sov. Phys. JETP* **1969**, *29*, 413.
28. Rychovsky, S. J.; Allan. G. R.; Venzke, C. H.; Smirl, Boggess. T. F. *Proc. SPIE* **1992**, *191*, 2274.
29. Lee, M. Proc. of the Int. Conf., Optics of Liquid Crystals, Italy, 1986.
30. Demartino, R. N.; Khanarian, G.; Leslie, T. M.; Sansone, M. J.; Stamatoff, J. B.; Yoon, H. N.; Mitchell, R. L. *Proc. SPIE* **1989**, *2*, 1105.
31. Ozaki, M.; Tagawa, A.; Hatai, T.; Sadohara, Y.; Ohmori, Y.; Yoshino, K. *Mol. Cryst. Liq. Cryst.* **1991**, *199*, 213.
32. Khoo, I. C.; Zhao, P.; Michael, R. R.; Lindquist, R. G.; Mansfield, R. *IEEE J. Quantum Elect.* **1989**, *25*, 1755.
33. Khoo, I. C.; Michael, R. R.; Finn, G. M. *Appl. Phys. Lett.* **1988**, *52*, 2108.
34. Yuan, H, J.; Li, L.; Palffy-Muhoray, P. *Proc. SPIE* **1990**, *1307*, 363.
35. Khoo, I. C.; Lindquist, R. G.; Michael, R. R.; Mansfield, R.; Zhao, P.; Lopresti, P. *Proc. SPIE* **1990**, *1307*, 336.
36. Pope, E. J. A.; Asmi, M.; Mackenzie, J. D. *J. Mater. Res.* **1989**, *4*, 1018.
37. Mukherjee, A. *Appl. Phys. Lett.* **1993**, *62*, 3423.
38. Ruland, G.; Givishi, R.; Prasad, P. N. *J. Am. Chem. Soc.* **1996**, *118*, 2985.
39. Bhawalkar, J. D.; Shih, A.; Pan, S. J.; Liou, W. S.; Swiatkiewicz, J.; Reinhardt, B. A.; Prasad, P. A.; Cheng, P. C. *Bioimaging* **1996**, *4*, 168.
40. Bhawalkar, J. D.;He, G. S.; Zhao, C. F.; Prasad, P. N. *IEEE J. Quant. Electron.* **1996**, *32*, 749.
41. Denk, W.; Strickler, J. H.; Webb, W. W. *Science* **1990**, *248*, 73.
42. See, e.g., Moerner, W. E.; Silence, S. M. *Chem. Rev.* **1994**, *94*, 127.
43. Belfield, K. D.; Chinna, C.; Najjar, O.; Sriram, S. *Polymer Preprints* **1997**, *38(1)*, 203. Belfield, K. D.; Chinna, C.; Najjar, O.*Macromolecules* (submitted).
44. Belfield, K. D.; Chinna, C.; Schafer, K. J. *Tetrahedron Lett.* **1997** (in press).
45. Zhao, M.; Samoc, M.; Prasad, P. N.; Reinhardt, B. A.; Unroe, M. R.; Prazak, M.; Evers, R. C.; Kane, J. J.; Jariwala, C.; Sinsky, M. *Chem. Mater.* **1990**, *2*, 670.
46. Xu, F.; Wang, Q. *Huaxue Shiji* **1986**, *8*, 374.
47. Haga, K.; Iwaya, K.; Kaneko, R. *Bull. Chem. Jpn.* **1986**, *59*, 803.
48. Belfield, K. D.; Najjar, O. unpublished results.
49. Spectra Group Limited, Inc., 1722 Indian Wood Circle, Suite H, Maumee, OH 43537.

A STUDY OF INTRA-CLASS VARIABILITY IN ATR SYSTEMS

Raj Bhatnagar
Associate Professor
ECECS Department

University of Cincinnati
Cincinnati, OH 45221

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

And

Wright Laboratory

August 1997

A Study of Intra-Class Variability in ATR Systems

Raj Bhatnagar, Associate Professor
Dax Pitts, Graduate Student
ECECS Department, University of Cincinnati
Cincinnati, OH 45221

Abstract

In this report we describe the results of our investigation into the intra-class variability of a vehicle class from the perspective of an automatic target recognition system. We examine the relevance of synthesized vehicle models for ATR systems and conclude that these models fall within the bounds of the vehicle class set by the intra-class variability of the vehicle. We also demonstrate the relevance of the mean-square-error between an image chip and a template when used as a measure of distance between the physical vehicles. We also show that it is feasible to intelligently merge chips from different vehicles of a class and construct classifiers that perform better than those designed with any individual member of the vehicle class.

A Study of Intra-Class Variability in ATR Systems

Raj Bhatnagar, Associate Professor
Dax Pitts, Graduate Student
ECECS Department, University of Cincinnati
Cincinnati, OH 45221

1 Introduction

An automatic target recognition system is usually trained completely with measured data obtained from target vehicles. However, such an effort, if done correctly, is extremely expensive from the perspective of computational resources. Collection of the templates requires an exhaustive data collection across all of the operating conditions of the ATR scenario. Storage of the templates would require extremely large amounts of storage space, and an algorithm based on comparisons with the numerous templates would require prohibitively large processing time. Such an effort would need to account for various possible articulations of target vehicles and the large intra-class variability of every vehicle class.

One solution for handling the above problems lies in using synthesized models of target vehicles and a model based ATR system. DARPA's MSTAR program, agented by the Wright Laboratory, WPAFB, Dayton, Ohio, follows this philosophy to a large degree. The performance of such a system depends on how well the synthesized vehicle-signatures match the actual measured signatures.

The objective of the research described in this paper is to examine the extent and consequences of variability within the class of T72 tanks from the perspective of MSTAR system developed by the Wright Laboratory in Dayton Ohio. In the context of this intra-class variability we then examine the closeness with which synthetic data identifies various instances of the T72 class.

The scope of our study includes the following tests with the measured and the synthetic data.

1. The performance of a classifier trained on a single member of the T72 class in discriminating the members of *T72-class* from the members of *confuser-vehicles* class. The problem is formulated as a two-class discrimination problem. An image chip from a target vehicle is sought to be classified as belonging to either the *T72* class or the *confuser-vehicles*

class.

2. The performance of a classifier trained on the synthetic signatures generated by DEMACO's XPATCH radar system prediction code and its comparison to the classifiers trained on the measured data.

3. Relevance of Mean-Square-Error distance measure between templates and test chips to physical features and articulations of the target vehicles.

2 The Data Set

Our study has been performed with data from the DARPA's MSTAR program. The measured data corresponds to 1ft. X-band SAR imagery. We have used the measured data for the following vehicles from the MSTAR collections:

Collection-1 T72s: T72-132, T72-812, and T72-s7. For each tank, the data from 31 and 32 degree elevations has been used.

Collection-2 T72s: T72-A04, T72-A05, T72-A07, T72-A10, T72-A32, T72-A62, T72-A63, and T72-A64. For each tank, the data from 30 and 31 degree elevations has been used.

Confuser Vehicles: M109, M110, BMP2, M2, M1. For these vehicles, data from 31 and 32 degree elevations has been used.

Confuser Vehicles M113, M35, and BTR70. For these vehicles, data from 31 degree elevation has been used.

In addition to this measured data, we used the synthetic data generated by DEMACO's XPATCH radar signature prediction code. The synthetic data models the same vehicle and articulation as has been measured for the T72-812 configuration mentioned above.

3 The Test Runs

Our target recognition tests consisted of a number of runs, each one of which contained the following main steps:

1. **Template Construction:** We constructed 72 templates of 5-degree width for each of the eleven measured T72s and also for the synthetic data from the XPATCH system.

We take all the available image chips from within the angular boundaries of a template and generate a mean template, and a mask. The mask is formed by thresholding the mean template and thus corresponds to the target's bright region. The mask specifies the pixels over which the distance computations are made during comparison with an unknown image chip. Each chip, used for constructing the templates, was typically of size 128 X 128 pixels. The number of image chips available for an elevation angle for any one vehicle ranged between 270 and 320. The image chips from two different elevation angles, as mentioned above, were included in the templates. A set of templates thus formed constituted one classifier for our study.

2. **ATR Tests:** We used a standard MSE classifier for comparing image chips from a target to the templates from a classifier. The classifier compensates for the unknown scale factor on the image chip and also for the unknown location of the target on the chip. Each test consisted of the following steps:

   (a) Select one vehicle as the *Target* vehicle and one T72 as the *classifier*. Obtain all image chips from the selected *Target* and all templates from the selected *classifier*.

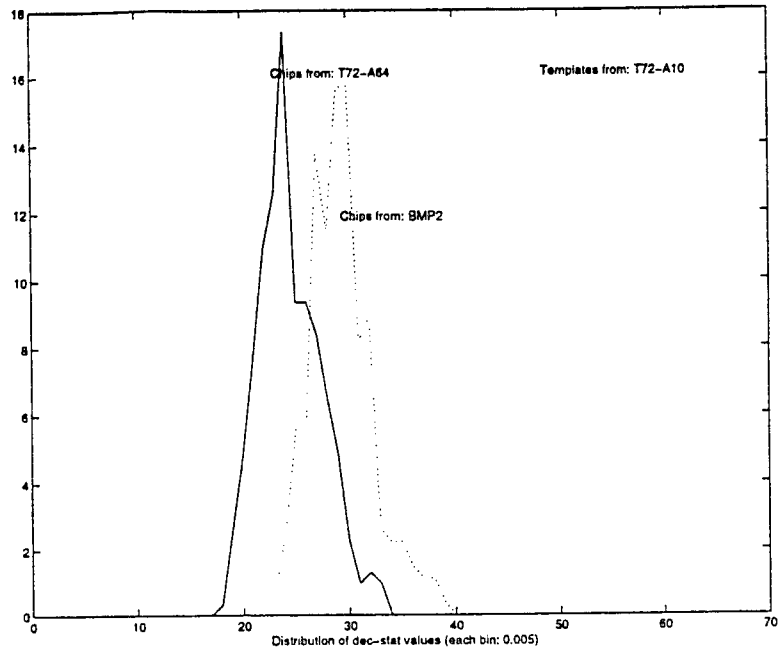
   (b) Compare each image chip from the *target* to each template from the *classifier* and determine the template closest to each chip. The *mean-square-error* between each chip and its closest template is also recorded.

   The output of a test is a list containing the closest matching template for each chip, the *mean-square-error* distance between the chip and the template, and the scale factor used with the chip. We have used eleven measured and one synthetic T72s as classifiers and there are nineteen different possible target vehicles; 11 T72s and 8 confusers. The number of test runs, therefore, was 228; one for each target-classifier combination.

4 Intra-Class Variability

A number of observations are made by examining the distributions of the MSE distances for each test. The following diagram shows the distributions of MSE values from two of the tests. The two curves in this plot correspond to the histograms for the MSE values from the two tests. The x-axis shows 70 bins for the histogram. Each bin corresponds to 0.005 units of MSE distance as determined by the MSE classifier. For both these tests the classifier templates have been used from the measured data for T72-A10. The solid curve on the left corresponds to the distribution for image chips from T72-A64 and the dotted curve to the right corresponds to the distribution for image chips from the confuser vehicle BMP2.

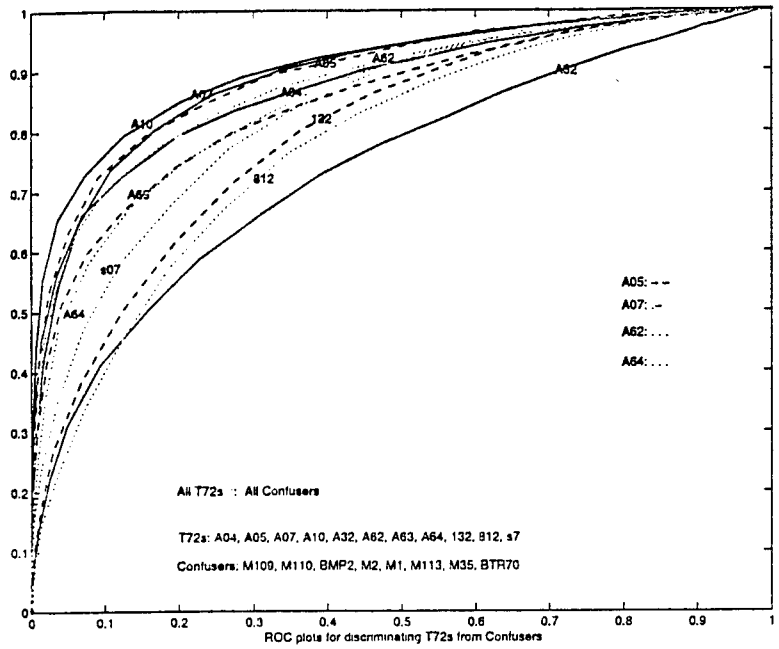
It can be seen from this plot that most of the MSE values for the T72-A64 are smaller than those for the confuser vehicle BMP2. A ROC plot can be constructed for any such pair of distributions.



4.1 T72 Class vs. Confusers Class

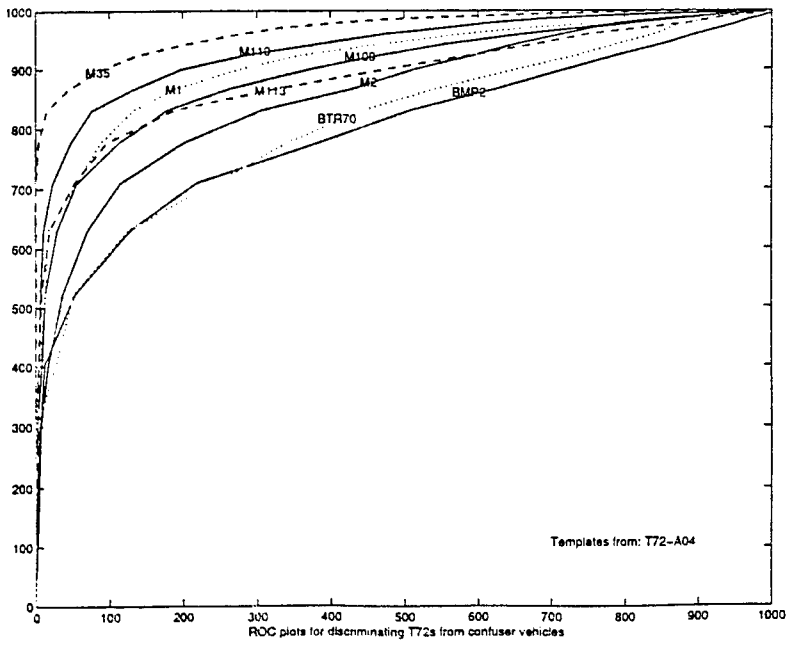
From our tests we accumulated data in such a way that for each classifier we obtained one distribution curve for the set of T72s and one distribution curve for all the confuser vehicles. When computing the distributions, the T72 being used as the classifier was excluded from the distribution curve for the set of T72 vehicles. This process was repeated for each of the eleven measured T72s acting as classifiers. For each classifier we constructed a ROC curve and the following plot shows all the ROC curves.

It is evident from this set of ROC curves that the performance of a classifier based on an individual member of the T72 class varies very widely with the selected member. According to this set of ROC curves it can be seen that T72-A10 acts as the best classifier and T72-A32 acts as the worst classifier. The difference between these two classifiers is very large. The statistical confidence level in each of the curves is significantly high because the number of chips used for computing each ROC curve is very high. For each curve we have used more than 5,500 image chips from the T72s and more than 4,000 image chips for the confuser vehicles.



4.2 T72 Class vs. Individual Confusers

Each of the above ROC curves can be examined in its decomposition into ROC curves where each individual confuser has been tested against the set containing all the eleven T72s. The set of ROC plots that decomposes the accumulated ROC plot for T72-A04 classifier is shown below.

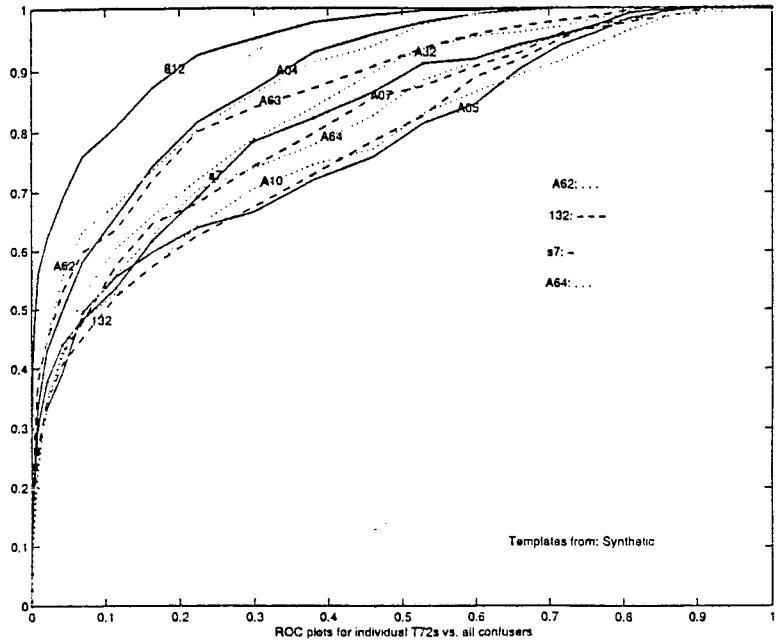


It is evident from the above set of ROC plots that the success of a classifier in discriminating

the T72s from a confuser vehicle varies very significantly with the confuser vehicles.

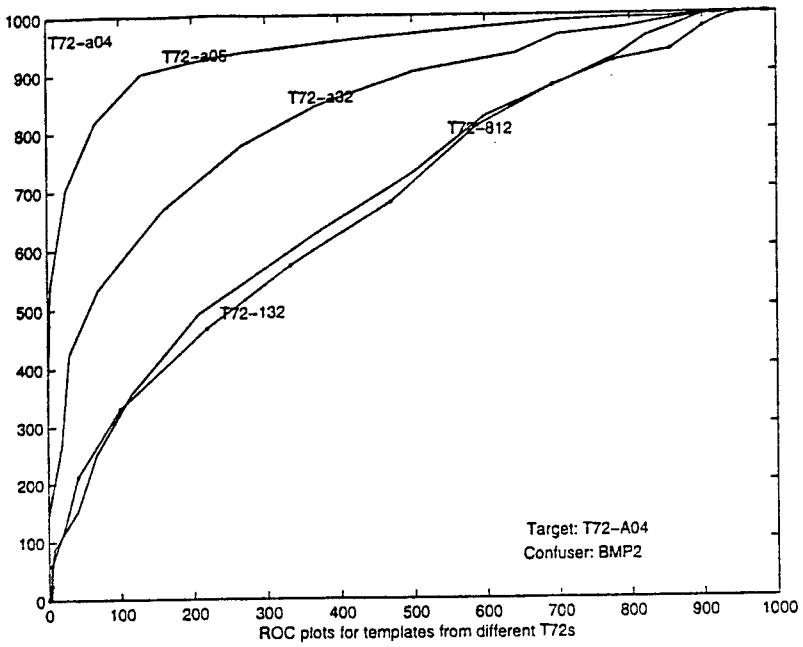
4.3 Individual T72s vs. Confusers Class

Another way to decompose a ROC curve from those described in Section 4.1 is to examine the performance of the classifier when individual T72s are to be discriminated against the complete set of confuser vehicles. In the following plot we have selected the classifier based on the synthetic T72-812 data and show the ROC plots for discriminating individual T72s from all confusers using this classifier. As expected, the synthetic classifier best discriminates the measured T72-812 from the confusers.



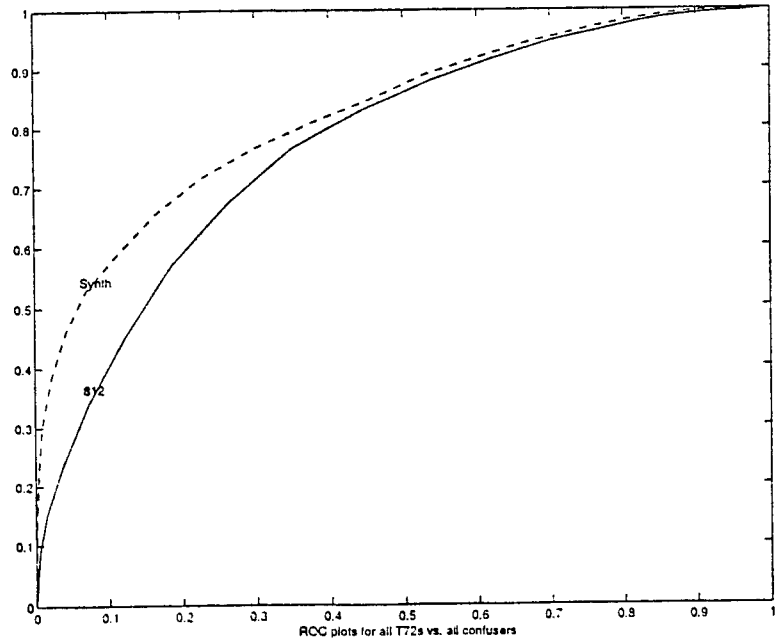
4.4 Individual T72s vs. Individual Confusers

For discriminating an individual T72 from an individual confuser vehicle the performance of various classifiers varies very significantly. This can be inferred from the above sets of ROC plots. We can explicitly pick a particular T72 and confuser and plot the ROC curves using different classifiers. One such set of ROC curves is shown below. Here T72-A04 has been selected as the target vehicle, BMP2 as the confuser vehicle, and ROC curves for five different classifiers have been shown.



5 Synthetic Data

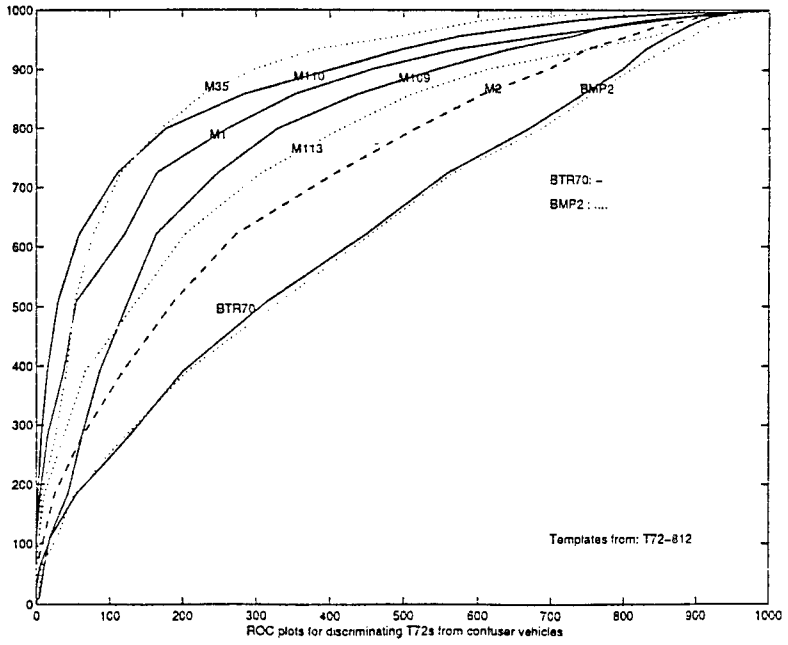
We now examine the performance of a classifier designed with the synthetic data. As mentioned above, the synthetic data available to us is for the same model and articulation of a T72 as the measured data for the T72-812. We show below the ROC plots for the classifiers designed with the measured and the modeled data for the same T72 vehicle.

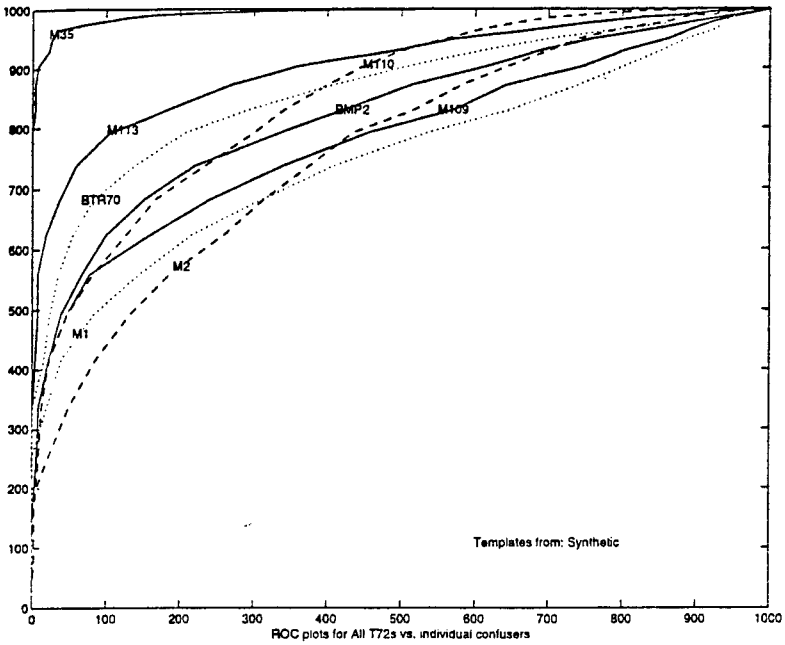


The ROC plots are for discriminating the class of all T72s from the set of all confuser vehicles. Two important conclusions can be derived from this set of plots. First, the classifier based upon the synthetic data has performed consistently better than the classifier based upon the measured data. Second, the ROC plot for the synthetic classifier falls well inside the range of ROC curves shown in Section 4.1 above. Both these observations are very significant from the perspective of usability of synthetic data for ATR systems.

A further comparison between the performance of classifiers based upon synthetic and measured data can be seen in the following two sets of ROC curves. In the first set we have used the classifier based upon the measured data for discriminating the set of all T72s from individual confuser vehicles. The second set of plots shows the same discrimination task being performed by the classifier based upon the synthetic data.

The ROC curves in the second plot show significantly better performance for classification against many confuser vehicles than that shown in the first plot. It is clear that the synthetic model is performing better as a classifier than its measured counterpart in several cases. A deeper analysis of the performance is needed to investigate the reasons for the order of various confuser vehicles not being the same in the two plots. Such an investigation will shed more light on the nature of the synthetic data.





6 A Notion of Separability: Quantified

We developed a notion of *distance* between two vehicles based upon the MSE values computed by the classifiers. For vehicles V_1 and V_2 we measure this distance as follows.

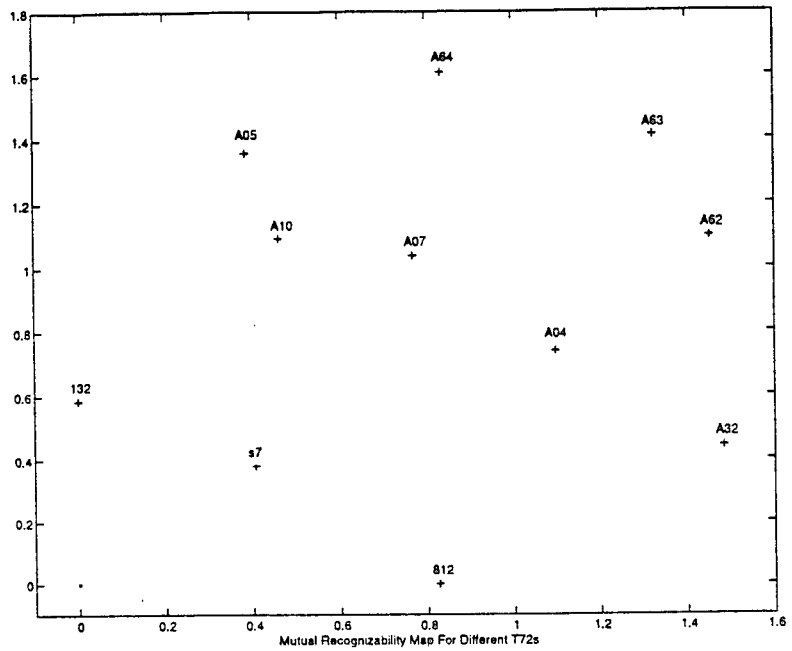
1. For each image chip from V_1, determine the minimum MSE distance from the templates for V_2.
2. For each image chip from V_2, determine the minimum MSE distance from the templates for V_1.
3. Take the average value of all the minimum distances from all the chips for V_1 and V_2. It was observed that typically, average distance from V_1 chips to V_2 templates remains very close in value to the average distance between V_2 chips and V_1 templates.

We use this average value as the distance between the vehicles V_1 and V_2 and call it the *MSE-Distance(1,2)*. We computed the MSE-Distance for every pair of T72s for which measured data is available.

We now seek to represent each T72 vehicle by a point on a 2-dimensional plane in such a way that the distance between points for V_1 and V_2 is as close as possible to the MSE-Distance between V_1 and V_2. If the actual distance on the 2-dimensional plane between points for V_1 and V_2 is $D(1,2)$ then the cumulative error for all the vehicles included on the plot is:

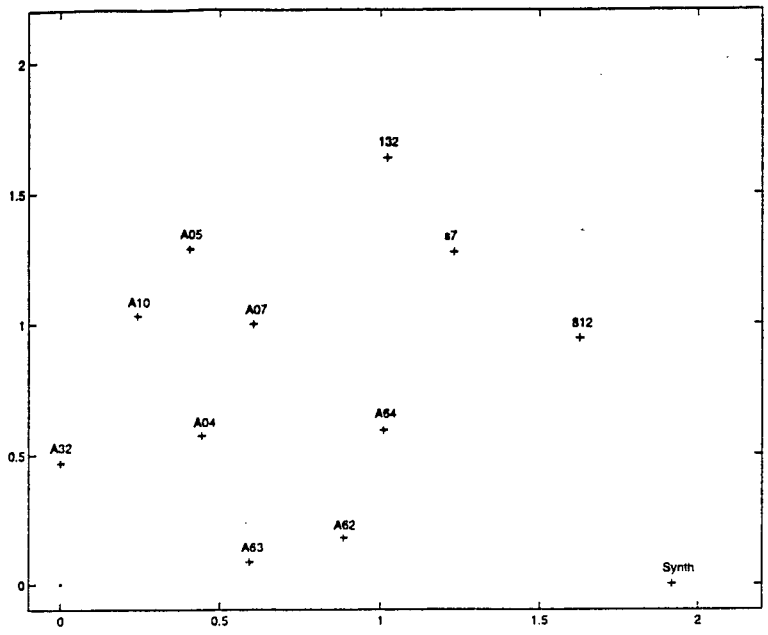
$$Error = \sum_{i=1}^{11} \sum j = i^{11} \ (abs(D(i,j) - MSEDistance(i,j))/MSEDistance(i,j).$$

We used a simple relaxation algorithm to place the points on the MSE distance plane in such a way that the mean square value of the above described error quantity is minimized. The result of placing the points for the eleven measured T72s is shown below.



It turns out that in this plot the T72s cluster according to their physical characteristics, articulations, and configurations. This, on one hand revalidates the use of MSE measure for classification purposes, and on the other hand provides a very good tool for visually examining the relative closeness and disparity among the various members of a vehicle class.

We now add to the set of T72s the synthesized model. The resulting plot is shown below.



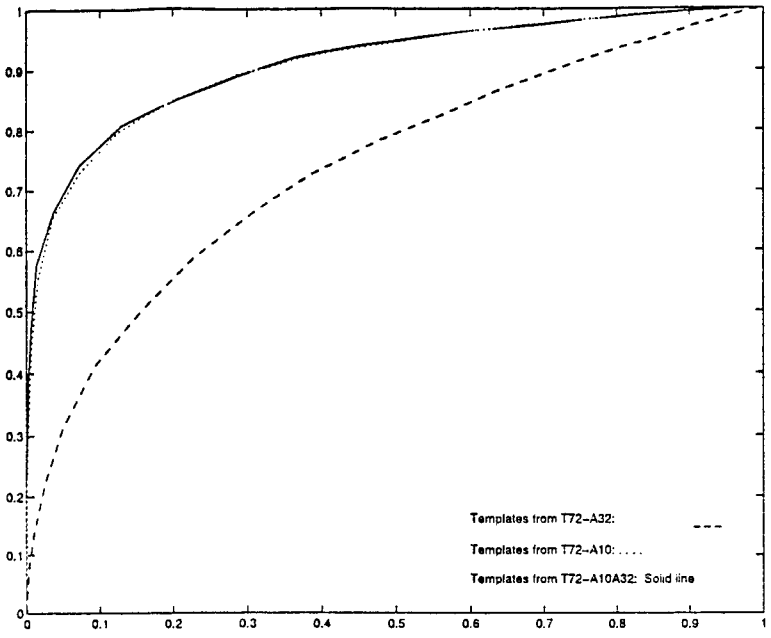
We see in this plot that the synthetic model lies far away from all the other T72s. Ideally, for a model to be a representative of the class of T72s, we would expect it to lie somewhere in the middle of the area populated by the T72s. Since a model emulates a unique T72, we can only expect it to be close to the point corresponding to the measured data for the same T72. In this case the point corresponding to the synthetic data is farther than expected from the point corresponding to the measured T72-812. Determining a synthetic model located at a point closer to the center of the population of T72s is a challenging problem and may involve creative use of chips from a number of T72 vehicles for creation of classifier templates.

The classifier trained on the synthetic model for a T72 performs better than its measured counterpart as evidenced by the ROC curves shown above but it itself appears to be somewhat distant from the measured T72-812 and also from the class of T72s. This requires further investigation and our current work is focused in that direction.

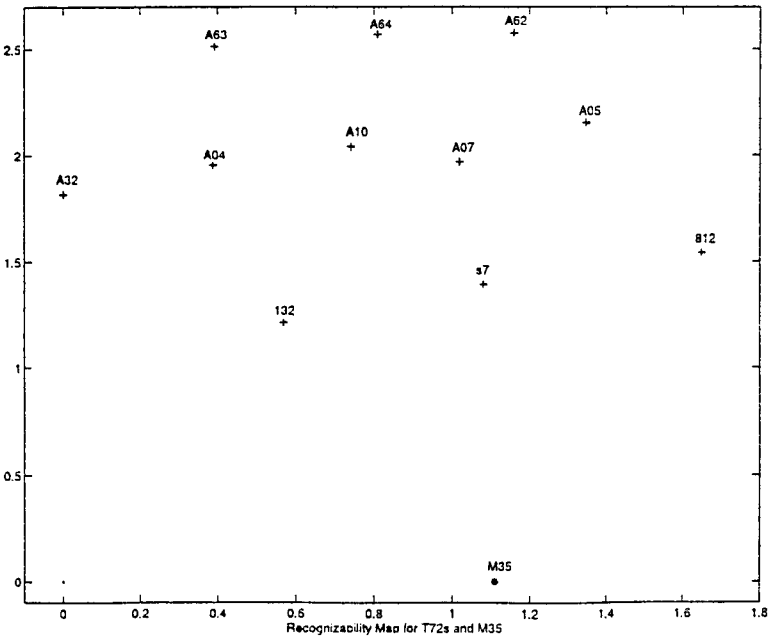
7 Construction Of Mixed Templates

We performed some preliminary tests to determine the feasibility of constructing templates for a classifier by mixing chips from a number of different T72s. An examination of ROC curves presented in Section 4.1 shows that T72-A10 is the best performing classifier and T72-A32 is the worst performing classifier. In our first test, we included chips from both these vehicles to construct a set of templates. We then performed the ATR tests with this new merged classifier.

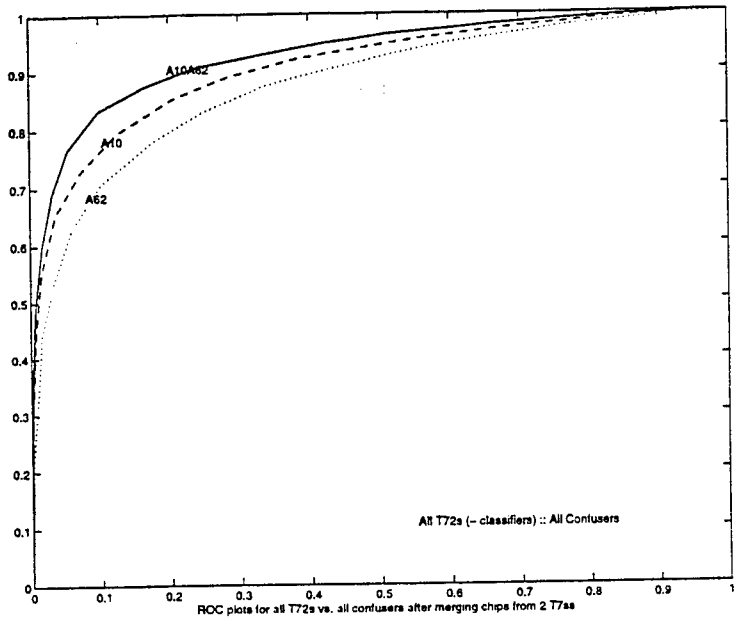
A comparison of its performance can be seen in the figure below containing the three ROC curves.



We can see from the above ROC curves that the mixed classifier performs almost as well as the best performing classifier. The chips from T72-A32 haven't made any apparent difference in performance. This is an interesting observation and a deeper investigation is needed to discover the reasons for this behavior. For our second test we looked at the following mutual distance map containing T72s and the M35 confuser vehicle.



We selected T72-A10 and T72-A62 for merging into a single classifier based on their positions in the above map. They belong to different clusters and are best performers among their respective clusters. (Performance, as discerned from the ROC curves in Section 4.1.) The performance of this mixed classifier, and its comparison with its constituting components, is shown below.



The composite classifier in this case has performed significantly better than its constituting components. This lends credence to the hypothesis that a better performing classifier can be constructed by using and intelligently merging data from different members of a vehicle class. How to select data to be merged is an important question and needs further testing and investigation. as discussed above, the T72s clustered in the MSE-distance map according to their physical features and articulations. The second test, therefore, can be seen to imply in a very general way that a classifier designed from data taken from vehicles with different physical features would perform better than individual members of the vehicle class. This is to be expected from an intuitive perspective.

8 Confusion Matrices

A new perspective on the performance of various classifiers can be obtained by constructing a confusion matrix. We show four different confusion matrices below.

For each image chip in our dataset we determined the template, across all classifiers, that comes closest to the chip. We marked the classifier corresponding to this template as the

selected classifier. However, each T72 almost always "recognizes" its own chips and this is evidenced in the table shown below.

Confusion Matrix for T72s
The **BEST** matching template for each chip.

	TEMPLATES FROM										
	A04	A05	A07	A10	A32	A62	A63	A64	132	812	s7
CHIP											
A04	589	0	0	4	1	0	2	1	1	0	0
A05	1	594	1	2	1	0	0	0	0	0	0
A07	0	3	591	4	0	1	0	0	0	0	0
A10	4	0	1	594	0	0	0	0	0	0	0
A32	0	0	0	0	599	0	0	0	0	0	0
A62	0	0	7	2	0	577	10	3	0	0	0
A63	1	2	0	1	0	1	592	1	0	1	0
A64	0	1	1	2	0	0	0	595	0	0	0
132	0	0	0	0	0	0	0	0	489	0	0
812	0	0	0	0	0	0	0	0	1	483	1
s7	0	0	0	0	0	0	0	0	2	0	474
M109	16	66	42	21	57	20	17	88	67	30	63
BMP2	18	35	59	18	54	23	40	79	66	36	61
M110	12	27	58	33	33	16	21	81	73	70	62
M2	48	34	43	21	34	24	25	75	65	27	87
M1	36	35	31	29	30	15	19	70	54	38	49
M113	24	32	56	30	40	32	32	66	118	32	78
M35	20	16	30	28	14	8	8	30	138	124	124
BTR70	42	28	42	24	52	14	18	72	74	56	86

In order to determine the T72 that is closest to a chip and is different from the T72 from which the chip emanates, we selected the second best template match for each image chip and considered its classifier as being the *2nd closest* to the chip. A confusion matrix based on this criterion is shown below.

Confusion Matrix for T72s
The **2nd BEST** matching template for each chip.

	TEMPLATES FROM										
	A04	A05	A07	A10	A32	A62	A63	A64	132	812	s7
CHIP											
A04	7	88	152	144	43	48	60	36	4	2	14
A05	51	4	133	256	13	22	30	65	11	2	12
A07	61	154	7	256	6	24	27	45	9	1	9
A10	69	199	206	1	9	35	23	36	8	3	10
A32	189	67	83	99	0	53	26	37	17	12	16
A62	46	30	52	48	3	16	270	112	1	10	11
A63	47	34	36	34	0	295	4	130	3	1	15
A64	23	97	76	91	4	132	110	3	20	1	42

132	24	42	71	35	6	5	7	9	0	85	205
812	46	35	20	22	15	33	13	17	165	2	117
s7	27	24	36	51	5	32	35	61	161	42	2
M109	17	58	44	31	52	30	27	59	79	37	53
BMP2	31	34	57	35	39	26	41	80	56	30	60
M110	24	63	47	29	31	22	24	60	74	43	69
M2	20	42	45	35	34	28	33	65	68	43	70
M1	25	38	50	32	45	17	16	35	57	34	57
M113	30	46	50	42	58	18	40	56	82	32	86
M35	50	20	34	20	26	12	22	50	106	78	122
BTR70	34	56	64	28	56	24	28	54	62	46	56

The above confusion matrix shows some patterns of affinity among subsets of T72s. Each T72 "looks" more like some other T72s and less like some others. This pattern is very similar to the one observed in the MSE-distance maps discussed above.

We further modified the structure of the confusion matrices by including the two mixed classifiers whose design has been discussed above, and the synthetic classifier designed with the XPATCH data. A repeat of the above two confusion matrices with this change resulted in the following two matrices.

Confusion Matrix for T72s+Mixed Classifiers
The best matching template for each chip.

	TEMPLATES FROM													
	A04	A05	A07	A10	A32	A62	A63	A64	132	812	s7	Syn	mx1	mx2
CHIP														
A04	585	0	0	2	0	0	2	1	1	0	0	0	4	3
A05	1	590	1	2	0	0	0	0	0	0	0	0	2	3
A07	0	3	587	3	0	0	0	0	0	0	0	0	1	5
A10	4	0	0	518	0	0	0	0	0	0	0	0	27	50
A32	0	0	0	0	565	0	0	0	0	0	0	0	34	0
A62	0	0	6	1	0	510	4	3	0	0	0	0	0	75
A63	0	2	0	1	0	1	590	1	0	1	0	0	1	2
A64	0	1	1	1	0	0	0	591	0	0	0	0	1	4
132	0	0	0	0	0	0	0	0	489	0	0	0	0	0
812	0	0	0	0	0	0	0	0	1	483	1	0	0	0
s7	0	0	0	0	0	0	0	0	2	0	474	0	0	0
M109	15	56	32	9	42	13	12	71	58	28	47	2	77	25
BMP2	17	27	54	12	43	18	30	69	58	30	55	0	42	34
M110	10	22	57	25	26	11	21	75	68	68	61	1	29	12
M2	45	32	37	15	26	22	21	69	58	26	77	0	33	22
M1	27	33	23	19	22	10	14	68	48	34	47	1	39	21
M113	18	26	48	20	24	28	30	60	112	32	76	2	52	12
M35	20	16	30	24	14	8	8	28	136	124	120	2	10	0
BTR70	38	24	38	14	38	12	18	72	66	54	72	2	40	20

1. *mx1* templates are formed by mixing chips from T72-A10 and T72-A32.
2. *mx2* templates are formed by mixing chips from T72-A10 and T72-A62.

Confusion Matrix for T72s+Mixed Classifiers
The **2nd BEST** matching template for each chip.

	TEMPLATES FROM													
	A04	A05	A07	A10	A32	A62	A63	A64	132	812	s7	Syn	mx1	mx2
CHIP														
A04	7	47	79	47	12	18	36	23	3	1	9	0	147	169
A05	19	7	96	116	1	6	16	38	9	0	4	0	108	179
A07	25	74	9	93	0	9	11	17	4	1	4	0	111	241
A10	2	12	24	63	1	0	1	1	1	0	0	0	200	294
A32	9	6	2	2	34	0	0	2	1	1	0	0	541	1
A62	5	0	4	0	0	72	52	7	0	0	0	0	2	457
A63	24	19	20	3	0	147	4	91	3	1	8	0	12	267
A64	10	62	39	19	1	30	56	6	12	1	24	2	33	304
132	17	34	59	16	3	3	2	7	0	78	186	2	56	26
812	35	29	14	10	13	17	5	8	143	2	103	27	36	43
s7	23	14	19	15	0	15	20	36	139	35	2	2	30	126
M109	10	37	35	17	38	23	18	57	57	29	48	3	69	46
BMP2	23	28	45	24	35	16	35	59	50	28	52	0	55	39
M110	17	52	39	27	28	21	15	52	69	37	54	0	37	38
M2	19	33	41	27	28	19	27	49	64	41	61	4	41	29
M1	24	24	38	25	34	15	10	24	45	26	41	6	69	25
M113	36	38	50	24	60	18	36	48	74	28	74	0	34	20
M35	42	16	30	18	20	12	20	52	94	78	120	2	20	16
BTR7 0	28	48	42	18	50	16	26	44	56	42	54	4	54	26

1. *mx1* templates are formed by mixing chips from T72-A10 and T72-A32.
2. *mx2* templates are formed by mixing chips from T72-A10 and T72-A62.

A number of interesting observations can be made from the above two matrices. The first is that a chip's affinity to its own classifier is very significant and is larger than that for the mixed classifiers. The second confusion matrix, however, reveals that the mixed classifiers are preferred overwhelmingly as the second choice by all the T72 vehicles in our set. This is a very interesting observation. When two good representatives of the class are merged to construct a classifier, even those T72s that are not part of the classifier prefer it over any other individual T72. This demonstrates that it is possible to capture features from multiple T72s into a single classifier and these classifiers then attract other T72s that may have affinity to any combination of features in the merged classifier. This is an interesting line of investigation and our results are only very primitive. A detailed investigation along this line is bound to result in deeper insight into the design of more representative classifiers for classes with large intra-class variability.

Another interesting observation that can be made from the last matrix is that the image chips from the measured T72-812 did not choose the synthetic classifier even as their second preference. This is a surprising observation. Synthetic classifier is designed for the same vehicle as represented by the T72-812. However, we have seen above that the synthetic classifier has performed better than its measured counterpart in all classification tasks. There is, therefore, need to investigate further the character of synthetic data, the reasons for its superior performance, and reasons for its greater-than-expected distance from its measured counterpart.

9 Conclusion

The main conclusions of the research described above can be summarized as follows.

1. There is a very significant amount of variability within the class of T-72 vehicles. A classifier designed with any one member of this class may perform very well for some members and very poorly for some other members.

2. The synthetic models of vehicles, generated by the XPATCH system, perform within the bounds of the performance set by the variability within the class of T72 vehicles.

3. The synthetic models of vehicles are not as close to their measured counterparts as expected. They perform better than the measured counterparts because the MSE-distance between synthetic models and confuser vehicles is even larger.

4. It is possible to mix data from two members of the T-72 class to design a classifier that performs better than a classifier designed with the members taken individually.

With the help of various tests we have demonstrated that there exists a large amount of variability within the class of T72s. This difference makes it harder for a classifier designed with any one member of the class to discriminate between the T72s and the confuser vehicles. Also, the classifiers designed based upon any one member of the class of T72s have very widely varying performance. The synthesized model of the T72 performs within the bounds of performance set by various members of the T72 class, but is still significantly different from all the measured T72s in some respects. This is an interesting observation and needs further investigations. Our tests towards designing *class-representative* classifiers by merging image chips from different members of the T72 class have shown very encouraging preliminary results. Further investigation along this line will yield beneficial insights and better performing classifiers.

10 Acknowledgment

We are thankful to Mark Minardi, WL/AACA, for hosting us at the Wright laboratory during the summer of 1997 and providing all the assistance and resources for conducting this study. We are thankful to Ron Dilsavor of Sverdrup Technology and based at Wright Laboratory for providing significant amount of help and assistance during the course of this study.

We express our thanks to the Wright Laboratory, Wright Patterson Air force Base, and Sverdrup Technology, Inc., for making available to us all the data and various pieces of software for making this study possible. We are also thankful to the AFOSR for supporting two authors from the University of Cincinnati during the summer of 1997 for performing this study.

THEORETICAL FOUNDATIONS FOR DETECTION OF POST-PROCESSING CRACKS IN CERAMIC MATRIX COMPOSITES BASED ON SURFACE TEMPERATURE

Victor M. Birman
Professor
Engineering Education Center

University of Missouri-Rolla
8001 Natural Bridge Road
St. Louis, MO 63121

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

July 1997

THEORETICAL FOUNDATIONS FOR DETECTION OF POST-PROCESSING CRACKS IN CERAMIC MATRIX COMPOSITES BASED ON SURFACE TEMPERATURE

Victor M. Birman
Professor
Engineering Education Center
University of Missouri-Rolla

Abstract

Ceramic matrix composites are processed at high temperatures and experience significant residual thermal stresses upon cooling to the room temperature. These stresses often result in cracking of the matrix, even prior to the application of external loads. It is important to detect these post-processing matrix cracks using a nondestructive technique. The method proposed in this report is based on measurements of the surface temperature of a ceramic matrix material subjected to cyclic stresses. The elevated surface temperature is due to friction between the fibers and the matrix that occurs in the presence of bridging matrix cracks. The solution presents a relationship between the surface temperature and the matrix spacing that can identify an extend of the damage.

THEORETICAL FOUNDATIONS FOR DETECTION OF POST-PROCESSING MATRIX CRACKS IN CERAMIC MATRIX COMPOSITES BASED ON SURFACE TEMPERATURE

Victor M. Birman

Introduction

The fact that ceramic matrix composites (CMCs) are processed at a high temperature implies significant thermally-induced residual stresses. These stresses can result in a damage, even before the material is subject to an external load. For example, Bischoff et al.[1] and Nishiyama et al.[2] observed post-processing cracking in CMCs. The cracks usually form a regular pattern with the spacing that can be assumed constant. Long cracks perpendicular to the fibers, similar to those observed by Marshall and Evans[3] and other investigators, are called "bridging cracks", because they "bridge" the fibers without breaking them. It is important to be able to determine an extend of post-processing damage that can be associated with the density of matrix cracks. This is because, although matrix cracks do not significantly degrade the strength and stiffness of the material in the fiber direction, they are detrimental to those in the transverse direction. In addition, in CMCs, matrix cracks serve as conductors of oxygen to the fibers. At high temperatures, this results in oxidation of the fiber-matrix interface and a dramatic embrittlement of the material[4-7].

In the presence of matrix cracks, the fibers slide relative to the matrix in the regions adjacent to the planes of the cracks. This sliding that occurs when the material experiences dynamic bending or tension is accompanied with an increase of temperature due to friction between the fibers and the matrix. This phenomenon was observed by

Holmes and his associates[8,9] in their experiments on carbon-fiber SiC matrix composites. Cho et al.[10] developed the solution that related the interfacial shear stress to the rise of temperature of the specimen. It was suggested that the interfacial shear stress along the fiber-matrix interface can be monitored as a function of temperature. In the present report, the analytical foundation is developed for prediction of the spacing of post-processing matrix cracks (and the interfacial shear stress) as a function of the surface temperature of a vibrating unidirectional CMC. This technique can be applied to a nondestructive testing of CMC components.

Analysis

The purpose of the present solution is to determine a relationship between the post-processing matrix crack spacing in a unidirectional CMC and its surface temperature during a nondestructive dynamic test. The amplitudes of cyclic stresses are assumed below the matrix cracking limit of the material so that cycling does not change the matrix crack spacing. An elevated temperature is due to frictional heating that is triggered by a relative movement of bridged fibers with respect to the matrix. The solution assumes the mode of cracking employed in the theories of Aveston-Cooper-Kelly[11] and Budiansky-Hutchnison-Evans[12], i.e. long, regularly spaced cracks. The formulation of the problem is shown in Fig. 1 that identifies the crack spacing (s), the length of the sliding distance (x_0), and a distribution of stresses in the fibers and the matrix.

The solution involves the following steps. First, the modulus of elasticity of the material is determined as a function of the interfacial shear stresses, matrix crack spacing and the residual thermal stresses in the fiber using a modified approach of Pryce and

Smith[13]. Residual thermal stresses are evaluated accounting for the effect of temperature on the properties of the constituent materials. Then the experimental findings of Karandnikar and Chou[14] are used to justify a simple relationship between the modulus of elasticity and the matrix spacing. Combining two solutions referred to above, the modulus of elasticity can be eliminated and a single equation relating the matrix crack spacing to the interfacial shear stress obtained. Subsequently, the balance between the rate of heat flow and the rate of dissipation of the frictional energy is employed, as suggested by Cho et al.[10], to obtain a relationship between the surface temperature and the interfacial shear stress. This procedure enables us to evaluate both the shear stress and the matrix crack spacing as functions of the surface temperature.

A distribution of stresses in a specimen subjected to an external stress σ_c is shown in Fig. 1. Note that this distribution corresponds to a partial slip, i.e., $2x_0 < s$. This case is considered in the present analysis because the full slip that occurs when the matrix cracks approach saturation is unlikely to be encountered immediately after the processing. The stresses in the fibers and the in the matrix can be evaluated based on their values at the points A and B:

$$\begin{aligned} \sigma_{fA} &= \frac{\sigma_c}{V_f} \\ \sigma_{fB} &= \frac{\sigma_c E_f}{E_c} + \sigma_f^T \\ \sigma_{mA} &= 0 \\ \sigma_{mB} &= \frac{\sigma_c E_m}{E_c} + \sigma_m^T \end{aligned} \qquad (1)$$

where the subscripts "f" and "m" refer to the fibers and matrix, respectively, σ_c is the stress applied to the composite, V_f is the volume fraction of the fibers, E_f, E_m and E_c are the moduli of elasticity of the fibers, matrix and undamaged composite, respectively, and σ_f^T and σ_m^T are the residual thermal post-processing stresses outside the slippage region. Note that during cycling the composite stress σ_c varies continuously. Therefore, the stresses given by eqns. (1) represent instantaneous values, although dynamic (viscous) effects are not included in the present analysis.

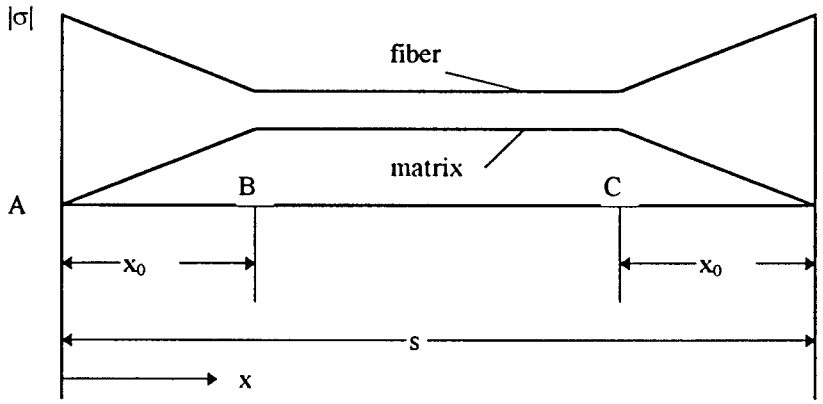


Fig. 1. Distribution of stresses in the fibers and in the matrix during cycling (not to scale). The stresses are shown along the fiber length between two parallel bridging cracks.

It can be immediately observed that the equilibrium of forces in the cross sections outside the slippage region is satisfied. The stresses within the slippage region are also balanced at each cross section. Within the slippage region, the stresses in the fiber and the

matrix are linear functions of the distance from the plane of the crack, i.e., x. The stress in the fiber is[12]:

$$\sigma_f = \frac{\sigma_c}{V_f} - \frac{2\tau x}{r} \tag{2}$$

where r is the fiber radius and τ is the interfacial shear stress that is assumed constant, as in the theories of Aveston-Cooper-Kelly and Budiansky-Hutchinson-Evans. Note that the finite element solution of Sorensen et al.[15] that showed that the maximum variation of the interfacial shear stress is 15% supports the theoretical solutions based on the assumption is that this stress is constant.

The stress in the matrix within the slippage region is

$$\sigma_m = \left(\sigma_c \frac{E_m}{E_c} + \sigma_m^T \right) \frac{x}{x_0} \tag{3}$$

where x_0 is the half-length of the slippage region that can be determined according to Pryce and Smith[13] as

$$x_0 = \frac{r}{2\tau} \left(\sigma_c \frac{V_m E_m}{V_f E_c} - \sigma_f^T \right) \tag{4}$$

V_m being the matrix volume fraction. The residual thermal stresses in the fiber and in the matrix can be found from the force equilibrium that should be preserved at an arbitrary cross section, i.e.,

$$E_f V_f (\varepsilon - \alpha_f \Delta T) + E_m V_m (\varepsilon - \alpha_m \Delta T) = 0 \tag{5}$$

where α_f and α_m are the coefficients of thermal expansion of the fibers and the matrix, respectively, and ΔT is a difference between the processing and operating temperatures.

The strain, ε, can be immediately evaluated from eqn. (5). However, considering the fact that the processing temperature of CMCs is usually above 1200^0C, it is necessary to account for an effect of the temperature on the properties of the materials of the fibers and the matrix. In this case, the strain will be found from

$$\varepsilon = \int_{T_p}^{T_0} \frac{E_f(T)V_f\alpha_f(T) + E_m(T)V_m\alpha_m(T)}{E_f(T)V_f + E_m(T)V_m} dT \tag{6}$$

where the integration is carried out from the processing (T_p) to the operating (T_o) temperature. An example of analytical expressions for the moduli of elasticity and the coefficients of thermal expansion of CMCs as functions of temperature can be found in the report of NASA TM 106789.

The residual stresses in the region that is not affected by the slip are found as

$$\begin{aligned} \sigma_f^T &= E_f(\varepsilon - \alpha_f\Delta T) = E_f(T_o)\int_{T_p}^{T_o} \frac{E_m(T)V_m(\alpha_m(T) - \alpha_f(T))}{E_f(T)V_f + E_m(T)V_m} dT \\ \sigma_m^T &= E_m(\varepsilon - \alpha_m\Delta T) = E_m(T_o)\int_{T_p}^{T_o} \frac{E_f(T)V_f(\alpha_f(T) - \alpha_m(T))}{E_f(T)V_f + E_m(T)V_m} dT \end{aligned} \tag{7}$$

where T_0 and T_p are the operating (room) and processing temperatures, respectively.

It is now necessary to evaluate an instantaneous modulus of elasticity of the material. This procedure follows the approach adapted by Pryce and Smith[13], although the problem considered in the present solution is different which dictates modifications outlined below.

The instantaneous mean strain in the fiber is found by averaging the strains over the spacing length, i.e.,

$$\bar{\varepsilon}_f = \frac{2x_0}{s}\varepsilon_{AB} + \frac{s-2x_0}{s}\varepsilon_{BC} \tag{8}$$

where the mean strain within the slippage region ε_{AB} and the strain outside the slippage region ε_{BC} can be expressed in terms of the fiber stresses. The substitution of these strains into eqn. (8) yields

$$\bar{\varepsilon}_f = \frac{x_0}{s}\left[\frac{\sigma_c(2V_f E_f + V_m E_m)}{V_f E_f E_c} + \frac{\sigma_f^T}{E_f}\right] + \frac{s-2x_0}{s}\left(\frac{\sigma_c}{E_c} + \frac{\sigma_f^T}{E_f}\right) \tag{9}$$

Note that the strain given by eqn. (9) includes the post-processing residual and the additional cycling components. The increase of the strain due to cyclic loading is represented by the latter component. The mean residual strain is found from an analog of eqn. (8) where the mean strain in the region AB is given by $\sigma_f^T/2E_f$, while the strain within the region BC is σ_f^T/E_f (see Fig. 2). Accordingly,

$$\bar{\varepsilon}_f^T = \frac{s-x_0}{s}\frac{\sigma_f^T}{E_f} \tag{10}$$

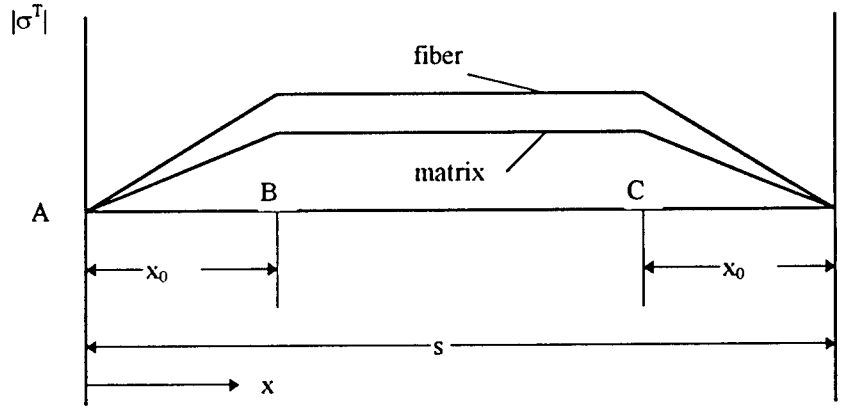


Fig. A2. Thermal residual stresses in a unidirectional material with post-processing cracks (not to scale).

The additional strain associated with cyclic loading can now be found as a difference of the strains given by eqns. (9) and (10), i.e.,

$$\bar{\varepsilon} = \left[1+\frac{r}{2s\tau}\frac{\left(\sigma_c\frac{V_mE_m}{V_fE_c}-\sigma_f^T\right)V_mE_m}{V_fE_f}\right]\frac{\sigma_c}{E_c} \tag{11}$$

The instantaneous modulus of elasticity is determined as

$$E = \frac{d\sigma_c}{d\bar{\varepsilon}} \tag{12}$$

yielding

$$E = \left[1+\frac{r}{2s\tau}\frac{E_mV_m}{E_fV_f}\left(2\sigma_c\frac{E_mV_m}{E_cV_f}-\sigma_f^T\right)\right]^{-1}E_c \tag{13}$$

Note that eqn. (13) presents the elastic modulus of a cracked material that is not affected by cycling. Accordingly, it is possible to compare this result to experimental findings of Karandikar and Chou [14] who showed that the change of the modulus of a unidirectional Nicalon fiber CAS matrix composite is a linear function of the matrix crack density (1/s):

$$\Delta E = E_c - E = k_1 + k_2(1/s) \tag{14}$$

where k_1 and k_2 are constants. For Nicalon fiber CAS matrix, k_1 = -6.7350 and k_2 = 6.2754 (the modulus of elasticity is measured in MPa and the crack density in number/mm). Note that the present approach to the solution remains valid as long as an arbitrary analytical relationship $\Delta E = f(s)$ is available.

A combination of eqns. (13) and (14) yields the relationship between the interfacial shear stress and the matrix crack spacing:

$$\tau = \left[\frac{E_c - k_1 - k_2/s}{k_1 + k_2/s}\right]\left[\frac{r}{2s}\frac{E_m V_m}{E_f V_f}\left(2\sigma_c \frac{E_m V_m}{E_c V_f} - \sigma_f^T\right)\right] \tag{15}$$

Now it is necessary to find a relationship between the interfacial shear stress and the surface temperature of the specimen. This problem can be addressed by considering the equilibrium between the rate of the steady state heat loss from a unit volume of the specimen and the rate of work performed by the interfacial friction within this volume[10]. The former quantity was presented in the paper of Cho et al.[10] for the general case where the flow loss occurs through conduction in the fiber direction and free convection and radiation from the surface. In the case of a uniform post-processing crack distribution and small-amplitude vibrations excited in the course of a nondestructive test prior to any additional fatigue loading, the temperature may be assumed independent of the axial coordinate oriented along the fibers. Therefore, no conduction takes place in the fiber direction. At the sea level the radiation is typically small compared to free convection (although the relative contribution of radiation increases with altitude). Therefore, radiation from the surface may be neglected, although this is not necessary for the solution. Retaining the radiation contribution, the rate of the heat loss from the element with the surface area A_S (including both surfaces of the specimen) and the volume V = $A_S t/2$ where t is the thickness becomes

$$\dot{q} = \left[h(T_s - T_a) + \varepsilon\beta_0(T_s^4 - T_a^4)\right]\frac{2}{t} \tag{16}$$

In eqn. (16), T_S and T_a are the surface and ambient air temperatures, respectively, h is the heat transfer coefficient, ε is the emissivity and $\beta_0 = 5.67*10^{-8}$ W/m^{2}K^4 is the Stefan-Boltzman constant. As indicated above, the second term in the square brackets in eqn. (16) may be negligible. Note that the heat transfer coefficient refers to the mean properties of the film adjacent to the surface of the specimen, i.e., its evaluation requires the knowledge of $T = (T_S + T_a)/2$.

The heat transfer coefficients from the surfaces of a representative element can be determined as

$$h_i = \frac{(Nu)(k)}{L} \tag{17}$$

where the subscript identifies the surface, Nu is the Nusselt number, L is the ratio of the surface area of the element to its perimeter and k is the thermal conductivity of the film. The Nusselt number can be found as a function of the Rayleigh number (see, for example, Incropera, F.P. and DeWitt, D.P., "Fundamentals of Heat and Mass Transfer," 4th edition, Wiley, New York, 1996). In particular, if the surfaces losing heat are horizontal, the following formulae apply:

Upper surface: $\quad Nu = 0.27Ra^{1/4} \quad$ for $10^5 < Ra < 10^{10}$

Lower surface: $\quad Nu = 0.54\,Ra^{1/4} \quad$ for $10^4 < Ra < 10^7$

$\qquad\qquad\quad 0.15Ra^{1/3} \quad$ for $10^7 < Ra < 10^{11}$ (18)

The Rayleigh number can be calculated by

$$Ra = \frac{g\beta(T_s - T_a)L^3}{\upsilon\alpha} \tag{19}$$

where g is the gravity acceleration, β is the volumetric thermal expansion coefficient, ν is the kinematic viscosity and α is the thermal diffusivity of the film at its average temperature. The total heat transfer coefficient is a sum of the coefficients of the opposite surfaces.

The instantaneous work produced by the interfacial friction on the slippage length of one fiber is obtained as[10]

$$W = 2\int_0^{x_0} (\pi d_f \tau)(u_f - u_m)dx \tag{20}$$

where u_f and u_m are dynamic components of the axial displacements of the fiber and the matrix due to cyclic loading that are functions of the x-coordinate. Note that the upper limit of integration given by eqn. (4) is affected by the magnitude of the interfacial shear stress. The difference between dynamic components of the axial fiber and matrix displacements can be found as

$$u_f - u_m = \frac{1}{2}[\delta\varepsilon_f(x) - \delta\varepsilon_m(x)](x_0 - x) \tag{21}$$

where $\delta\varepsilon_f$ and $\delta\varepsilon_m$ are the dynamic strains at the cross section x. This equation reflects the fact that the fiber and the matrix experience identical axial displacements at $x = x_0$, and the change of the length of the element (x_0 - x) can be found as the mean strain within this element multiplied by its length.

The dynamic strain in the fiber is determined as a difference between the total and residual strains. The former can be found from eqn. (2) but it is more convenient to use the following expression that immediately follows from Fig. 1:

$$\varepsilon_f(x) = \left[\frac{\sigma_c}{V_f} - \left(\frac{\sigma_c}{V_f} - \frac{\sigma_c E_f}{E_c} - \sigma_f^T\right)\frac{x}{x_0}\right]E_f^{-1} \tag{22}$$

The latter strain is of course (see Fig. 2), $(\sigma_f^T/E_f)(x/x_0)$.

Now the dynamic components of the strains in the fibers and in the matrix can be found as

$$\begin{aligned} \delta\varepsilon_f(x) &= \left[\frac{1}{V_f} - \left(\frac{1}{V_f} - \frac{E_f}{E_c}\right)\frac{x}{x_0}\right]\frac{\sigma_c}{E_f} \\ \delta\varepsilon_m(x) &= \frac{x}{x_0}\frac{\sigma_c}{E_c} \end{aligned} \tag{23}$$

Substituting dynamic strains given by eqns. (23) into eqn. (21) and subsequently, integrating eqn. (20) one obtains

$$W = \frac{\pi d_f \tau \sigma_c}{3V_f E_f} x_0^2 \tag{24}$$

Note that the composite stress in eqn. (24) represents the maximum stress during the cycle, while the minimum stress is assumed equal to zero. This is probably the most practical case for a nondestructive dynamic testing where a unidirectional load is periodically applied to the component (like in the case of acoustic pressure). In the case of a stress ratio different from zero, the composite stress should be replaced with the stress range.

The rate of the frictional energy dissipation per unit volume can be found as recommended by Cho et al.[10], i.e.,

$$\dot{w} = 2fW / (\pi r^2 s / V_f) \tag{25}$$

where f is the frequency of loading and the factor "2" in the numerator accounts for the

fact that equal amounts of energy are generated during the loading and unloading phases of each cycle.

The solution can be obtained by prescribing the surface temperature. Then a relationship between the matrix crack spacing and the interfacial shear stress can be obtained from the requirement that the rate the heat loss given by eqn. (16) must be equal to the rate of the frictional energy dissipation according to eqn. (25). This relationship should be considered together with eqn. (15) to specify both the interfacial shear stress as well as the matrix crack spacing.

Discussion and Conclusions

The solution presented in this report outlines theoretical backgrounds for a nondestructive detection of the presence and density of post-processing cracks in ceramic matrix composite materials. The density of the matrix cracks and the interfacial shear stress can be evaluated using this solution, based on the measurements of the surface temperature of the component subjected to forced periodic vibrations. The solution is obtained in a closed-form, i.e., it is accurate as long as the basic assumptions incorporated into the analysis are valid. In particular, these assumptions include the form of matrix cracking, i.e. long bridging cracks, and the presumed analytical relationship between the change of the modulus of elasticity and the matrix crack spacing. However, these assumptions are justified by available experimental evidence. Note that vibrations of the component are assumed to take place in such manner that the stresses are uniform throughout the thickness. This implies a membrane state of stresses and in thin ceramic matrix components subjected to a small amplitude dynamic pressure during a nondestructive test such assumption is acceptable. Another factor that will be considered

in the future research is an effect of viscosity of the matrix (and possibly the fibers) on the frictional heating.

References

1. Bischoff, E., Ruhle, M., Sbaizero, O. and Evans, A.G., Microstructural studies of the interface zone of a SiC-fiber-reinforced lithium aluminum silicate glass-ceramic. *Journal of the American Ceramic Society*, 1989, **72**, 741-745.
2. Nishiyama, K., Umekawa, S.,Yuasa, M. and Fukumoto, I., Effect of fiber coating on interfacial phenomena in tungsten-fiber/boron carbide composites. *Proc. Fourth Japan-US Conf. Composite Mater.*, 1988, Technomic, Lancaster, PA, pp. 776-788.
3. Marshall, D.B. and Evans, A.G., Failure mechanisms in ceramic-fiber/ceramic-matrix composites. *Journal of the American Ceramic Society*, 1985, **68**, 225-231.
4. Wetherhold, R.C. and Zawada, L.P., Fractography of glasses and ceramics. Eds. Frechete, V.D. and Varner, J.P., *Ceramic Transactions*, 1991, Vol. 17, American Ceramic Society, Westerville, Ohio, p. 391.
5. Zawada, L.P. and Wetherhold, R.C., The effect of thermal fatigue on a SiC fibre-reinforced nitrogen glass matrix composites. *Journal of Materials Science*, 1991, **26**, 648-654.
6. Kahraman, R., A microdebonding study of the high-temperature oxidation embrittlement of a cross-ply glass-ceramic/SiC composite. *Composites Science and Technology*, 1996, **56**, 1453-1459.
7. Westwood, M.E,, Webster, J.D., Day, R.J., Hayes, F.H. and Taylor, R., Review. Oxidation protection for carbon fibre composites. *Journal of Material Science*, 1996, **31**, 1389-1397.
8. Holmes, J.W. and Shuler, S.F., Temperature rise during fatigue of fibre-reinforced ceramics. *Journal of Materials Science Letters*, 1990, **9**, 1290-1291.

9. Holmes, J.W. and Cho, C., Experimental observations of frictional heating in fiber-reinforced ceramics. *Journal of the American Ceramic Society*, 1992, **75**, 929-938.

10. Cho, C., Holmes, J.W. and Barber, J.R., Estimation of interfacial shear in ceramic composites from frictional heating measurements. *Journal of the American Ceramic Society*, 1991, **74**, 2802-2808.

11. Aveston, J. and Kelly, A., Theory of multiple fracture of fibrous composites. *Journal of Material Science*, 1973, **8**, 352-362.

12. Budiansky, B., Hutchinson, J.W. and Evans, A.G., Matrix fracture in fiber-reinforced ceramics. *Journal of Mechanics and Physics of Solids*, 1986, **34**, 167-189.

13. Pryce, A.W. and Smith, P.A., Matrix cracking in unidirectional ceramic matrix composites under quasi-static and cyclic loading. *Acta Mettalurgica et Materialia*, 1993, **41**, 1269-1281.

14. Karandikar, P.G. and Chou, T.-W., Microcracking and elastic moduli reductions in unidirectional Nicalon-CAS composite under cyclic fatigue loading. *Ceramic Engineering & Science Proceedings*, 1992, **13**, 882-888.

15. Sorensen, B.F., Talreja, R. and Sorensen, O.T., Micromechanical analysis of damage mechanics in ceramic-matrix composites during mechanical and thermal cycling, *Composites*. 1993, **24**, 129-140.

A REVIEW OF BENCHMARK FLOWS FOR LARGE EDDY SIMULATION

G. A. Blaisdell
Associate Professor
School of Aeronautics and Astronautics

Purdue University
West Lafayette, IN 47907

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

And

Wright Laboratory

August 1997

A REVIEW OF BENCHMARK FLOWS FOR LARGE EDDY SIMULATION

G. A. Blaisdell
Associate Professor
School of Aeronautics and Astronautics
Purdue University

Abstract

The purpose of this report is to provide an overview of the standard references on large eddy simulation (LES) of turbulent flows and to outline a series of benchmark flows which can be used to test and validate a new LES code. A list of review articles on LES is given, including several recent papers which show the current state-of-the-art. References for the standard subgrid-scale (SGS) models are provided. An outline is given of flows which can be used as test cases, and references are listed for previous direct and large eddy simulations of these flows, so that comparisons can be made. The emphasis here is on compressible flows; however, many of the test cases included are incompressible. Lastly, several open issues in LES are discussed and some relevant papers are cited.

A REVIEW OF BENCHMARK FLOWS FOR LARGE EDDY SIMULATION

G. A. Blaisdell

1 Introduction

Most flows of engineering interest are turbulent, and turbulence can have a large effect on the performance of engineering systems. For instance the drag on a body is generally higher for turbulent flow; turbulence greatly increases heat transfer rates, and turbulent mixing is important to combustion systems. Numerical simulation of turbulent flows has become a valuable tool for gaining insight into the complicated flow physics of turbulence and it has provided data for evaluation of engineering turbulence models.

Direct numerical simulation (DNS) of turbulent flows involves solving for the time dependent motion of all the relevant length scales within a turbulent flow. Because the range of length scales increases with Reynolds number, it becomes prohibitively expensive in terms of computer resources to compute DNS of high Reynolds number flows. A method for overcoming this limitation is to directly solve for the large scale motions and to model the effect of the small scale motions. This approach is called large eddy simulation (LES).

In LES the small scales are averaged out using a spatial filter. As a result of the averaging procedure some information is lost which must be replaced with a subgrid-scale (SGS) model. The SGS model provides a closure for the subgrid-scale stresses, much the same as the way an engineering turbulence model provides information on the Reynolds stresses. However, it is believed that the large scales vary widely from one flow to another while the small scales are more universal. Therefore, in LES one can use a simple SGS model and, since the large scales are solved for directly, the results of an LES should be more robust than those of an engineering turbulence model.

The purpose of this report is to provide an overview of the standard references on large eddy simulation (LES) of turbulent flows and to outline a series of benchmark flows which can be used to test and validate a new LES code. The scope of this report is limited in that it is assumed that the LES code to be developed is a compressible flow finite difference code meant to simulate flows of engineering interest. Although it is assumed that the compressible equations are solved, many of the benchmark examples are incompressible. Therefore, it will be necessary to run the code at low Mach numbers for some of the test cases.

The report is organized into three sections. In the first section a list of general references on LES is given, followed by a discussion of SGS models in current use. Section 2 contains a list of possible test cases for validating an LES code. Then in the third section a list is given of current open issues in LES. An extensive reference list is provided. Copies of most of the references will be given to the technical contact at Wright Lab (Dr. Miguel Visbal); however, they are not included as part of this report.

1.1 LES Review Articles

There are several good review papers on LES. It is helpful to have some historical perspective in looking at these. LES was first developed in the meteorological community in the 1960's, and it has been used extensively in weather prediction. It was adopted for basic studies of turbulence by the engineering community during the 1970's. Because of computer limitations during that time, DNS studies were very limited and LES was relied upon for turbulence simulations. The early work in engineering is reviewed in Ferziger[18], where a hierarchy of turbulence simulations is outlined which provides a useful framework for understanding

turbulence models and turbulence simulations. A review more specific to DNS and LES and which includes the advances made in the early 1980's is given in Rogallo & Moin[70].

During the 1980's and early 1990's computer resources became sufficient that DNS could be used to study basic turbulent flows. This avoided the ambiguity of relying on an SGS model. As computer capability grew, especially with the proliferation of large parallel computer systems, people began to consider LES as a possible engineering tool, or at least as a means to study turbulence at higher Reynolds numbers and in more complex flows. A paper which considers the future prospects of LES is given by Reynolds[68]. It is argued that although LES may become a useful tool for solving the most difficult problems, traditional engineering turbulence models will continue to be relied upon for routine design work. At the current time this is not a consensus viewpoint.

The current state-of-the-art in LES is provided in several recent review articles. An extensive review which provides an emphasis on some of the work done in France is given in Lesieur & Métais[40] and Lesieur & Compte[39]. A thorough review of LES and DNS simulations with an emphasis on the work done by researchers from Stanford is given in Piomelli[59, 60]. An article written for the general public is Moin & Kim[52], while a more technical discussion of progress in LES and some outstanding issues is given in Moin[54]. Although it is not a review of LES, the paper by Crawford *et al.* [16] explains some of the issues in doing turbulence simulations on parallel computers. Lastly, LES is also used in the wind engineering community, and a good overview from that perspective is given by Murakami[57].

There have been a several conferences or special forums which have focused on LES. One is the recent FAICDL conference from which several of the references in this report are taken. Others include the ASME symposium organized by Ragab & Piomelli[66] and the AGARD sponsored symposium[2] held in Crete in 1994.

1.2 SGS Model Formulations

With LES a spatial filter is applied to the Navier-Stokes equations in order to remove the smaller length scales. As with traditional Reynolds averaging, this process results in a closure problem. For LES the unclosed terms are the called the subgrid-scale stresses, and a subgrid-scale (SGS) model is needed. The first SGS model developed was the eddy viscosity model of Smagorinski[71]. As with engineering turbulence models, a velocity scale and a length scale are needed to form the eddy viscosity. The Smagorinski model uses a mixing length hypothesis to obtain the velocity scale from the length scale and the resolved strain rate. Unlike an engineering turbulence model, the length scale is set by the filter width, which is usually assumed to be the same as the grid spacing, although for finite difference calculations Moin[54] and Spyropoulos[73] have recently advocated applying a filter with a width that is larger than the grid spacing. A difficulty with the Smagorinski model is that the model constant in the formula for the eddy viscosity must be changed from one flow to another to get optimal results. Also, the formulation assumes that the SGS stress is aligned with the resolved scale strain rate, which is not always true.

A model which avoids the assumptions inherent in an eddy viscosity model is the scale-similarity model of Bardina *et al.* [5]. However, it was found that this SGS model does not dissipate any energy and so leads to a numerical instability. An ad-hoc fix suggested by Bardina is to use a linear combination of the scale-similarity model and Smagorinski's model — the so-called mixed model. The eddy viscosity part of the models provides the needed dissipation.

A third modeling approach is the structure function model of Lesieur[40, 39]. Structure functions involve correlations of velocity difference and for homogeneous turbulence are related to spectra. The models are

based on theoretical work on homogeneous turbulence; however, the use of structure functions, which are computed in physical space, allows the model to be extended to inhomogeneous flows. This model is not widely used outside of France.

As noted above, one of the difficulties with Smagorinski's model is one must change the model constant depending on the flow considered. However, a new development which has overcome this limitation is the dynamic model of Germano *et al.* [19]. The dynamic model is more properly referred to as the dynamic procedure, because it is a procedure to determine the model constant (coefficient) as a function of time and space in a dynamic way using information from the resolved scales of the LES. The formulation is complicated and the interested reader is referred to the review articles cited above and the original references. The dynamic model has many nice properties. For instance, when Samgorinski's model is used in wall bounded flows it is necessary to use a damping function to modify the eddy viscosity. However, the dynamic model automatically gives the correct limiting behavior in the near wall region without the need for any ad-hoc damping functions. Also, the dynamic model turns itself off in laminar flows and so can simulate transitional flows without any special modification. There are some problems yet to be completely resolved in using the dynamic model, and these are discussed in section 3.

Several modifications to the dynamic model have been proposed. The most widely used is that due to Lilly[42]. The dynamic model makes use of the Germano identity, which is a tensor equation. In order to use this relation to determine the scalar model coefficient, Lilly proposed to use a least squares approach rather than the method employed in the original formulation.

The dynamic procedure was extended to the mixed model by Zang[87]. (See also Vreman *et al.* [79].) A dynamic version of the structure function model has also been created.

Large eddy simulation of compressible turbulent flows was first developed by Erlebacher *et al.* [17]. They used an extension of the Smagorinski model based on mass-weighted filtering, similar to the way incompressible turbulence models are extended to compressible flows using mass-weighted averaging. The dynamic model was extended to compressible flows in Moin *et al.* [55].

2 Benchmark Flows

In this section an outline of flows is given which can be used to test an LES code. The flows considered are ones for which DNS or LES have already been computed and, therefore, comparison can be made with results from previous simulations. It will be obvious that the list of references does not include experimental investigations. However, it should be emphasized, that in validating an LES code, comparison should be made to both other simulation results and experimental data. This is especially true when extending the simulations to higher Reynolds numbers, where DNS data is not available. However, a complete reference list including relevant experimental work would require a large amount of time to compile and would be excessively long. Therefore, the reference list included here is limited to DNS and LES simulations. References to relevant experimental research and other simulations should be available in the papers cited or by searching on-line data bases.

The outline of simulations provided here is not a critical review. Some of the simulations may be better suited as benchmark cases than others. Also, no assessment is provided as to the quality of the simulation results. Quality of the results is, however, an important issue. Simulations should not be trusted merely because no turbulence model was used and the label "DNS" applied. Simulation results should be checked for accuracy based on computing spectra (when possible) and on grid resolution studies. Also, comparison with experimental data should be included when available. Even with the above limitations, it is hoped that

the following list will provide a useful set of flows with which to test a new LES code. Again, the emphasis here is on compressible flows amenable to solution by finite difference methods.

2.1 Compressible Isotropic Turbulence

The simplest turbulent flow is decaying isotropic turbulence. LES of compressible isotropic turbulence were used by Moin *et al.* [55] to investigate their compressible formulation of the dynamic model. Similar simulations were done by Spyropoulos & Blaisdell[74] and Spyropoulos[73] to more thoroughly investigate the ability of the dynamic model to capture compressibility effects and to examine several other issues in applying the dynamic SGS model. The above LES and previous DNS of compressible homogeneous turbulence all used initial conditions which created artificial acoustic waves. Ristorcelli & Blaisdell[69] have devised a method of producing initial conditions which are more consistent and which do not produce strong acoustic waves. They have produced very clean DNS results; however, further refinement of the initial conditions is continuing. The previous LES and DNS results can be used for comparison; however, investigations into compressibility effects on isotropic turbulence should use the updated initial condition method to avoid contamination by acoustic waves generated by strong transient behavior.

2.2 Channel Flow

The most widely studied inhomogeneous flow is fully developed turbulent channel flow. This flow is inhomogeneous in the wall normal direction, but is homogeneous in planes parallel to the wall. The first high-quality LES of this flow was done by Moin & Kim[53]. This was later followed by a DNS by Kim *et al.* [32], which has been widely used as a benchmark simulation. Detailed statistics from this simulation and comparisons with Reynolds stress model formulations are available in Mansour *et al.* [46]. A DNS at higher Reynolds number has been recently completed (Mansour, private communication). Piomelli *et al.* [63] performed LES of a channel flow and pointed out the need for consistency between the filter and the model formulation. A very high Reynolds number LES has been done by Kravchenko *et al.* [35] who used a B-spline spectral method with embedded grids, which allows a very fine mesh near the wall and a coarser mesh in the center of the channel.

The above simulations are incompressible. They would make suitable test cases for a compressible code if a low Mach number is used. However, there is a difficulty in using a compressible code to simulate fully developed channel flow. For incompressible flow the pressure only occurs inside a gradient. Therefore, a constant pressure gradient can be applied and it is still possible to use periodic boundary conditions in the streamwise direction. (Periodic boundary conditions are natural to use in homogeneous directions.) However, in the compressible equations the absolute pressure is important, and periodic boundary conditions cannot be used if the pressure has a gradient in the streamwise direction. Therefore, the compressible flow either cannot be fully developed (*i.e.* homogeneous) or it cannot be pressure driven. A way around this is to drive the flow with an artificial body force. This was done by Coleman *et al.* [15] whose results were further analyzed by Huang *et al.* [24]. A similar approach was also taken by Wang & Pletcher[83]. There remains some difficulty in interpreting the results, however, because of mean density variations in the wall normal direction.

Channel flow simulations have also been used to study transition. This has been done by Piomelli & Zang[64] for incompressible flow.

2.3 Square Duct

An added complication over channel flow is provided by flow in a square duct. In addition to having two inhomogeneous directions, this flow develops a secondary mean flow that is difficult for turbulence models to capture and which is important to heat transfer. Flow in a square duct has been simulated by Huser & Biringen[25, 26] and Balaras *et al.* [7].

2.4 Curved Channel

Another complication is to consider flow in a curved channel. Incompressible DNS have been done by Moser & Moin[56]. Because of the curved geometry, this flow could provide a test of using generalized coordinates.

2.5 Couette Flow

Couette flow provides a test case similar to channel flow; however, it avoids the ambiguity involved with the pressure, since it is driven by a moving wall. One difficulty with turbulent Couette flow, however, is that very long streamwise vortical structures develop. An extremely long computational domain is required in order to properly capture these structures and to ensure the periodic boundary conditions do not affect the solution. A simulation has been done by Komminaho & Johansson[33].

2.6 Boundary Layer

For incompressible boundary layers on a flat plate the standard for comparison is the DNS of Spalart[72]. Again a compressible simulation of this flow could be done at low Mach numbers.

The DNS by Spalart is of a spatially developing boundary layer. Less computer resources are required to simulate a time-developing boundary layer, and this has been done for compressible flow by Hatay & Biringen[23] and Childs & Reisenthel[10]. A DNS of a supersonic ($M_\infty = 2.25$) spatially developing boundary layer has been done by Rai *et al.* [67]. This simulation gives a descent mean velocity profile; however, the Reynolds number is lower than that of any of the experiments available for comparison. An LES of this same flow was performed by Spyropoulos[73] and Spyropoulos & Blaisdell[75] using a modified version of the code of Rai. They found that the numerical errors associated with the upwind-biased scheme of Rai suppressed the smaller resolved scales and caused the skin friction to be underpredicted. Even using a total number of grid points that was only a factor of 3 less than that of the DNS, unsatisfactory results were obtained. The numerical method may be useful for DNS if a sufficient number of grid points is used to ensure the solution is well resolved. However, the flow field in an LES is by definition not well resolved and, therefore, numerical errors can have a large impact. This is discussed in further detail in section 3.

It is very difficult to perform simulations at higher Mach numbers. However, DNS of boundary layers up to Mach 6 have been done by Adams[1].

2.7 Backward Facing Step

One of the standard test cases for Reynolds averaged turbulence models is the backward facing step. This flow may not be an appropriate test case for a single-block generalized geometry code, because of the geometric singularity at the step edge; however, it would make an excellent test case for a multi-block code.

A very large DNS of this flow has been done by Le & Moin[36] and Le *et al.* [37]. An LES on a coarser grid has been performed by Akselvoll *et al.* [3].

2.8 Cylinder Wake

Flow over a circular cylinder seems to be becoming a standard test case for LES codes. It also would be a good test case for testing formulations for generalized geometries.

An LES was done by Beaudan & Moin[8] who used the up-wind biased scheme of Rai *et al.* [67]. They also reported difficulty due to numerical errors suppressing the turbulence. They compared use of the dynamic model and the Smagorinski model and found that the dynamic model was able to produce large eddy viscosity in the correct regions of the flow, while the Smagorinski model put large eddy viscosity in regions of large shear, even if the flow there was laminar. Mittal[49] performed LES using a second order central scheme and obtained better results than those of Beaudan & Moin who used a fifth order upwind-biased scheme. Xia & Karniadakis[86] used a finite element method and make comparisons with the results of Beaudan & Moin and Mittal. Karniadakis has pointed out that in comparing with experiments the time history of the base pressure is a more discriminating quantity than integrated quantities such as the drag coefficient. He also has presented results for supersonic flow over a cylinder.

Li & Dalton[41] have done LES of oscillating flow over a cylinder. This has applications to ocean engineering.

2.9 Sphere Wake

Few simulations over a sphere have been done. Two such studies can be found in Karniadakis[31] and Mittal[50].

2.10 Mixing Layer

The compressible mixing layer is a standard test case for compressibility effects in Reynolds averaged turbulence models. There have been several simulations of time developing compressible mixing layers; however, they mostly have been used to understand turbulent structure, but have not contained sufficient sample size to obtain statistical information. An exception is the recent work of Vreman[81] and Vreman *et al.* [80], who have been able to compute mean profiles and turbulent statistics.

2.11 Jets

Several researchers have performed LES of a turbulent jet. This has mostly been done in the computational aeroacoustics (CAA) community. Examples are Chyczewski & Long[14], Mankbadi[45], and Wang *et al.* [82]. The more complicated case of a confined co-annular jet has been simulated by Akselvoll[4]. Here, because of the large separated flow, the grid spacing requirement in the near wall region is not as stringent as one might expect.

Voke *et al.* [78] have performed an LES of a jet impinging on a solid wall. This flow is important in combustion systems.

2.12 Axial Vortex

Strong rotation generally suppresses turbulence, and one context in which to study this effect is in a vortex. DNS and LES of a time developing axial vortex have been done by Sreedhar & Ragab[76] using a finite difference scheme in a finite sized domain. A higher quality DNS has been done by Qin[65] using a spectral method in an infinite domain. A simulation of an axial vortex embedded in a turbulent boundary layer has been done by Liu *et al.* [43].

2.13 Flow Over General Geometries

The ultimate goal of LES is to be able to simulate complex flows of engineering interest. There are several examples of simulations that have taken steps in that direction. DNS of turbulent channel flow with riblets mounted on one of the walls have been performed by Chu & Karniadakis[13] and Choi *et al.* [12]. Wu & Squires[84] have performed LES of a boundary layer encountering a bump using generalized coordinates. Adams[1] has simulated a supersonic compression ramp, although the grid used is coarse and the computational domain is small. Flow over an airfoil with separation has been simulated by Jansen[27] using a finite element LES code. This study has shown how important it is to match the exact conditions of the experiment, including the effect of wind tunnel walls and surface roughness on the airfoil.

These latter cases would be of benefit to testing an LES code for use in generalized geometries. However, simulating these flows can require very significant computer resources.

3 Open Issues in LES

There are several open issues which remain unresolved in LES. They range from fundamental questions regarding the basic formulation of the LES equations to practical matters of applying LES to high Reynolds number wall bounded flows.

3.1 Effect of Numerical Errors

Within the past two years it has become apparent that errors from the differencing scheme used to compute a large eddy simulation can have an overwhelming effect. This was not the case with early LES because researchers tended to use spectral methods, which are very accurate for a wide range of length scales. However, with the recent interest in more complex flows, simulations have been done using finite difference, finite volume and finite element methods. The emphasis in the current report is on finite difference methods. A finite difference method does a better job of representing smooth functions than functions with high-wavenumber oscillations. Therefore, the smaller resolved scale eddies within an LES using a finite difference scheme are inaccurate. However, the LES models that are used assume that the simulation is accurate up to the cut-off wavenumber representative of the grid spacing. The truncation error of a finite difference scheme can be larger than the SGS stress predicted by the model and, therefore, overwhelm the model. One way to test for this is to run the simulation without using the SGS model. If the results differ little, then the SGS model is not active. This either means that the flow is well resolved — essentially a DNS — or that numerical errors are dominating the model. An additional concern comes up when using the dynamic procedure to compute the model coefficients. Since the dynamic procedure uses information from the smallest resolved scale eddies, if these are contaminated by large numerical errors, the coefficients computed will be incorrect.

Ghosal[22] performed a theoretical analysis of the effect of truncation and aliasing errors on LES calculations. It was determined that low order differencing schemes would be almost completely dominated by truncation errors. For higher order methods aliasing errors become a concern. Aliasing errors are caused by nonlinear terms or terms with nonconstant coefficients. For instance, a product of two terms can create a function that oscillates with a spatial wavenumber that is higher than what the grid can support (corresponding to the Nyquist frequency). When a high wavenumber mode is discretized it mimics or is aliased to a lower wavenumber mode. This causes an error which can lead to nonlinear instabilities. Aliasing errors are traditionally thought of in the context of spectral methods; however, they also occur for finite difference

schemes. They are less important for low-order schemes because the high wavenumber modes which give rise to aliasing errors are suppressed.

Kravchenko & Moin[34] have investigated the effect of numerical errors on LES calculations of a channel flow. They also give a nice analysis of aliasing errors and the relation between energy conservation properties and stability for incompressible flow. The issue of numerical errors is also pointed out in the review of Moin[54].

Spyropoulos[73] and Spyropoulos & Blaisdell[75] performed LES of a supersonic boundary layer on a flat plate using an upwind-biased finite difference scheme. They found that the upwind-biased scheme artificially suppressed the smaller turbulent scales, resulting in reduced skin friction. The SGS model was ineffective and no computational savings could be realized relative to a DNS.

In a recent paper Mittal & Moin[51] discuss the use of upwind schemes in LES. However, there is not a consensus view on this issue in the LES community. Some researchers hold to the view that LES can be done with no SGS model, in which case the numerical error of the differencing scheme acts as a type of SGS model. The argument is that all a model must do is provide a mechanism to dissipate energy. From the flows that have been simulated using this approach, it seems that this might work for free shear flows, in which the large scales dominate and the small scales are passive. However, for wall bounded flows, in which the small scale eddies in the near wall region are important, this approach is like to give poor results. It would be useful to have a systematic comparison of different methods on a variety of flows so that their ranges of validity can be established.

For complex geometry flows, traditional CFD methods employ implicit time advancement schemes. The issue of explicit versus implicit schemes for turbulence simulations is addressed in Choi & Moin[11].

3.2 Generation of Turbulent Inlet Conditions

One large difference between LES and Reynolds averaged calculations is that at an inflow boundary with turbulent flow the boundary conditions for LES must be time dependent and should represent the dynamics of the large scale eddies coming into the domain. One approach which has been widely used is to generate artificial turbulence at the inlet and to allow it to develop into realistic turbulence downstream. Examples of this approach is given in Le & Moin[36], Le *et al.* [37], and Lund *et al.* [44]. The disadvantage of the method is that it can require a significant distance for the nonlinear turbulent processes to become established and to develop realistic turbulence.

A second approach is to carry out two simulations simultaneously. One simulation is of a simple flow, such as a fully developed channel flow. Data from a plane in the simple simulation is then fed into the inlet of the complex simulation. In this way realistic turbulence with the correct spatial and temporal correlations is supplied. This approach is discussed by Moin[54] for the cases of a co-annular dump combustor and a two-dimensional diffuser. It has also been used for a boundary layer by Wu & Squires[85]. As an additional problem, Moin points out that in comparing between LES and experiments it is very important that the boundary conditions are the same. Small difference in inflow conditions or differences in outflow geometry can have significant effects on the results.

As an alternative to providing turbulent inflow conditions, one may simulate the transition process. This can either be natural transition or by-pass transition. For natural transition inflow disturbances still have to be specified, but it is easier to specify small disturbances rather than fully nonlinear turbulent flow. For by-pass transition a mechanism for tripping the flow must be provided, such as simulating a blowing and suction slot. Simulating the transition process has the advantage of not having to specify turbulent inflow

conditions; however, it can require a long development distance to obtain fully turbulent flow and for some flow situations it may not be appropriate.

3.3 Approximate Boundary Conditions

Another issue with boundary conditions arises for high Reynolds number simulations of wall-bounded flows. The production of turbulence within a boundary layer is controlled by the near-wall dynamics. For high Reynolds numbers this near-wall region can be a tiny fraction of the boundary layer thickness. Therefore, it becomes prohibitively expensive to compute the details of the eddies in the near-wall region. A way to get around this is to provide an approximate boundary condition at some distance away from the wall. This is similar to using wall functions in a Reynolds averaged calculation, except that the approximate boundary conditions, as they are called, must be time dependent.

Examples of approximate boundary conditions are given in Piomelli *et al.* [61] and Balaras *et al.* [6, 7]. Simple approximate boundary conditions have been widely used in wind engineering and meteorology where the Reynolds numbers are so large that computing the near-wall region of a boundary layer is infeasible. It has been pointed out that this issue is very important to the future application of LES to engineering flows at high Reynolds numbers.

3.4 Formulation for Complex Geometry

LES has gained a lot of attention as a possible tool for analyzing complex flows for which engineering turbulence models fail or for which the time dependent motion is important. However, there are several issues in applying LES to complex flows.

3.4.1 Dynamic Model Localization

The dynamic procedure provides a means of computing model coefficients which are functions of time and space. This is very desirable for complex flows, since it makes sense to have the model adjust to the local dynamics. However, the original derivation of the relations for the model coefficients has some mathematical inconsistencies and difficulties with numerical stability. In Germano's identity the model coefficient occurs in two locations, one inside a filter and one outside. The original formulation simply moved the coefficient outside the filtering operation; however, if the coefficient is a function of space and time, this is inconsistent. Another difficulty is that the relations for the model coefficients involve the ratio of two quantities, and it can happen that the quantity in the denominator may approach zero locally. In the original formulation it was recognized that this ill-conditioned behavior could give rise to a numerical instability, and so an averaging procedure was employed over homogeneous directions to make the model coefficients better behaved. However, even with the averaging procedure it is still possible to obtain eddy viscosities that are large and negative. Therefore, the model coefficients are often clipped to ensure the eddy viscosity or at least the sum of the eddy and molecular viscosities is positive. However, for a complex geometry problem there are no homogeneous directions and so a method must be found to make the local values of the model coefficients well conditioned.

Kim & Menon[48] have used a simple localization procedure based on local spatial averaging. A rigorous approach to localization based on solving a Fredholm integral equation was developed by Ghosal *et al.* [20]. However, because it is complicated and computationally expensive it has not been widely used. A simpler, approximate method was developed by Piomelli & Liu[62] in which the local coefficient inside the filter in Germano's identity is time lagged from the coefficient outside the filter. Also, the denominator in the relation

for the model coefficient is made to be positive definite, which helps with the numerical stability problem. A method which averages in time over Lagrangian particle paths was proposed and used by Meneveau *et al.* [47]. It was also used in the boundary layer calculations of Wu & Squires[85]. Although several localization procedures have been successfully used, it remains to be seen which method is best for use in complex geometry flows.

An issue related to localization is that in uniform flow the dynamic procedure produces an indeterminate form (0/0). Piomelli & Liu[62] state that any spurious values of the model coefficient determined in this manner will have little effect because in a uniform flow the computed strain rates are small giving a negligible eddy viscosity. However, Spyropoulos & Blaisdell[75] found it necessary to set the coefficient to zero in the freestream of their boundary layer simulation.

3.4.2 Filtering in Generalized Coordinates

When the SGS models are formulated they are usually put in terms of Cartesian tensors. Most researchers who have applied LES to problems in curvilinear coordinates have not provided any details on how the LES equations or the models are implemented. An issue to address is how the filter is applied to the equations when there are mesh metrics that arise from the coordinate transformation. This issue has been considered in a series of papers by Jordan & Ragab[30] and Jordan[28, 29]. They recommend filtering the equations after the coordinate transformation is performed, so that the filter is defined in terms of computational coordinates rather than physical coordinates. They also recommend bringing the mesh metric factors outside the filtering process based on the fact that they are smooth functions. These seem to be good recommendations and they have been applied to flow over a cylinder; however, further investigation is needed.

3.4.3 Commutation of Filtering and Differentiation

A more fundamental questions regarding LES was raised by Ghosal & Moin[21] who pointed out that for nonuniform grids the filtering operation does not commute with differentiation as is generally assumed when the LES equations are derived. They were able to devise a filtering method that commutes with differentiation up to second order in the filter width. However, this is a difficulty if one wishes to use high-order finite difference schemes, such as the compact differencing schemes of Lele[38]. A filter which commutes to arbitrary order was developed by van der Ven[77]; however, it is limited to infinite domains with no boundaries. A survey of the issue and an attempt to create discrete filters which commute with discrete derivatives is given by Blaisdell[9]. The above investigations have considered only Cartesian grids (with possible grid stretching) and the issue of commutation of filters for generalized geometries has not even been addressed.

3.5 Formulation for Compressible Flows

One other issue is that for compressible turbulent flows there are several terms which are neglected in the formulation of Erlebacher *et al.* [17]. Moin *et al.* [55] found that these terms were small for isotropic turbulence; however, no investigation of their significance for inhomogeneous flows has been done and this remains an open issue.

References

[1] Adams, N. A., "Direct Numerical Simulation of Turbulent Supersonic Boundary Layer Flow," Paper I-03, Proceedings of the First AFOSR International Conference on Direct Numerical Simulation and Large Eddy Simulation, Louisiana Tech Univ., Ruston, LA, USA, August 4-8, 1997.

[2] AGARD, *Application of Direct and Large Eddy Simulation to Transition and Turbulence*, AGARD-CP-551, Proceedings of the 74th Fluid Dynamics Symposium held at Chania, Crete, Greece, April 1994.

[3] Akselvoll, K. and Moin, P., "Large Eddy Simulation of Turbulent Confined Coannular Jet and Turbulent Flow over a Backward-Facing Step," Report No. TF-63, Dept. Mech. Eng., Stanford Univ., Stanford, CA 94305, 1995.

[4] Akselvoll, K. and Moin, P., "Large-Eddy Simulation of Turbulent Confined Coannular Jets," *J. Fluid Mech.*, Vol. 315, pp. 387-411, 1996.

[5] Bardina, J. H., Ferziger, J. H., and Reynolds, W. C., "Improved turbulence models based on large-eddy simulation of homogeneous, incompressible, turbulent flows," Report TF-19, Department of Mechanical Engineering, Stanford Univ., Stanford, CA, 1983.

[6] Balaras, E., Benocci, C. and Piomelli, U., "Finite-Difference Computations of High Reynolds Number Flows Using the Dynamic Subgrid-Scale Model," *Theoret. Comp. Fluid Dyn.*, Vol. 7, pp. 207-216, 1995.

[7] Balaras, E., Benocci, C. and Piomelli, U., "Two-Layer Approximate Boundary Conditions for Large-Eddy Simulations," *AIAA Journal*, Vol. 34, pp. 1111-1119, June 1996.

[8] Beaudan, P. and Moin, P., "Numerical Experiments on the Flow Past a Circular Cylinder at Subcritical Reynolds Numbers," Report No. TF-62, Dept. Mech. Eng., Stanford Univ., Stanford, CA 94305, 1994.

[9] Blaisdell, G. A., "Commutation of Discrete Filters and Differential Operators for Large-Eddy Simulation," Paper L-07, Proceedings of the First AFOSR International Conference on Direct Numerical Simulation and Large Eddy Simulation, Louisiana Tech Univ., Ruston, LA, USA, August 4-8, 1997.

[10] Childs, R. E. and Reisenthel, P. H., "Simulation study of compressible turbulent boundary layers," AIAA Paper 95-0582, 33rd AIAA Aerospace Sciences Meeting, Reno, NV, 1995.

[11] Choi, H. and Moin, P., "Effects of the Computational Time Step on Numerical Solutions of Turbulent Flow," *J. Comp. Phys.*, Vol. 113, pp. 1-4, 1994.

[12] Choi, H., Moin, P. and Kim, J., "Direct Numerical Simulation of Turbulent Flow Over Riblets," *J. Fluid Mech.*, Vol. 255, pp. 503-539, 1993.

[13] Chu, D. C. and Karniadakis, G. E., "A Direct Numerical Simulation of Laminar and Turbulent Flow Over Riblet-Mounted Surfaces," *J. Fluid Mech.*, Vol. 250, pp. 1-42, 1993.

[14] Chyczewski, T. S. and Long, L. N., "Numerical Prediction of the Noise Produced by a Perfectly Expanded Rectangular Jet," AIAA Paper 96-1730, 2nd AIAA/CEAS Aeroacoustics Conference, May 6-8, 1996, State College, PA.

[15] Coleman, G. N., Kim, J., and Moser, R. D., "A Numerical Study of Turbulent Supersonic Isothermal-Wall Channel Flow," *J. Fluid Mech.*, Vol. 305, pp. 159-183, 1995.

[16] Crawford, C. H., Constantinos, E., Newman, D. and Karniadakis, G. E., "Parallel Benchmarks of Turbulence in Complex Geometries," *Computers & Fluids*, Vol. 25, pp. 677-698, 1996.

[17] Erlebacher, G., Hussaini, M. Y., Speziale, C. G., and Zang, T. A., "Toward the Large-Eddy Simulation of Compressible Turbulent Flows," *J. Fluid Mech.*, Vol. 238, pp. 155-185, 1992.

[18] Ferziger, J. H., "Higher-Level Simulations of Turbulent Flows," Report TF-16, Thermosciences Division, Department of Mechanical Engineering, Stanford Univ., Stanford, California, March 1981.

[19] Germano, M., Piomelli, U., Moin, P., and Cabot, W., "A dynamic subgrid-scale eddy-viscosity model," *Phys. Fluids A*, Vol. 3, pp. 1760–1765, 1991.

[20] Ghosal, S., Lund, T. S., Moin, P., and Akselvoll, K., "A Dynamic Localization Model for Large-Eddy Simulation of Turbulent Flows," *J. Fluid Mech.*, Vol. 286, pp. 229-255, 1995.

[21] Ghosal, S. and Moin, P., "The basic equations for the large eddy simulation of turbulent flow in complex geometry," *J. Comp. Phys.*, Vol. 118, pp. 24–37, 1995.

[22] Ghosal, S., "An analysis of numerical errors in large-eddy simulations of turbulence," *J. Comp. Phys.*, Vol. 125, pp. 187–206, 1996.

[23] Hatay, F. F. and Biringen, S., "Direct numerical simulation of low-Reynolds number supersonic turbulent boundary layers," AIAA Paper 95-0581, 33rd AIAA Aerospace Sciences Meeting, Reno, NV, 1995.

[24] Huang, P. G., Coleman, G. N., and Bradshaw, P., "Compressible Turbulent Channel Flows: DNS Results and Modeling," *J. Fluid Mech.*, Vol. 305, pp. 185-218, 1995.

[25] Huser, A. and Biringen, S., "Direct Numerical Simulation of Turbulent Flow in a Square Duct," *J. Fluid Mech.*, Vol. 257, pp. 65-95, 1993.

[26] Huser, A., Biringen, S., and Hatay, F. F., "Direct Simulation of Turbulent Flow in a Square Duct: Reynolds Stress Budgets," *Phys. Fluids*, Vol. 6, pp. 3144-3152, September 1994.

[27] Jansen, K. E., "Large-Eddy Simulation Using Unstructured Grids," Paper I-10, Proceedings of the First AFOSR International Conference on Direct Numerical Simulation and Large Eddy Simulation, Louisiana Tech Univ., Ruston, LA, USA, August 4-8, 1997.

[28] Jordan, S. A., "On the formulation of the large-eddy simulation for turbulent vortical flows in complex domains," American Society of Mechanical Engineers, Fluids Engineering Division (Publication) FED Volume 238, Number 3, pp. 141-148. Proceedings of the 1996 ASME Fluids Engineering Division Summer Meeting, Part 3, San Diego, CA, July 7-11, 1996.

[29] Jordan, S. A., "Dynamic Subgrid-Scale Modeling in Generalized Curvilinear Coordinates," Paper L-03, Proceedings of the First AFOSR International Conference on Direct Numerical Simulation and Large Eddy Simulation, Louisiana Tech Univ., Ruston, LA, USA, August 4-8, 1997.

[30] Jordan, S. A. and Ragab, S. A., "A Large-Eddy Simulation of the Near Wake of a Circular Cylinder," in *Turbulence in Complex Flows*, American Society of Mechanical Engineers, Fluids Engineering Division (Publication) FED-Vol. 203, pp. 1-9, Proceedings of the 1994 International Mechanical Engineering Congress and Exposition, Chicago, IL, November 6-11, 1994.

[31] Karniadakis, G. E. and Orszag, S. A., "Nodes, Modes And Flow Codes," *Physics Today*, Vol. 46, No. 3, p. 34, March 1993.

[32] Kim, J., Moin, P., and Moser, R., "Turbulence Statistics in Fully Developed Channel Flow at Low Reynolds Number," *J. Fluid Mech.*, Vol. 177, pp. 133-166, 1987.

[33] Komminaho, J., Lundbladh, A., and Johansson, A. V., "Very large structures in plane turbulent Couette flow," *J. Fluid Mech.*, Vol. 320, p. 259, 1996.

[34] Kravchenko, A. G. and Moin, P., "On the Effect of Numerical Errors in Large Eddy Simulations of Turbulent Flows," *J. Comp. Phys.*, Vol. 131, pp. 310-322, 1997.

[35] Kravchenko, A. G., Moin, P., and Moser, R., "Zonal Embedded Grids for Numerical Simulations of Wall-Bounded Turbulent Flows," *J. Comp. Phys.*, Vol. 127, pp. 412-423, 1996.

[36] Le, H. and Moin, P., "Direct Numerical Simulation of Turbulent Flow over a Backward-Facing Step," Report No. TF-58, Dept. Mech. Eng., Stanford Univ., Stanford, CA 94305, December 1994.

[37] Le, H., Moin, P., and Kim, J., "Direct Numerical Simulation of Turbulent Flow over a Backward-Facing Step," *J. Fluid Mech.*, Vol. 330, p. 349, 1997.

[38] Lele, S. K., "Compact finite difference schemes with spectral-like resolution," *J. Comp. Phys.*, Vol. 103, pp. 16-42, 1992.

[39] Lesieur, M. and Comte, P., "Large-Eddy Simulations of Compressible Turbulent Flows," in *Turbulence in Compressible Flows*, AGARD-R-819, pp. 4–1-39, Proceedings of the AGARD FDP Special Course on "Turbulence in Compressible Flows," held at the von Kármán Institute for Fluid Dynamics (VKI) in Rhode-Saint-Genèse, Belgium, June 2-6, 1997, and in Newport News, Virginia, USA, October 20-24, 1997.

[40] Lesieur, M. and Métais, O., "New Trends in Large-Eddy Simulations of Turbulence," *Annu. Rev. Fluid Mech.*, Vol. 28, pp. 45-82, 1996.

[41] Li, G. and Dalton, C., "Computation of Oscillating Flow Past a Circular Cylinder Using LES/DSGS," Paper L-14, Proceedings of the First AFOSR International Conference on Direct Numerical Simulation and Large Eddy Simulation, Louisiana Tech Univ., Ruston, LA, USA, August 4-8, 1997.

[42] Lilly, D. K., "A Proposed Modification of the Germano Subgrid-Scale Closure Method," *Phys. Fluids A*, Vol. 4, pp. 633-635, March 1992.

[43] Liu, J., Piomelli, U. and Spalart, P. R., "Interaction Between a Spatially Growing Turbulent Boundary Layer and Embedded Streamwise Vortices," *J. Fluid Mech.*, Vol. 326, pp. 151-179, 1996.

[44] Lund, T. S., Wu, X., and Squires, K. D., "Generation of turbulent inflow data for spatially-developing boundary layer simulations", accepted for publication in J. Comp. Physics, 1997.

[45] Mankbadi, R., "Computational Aero-Acoustics in Propulsion Systems," Paper I-07, Proceedings of the First AFOSR International Conference on Direct Numerical Simulation and Large Eddy Simulation, Louisiana Tech Univ., Ruston, LA, USA, August 4-8, 1997.

[46] Mansour, N. N., Kim, J., and Moin, P., "Reynolds-Stress and Dissipation-Rate Budgets in a Turbulent Channel Flow," *J. Fluid Mech.*, Vol. 194, pp. 15-44, 1988.

[47] Meneveau, C., Lund, T. S., and Cabot, W. H., "A Lagrangian Dynamic Subgrid-Scale Model of Turbulence," *J. Fluid Mech.*, Vol. 319, p. 353, 1996.

[48] Kim, W-W. and Menon, S., "A New Dynamic One-Equation Subgrid-Scale Model for Large Eddy Simulations," AIAA paper no. 95-0356, presented at the 33rd Aerospace Sciences Metting, Reno, NV, January 1995.

[49] Mittal, R., "Progress on LES of Flow Past a Circular Cylinder," *Annual Research Briefs*, Center for Turbulence Research, Stanford Univ., Stanford, CA 94305, pp. 233-241, December 1996.

[50] Mittal, R., "Direct Numerical Simulation of Flow Past Spheres and Spheroids," Paper D-02, Proceedings of the First AFOSR International Conference on Direct Numerical Simulation and Large Eddy Simulation, Louisiana Tech Univ., Ruston, LA, USA, August 4-8, 1997.

[51] Mittal, R. and Moin, P., "Suitability of Upwind-Biased Finite Difference Schemes for Large-Eddy Simulation of Turbulent Flows," *AIAA J.*, Vol. 35, pp. 1415-1417, August 1997.

[52] Moin, P., "Progress in large eddy simulation of turbulent flows," AIAA Paper 97-0749, 35th Aerospace Sciences Meeting, Reno, NV, January 6-10, 1997.

[53] Moin, P. and Kim, J., "Numerical Investigation of Turbulent Channel Flow," *J. Fluid Mech.*, Vol. 118, pp. 341-377, 1982.

[54] Moin, P. and Kim, J., "Tackling Turbulence with Supercomputers," *Scientific American*, Vol. 276, pp. 62-68, January 1997.

[55] Moin, P., Squires, K., Cabot, W., and Lee, S., "A Dynamic Subgrid-Scale Model for Compressible Turbulence and Scalar Transport," *Phys. Fluids A*, Vol. 3, pp. 2746-2757, November 1991.

[56] Moser, R. D. and Moin, P., "The Effect of Curvature in Wall-Bounded Turbulent Flows," *J. Fluid Mech.*, Vol. 175, pp. 479-510, 1987.

[57] Murakami, S., "Overview of Turbulence Models Applied in CWE-1997," in *2 EACWE*, Vol. 1, pp. 3-24, Proceedings of the 2nd European & African Conference on Wind Engineering held at Genova, Italy, June 22-26, 1997, ed. G. Solari, SGE Ditoriali Padova.

[58] Piomelli, U., "High Reynolds Number Calculations Using the Dynamic Subgrid-Scale Stress Model," *Phys. Fluids A*, Vol. 5, pp. 1484-1490, June 1993.

[59] Piomelli, U., "Large-Eddy Simulation of Turbulent Flows," TAM Report No. 767, Department of Theoretical and Applied Mechanics, Univ. of Illinois Champaign-Urbana, September 1994.

[60] Piomelli, U., "Large-Eddy and Direct Simulation of Turbulent Flows," Lecture notes for the course *Introduction to Turbulence*, Von Karman Institute, Rhode Saint Genese, Belgium, March 17-21, 1997.

[61] Piomelli, U., Ferziger, J., and Moin, P., "New Approximate Boundary Conditions for Large Eddy Simulations of Wall-Bounded Flows," *Phys. Fluids A*, Vol. 1, pp. 1061-1068, June 1989.

[62] Piomelli, U. and Liu, J., "Large-Eddy Simulation of Rotating Channel Flows Using a Localized Dynamic Model," *Phys. Fluids*, Vol. 7, pp. 839-848, April 1995.

[63] Piomelli, U., Moin, P., and Ferziger, J., "Model Consistency in Large Eddy Simulation of Turbulent Channel Flow," *Phys. Fluids*, Vol. 31, pp. 1884-1891, July 1988.

[64] Piomelli, U. and Zang, T. A., "Large-Eddy Simulation of Transitional Channel Flow," *Comp. Phys. Comm.*, Vol. 65, pp. 224-230, 1991.

[65] Qin, J. H., "Numerical Simulation of a Turbulent Axial Vortex," Prelim Report and Thesis Proposal, School of Aero. & Astro., Purdue Univ., West Lafayette, IN 47907, April 1997.

[66] Ragab, S. A. and Piomelli, U., (eds.), *Engineering Applications of Large Eddy Simulations*, The American Society of Mechanical Engineers, Fluids Engineering Division Publication FED-Vol. 162, Proceedings of the Fluids Engineering Conference, Washington, D.C., June 20-24, 1993.

[67] Rai, M. M., Gatski, T. B., and Erlebacher, G., "Direct simulation of spatially evolving compressible turbulent boundary layers," AIAA Paper 95-0583, 33rd AIAA Aerospace Sciences Meeting, Reno, NV, 1995.

[68] Reynolds, W. C., "The Potential and Limitations of Direct and Large Eddy Simulations," in *Whither Turbulence? Turbulence at the Crossroads*, Lecture Notes in Physics, Vol. 357, pp. 313-343 (also see comments following this paper), Springer-Verlag, 1990. Proceedings of a conference held at Cornell Univ., Ithaca, NY, USA, March 22-24, 1989.

[69] Ristorcelli, J. R. and Blaisdell, G. A., "Consistent Initial Conditions for the DNS of Compressible Turbulence," *Phys. Fluids*, Vol. 9, pp. 4-6, January 1997.

[70] Rogallo, R. S. and Moin, P., "Numerical Simulation of Turbulent Flows," *Ann. Rev. Fluid Mech.*, Vol. 16, pp. 99-137, 1984.

[71] Smagorinski, J. S., "General circulation experiments with the primitive equations. I. The basic experiment," *Monthly Weather Review*, Vol. 91, pp. 99-164, 1963.

[72] Spalart, P. R., "Direct simulation of a turbulent boundary layer up to $Re_\theta = 1410$," *J. Fluid Mech.*, Vol. 187, p. 61, 1986.

[73] Spyropoulos, E. T., "On Dynamic Subgrid-Scale Modeling for Large-Eddy Simulation of Compressible Turbulent Flows," Ph.D. Thesis, School of Aeronautics and Astronautics, Purdue Univ., West Lafayette, Indiana, December 1996.

[74] Spyropoulos, E. T. and Blaisdell, G. A., "Evaluation of the Dynamic Model for Simulations of Compressible Decaying Isotropic Turbulence," *AIAA J.*, Vol. 34, pp. 990-998, May 1996.

[75] Spyropoulos, E. T. and Blaisdell, G. A., "Large-eddy simulation of a spatially evolving compressible boundary layer flow," AIAA Paper 97-0429, 35th Aerospace Sciences Meeting, Reno, NV, January 6-10, 1997. Submitted to the AIAA Journal.

[76] Sreedhar, M. and Ragab, Saad, "Large Eddy Simulation of Longitudinal Stationary Vortices," *Phys. Fluids*, Vol. 6, pp. 2501-2514, July 1994.

[77] van der Ven, H., "A family of large eddy simulation (LES) filters with non-uniform filter widths," *Phys. Fluids*, Vol. 7, pp. 1171–1172, 1995.

[78] Voke, P. R., Gao, S., and Leslie, D., "Large-eddy Simulations of Plane Impinging Jets," *Int. J. Num. Meth. Eng*, Vol. 38, p. 489, 1995.

[79] Vreman, B., Geurts, B., and Kuerten, H., "On the Formulation of the Dynamic Mixed Subgrid-Scale Model," *Phys. Fluids*, Vol. 6, pp. 4057-4059, December 1994.

[80] Vreman, A. W., Sandham, N. D., and Luo, K. H., "Compressible mixing layer growth rate and turbulence characteristics," *J. Fluid Mech.*, Vol. 320, p. 235, 1996.

[81] Vreman, B. (A. W.), "Direct and Large-Eddy Simulation of the Compressible Turbulent Mixing Layer," Ph.D. Thesis, Dept. Appl. Math., Univ. Twente, The Netherlands, December 1995.

[82] Wang, Q., Morris, P. J., and Long, L.N., "Supersonic Jet Noise Prediction Using Large Eddy Simulation of Parallel Computers," Paper N-09, Proceedings of the First AFOSR International Conference on Direct Numerical Simulation and Large Eddy Simulation, Louisiana Tech Univ., Ruston, LA, USA, August 4-8, 1997.

[83] Wang, W.-P. and Pletcher, R. H., "Large Eddy Simulation of a Low Mach Number Channel Flow with Property Variations," Proceedings of the Tenth Symposium on Turbulent Shear Flows, Penn State Univ., August 1995.

[84] Wu, X. and Squires, K. D., "Large eddy simulation of the turbulent flow over a bump", Boundary Layers and Free Shear Flows, edited by J.F. Donovan and M.W. Plesniak, ASME-FED 237, pp. 583-588, 1996.

[85] Wu, X. and Squires, K. D., "Large eddy simulation of an equilibrium three-dimensional turbulent boundary layer", AIAA Journal, 35(1), pp. 67-74, 1997.

[86] Xia, M. and Karniadakis, G. E., "The Spectrum of the Turbulent Near-Wake: A Comparison of DNS and LES," Paper I-11, Proceedings of the First AFOSR International Conference on Direct Numerical Simulation and Large Eddy Simulation, Louisiana Tech Univ., Ruston, LA, USA, August 4-8, 1997.

[87] Zang, Y., Street, R., and Koseff, J. R., "A Dynamic Mixed Subgrid-Scale Model and Its Application to Turbulent Recirculating Flows," *Phys. Fluids A*, Vol. 5, pp. 3186-3196, December 1993.

MDL Texture Segmentation of Compressed Images

Octavia I. Camps
Assistant Professor
Department of Computer Science and Electrical Engineering

The Pennsylvania State University
University Park, PA 16802

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

And

Wright Laboratory

August 1997

MDL Texture Segmentation of Compressed Images

Octavia I. Camps
Assistant Professor
Department of Electrical Engineering
Department of Computer Science and Engineering
The Pennsylvania State University
University Park, PA 16802,
email: camps@whale.ee.psu.edu

Abstract

The design of efficient, real-time solutions for the problem of segmenting images into "homogeneous" regions is critical for the success of computer vision applications such as automatic target recognition, medical imaging and industrial automation. When images present texture regions, image segmentation becomes one of the most challenging problems in computer vision. In this research we developed a texture segmentation technique based on the minimum description length principle applied to multiband images, that works *directly on semi-compressed data* obtained using a wavelet decomposition. The proposed algorithm saves time and space by operating on compressed data. Furthermore, since it is based on the MDL principle, it does not require ad hoc parameters and it is efficiently implemented using an incremental approach.

Contents

1 Introduction **9–4**

2 Texture Segmentation of Compressed Images **9–4**

2.1 Image Compression using Wavelets 9–5

2.1.1 Image Pyramids using Wavelets from filters 9–5

2.2 MDL based Segmentation 9–6

2.2.1 Encoding the boundaries 9–8

2.2.2 Encoding the parameters 9–8

2.2.3 Encoding the residuals 9–9

2.2.4 Incremental Implementation 9–9

3 Examples **9–11**

4 Conclusions and Future Work **9–11**

1 Introduction

The design of efficient, real-time solutions for the problem of segmenting images into "homogeneous" regions is critical for the success of computer vision applications such as automatic target recognition, medical imaging and industrial automation. One of the most challenging aspects of this problem is that most natural scenes present texture. Although texture is an important characteristic for the analysis of images, it has not been formally defined to this day, despite its importance and ubiquity. As a result, most techniques using texture are for the most part *ad hoc*.

In an effort to understand texture, they have been classified into one of four classes [7]: *strongly ordered*, such as a brick wall, *weakly oriented*, such as wood grain, *disordered*, such as grass, and *compositional*, such as a human-made net. A better understood, related problem to texture segmentation, is automatic texture classification. Most algorithms for automatic texture classification, rely on extracting quantitative measures derived from co-occurrence matrices of a given region. However, most of these quantitative measures are not reliable, unless they are computed within a *single* texture region and therefore cannot be directly applied to texture segmentation. Thus, most existing texture segmentation techniques have been designed as extensions of border-based or region-based algorithms to segment *homogeneous* gray-level regions [2, 8, 3]. Some of the problems that these approaches must face are 1) the selection of which features should be used to describe the texture; 2) the fact that texture is only well defined at a region level, but not at the pixel level; and 3) the dimensionality of the feature space is in general large.

Recently, multiresolution approaches [11, 10] have been proposed to address the above problems, since the resolution at which an image is observed changes the scale at which the texture is perceived. In this research we explored a new multiresolution approach combining a pyramid-structure obtained using a wavelet image decomposition with an minimum description length (MDL) based segmentation. The proposed algorithm provides several advantages over previous approaches. In particular, it works directly on semi-compressed data, thus significantly reducing time and space requirements. Furthermore, since it is based on the MDL principle, it does not depend on *ad hoc* parameters. Finally, it can be efficiently implemented.

This report is organized as follows. In section 2 we present the new algorithm. In section 2.1 we give a review of image compression using wavelets. In section 2.2 we describe the MDL-based segmentation algorithm. The algorithm is illustrated with two examples in section 3. Finally, in section 4 we summarize our results an indicate directions for future research.

2 Texture Segmentation of Compressed Images

In this research effort we explored a hierarchical approach to texture segmentation using semi-compressed data. Image compression is concerned with minimizing the number of bits required to represent images. Traditional applications of image compression are storage and transmission of images. It is only recently, that attention has been dedicated to its use for the development of fast algorithms that work directly on compressed data. The obvious advantages of using compressed data are that decompression of the data is avoided and that the algorithms process a reduced amount of data. A less obvious advantage is that most compression techniques are based on

transformations into the frequency domain that make frequency information, such as texture, more explicit.

2.1 Image Compression using Wavelets

Most image compression techniques fall into one of two main categories, predictive coding and transform coding, or a combination of these two. Predictive coding techniques operate directly on pixels to exploit data redundancy. Transform coding techniques use a reversible, linear transform to pack information into semi-compressed data as a small number of samples which are then fully-compressed using quantizing and coding. When the channels of a filter are used to perform coding, the technique is called *subband coding.* An important property of subband coding is that in the transform domain the transformed coefficients are not correlated, thus resulting on "decorrelation" of the original data. Although the only transform that achieves exact decorrelation or diagonalization is the Karhunen-Loeve transform, several other transforms achieve near diagonalization, such as the discrete cosine transform and the wavelet transform.

A simple and yet powerful compression technique is based on using a pyramid image representation. From an original image, a coarse approximation is derived, for example by using a lowpass filter and a downsampler. The coarse image and the predicted error (the difference between the original image and the upsampled and filtered coarse image) can then be compressed. Reconstruction is then accomplished by adding the coarse image and the predicted error. This process can then be iterated over an over, forming a pyramid. One way of building this pyramid is to use a Wavelet decomposition. The wavelet transform has the advantages that it provides good localization both in frequency and space domain and that its window size changes with the frequency content of the image. Furthermore, the wavelet transform provides orientation sensitive information at various resolutions, that is specially well suited for texture segmentation. Thus, we chose this compression technique as the basis for our segmentation algorithm. The pyramid obtained this way is then segmented using a multi-band multi-resolution segmentation technique based on the minimum description length. The pyramid construction as well as the segmentation algorithm used are described in more detail next.

2.1.1 Image Pyramids using Wavelets from filters

To implement the wavelet transform we used the same set of Quadrature Mirror Filters (QMF) that have been successfully used by Franques and coworkers [4]. These QMF filters have eight taps as follows:

$$
\begin{aligned}
h_o(1) &= \quad 1.2075224e-02 \quad &= h_o(8) \\
h_o(2) &= -9.3241559e-02 \quad &= h_o(7) \\
h_o(3) &= \quad 9.1663910e-02 \quad &= h_o(6) \\
h_o(4) &= \quad 6.9660920e-01 \quad &= h_o(5)
\end{aligned}
$$

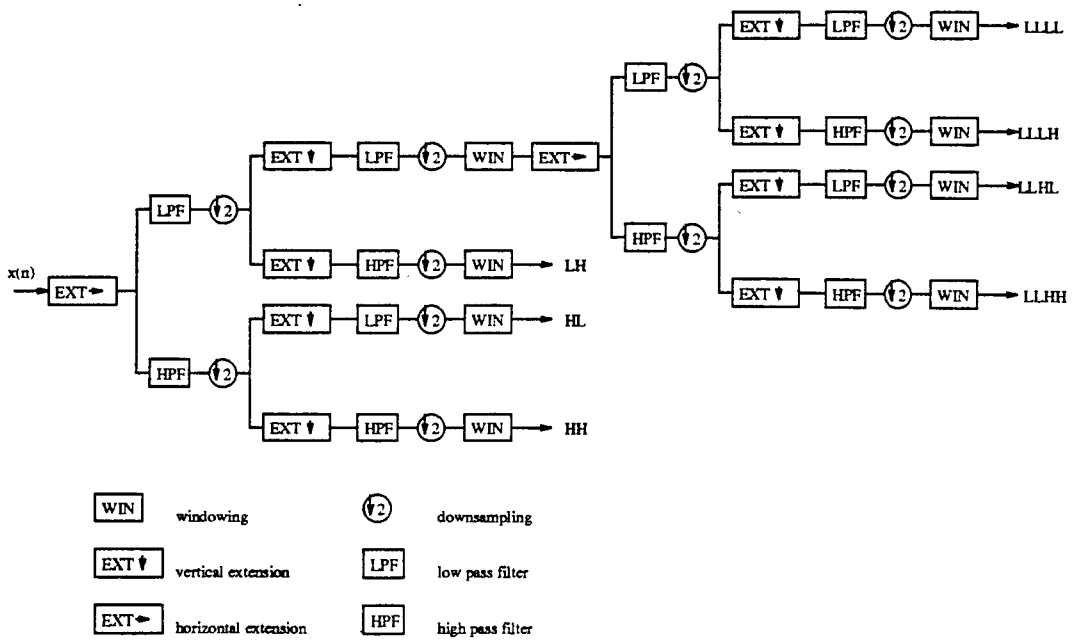


Figure 1: Decomposition of an image using a two-level dyadic tree.

and are such that

$$\sum_k h_o(k) = \sqrt{2}$$

$$\sum_k h_o(k)h_o(k+2l) = \delta(l)$$

$$h_1(k) = \sum_k (-1)^k h_o(-k+1)$$

The octave-band tree split was then obtained by splitting the lower half of the spectrum into two equal bands in the horizontal and vertical directions, at each level of the tree, as shown in Figure 1. The image was symmetrically extended before filtering, to reduce distortions due to boundary effects. This process generates at each level a coarse approximate of the input image and three orientation selective detail images, which are very important for texture segmentation. Figure 2 shows a three-level hierarchical decomposition of an image.

2.2 MDL based Segmentation

The Minimum Description Length (MDL) principle is based on Occam's Razor, the principle which says that one should prefer the simpler of two theories explaining some data, if everything else is being equal. For MDL, Occam's Razor is applied in a coding sense, by fitting models to the given data, encoding the model parameters and the data using these models, and selecting the model

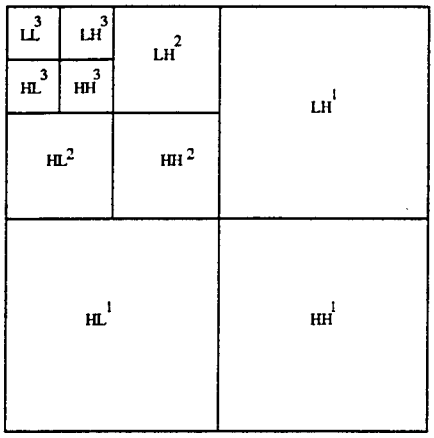


Figure 2: Wavelet representation of an image using three levels of detail and coarse images.

that results in the smallest code length. This approach exploits the tradeoff existing between the complexity of a model and how well it fits the data.

The MDL principle was first used for image segmentation by Leclerc [6], and more recently by Kanungo et al [5] and Zhu and Yuille [12]. In this section we summarize the segmentation algorithm proposed in [5], since it is fast and suitable to process the multiband compressed images obtained using the wavelet decomposition. The main advantages of this approach are that it guarantees closed regions, it does not require the use of arbitrary parameters, thus providing consistent results, and it can be efficiently implemented.

The idea behind the algorithm is to optimize a cost function representing the length of encoding a segmentation of the given image, into a set of non-overlapping regions that are homogeneous in a statistical sense. In particular, the algorithm encodes an image segmentation as a collection of regions modeled as polynomial surfaces of variable degree, perturbed by zero mean Gaussian noise and whose boundaries are described using a chain code representation. Thus, the algorithm will select the image regions, as well as the degree and coefficients of polynomials that best fit the given data in each region.

Formally, consider an image with d bands, let $\Omega = \{\omega_j\}$ denote the image segmentation into regions $\{\omega_j\}$ and let Y represent the image data and Y_j represent the image data within region ω_j. Further, assume that the image comes from a stochastic process that can be characterized as polynomial gray scale surfaces of unknown degree plus Gaussian noise described by a vector of parameters $\beta = \{\beta_j\}$. Then, the MDL objective function to optimize is given by:

$$L(Y, \Omega, \beta) = L(\Omega) + L(\beta|\Omega) + L(Y|\Omega, \beta). \tag{1}$$

where the first term is the length of encoding the region boundaries, the second term is the length of encoding the parameters and the last term is the length of encoding the residuals. Each of these terms are described in more detail next.

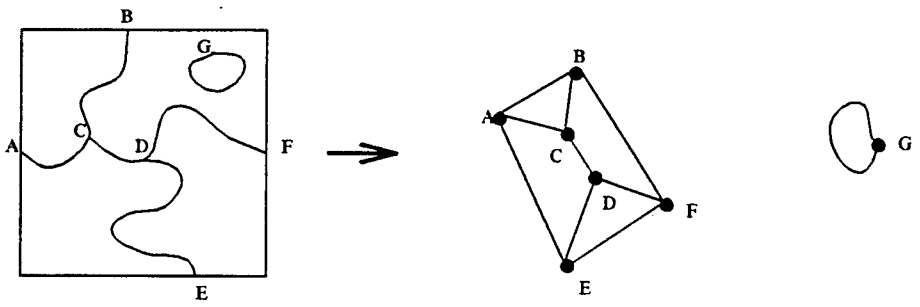


Figure 3: Graph representation of an image segmentation.

2.2.1 Encoding the boundaries

The region boundaries are encoded by representing the given segmentation using a graph where the nodes correspond to boundary intersections and the links correspond to the boundary branches between the intersections, as shown in Figure 3. In turn, this graph can be described in terms of its number of connected components, a reference node for each component, followed by the number of branches out of that node, followed by the lengths of the boundary branches and chain codes representing the path of the boundaries along the pixel grid.

Thus, assuming that at each point the number of possible directions is 3 (i.e. the number of adjacent grid points, excluding the current one), and neglecting the cost of encoding the number of connected components and the reference points, the first term of the encoding cost can be approximated by [9]:

$$L(\Omega) = \sum_i (l_i \log 3 + \log^*(l_i) + \log(2.865064))$$

where l_i is the length of the boundary i and $\log^*(x) = \log x + \log\log x + \log\log\log x + \ldots$ up to all positive terms.

2.2.2 Encoding the parameters

The second cost term, $L(\beta|\Omega)$, encodes the parameters describing the regions and it can be expressed using Rissanen's [9] expression for optimal-precision analysis that says that K independent real-valued parameters characterizing n data points can be encoded using $(K/2)\log n$ bits. Thus,

$$L(\beta|\Omega) = \frac{1}{2}\sum_j K_{\beta_j} \log n_j$$

where n_j is the number of pixels in region j, and K_{β_j} is the number of free parameters describing region j. K_{β_j} is a function of the number of bands d and the number of polynomial coefficients per band in region j, m_j:

$$K_{\beta_j} = \frac{d(d+1)}{2} + dm_j$$

$$m_j = \frac{1}{2}(g_j + 2)(g_j + 1)$$

where g_j is the degree of the polynomial in region j.

2.2.3 Encoding the residuals

The first two terms of the cost correspond to the length of describing the boundaries of the regions and the ideal polynomial surfaces and the noise parameters. The third term, on the other hand, corresponds to the length of encoding the data, given the model. In another words, the third cost term $L(Y|\Omega,\beta)$ models the *residuals* between the ideal polynomial surface and the actual data.

$L(Y|\Omega,\beta)$ can be written using Shannon's theorems [1] as:

$$L(Y|\Omega,\beta) = -\log p(Y|\Omega,\beta)$$

Assuming that the conditional probability distribution can be written as the product of the individual probability distributions for all the image regions in Ω, which in turn can be written as the product of the individual probability distributions for all the pixel values in each region, we have:

$$L(Y|\Omega,\beta) == -\sum_j \log p(Y_j|\beta_j) = -\sum_j \sum_i \log p(y_i|\beta_j)$$

Since the individual pixel conditional distributions are Gaussian distributions, such that:

$$p(y_i|\beta_j) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_j|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}[y_i - \mu_j(x_i)]^t \Sigma_j [y_i - \mu_j(x_i)]\right\}$$

where μ_j is a point on a polynomial surface, we have:

$$p(Y_j|\beta_j) = \frac{1}{(2\pi)^{\frac{dn_j}{2}}|\Sigma_j|^{\frac{n_j}{2}}} \exp\left[-\frac{n_j}{2} trace\left\{\Sigma_j^{-1} S_j\right\}\right]$$

where S_j is the sample covariance matrix in region j.

Furthermore, using the fact that the distributions are Gaussian and that the maximum likehood estimate for the covariance matrix Σ_j is S_j we have that

$$trace(\Sigma_j^{-1} S_j) = d$$

and thus

$$p(Y_j|\beta_j) = \frac{1}{(2\pi)^{\frac{dn_j}{2}}|\Sigma_j|^{\frac{n_j}{2}}} \exp\left[-\frac{n_j}{2} d\right]$$

2.2.4 Incremental Implementation

Using the results described above, the objective function to be minimized is given by:

$$L(Y,\Omega,\beta) = \sum_i (l_i \log 3 + L^o(l_i)) + \sum_j \frac{1}{2} K_{\beta_j} \log n_j + \sum_j \frac{1}{2} n_j [d \log 2\pi + \log|\Sigma_j| + d]$$

Unfortunately, the minimum of this objective function cannot be found analytically. Thus, we use a steepest decent approach that searches for local minimums. This is done by starting with

an initial segmentation of small regions (1 or 2 pixels) and iteratively merging the two adjacent regions that decrease the objective function the most.

Let ω_t and ω_v be two adjacent regions. If these two regions are merged the overall objective cost change δ_{tv} is given by:

$$\begin{aligned}\delta_{tv} &= l_{tv}\log 3 + \log l_{tv} \\ &+ \frac{1}{2}[n_t \log|S_t| + n_v \log|S_v| - (n_t+n_v)\log|S_{tv}|] \\ &+ \frac{1}{2}[K_{\beta_t}\log n_t + K_{\beta_v}\log n_v - K_{beta_{tv}}\log(n_t+n_v)]\end{aligned}$$

where S_tv is the sample covariance matrix of the combined region $\omega_t \cup \omega_v$. Thus, computing δ_{tv} requires estimating the covariance of the merged region. This can be an expensive operation if the size of the regions is large. Furthermore, in order to decide which two regions need to be merged next, the change in codelength must be computed for every pair of adjacent regions. Fortunately, this computation can be done incrementally as follows.

Let Y_t be an $n_t \times 1$ column vector with the gray scale pixel values in part ω_t and Y_v be an $n_v \times 1$ column vector with the gray scale pixel values of part ω_v. Let g be the order of the polynomials used to fit the parts, and $m = (g+1)(g+2)/2$ be the number of polynomial coefficients. Let Φ_t and Φ_v be an $n_t \times m$ and an $n_v \times m$ matrix of m basis functions spanning the polynomials, evaluated at each of the n_t and n_v pixels – i.e. products of powers of pixel row and column coordinates – respectively. Finally, let Θ_t and Θ_v be two $m \times 1$ column vectors with the *optimal* regression coefficients for ω_t and ω_v, respectively. Using these definitions, we have:

$$Y_i = \Phi_i\Theta_i + \Psi_i \; i = t, v$$

where Ψ_t and Ψ_v are vectors of zero mean Gaussian noise with covariance $\sigma^2 I$, and Θ_t and Θ_v are estimated by minimizing the expected fitting error:

$$\epsilon_i = \|Y_i - \Phi_i\Theta_i\| \; i = t, v$$

The solution to this problem is given by:

$$\hat{\Theta}_i = [\Phi_i^t\Phi_i]^{-1}\Phi_i^t Y_i \; i = t, v$$

and

$$n_i S_i = Y_i^t Y_i - \hat{\Theta}_i^t[\Phi_i^t Y_i] \; i = t, v$$

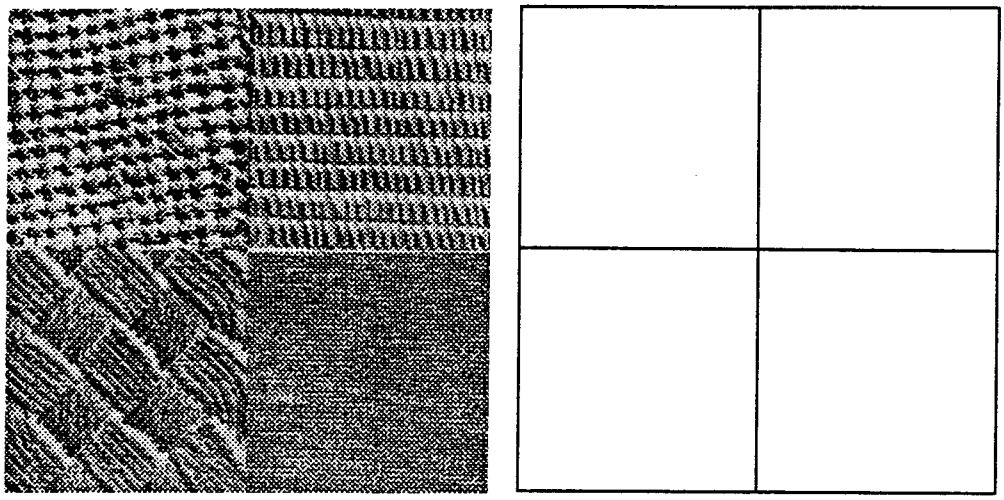
Then, $\hat{\Theta}_{tv}$ and S_{tv} can be incrementally computed by using the fact that

$$Y_{tv}^t Y_{tv} = Y_t^t Y_t + Y_v^t Y_v$$

and

$$\Phi_{tv}^t Y_{tv} = \Phi_t^t Y_t + \Phi_v^t Y_v$$

A further advantage of this approach is that the matrices used in this computation are all of fixed dimensions, regardless of the size of the regions involved.



(a) (b)

Figure 4: Example 1. (a) Original 256×256 image composed of four textures. (b) Ideal segmentation of (a).

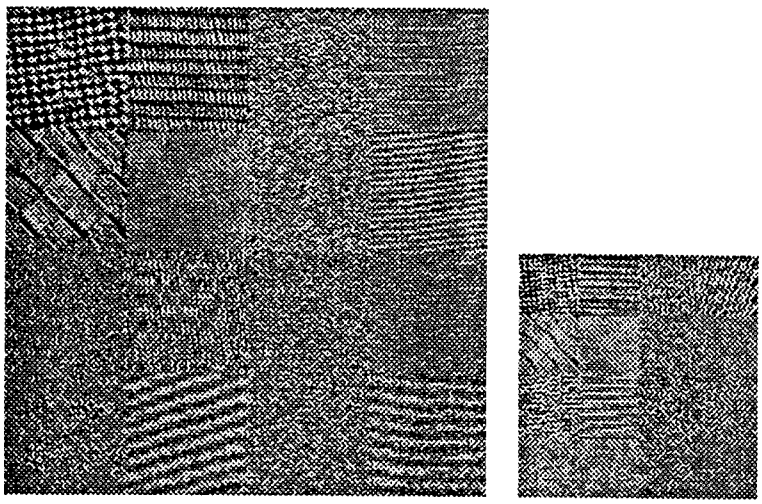
3 Examples

Next we show two examples of texture segmentation using the proposed algorithm. First, we show the results obtained segmenting an image artificially made by mosaicing four textures. This example is important, since the true segmentation is known by construction, and can be objectively compared with the obtained result. Figure 4(a) shows the original image, while Figure 4(b) shows the true segmentation. Figure 5 shows its wavelet decomposition. The result of segmenting the lowest resolution image (64×64) using its four bands is shown in Figure 6. It is seen that the four regions are correctly segmented except for a few small regions.

The second example shows the segmentation of a ladar image of a real vehicle among heavy clutter. Figures 7(a) and (b) show the intensity and range images, respectively. Figures 8 and 9 show their wavelet decomposition. Figures 10, 11 and 12 show the results obtained segmenting the lowest resolution images using the four bands of the intensity image, the four bands of the range image, and using all eight bands respectively. It is seen that the best segmentation is obtained when all bands are used. In this case different parts of the target truck, such as its roof, hood, wheels, side door, cargo bed, etc. are all differentiated. Also the different textures of the background and floor are segmented.

4 Conclusions and Future Work

An algorithm for texture image segmentation, combining a wavelet image compression technique with an MDL based segmentation approach was presented. The algorithm operates on semi-compressed data, thus it significantly reduces the time and memory requirements for image seg-



(a) (b)

Figure 5: Wavelet decomposition for Example 1. (a) 128×128. (b) 64×64.

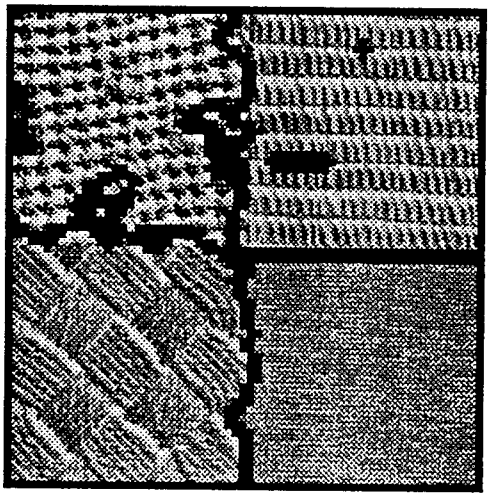
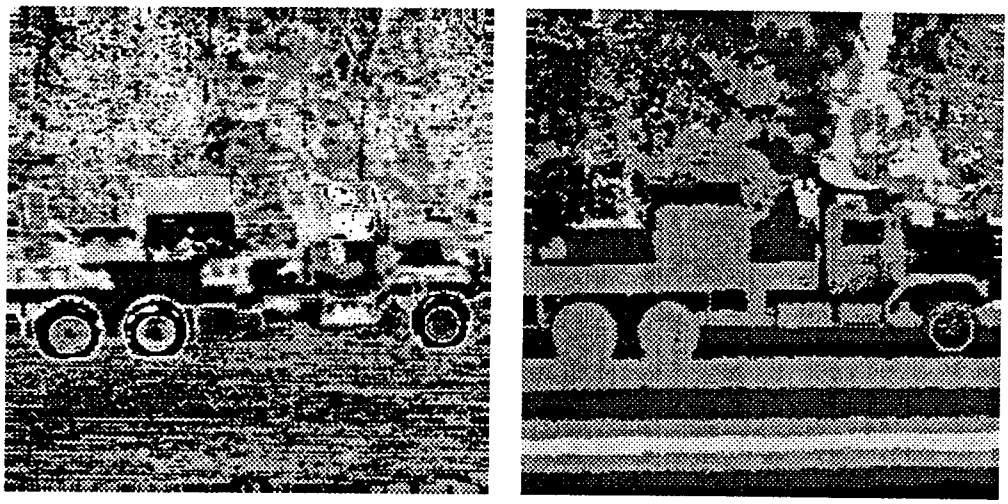
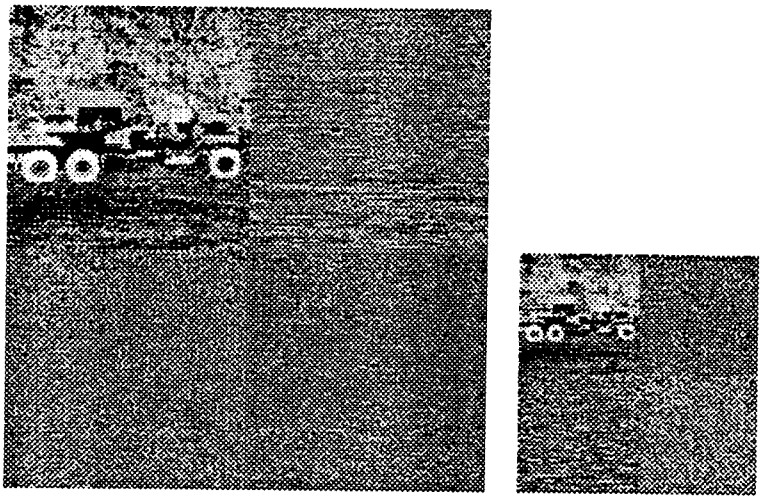


Figure 6: Segmentation result for Example 1.



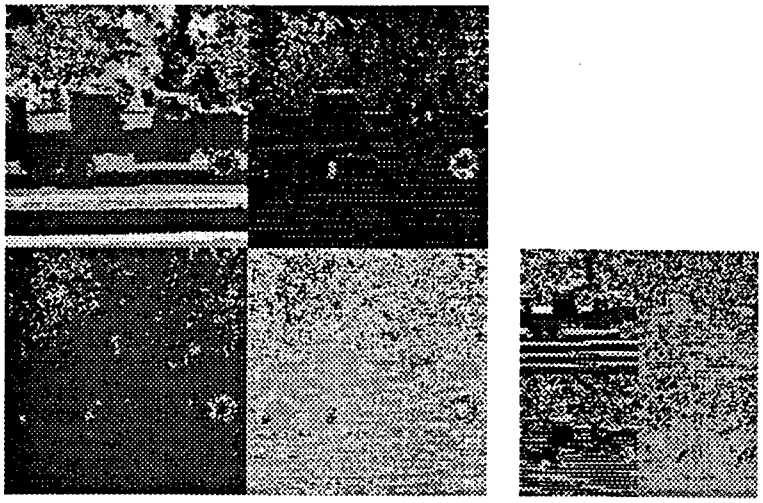
(a) (b)

Figure 7: Example 2. (a) Original intensity image. (b) Original range image.



(a) (b)

Figure 8: Wavelet decomposition for Example 2, intensity data. (a) 128×128. (b) 64×64.



(a) (b)

Figure 9: Wavelet decomposition for Example 2, range data. (a) 128×128. (b) 64×64.

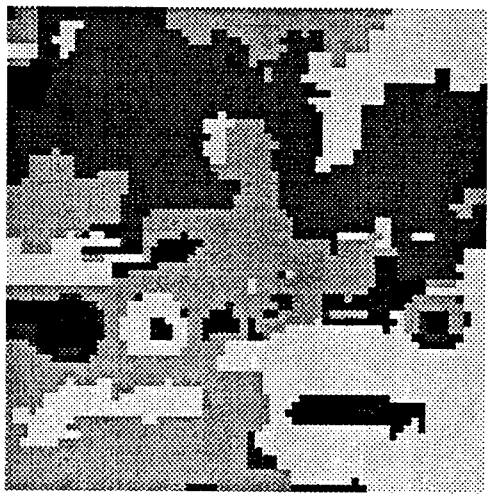


Figure 10: Segmentation result for Example 1, using intensity bands.

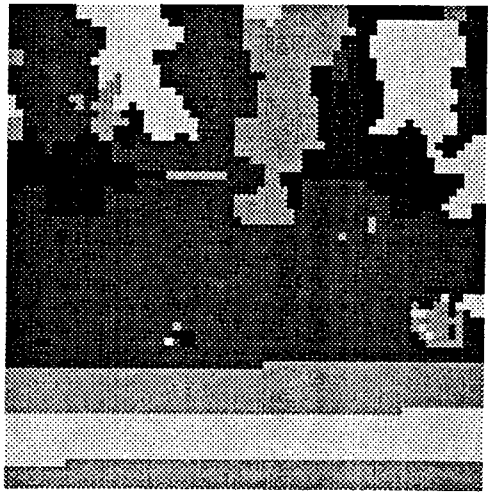


Figure 11: Segmentation result for Example 1, using range bands.

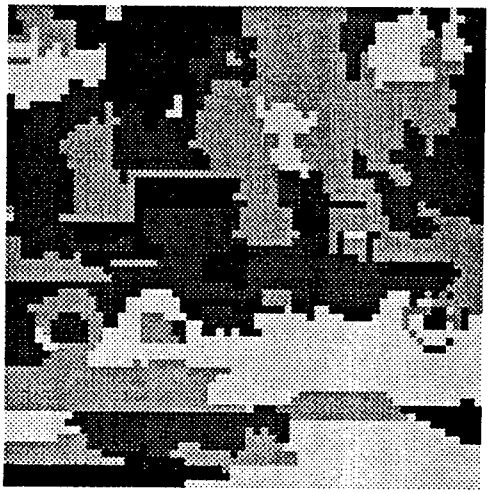


Figure 12: Segmentation result for Example 1, using intensity and range bands.

mentation. Furthermore, the multibands obtained using a wavelet decomposition present selective orientation data that is useful to segment texture. Finally, using an MDL approach eliminated the need for the selection and tuning of ad hoc parameters.

As part of our future research, we will explore the use of this algorithm to train an automatic target recognition system capable of detecting and recognizing targets in the presence of occlusion and clutter. This will be accomplished by training the system using segmented regions as opposed to training the system using whole objects. In this way, the system will be able to recognize targets even if parts of it are occluded.

Acknowledgements

The author is indebted to Dr. Victoria Franques (USAF Armament Directorate, Eglin AFB) for many suggestions and discussions during the course of this research and for providing the code for the wavelet decomposition.

References

[1] N. Abramson. *Information Theory and Coding.* McGraw Hill, 1963.

[2] G. B. Coleman and H. C. Andrews. Image segmentation by clustering. In *IEEE*, volume 67, pages 153–172, May 1979.

[3] D. Dunn, W. Higgins, and J. Wakeley. Texture segmentation using 2-d gabor elementary functions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(2):130–149, 1994.

[4] V. T. Franques and D. A. Kerr. Wavelet-based rotationally invariant target classification. In *SPIE Conference*, 1997.

[5] T. Kanungo, B. Dom, W. Niblack, and D. Steele. A Fast Algorithm for MDL-based Multi-band Image Segmentation. In *Proc. IEEE Computer Vision and Pattern Recognition*, pages 609–616, Seattle, Washington, June 1994.

[6] Y. G. Leclerc. Region grouping using the minimum-description-length principle. In *DARPA Image Uderstanding Workshop*, 1990.

[7] A. R. Rao. *Taxonomy for Texture Description and Identification.* Springer-Verlag, New York, 1990.

[8] T. R. Reed and J. M. Hans Du Buff. A review of recent texture segmenation and feature extraction techniques. *CVGIP: Image Understanding*, 53(3):359–372, 1993.

[9] J. Rissanen. A universal prior for integers and estimation by minimum description length. *The Annals of Statistics*, 11(2):211–222, 1983.

[10] E. Salari and Z. Ling. Texture segmentation using hierarchical wavelet decomposition. *Pattern Recognition*, 28(12):1819–1823, 1995.

[11] M. Unser. Texture classification and segmentation using wavelet frames. *IEEE Trans. Image Processing*, 4:1549–1560, 1995.

[12] S. C. Zhu and A. Yuille. Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(9):884–900, September 1996.

A FEASIBILITY STUDY OF TURBINE DISK COOLING BY EMPLOYING RADIALLY ROTATING HEAT PIPES

Yiding Cao
Assistant Professor
Department of Mechanical Engineering

Florida International University
Miami, Florida 33199

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

And

Wright Laboratory

August 1997

Abstract

In modern aero-engines, a turbine disk is normally cooled by compressed air. A literature survey regarding the turbine disk cooling reveals that although the average air-cooling heat transfer coefficient is generally high, the local heat-transfer coefficient at the disk rim is low. Jet cooling could be used to enhance the heat transfer at the rim, but its implementation is costly. It is believed that one of the major causes of the high temperature at the rim is the low thermal conductivity associated with the turbine disk material. Based on this understanding, a turbine disk that incorporates radially rotating heat pipes is introduced. The incorporation of the heat pipe would significantly increase the effective thermal conductance of the disk and spread the heat from the disk rim to a much large surface area. A unique disk design that employs interconnected heat pipe branches is also proposed for the purpose of cost reduction. To evaluate the effectiveness of the new turbine disk, a simplified analysis based on the one-dimensional and steady-state assumptions is made. The analytical results indicate that the disk that incorporates the heat pipe could reduce the disk rim temperature by more than 300 °C, with only a moderate increase in the disk base temperature. In conclusion, a turbine disk that employs rotating heat pipes is very effective for the disk rim temperature reduction, and it could find important applications in high-pressure gas turbines, such as those being developed under the Air Force program of the Integrated High Performance Turbine Engine Technology.

Nomenclature

A_2	circumferential surface area at disk rim
a_0, a_1	constants in the general solution of the modified Bessel differential equation
h	heat transfer coefficient between the cooling air and disk surfaces
h_1	heat transfer coefficient at the inner surface of the disk
I_0, I_1	the modified Bessel functions of the first kind
K_0, K_1	the modified Bessel functions of the second kind
k	thermal conductivity of the disk material
k_{hp}	effective thermal conductance of the heat pipe
k_{eff}	effective thermal conductance of the disk incorporating the rotating heat pipe
q	heat transfer rate at the disk rim
Re_r	rotational Reynolds number ($= \omega r^2/\nu$)
r_1	disk inner radius
r_2	disk rim radius
r	radial location
r^*	dimensionless radial location, $(r - r_1)/(r_2 - r_1)$
T	disk temperature
T_c	cooling air temperature
T_d	disk rim temperature
z	axial location

δ disk thickness

δ_1 disk thickness at disk inner radius

φ ratio of the volume occupied by the heat pipe to the total disk volume

ν kinematic viscosity of cooling air

ω disk rotational speed

θ disk excess temperature ($= T - T_c$)

Background Information

Temperature limitations are the most crucial limiting factors to the efficiency of a gas turbine engine. An increased turbine inlet temperature decreases both specific fuel and air consumption while increasing efficiency. This desired high temperature, however, is often in conflict with available materials that can withstand the high temperature. Like turbine blades and nozzle guide vanes of an aero-engine, turbine disks are cooled by compressed air that is bled from the compressor (Fig. 1). However, as the turbine is working at an increasingly high temperature, the disk rim temperature may also reach an unbearably high level. For example, the disk rim temperature of a high-pressure turbine in a development engine, under the Air Force program of the Integrated High Performance Turbine Engine Technology (IHPTET), has exceeded 1000 °C that is approaching the creep limitation of the disk material.

The disk cooling through compressed air has been studied extensively in the past. Notable studies in this area include those by Metzger (1970), Metzger et al. (1979), Owen (1988, 1992), and Nakata et al. (1992). Metzger (1970) studied the heat transfer and pumping on a rotating disk with freely induced and forced cooling. The experimental configurations included an unshrouded disk that simulated freely induced disk cooling condition, disk with back shroud, disk with both front and back shrouds, and front and back shrouds with shroud rim. The experiments showed that with a relatively small space ratio, $z_0/r_0 = 0.167$, where z_0 is the disk-to-shroud space and r_0 is the disk radius, the heat transfer results in terms of Nusselt number, Nu, approached those for an unshrouded disk rotating in an infinite medium. The correlations for the unshrouded disk are available under both laminar and turbulent flow conditions. The heat transfer rate under this unshrouded condition is a function of rotational Reynolds number Re_r, based on the rotational speed ω and the disk radius r_0. When the rotating speed is high, the average heat transfer coefficient as calculated using the available correlation would be very high. For instance, the average heat transfer coefficient may exceed 1,000 W/m^2-°C when the rotational Reynolds number is on the order of 10^7. From the perspective of disk cooling, however, it is the local heat transfer coefficient at the disk rim that is important for the reduction of the maximum disk temperature.

Owen (1992) described several rotating disk systems that more resemble the working condition of a turbine disk. These systems include rotor-stator systems, rotating cavities with radial outflow, rotating cavities with radial inflow, rotating cavities with axial through-flow, and contra-rotating disks. For the rotating cavity related to the disk-cooling with radial outflow, some experimental and numerical results were presented by Owen (1992). These results were in terms of local Nusselt number versus dimensionless disk radial location, $x = r/b$, where b is the outer radius of the disk. The variation of local Nusselt number with different rotational Reynolds numbers was also considered. The experiments indicated that for small values of x, Nu increased radially as air is entrained into the boundary layers.

For a larger x, Nu decreased as the temperature of the air inside the none-entraining Ekman-type boundary layer increased with radius r. Near the disk rim, the value of Nusselt number dropped down to almost zero when the rotational Reynolds number was relatively large. The same behavior of the local Nusselt number was also reported by Nakata et al. (1992) when the laminar flow over the disk was numerically analyzed.

Based on the aforementioned discussion, the local heat transfer rate at the disk rim needs to be enhanced if the turbine disk rim temperature is to be significantly reduced. A common practice for this enhancement is to utilize a jet impingement at the rim. This approach usually involves the impingement of cooling jets onto the blade attachment region at the rim. Metzger et al. (1979) experimentally studied the jet cooling at the rim of a rotating disk for the enhancement of the heat transfer in that region. The experimental results indicated that the heat transfer rates were unaffected by impingement for small jet flowrates. To gain significant enhancement in heat transfer rates, flowrates of at least one-tenth the disk pumping flow capacities is required for a single jet. When a multiple jet array is used, this is a relatively large amount air flow from the compressor. The authors further concluded that many multiple jet rim cooling configurations were probably not very effective in raising the rim cooling rates although they may help to reduce the radial inflow of hot combustion gases. This phenomenon can be explained by the theory of rotationally dominant zones and impingement dominant zones (Metzger and Grochowsky, 1977). When the cooling jet flowrate is low, the jet is apparently swept away by the pumped boundary layer on the disk surface before it has any significant effect on the disk surface. In summary, the jet cooling may be used for the reduction of the maximum temperature at the rim, but it is costly and would consume a relatively large amount of compressed air.

It is believed that one of the major causes of the high temperature at the disk rim is the low thermal conductivity of the disk material under a high heat transfer rate from the rotor blade or the hot combustion gas into the disk. Typical thermal conductivity of commonly used materials for an aero-engine disk is on the order of 10 W/m-°C (6.4 Btu/h-ft-°F) to 25 W/m-°C (14.4 Btu/h-ft-°F). With this low thermal conductivity, most of the heat that is transferred into the disk is accumulated in the disk rim region. Even with a high heat transfer coefficient between the disk surface and the cooling air, the heat dissipation is still not effective due to the small actual heat-transfer surface area near the disk rim. It is believed that if the heat could be spread from the disk rim to the inner portion of the disk, the temperature at the disk rim could be substantially reduced under the same air cooling condition, due to an increase in the actual heat transfer area. To achieve this goal, the thermal conductivity of the disk material or the thermal conductance of the disk must be increased. The selection of the turbine disk material is based on many important and often conflicting factors, such as stress, temperature, and corrosion conditions. At the present time, the selection of a disk material that could satisfy turbine working requirements and at the same time has a sufficiently high thermal conductivity is virtually impossible. As a result, the effort in this study focuses on the improvement of the thermal conductance of the disk.

In this study, radially rotating heat pipes are introduced as an alternative means for cooling turbine disks through the increase of disk effective thermal conductance. For this application, miniature heat pipes that have a radius on the order of 1 to 2 mm are expected to be employed. Heat pipes in general are passive heat transfer devices that may have an effective thermal conductance hundreds of times higher than the thermal conductivity of copper (380-400

W/m-°C). The effective thermal conductance of the low-temperature miniature heat pipes with water as the working fluid is on the order of 100 to 200 times that of copper, as evaluated by Cao et al. (1997) from their experimental dada based on the cross-sectional area of the vapor space. The miniature heat pipes to be used in the disk cooling are high-temperature heat pipes with a liquid metal as the working fluid, which in general have a much higher heat transfer capacity compared to that of low-temperature heat pipes. The high-temperature miniature heat pipe has also been proposed for the cooling of turbine rotor blades (Cao, 1996). The effective thermal conductance of the high-temperature miniature heat pipe excluding the shell effects, as evaluated from the results of Cao and Chang (1997), is on the order of 500 to 1,000 times that of copper. This in turn is more than 5,000 to 10,000 times the thermal conductivity of commonly used disk materials. The incorporation of the heat pipe that has such a high thermal conductance is expected to increase the thermal conductance of the disk dramatically while occupying a reasonably small amount of volume in the disk.

Proposed Turbine Disk Configurations Incorporating Radially Rotating Heat Pipes

Figure 2 schematically illustrates a turbine disk with a number of radially rotatimng heat pipes embedded in the disk. The heat pipes are circumferentially arranged and extend radially from the disk rim towards the inner radius of the disk. The heat pipe, however, is not required to extend to the inner radius of the disk; its length is determined by the heat transfer requirement from the disk rim to the inner portion of the disk. The cross-sectional area of the heat pipe should be as small as possible for the disk strength consideration ($d \sim 3$ mm). For a gas turbine employing the traditional dovetail attachment, the maximum disk temperature would occur near the tip of the dovetail. As a result, the heat pipe should extend as closely to the dovetail tip as possible provided that the strength consideration is satisfied. A heat pipe arrangement under this consideration is elaborated in detail A of Fig. 2. The spaces between the individual heat pipes could be reserved, among other considerations, for the passages that supply the cooling air to the rotor blades.

The heat pipes employed in this application are radially rotating heat pipes without any requirement for a wick structure. The heat pipe structure is very simple. It consists of an elongated cavity that is air-evacuated and filled with an amount of working fluid. Because of this simple structure, the reliability of the heat pipe should be very high. The liquid return mechanism from the condenser section to the evaporator section is provided by the centrifugal force due to the rotating motion of the disk. As mentioned earlier, the study of Cao and Chang (1997) indicated that the heat transfer capacity of this type of heat pipe is very high due to the high rotating speed of the turbine rotor. The heat pipe can be directly fabricated in the disk and the disk material is also the shell of the heat pipe. In this case, the incorporation of the heat pipe into the disk would not render any significant weight penalty. Only a small volume penalty would be involved. For a turbine disk that is traditionally forged, the elongated heat pipe cavities could be formed through the cores. The heat pipe cavities can also be formed through drilling after the disk is forged. In case that the disk could be cast, the heat pipe cavity could be more conveniently formed and the heat pipe having an irregular and complex geometry could be employed. Another important feature for the present application is that the exact boundary between the heat pipe evaporator section in the outer portion of the disk and the heat pipe condenser section in the inner portion of the disk is not prescribed. The evaporator length or condenser length would be determined by disk working conditions that include the heat input near the disk rim, the air cooling

conditions over the disk surface, and the geometry of the disk and heat pipe. For the present application, sodium can be used as the heat pipe working fluid, which has a working temperature range of about 600-1200 °C. The compatibility of sodium with Inconel, which is the commonly used disk material for aero-engines, has been documented (Faghri, 1995; Dunn and Reay, 1994). The compatibility of some other nickel-based alloys with sodium, such as Hastelloy X, has also been indicated in the literature (Dunn and Reay, 1994).

For the disk shown in Fig. 2, a number of individual heat pipes are employed. The independence of each heat pipe would definitely increase the reliability of the whole disk system. However, the heat pipe filling and processing may be costly if the number of the heat pipes in a disk is large. Fig. 3 presents a new heat pipe design in the disk, which includes a circumferentail slot and a number of heat pipe branches. As indicated in the figure, the circumferential slot is provided near the inner radius of the disk. Through this slot, the individual heat pipe branches are interconnected. As a result of this interconnection, the whole disk becomes a single heat pipe that can be processed and filled with working fluid only once, and the fabrication cost of the disk could be substantially reduced. With this configuration, however, a relatively uniform liquid distribution among the individual heat pipe branches should be guaranteed. It is believed that through optimizing the working fluid charge ratio and the slot configuration, the potential liquid distribution problem could be solved.

Simplified Analyses

In the previous section, the rotating heat pipe has been proposed to achieve a higher thermal conductance for the turbine disk. To justify the application of the heat pipe, the effectiveness of the disk that incorporates the heat pipe should be evaluated. A comprehensive analysis would involve liquid-vapor two-phase flow in the heat pipe, three-dimensional heat conduction in the disk material, flow field of the cooling air surrounding the disk, and the heat input condition at the disc rim. An analysis of this level would necessitate a detailed numerical approach, which is beyond scope of the present study.

In this report, a simplified analytical model that seeks a closed-form solution is adopted. Although the analysis is simplified, it would reflect the major performance characteristics of the turbine disk that incorporates the heat pipe cooling technique. The emphasis in this analysis is placed on the performance comparison between the disk with the heat pipe and the disk without the heat pipe. The reduction of the maximum disk temperature at the rim is considered as the major criteria for the performance comparison.

To obtain a closed-form analytical solution, the disk is isolated from other turbine components with a simplified geometry as shown in Fig. 4. The air cooling condition is represented by an average heat transfer coefficient h in association with the cooling air temperature T_c. At the disk rim, either a heat transfer rate q or a rim temperature T_d is specified. As indicated in the figure, the normally contoured turbine disk has been simplified as a plane one. As a result, the disk thickness δ shown in Fig. 4 should be interpreted as an effective disk thickness. For a conservative analysis, however, it can be simply taken to be the disk thickness at the rim. As indicated in Fig. 1, a turbine disk is usually much thicker at the inner radius, $r = r_1$, than at the rim. Also, the cooling condition at the inner surface of the disk may be different from that at the side surfaces. To take these differences into account, an effective heat transfer coefficient, h_1, which may be different from h, is specified at $r = r_1$.

Since the disk thickness is usually much smaller than the disk radius, the disk is thermally lumped in the turbine axial direction, z, and only the temperature variation in the radial direction, r, is considered. To further simplify the problem, the heat pipe in the disk is considered as a thermal conductor with an effective thermal conductance of k_{hp}. The value of the heat pipe effective thermal conductance can be evaluated using the heat pipe testing or analytical results from the literature. The order of the magnitude of the heat pipe effective thermal conductance has also been discussed in the background information of this report. Once k_{hp} is obtained, the effective thermal conductance of the disk can be evaluated. Since the heat transfer has been lumped in the axial direction, it can be assumed that the radial heat conduction in the disk material and the heat transfer in the heat pipe conductor are parallel. The effective thermal conductance of the disk, k_{eff}, based on this parallel conduction model can be evaluated by the following relation:

$$k_{eff} = (1-\varphi)k + \varphi k_{hp} \tag{1}$$

where k is the thermal conductivity of the disk material, φ is the ratio of the cross-sectional area occupied by the heat pipe to the total cross-sectional area of the disk.

Having introduced the concept of the effective disk thermal conductance, the heat transfer problem as illustrated in Fig. 4 can be treated as conduction problem under a cylindrical system with appropriate boundary conditions. As mentioned earlier, the major purpose of this analysis is to evaluate the effectiveness of the disk incorporating the heat pipe in comparison with the disk without the heat pipe. A turbine disk without the heat pipe is first analyzed. The governing equation based on the simplified one-dimensional conduction model under the steady-state condition is a Bessel's modified differential equation of the following form:

$$r\frac{d}{dr}(r\frac{d\theta}{dr}) - m^2r^2\theta = 0 \tag{2}$$

where $\theta = T\text{-}T_c$ is the excess temperature, m is a parameter defined by the following equation:

$$m^2 = \frac{2h}{\delta k} \tag{3}$$

A temperature boundary condition is specified at the outer radius (disk rim), $r = r_2$, and a convection boundary condition is given at the inner radius, $r = r_1$:

$$\frac{d\theta}{dr} = \frac{h_1}{k}\theta, \qquad at \quad r = r_1 \tag{4}$$

$$\theta = T_d - T_c = \theta_d \qquad at \quad r = r_2 \tag{5}$$

As mentioned earlier, the boundary condition at $r = r_2$ could be specified in terms of a given heat transfer rate q. Practically, however, the heat transfer rate into the disk at the rim is difficult to estimate accurately. On the other hand, the temperature at the rim, T_d, can be more conveniently measured. T_d in the present study also represents the maximum temperature of the disk that should be limited to an acceptable level. The general solution of Eq. (2) is:

$$\theta(r) = a_0 I_0(mr) + a_1 K_0(mr) \tag{6}$$

where I_0 is the modified Bessel function of the first kind, of order zero, and K_0 is the modified Bessel function of the second kind, of order zero. a_0 and a_1 are constants to be determined by the given boundary conditions. These two constants, a_0 and a_1, as determined by the two boundary conditions, Eqs. (4) and (5) are:

$$a_1 = \frac{[I_0(mr_1)\frac{h_1}{k} - mI_1(mr_1)]\theta_d}{[I_0(mr_1)K_0(mr_2) - K_0(mr_1)I_0(mr_2)]\frac{h_1}{k} - [K_1(mr_1)I_0(mr_2) + K_0(mr_2)I_1(mr_1)]m} \tag{7}$$

and

$$a_0 = \frac{\theta_d - a_1 K_0(mr_2)}{I_0(mr_2)} \tag{8}$$

Once the temperature distribution in the radial direction is obtained, the heat transfer rate into the disk rim can be found by the following relation in connection with Eq. (6):

$$q = kA_2\left(\frac{d\theta}{dr}\right)_{r=r_2} = kA_2[a_0 mI_1(mr_2) - a_1 mK_1(mr_2)] \tag{9}$$

where $A_2 = 2\pi r_2 \delta$ is the circumferential surface area at $r = r_2$.

Having obtained the solution for a conventional disk, the disk that incorporates the heat pipe is then considered. In this case, the heat transfer rate from Eq. (9) is specified at the disk rim, and the disk rim temperature becomes an unknown that is to be found from the solution of the problem. With this arrangement, the disk rim temperatures for the two disks can be compared with the same heat input rate at the rim and the same air cooling conditions at the disk surfaces. The governing equation for this case is similar to Eq.(2) with k replaced by k_{eff}. The general solution that contains two constants to be determined is as follows:

$$\theta(r) = a_0 I_0(mr) + a_1 K_0(mr) \tag{10}$$

where m is a parameter defined by the following equation:

$$m^2 = \frac{2h}{\delta k_{eff}} \tag{11}$$

The two boundary conditions for this problem are given by the following relations:

$$\frac{d\theta}{dr} = \frac{h_1}{k_{efff}}\theta, \qquad at \quad r = r_1 \tag{12}$$

$$\frac{d\theta}{dr} = \frac{q}{A_2 k_{efff}}, \qquad at \quad r = r_2 \tag{13}$$

The two constants in the general solution (Eq. (10)), as determined by the above two boundary conditions, are:

$$a_1 = \frac{[I_0(mr_1)\frac{h_1}{k_{eff}} - mI_1(mr_1)]\ \frac{q}{A_2 k_{eff}}}{m^2[K_1(mr_2)I_1(mr_1) - K_1(mr_1)I_1(mr_2)] - [K_1(mr_2)I_0(mr_1) + K_0(mr_1)I_1(mr_2)]\frac{h_1 m}{k_{eff}}}$$

and (14)

$$a_0 = \frac{\frac{q}{A_2 k_{eff}} + a_1 mK_1(mr_2)}{mI_1(mr_2)} \tag{15}$$

Results and Discussion

Once the analytical solutions are obtained, the calculations are then made for a typical disk configuration and the air cooling condition as follows:

k = 24 W/m-K; h = 500 W/m^2-K; $\theta_d = T_d - T_c$ = 1010 – 525 = 485 °C; r_1 = 0.07 m; r_2 = 0.3 m; δ = 0.033 m; δ_1 = 0.1155 m; and A_2 = 0.062 m^2.

Notice that the rim temperature T_d is taken to be 1010 °C, which is very high for many commonly used turbine disk materials. The heat pipe radius is taken to be 1.5 mm and the number of the heat pipes, or the number of the

heat pipe branches when the disk design in Fig. 3 is considered, is 58. The effective thermal conductance of the disk incorporating the heat pipe, as evaluated based on Eq (1), is k_{eff} = 1670.W/m-K, where a moderate value of k_{hp}/k = 6,250 for the heat pipe effective thermal conductance has been used. It should be pointed out that the calculation of the heat-pipe effective thermal conductance in the present case should base on the cross-sectional area of the vapor space because the disk material is actually the shell of the heat pipe. The calculated volume fraction of the heat pipes, φ, is about 1.1 to 1.5 percent. The value of h_1 is related to h through $h_1 = (\delta_1/\delta)h = 3.5h$, where $\delta_1$is the disk thickness at the inner radius and h_1 is the corresponding adjusted heat transfer coefficient at that location. As mentioned earlier, the purpose of adjusting h_1 is to take into account the effect of the much larger disk thickness at $r = r_1$.

The calculation procedure has been mentioned in the previous section and is repeated here. The temperature distribution for the case of a conventional disk with T_d specified at the disk rim is determined by using Eqs. (6), (7) and (8). The heat transfer rate, q, into the disk rim is then calculated by using Eq. (9). This heat transfer rate is used as the boundary condition at the rim for the disk that incorporates the heat pipe. The cooling conditions in terms of the heat transfer coefficient and cooling air temperature are kept the same. The solutions for the case of incorporating the heat pipe are then found through Eqs. (10), (14) and (15).

Figure 5 shows the temperature distributions in the radial direction for the conventional disk and the disk that incorporates the heat pipe. The horizontal axis is the dimensionless radial location, $r^* = (r-r_1)/(r_2-r_1)$. The vertical axis is the excess temperature, $\theta = T - T_c$, where T_c is the cooling air temperature, which is considered to be a constant and taken to be 525 °C. As indicated by the results in the figure, the temperature of the conventional disk drops sharply from the disk rim, $r^* = 1$ ($r = r_2$), towards the inner region of the disk. For the majority of the disk surface area, the temperature is very close to the cooling air temperature with a very small heat dissipation rate into the air. In contrast, the temperature distribution for the disk that incorporates the heat pipe is much more uniform in the radial direction. The disk rim temperature in this case is only about 103 °C, which is a reduction of 382 °C compared to the rim temperature without the heat pipe. The temperature at the disk base, $r = r_1$, is about 69.5 °C higher than that of the conventional disk. Since the comparison is based on the same heat input into the disk, cooling conditions at the disk surface, and disk geometry, the analytical results indicate that the heat pipe is very effective; it reduces the rim temperature considerably while increasing the disk base temperature only moderately.

The results presented in Fig. 5 are for disks with air cooling at the disk base, $r = r_1$. If the cooling effect at the base is negligible, an adiabatic boundary condition should be specified at that location. The temperature distribution under this condition can be readily obtained by simply setting $h_1 = 0$ in the previous solutions. The analytical results with an adiabatic disk base are presented in Fig. 6. For the conventional disk, the change in the boundary condition has a negligible effect on the temperature distribution. This conclusion is not surprising if the result presented in Fig. 5 is reviewed. The heat transfer rate at the disk base is already nearly zero when a convection boundary condition is specified at the base. For the disk with the heat pipe, the disk rim temperature increases from 103 °C to 109.5 °C, and the disk base temperature increases from 69.5 °C to 81.2 °C. This indicates that the cooling condition at the base has only a small effect on the temperature distribution. It should be pointed out that this conclusion may be applicable only to a relatively large disk. For a small disk having a small heat transfer surface area, the air cooling at

the base may be a necessity to bring the disk temperature down for the disk that incorporates the heat pipe. In the disk configurations shown in Figs. 2 and 3, the heat pipe extends towards the base of the disk. In some cases, however, the heat pipe may be allowed to extend only to the middle section of the disk for some practical considerations. In this case, an adiabatic boundary condition can be specified at the lower end of the disk portion that contains the heat pipe, due to the small thermal conductivity of the disk material. For this reason, the solutions in Fig. 6 can also be used for analyzing the performance of a disk with such a heat pipe disposition.

The heat transfer coefficient h in the analytical solution is the average heat transfer coefficient over the disk surface. This heat transfer coefficient would change with a different disk rotating speed or a different cooling air flow rate. To illustrate the effect of h on the temperature distribution, calculations are made by using different values of h. The other cooling conditions are kept the same as those used for the solutions in Fig. 5. Figure 7 shows the temperature distributions in the radial direction for h = 1,000 W/m^2-°C, which represents a strong cooling condition when the disk rotates at a high speed. As can be seen from the figure, the disk rim temperature is further reduced for the disk that incorporates the heat pipe. This indicated that the heat pipe cooling technique is more effective for a disk that rotates at a higher speed. This conclusion can be further confirmed by examining the temperature distribution with a small heat transfer coefficient. Figure 8 shows the temperature distributions in the radial direction for h = 100 W/m^2-°C , which represents a weak cooling condition when the disk rotates at a very low speed. The temperature distribution curve for the disk without the heat pipe becomes flatter, and the reduction of the rim temperature for the disk having the heat pipe is decreased. Still, the rim temperature of the disk that incorporates the heat pipe is about 300 °C lower than that without the heat pipe.

The foregoing performance comparison between the disks with and without the heat pipe is based on the same heat input at the disk rim. Practically, when the disk rim temperature is changed, the heat transfer from the rotor blade or the combustion gas into the disk would also change accordingly. It is understandable that for a given turbine inlet temperature and a designed average rotor blade temperature, the total heat transfer rate into the rotor system is fixed. If more heat is transferred into the disk, the heat removal requirement for the blade-cooling air is reduced. On the other hand, if the heat removal rate of the blade-cooling air is the same, the turbine blade could work at a lower temperature. In this case, the disk cooling technique presented in this report is also a new cooling approach for the rotor blade. Because of the incorporation of the heat pipe, the heat dissipation capacity of the disk is substantially increased. As a result, a relatively large amount of heat can be transferred from the rotor blade into the disk to be dissipated, and the rotor blade can be more effectively cooled.

Conclusions and Future Research

1. A literature survey regarding the turbine disk cooling using compressed air was given. It indicated that although the average heat transfer coefficient for a disk rotating at high speeds is high, the local heat transfer coefficient at the disk rim is generally low. In addition, the jet impingement at the disk rim can be used but the implementation is costly.
2. It is believed that one of the major causes of the high temperature at the disk rim is the low thermal conductivity of the disk material. Based on this understanding, two disk designs that incorporate the rotating heat pipe

technique have been proposed. The turbine disk that incorporates the heat pipe would have a much higher effective thermal conductance for the thermal spreading purpose.

3. A simplified analysis indicates that the disk rim temperature can be reduced by more than 300 °C after incorporating the heat pipe, with only a moderate increase in the disk base temperature. The penalty in weight increase for the disk is negligible due to the hollow structure of the heat pipe. The volume penalty for the present configuration is less than 1.5%. In summary, a turbine disk that incorporates the rotating heat pipe cooling technique is feasible and it could help to push the turbine inlet temperature to a new high level.
4. The disk cooling technique presented in this report is also a new cooling technique for rotor blades. Since the heat dissipation capacity of the disk is substantially increased by employing the heat pipe, a large amount of heat can be transferred form the rotor blade into the disk to be dissipated. As a result, the rotor blade can be kept at a lower temperature.

With regard to the future study, the following discussion and suggestions are made:

1. The evaluation of the heat-pipe effective thermal conductance is based on the results from the literature that have a different application background. An analysis on the rotating heat pipes that would be specifically used for the disk cooling should be made. This analysis is very important for the transition of the present study to the real product.
2. From the cost-reduction point of view, this author believes that the disk configuration shown in Fig. 3, which features interconnected heat pipe branches, is more practical. However, the heat pipe structure proposed in the figure is relatively new. Fundamental study regarding the performance of this type of heat pipe should be conducted.
3. The analysis presented in this report is based on the one-dimension and steady-state simplifications. A three-dimensional modeling employing a numerical technique would be appropriate for more detailed and accurate solutions.
4. As mentioned earlier, the disk cooling technique presented in this report is also a new approach for the cooling of rotor blades. To understand the cooling effect on the rotor blade quantitatively, a coupled analysis between the rotor blade and the disk is needed.
5. A stress analysis needs to be performed for the disk that incorporates the heat pipe. This stress analysis would determine the optimum heat pipe size under various performance constraints.
6. For more comprehensive analyses, the cooling air flow field may be analyzed using a CFD code. Also, instead of simplifying the heat pipe as a thermal conductor, the liquid-vapor two-phase flow in the heat pipe could be analyzed by employing an appropriate numerical techniques, such as that by Cao and Faghri (1990).
7. The heat pipes employed in this application are wickless rotating heat pipes. The reliability and heat transfer capacity should be very high. Still, a reliability study on this type of heat pipe should be conducted for aero-engine applications, where the reliability and safety are one of the most important considerations.
8. Finally, the heat pipe structures proposed in Figs. 2 and 3 are only two possible configurations and they could be modified based on available manufacturing capabilities. Heat pipe designs with different configurations to satisfy specific requirements are also possible in the future research.

Acknowledgements

This project is sponsored by Air Force Office of Scientific Research with Dr. Won Chang at Wright Laboratory as the focal point. The author would like to thank his hospitality and both technical and logistic support during my tour at Wright Lab. The author would also like to thank Dr. Charlie MacArthur, the chief of the turbine branch, for this technical help towards this project. Despite his busy schedule, he was always available to answer my questions and help me to understand the turbine cooling problem. Finally, the author would like to thank Dr. Bob Gray for his help in defining the objective of this project.

References

Cao, Y., 1996, "Rotating Micro/Miniature Heat Pipes for Turbine Blade Cooling Applications," *AFOSR Contractor and Grantee Meeting on Turbulence and Internal Flows,* September 1996, Atlanta, GA.

Cao, Y. and Chang, W.S., 1997, "Analyses of Heat Transfer Limitations of Radially Rotating Heat Pipes for Turbomachinery Applications," AIAA 97-2542, *32nd Thermophysics Conference*, Atlanta, GA.

Cao, Y., Gao, M.C., Beam, J.E., and Donovan, B., 1997, "Experiments and Analyses of Flat Miniature Heat Pipes," *AIAA Journal of Thermophysics and Heat Transfer*, Vol. 11, No. 2, pp. 158 – 164.

Cao, Y. and Faghri, A., 1990, "A Transient Two-Dimensional Compressible Analysis for High Temperature Heat Pipes with a Pulsed Heat Input," *Journal of Numerical Heat Transfer*, Part A, Vol. 18, pp. 483 – 502.

Dunn, P.D. and Reay, D.A., 1994, *Heat Pipes*, Fourth Edition, Pergamon.

Faghri, A., 1995, *Heat Pipe Science and Technology*, Taylor & Francis.

Metzger, D.E., Mathis, W.J., and Grochowsky, L.D., 1979, "Jet Cooling at the Rim of a Rotating Disk," *Journal of Engineering for Power*, Vol. 101, pp. 68-72.

Metzger, D.E. and Grochowsky, L.D., 1977, "Heat Transfer Between an Impinging Jet and a Rotating Disk," *ASME Journal of Heat Transfer*, Vol. 99, No. 4.

Metzger, D. E., 1970, "Heat Transfer and Pumping on a Rotating Disk With Freely Induced and Forced Cooling," *Journal of Engineering for Power*, July, 1970, pp. 342-348.

Nakata, Y., Murthy, J.Y., and Metzger, D.E., 1992, "Computation of Laminar Flow and Heat Transfer Over an Enclosed Rotating Disk With and Without Jet Impingement," *Journal of Turbomachinery*, Vol. 114, pp. 881-890.

Owen, J..M., 1988, "Air-Cooled Gas-Turbine Discs: a Review of Recent Research," *Int. J. Heat and Fluid Flow*, Vol. 9, No. 4, pp. 354-365.

Owen, J.M., 1992, "Recent Developments in Rotating-Disc Systems," *ImechE*, pp. 83-92.

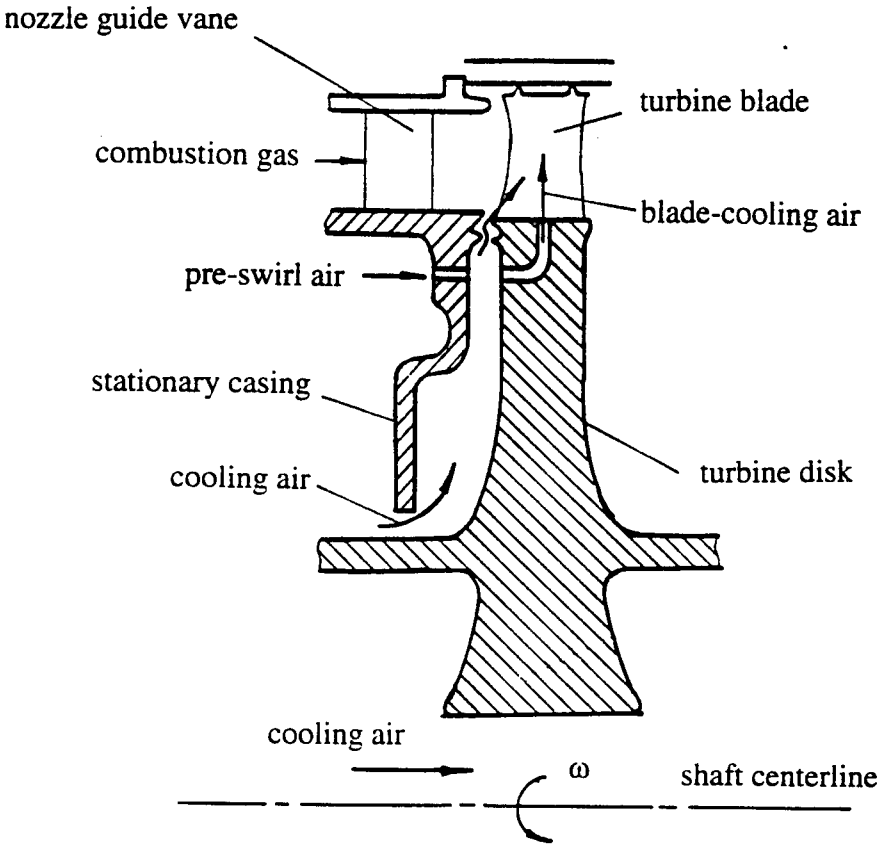


Figure 1 Schematic representation of a high-pressure gas-turbine disk cooled by compressed air.

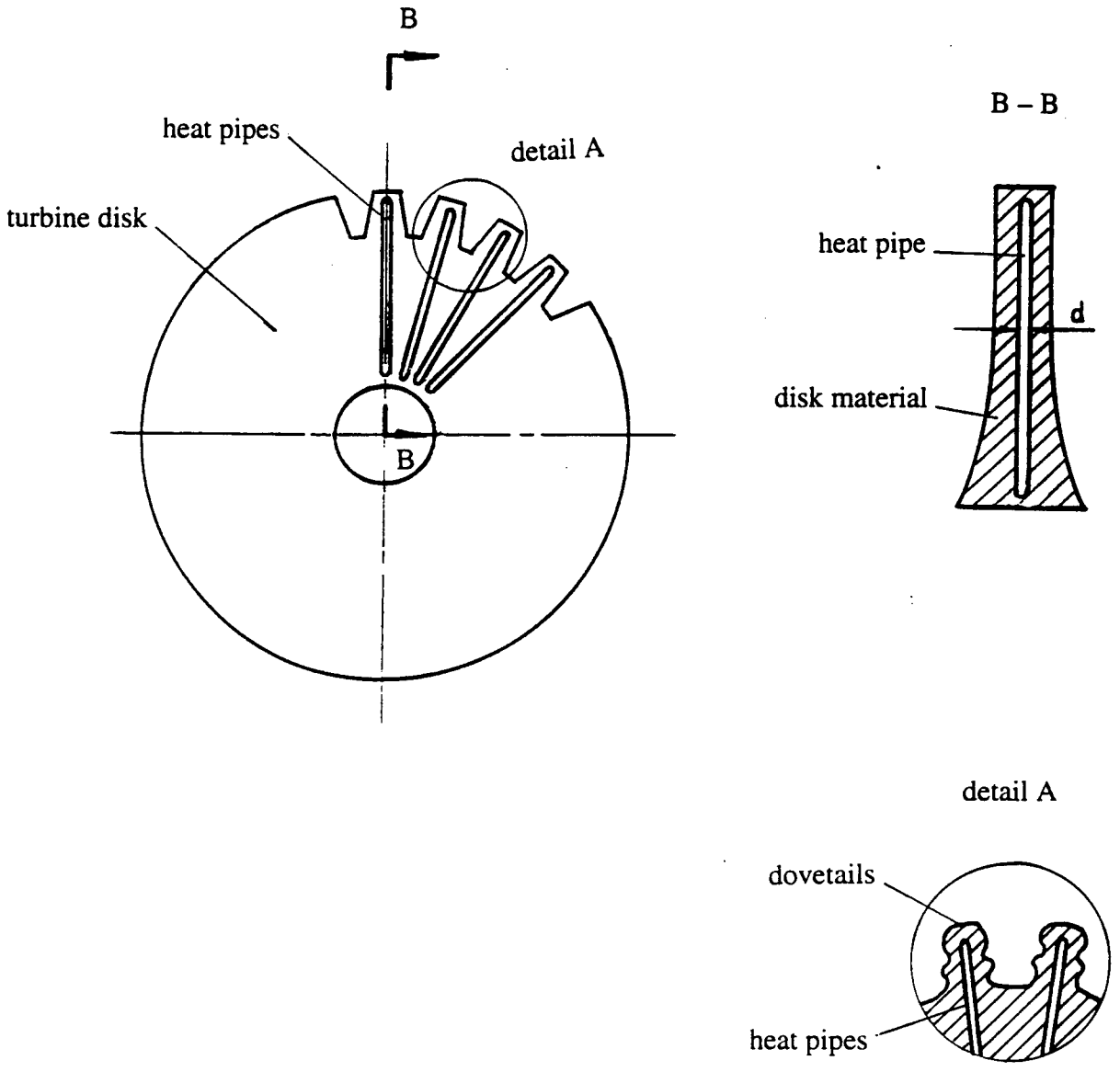


Figure 2 Schematic of a turbine disk incorporating a number of individual radially rotating heat pipes.

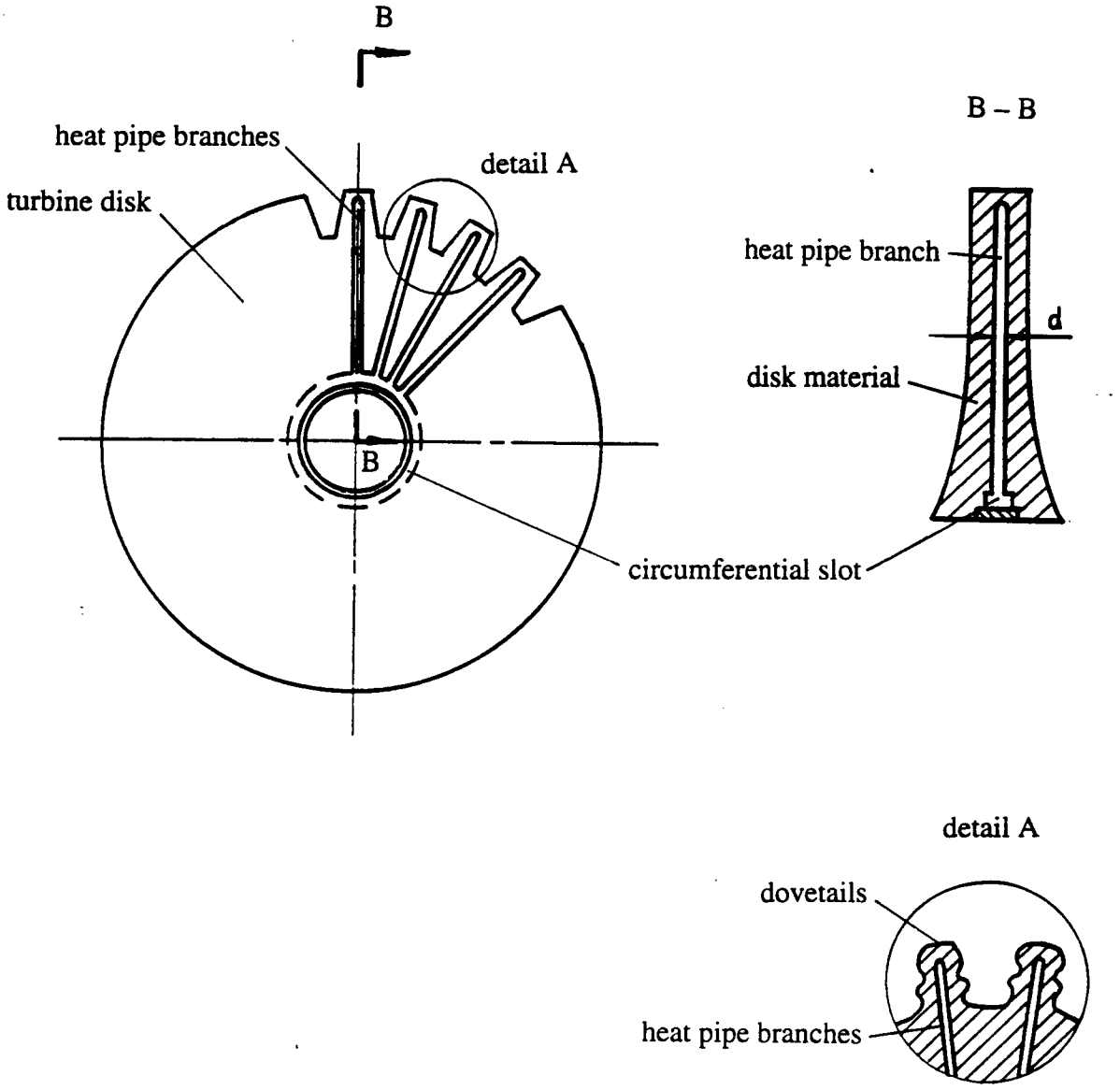


Figure 3 Schematic of a turbine disk incorporating a single heat pipe with a number of interconnected heat pipe branches.

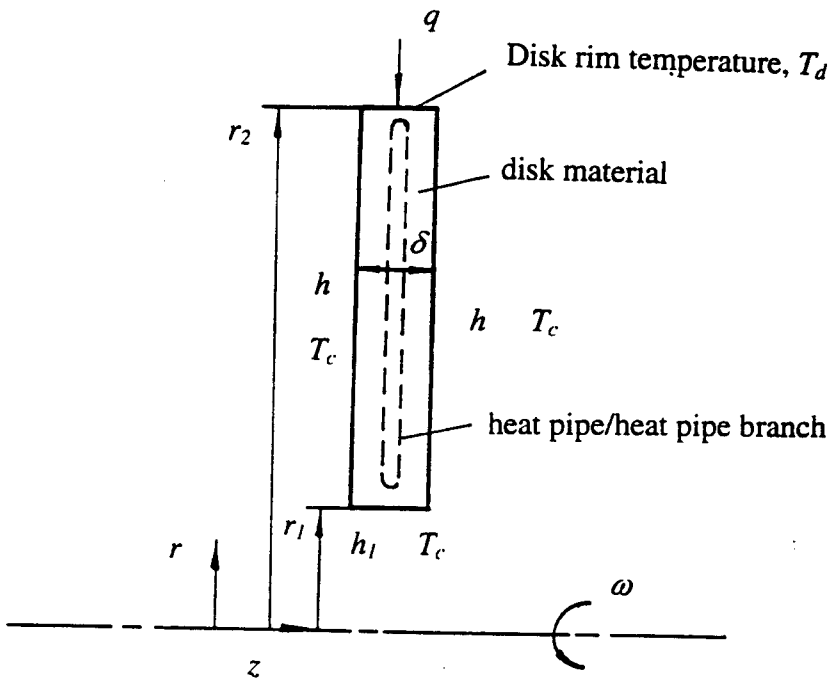


Figure 4 Simplified configuration of a turbine disk with specified boundary conditions.

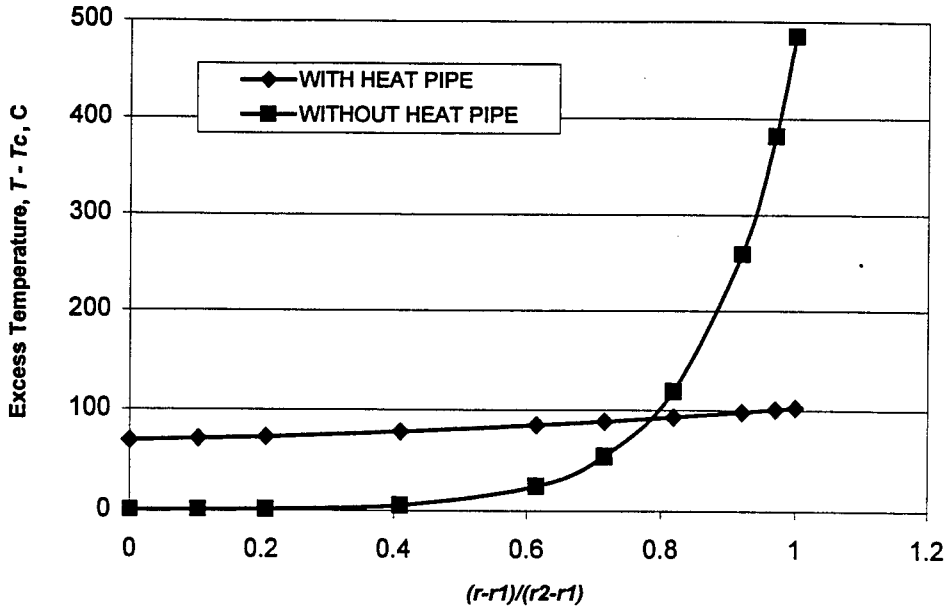


Figure 5 Temperature distributions in the radial locations for the disks with and without heat pipes (*h* = 500 W/(m2-C)

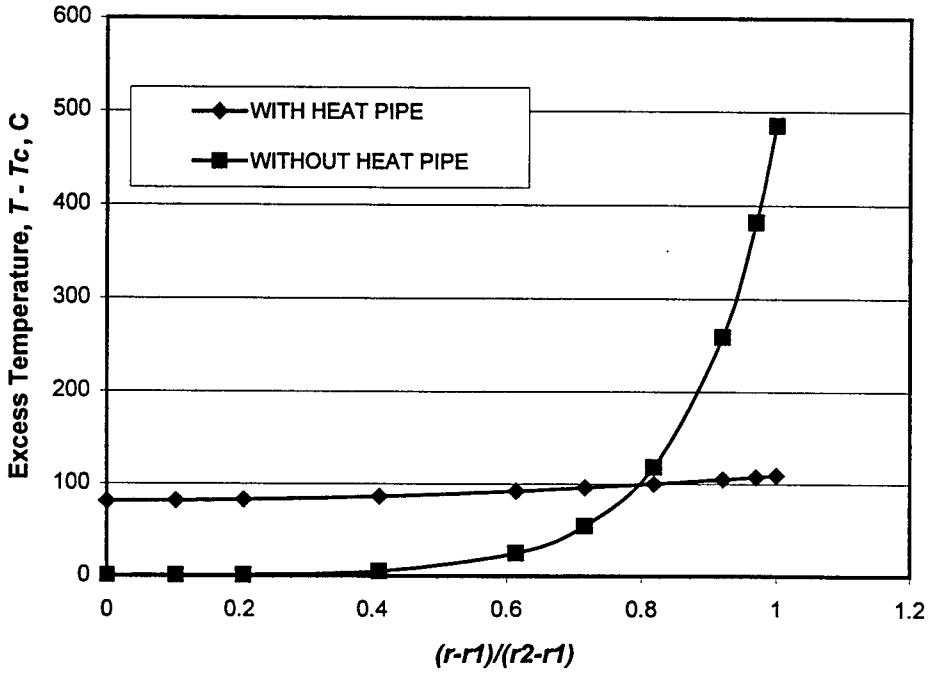


Figure 6 Comparison of the temperature distributions in the radial locations with an adiabatic boundary condition at the disk base (h = 500 W/m2-C, h1 = 0)

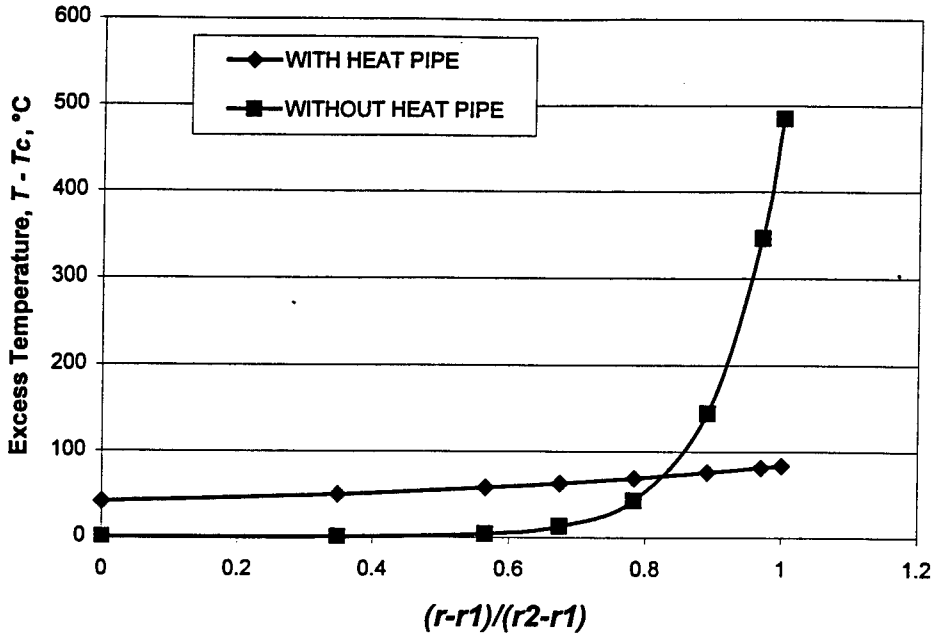


Figure 7 Temperature distributions in the radial direction with h = 1,000 W/(m^2-°C).

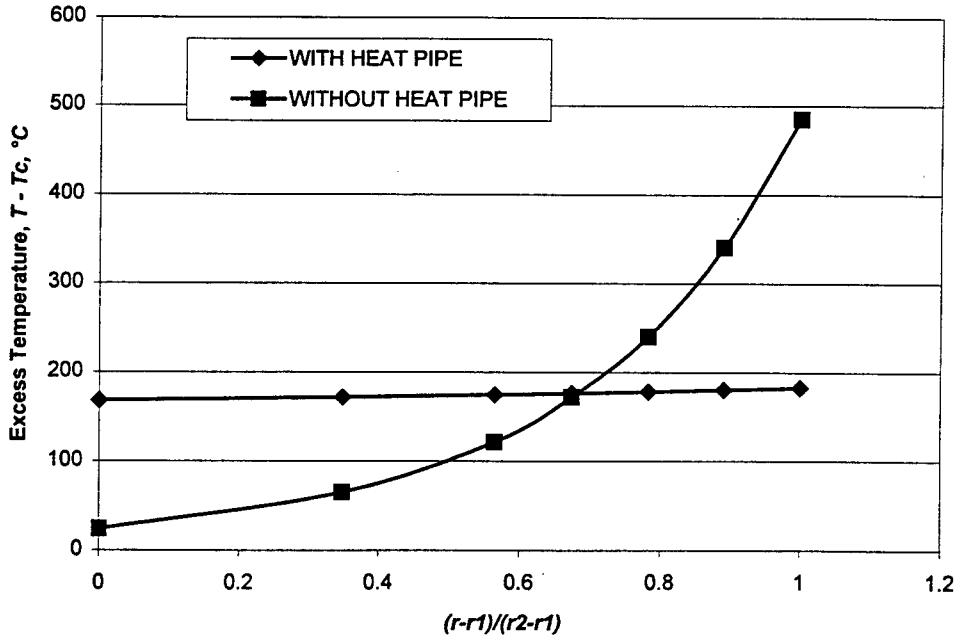


Figure 8 Temperature distributions in the radial direction with h = 100 W/(m^2-°C)

A NOVEL COMPATIBILITY/EQUILIBRIUM BASED INTERATIVE POST-PROCESSING APPROACH FOR AXISYMMETRIC BRITTLE MATRIX COMPOSITES

Reaz A. Chaudhuri
Assistant Professor
Department of Materials Science & Engineering

University of Utah
304 EMRO Building
Salt Lake City, UT 84112

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

And

Wright Laboratory

August 1997

A NOVEL COMPATIBILITY/ EQUILIBRIUM BASED ITERATIVE POST-PROCESSING APPROACH FOR AXISYMMETRIC BRITTLE MATRIX COMPOSITES

Reaz A. Chaudhuri, Associate Professor

Department of Materials Science & Engineering, University of Utah

Abstract

A semi-analytical iterative approach for enhancing the existing two-dimensional quasi-continuous axisymmetric stress field for a brittle matrix micro-composite (i. e., a single fiber surrounded by a concentric matrix cylinder), is presented. The existing solution employs Reissner's variational theorem in conjunction with an equilibrium stress field in which the radial (r-) dependence is assumed *a priori.*

In the present approach, the stress distribution in the radial direction obtained from the afore-cited variational model is improved *a posteriori* through an iterative approach that involves successive substitution of the previously computed strains (or stresses) into the equations of compatibility and equilibrium. The equations of compatibility are selected such that they form Euler equations corresponding to appropriate variational principle, such as the principle of minimum complementary potential energy, etc. The boundary/interface conditions at r = constant and z = constant surfaces/interfaces are satisfied in the pointwise sense. The expressions for the improved axisymmetric displacement and stress fields are derived using the symbolic language, MAPLE. An illustrative thermal stress problem is currently being solved, and will be used to compare with the existing variational solution.

1. Introduction:

Studies of the behavior of unidirectional and laminated composites made from stiff elastic matrix materials which may develop imperfect interfaces with the fibers have enjoyed a revival after the early classical work of Aveston, Cooper, and Kelly (1971). Pagano (1991) refers to these materials as brittle matrix composites (BMC). The aforementioned ACK modeling, as well as the more recent development proposed by Budiansky, Hutchinson, and Evans (1986), are based upon primitive approximations of the stress field developed within a concentric cylinder, i.e., a circular cylindrical body of one material surrounded by a concentric annulus or ring of a second material. There exists a considerable body of literature associated with the elasticity problem of a concentric cylinder, where modern interest is focused on its use as a representative volume element (RVE) of a unidirectional composite (Hashin and Rosen, 1964; Pagano and Tandon, 1988). Additionally, a significant segment of the composite literature is based upon the one-dimensional shear lag analysis, which was apparently originated by Cox (1952). The results obtained using this kind of analysis are too inaccurate to merit further attention in this report.

Sternberg (1970) solved several axisymmetric load diffusion problems within a concentric cylindrical domain, in which the elasticity formulations are simplified by the assumption that the core material (fiber) can be modeled as a one-dimensional bar while the annular region (matrix) in all cases extends to infinity. An exact solution was derived for the case in which the bar was circular in cross-section, and was fully immersed within the unbounded matrix. A similar approach, with the difference of the fiber being assumed rigid, was employed by Luk and Keer (1979). This assumption is more appropriate for resin matrix composites, where the fiber to matrix modulus ratio is very high. The axisymmetric elasticity problem of a broken fiber embedded in an infinite matrix was treated by Pickett and Johnson (1967). In that work, the fiber is represented as a three-dimensional elastic medium; however, the report contained no numerical results for the stress field. Smith and Spencer (1970) also formulated an axisymmetric elasticity solution by a semi-inverse method for a class of boundary value problems in which the radius and length of the concentric cylinder are both finite. The solution is expressed in the form of a series of non-orthogonal functions which satisfy the field equations exactly. Homogeneous boundary conditions on the radial surface of the body are satisfied exactly while realistic end conditions can be approximated. The singularities predicted by Zak (1964) are smoothed out in this approach. A very formidable study, which includes correlation with experimental observations, is that by Atkinson et al (1982). In this work, the pullout of a single fiber from a matrix cylinder is treated. The fiber extends only partially along its length into the matrix. Results are provided for a perfectly bonded fiber-matrix interface as well as for states in which lateral (curved) surface debonding or fiber end plane debonding take place. The problem is solved by "patching" the asymptotic

singular stress field to that given by finite elements, although this method is not always reliable. An interesting conclusion is reached that, at least in a qualitative sense, the interface failure response can be anticipated from the stress field within the uncracked rod. In a model similar to Pagano's (1991), McCartney (1990) treated a class of concentric cylinder problems in which matrix cracking or debonding with or without friction are present. In that work, the functional r-dependence of the stress components is assumed which leads to a system of ordinary differential equations in z. All appropriate field equations are satisfied with the exception of two of the constitutive relations while some of the boundary/interface conditions could only be satisfied in an average sense. The ease and potential effectiveness of this model for composite analysis will demand its careful consideration in comparison with exact solutions and solutions given by competitive approaches. For example, Kurtz and Pagano (1991) formulated an infinite series solution of the axisymmetric elasticity problem in which a fiber is being pulled from the matrix. The length of the body as well as the outer radius are finite. Although the singularity is not explicit in the solution, Cesaro summation has been employed to improve the convergence of the stress field within the singular region.

The solution for the stated problem has been obtained by employing a modified version of the variational model (Pagano, 1991) of an axisymmetric concentric cylinder, which was successfully implemented earlier in the case of a flat laminate by Pagano (1978). The model is generated by subdividing the body into regions consisting of a core and a number of shells of constant thickness and length and satisfying the Reissner variational equation (1950) with an assumed stress field in each region. The number of regions, in particular in the r direction, can be increased in order to improve solution accuracy. The regions are selected such that the thermoelastic properties are constant and the boundary conditions do not change character on any of the bounding surfaces within each region. Pagano (1991) has sought to predict the influence of various kinds of damage (see Figure 1) and their interactions by accurately modeling the micro-mechanical stress field in their presence by using the afore-cited variational approach.

Strengths and Weaknesses of the Existing Approach (Pagano, 1991)

Strengths:

1. This assumed stress based approach insures satisfaction of the axisymmetric equilibrium equations;
2. The existing method utilizes a reasonably accurate non-singular axial (z-direction) variation of the computed stress field that also satisfies the end (z = const.) boundary conditions (i. e., no artificial discontinuity due to sectioning).

Weaknesses:

1. Layering in the radial direction, that introduces artificial discontinuities in some stress components at a layer interface within the same material;

2. As the number of layer increases, the computed eigenvalues become numerically very large, thus limiting the number of layers required for accurate stress field in the vicinity of a stress singularity point.

Objectives of the Current Research:

The present method seeks to alleviate the afore-mentioned weaknesses of the existing approach, while preserving its inherent strengths. The specific goals of the present investigation includes

- Improvement of the radial (r-) variation of the stresses so that subdivision into very thin layers and the associated blow-up of the computed eigenvalues can be avoided;
- Artificial discontinuities of the stress components, σ_θ and σ_z, at a layer interface within the same material can be avoided;
- Determination of the "stress intensity factor" by way of matching the local asymptotic stress field with the afore-cited improved solution at an arbitrarily close distance from the point of stress singularity (i. e., fiber-matrix interface located at a free edge).

As a first step, Chaudhuri (1996) presented an approximate "plane strain" version of the present solution, wherein the equations were simplified by dropping 1/r terms in the equilibrium equations. Additionally, the constitutive relations used were those due to the plane strain. The primary reason for resorting to this approximation was the fact that Chaudhuri's (1996) previously "improved" axisymmetric solution generated 10 "constants" of integration for a 2-layer fiber-matrix cylinder model against 8 interface/boundary conditions in the radial direction. Plane strain approximation removed 2 constants of integration, thus rendering the boundary-value problem under investigation solvable in closed form. However, although the preliminary results thus obtained demonstrated that the plane strain based approach can be implemented in the existing FORTRAN code due to Pagano and co-workers (1991), the accuracy of the results were far from encouraging. This discrepancy was due to the fact that although the plane strain condition prevails in the vicinity of the point of stress singularity (i. e., fiber-matrix interface located at a free edge) as was shown by Zak (1964), the same is not true in the far field where the boundary conditions are applied. The present (summer, 1997) research has corrected this situation by rederiving the correct axisymmetric solution with 4 "constants" (i.e., functions of z) of integration per layer. This is currently being implemented in the framework of symbolic language MAPLE software, for the purpose of solving a system of simulaneous ODE's in terms of the integration "constants" (functions of z).

Secondly, although the boundary/interface conditions at r = constant surfaces were satisfied in the pointwise sense, the past (summer, 1996) approach left the end boundary conditions in the axial direction (i.e., at surfaces z = constant) undefined thus rendering the boundary-value formulation ill-posed. This ill-posedness was removed by introducing boundary error terms that helped satisfy the end boundary conditions at surfaces z = constant in

a somewhat ad hoc fashion. The present (summer, 1997) research has identified a novel approach wherein the z-boundary conditions, that can be satisfied in a pointwise sense, are generated using the equations of the theory of elasticity within a particular cycle or iteration. However, this "discovery" came too late to be implemented in the framework of MAPLE during the summer of 1997. Once the afore-mentioned boundary problem is solved using the MAPLE, the solution will be implemented into the existing ADM Code of Pagano and co-workers (1991).

2. Solution Strategy

Starting point of the present research is Pagano's (1991) layerwise axisymmetric solution for fiber-matrix concentric cylinder model, based on Reissner's variational theorem (1950) in conjunction with an equilibrium stress field, in which the radial (r-) dependence is assumed *a priori*. An approximate model was formulated to define the thermoelastic response of a concentric fiber-matrix cylindrical body under axisymmetric boundary conditions. The interfaces between continguous cylinders may be either continuous or subjected to mixed traction and displacement boundary conditions. The external surfaces may be subjected to mixed boundary conditions that are consistent with the model assumptions but otherwise arbitrary.

In what follows, an improved stress field within a layer is derived starting from Pagano's layerwise axisymmetric fiber-matrix concentric cylinder solution. The stress distribution in the radial direction obtained from the afore-cited layerwise fiber-matrix concentric cylinder model (Pagano, 1991) is improved *a posteriori* through an iterative approach that involves successive substitution of the previously computed strains (or stresses) into the equations of compatibility and equilibrium. A similar procedure was implemented in the post-processing part of a layerwise finite element code for analysis of quasi-three-dimensional laminated plates/shells to obtain a more accurate through-thickness distribution of interlaminar shear stresses (see Chaudhuri and Seide, 1987; and Chaudhuri, 1990). The boundary/interface conditions at r = constant surfaces are to be satisfied in the pointwise sense, thus eliminating artificial discontinuity in computed $\sigma_\theta(r,z)$ and $\sigma_z(r,z)$ across a layer-interface within the same material.

Pagano's (1991) Model

Pagano (1991) considered an arbitrary region within the body defined by inner and outer radii r_1^k and r_2^k, respectively and end planes $z = z_1, z_2$ as shown in Figure 2. He introduced a right-handed cylindrical coordinate system z, θ, r, and employed a contracted notation in the representation of the stress and strain components, i.e.,

$$\sigma_1 = \sigma_{zz},\ \sigma_2 = \sigma_{\theta\theta},\ \sigma_3 = \sigma_{rr},\ \sigma_5 = \sigma_{rz} \qquad (1)$$

and the analogous relation for the engineering strain components ε_i (i=1, 2, 3, 5). The r, z components of displacement are designated as u, w, respectively.

The form of the stress distribution within the annular region is assumed to be given by

$$\sigma_i = p_{iJ} f_J^{(i)} \quad (i = 1, 2, 3, 5; J = 1, 2, ---5) \tag{2}$$

where p_{iJ} are functions of z only. In order to avoid confusion with layer superscript k, the index i in $f_J^{(i)}$ is placed in parentheses. For a region in which $r_1 \neq 0$, the functions $f_J^{(i)}$ are defined by

$$\begin{aligned} f_1^{(1)} &= f_1^{(2)} = f_1^{(3)} = \frac{r_2 - r}{r_2 - r_1} \\ f_2^{(1)} &= f_2^{(2)} = f_2^{(3)} = \frac{r - r_1}{r_2 - r_1} \\ f_1^{(5)} &= \frac{r}{r_1} f_1^{(1)}, \quad f_2^{(5)} = \frac{r}{r_2} f_2^{(1)} \qquad (r_1 \neq 0) \\ f_3^{(3)} &= r^3 - (r_1^2 + r_1 r_2 + r_2^2) r + r_1 r_2 (r_1 + r_2) \\ f_4^{(3)} &= r^2 - (r_1 + r_2) r + r_1 r_2, \qquad f_5^{(3)} = \frac{1}{r_1 r_2 r} f_4^{(3)} \\ f_3^{(5)} &= \frac{(r_1 + r_2) r^2 - (r_1^2 + r_1 r_2 + r_2^2) r}{r_1^2 r_2^2} + \frac{1}{r} \end{aligned} \tag{3}$$

with

$$p_{iJ} = f_J^{(i)} = 0 \quad (r_1 \neq 0; \text{ i = 1, 2 and J = 3, 4, 5 or i = 5 and J = 4, 5}) \tag{4}$$

In other words, the functions $f_J^{(i)}$ and the corresponding p_{iJ} not displayed in (3) all vanish. In the event that $r_1 = 0$ (fiber core), eqn (3) is replaced by

$$\begin{aligned} f_1^{(1)} &= f_1^{(2)} = f_1^{(3)} = \frac{r_2 - r}{r_2} \\ f_2^{(1)} &= f_2^{(2)} = f_2^{(3)} = f_2^{(5)} = \frac{r}{r_2} \qquad (r_1 = 0) \\ f_3^{(3)} &= (r^2 - r_2^2) r \quad , \quad f_4^{(3)} = f_3^{(5)} = (r - r_2) r \end{aligned} \tag{5}$$

with

$$p_{iJ} = f_J^{(i)} = 0 \quad (r_1 = 0; \text{ i = 1, 2 and J = 3, 4 or i = 5 and J = 1, 4 or J = 5}) \tag{6}$$

It may be noted that the superscripts k have been omitted in eqns (2) - (6) to avoid unnecessary congestion. The relations (2) - (6) arise by assuming that σ_1 and σ_2 are linear functions of r in the region and then determining the form of the remaining stress components from the equations of equilibrium of axisymmetric elasticity subjected to the following conditions

$$p_{i\alpha}(z) = \sigma_i(r_\alpha, z) \qquad (i = 1, 2, 3, 5;\ \alpha = 1, 2) \tag{7}$$

Thus, the p functions are equal to actual stresses at $r = r_1, r_2$.

The general form of Pagano's (1991) solution for any of the dependent variables P(z) is expressed by

$$P(z) = \sum_i A_i e^{\lambda_i z} + P_p(z) \tag{8}$$

within each constituent where A_i are constants, λ_i are eigenvalues of a determinant, and $P_p(z)$ is a particular solution, which in the present case is a simple polynomial. Further details of the solution procedure including the method for determining the higher order eigenvector and higher order particular solution are discussed by Brown (1992).

Improved Stress Field in a Fiber or Matrix Layer

An examination of Pagano's (1991) axisymmetric fiber-matrix concentric cylinder model reveals that he has assumed a linear variation of axial stress, $\sigma_z(r,z)$ and hoop stress, $\sigma_\theta(r,z)$, with respect to r within a fiber or matrix layer. This is consistent with the assumption of Love-Kirchhoff's thin shell theory. However, he derived the interlaminar shear stress, $\tau_{rz}(r,z)$, and the radial stress, $\sigma_r(r,z)$, using the equations of equilibrium in line with his earlier work (see Pagano, 1969) before substituting these stresses into Reissner's (1950) variational principle. In what follows, improved deformation and stress fields based on successive use of compatibility and equilibrium equations of axisymmetric elasticity theory is derived in the annular and core regions, respectively.

Annular Layer:

Pagano's stress field for an annular (fiber or matrix) layer k = 1, 2, ... is as shown below:

$$\bar{\sigma}_z^{(k)} = p_{11}^{(k)}(z)\left(\frac{r_2 - r}{r_2 - r_1}\right) + p_{12}^{(k)}(z)\left(\frac{r - r_1}{r_2 - r_1}\right) \tag{9a}$$

$$\bar{\tau}_{rz}^{(k)}(r,z) = p_{51}^{(k)}(z)\frac{r(r_2 - r)}{r_1(r_2 - r_1)} + p_{52}^{(k)}(z)\frac{r(r - r_1)}{r_2(r_2 - r_1)} + p_{53}^{(k)}(z)\left(\frac{(r_1 + r_2)r^2 - (r_1^2 + r_1 r_2 + r_2^2)r}{r_1^2 r_2^2} + \frac{1}{r}\right) \tag{9b}$$

$$\bar{\sigma}_r^{(k)}(r,z) = p_{31}^{(k)}(z)\left(\frac{r_2 - r}{r_2 - r_1}\right) + p_{32}^{(k)}(z)\left(\frac{r - r_1}{r_2 - r_1}\right) + p_{33}^{(k)}(z)\left(r^3 - (r_1^2 + r_1 r_2 + r_2^2)r + r_1 r_2(r_1 + r_2)\right)$$
$$+ p_{34}^{(k)}(z)\left(r^2 - (r_1 + r_2)r + r_1 r_2\right) + p_{35}^{(k)}(z)\left(r^2 - (r_1 + r_2)r + r_1 r_2\right)\frac{1}{r r_1 r_2} \tag{9c}$$

It is worthwhile to point out here that although the barred stresses satisfy the equations of equilibrium for an axisymmetric elastic body, the corresponding strains fail to satisfy the equations of compatibility in the pointwise sense. It is noteworthy that although both the plane strain and axisymmetric deformations represent two-dimensional states, the latter case requires

four compatibility equations to be satisfied by the strains computed using Pagano's (1991) axisymmetric variational model. This is in contrast to the case of plane strain, wherein only one compatibility equation out of the six is not an identity. However, if the hoop stress, σ_θ, is derived using the exact axisymmetric elasticity based kinematic relations and Hooke's law, then three of the four compatibility equations required by the axisymmetric elasticity theory become identities. The combined kinematic and stress-strain relations for axisymmetric elasticity theory are given as follows (Timoshenko and Goodier, 1959):

$$\tilde{u}(r,z) = r\tilde{\varepsilon}_\theta(r,z) = \frac{r}{E_k}[\tilde{\sigma}_\theta(r,z) - \nu_k\{\overline{\sigma}_r(r,z) + \overline{\sigma}_z(r,z)\}] \tag{10a}$$

$$\frac{\partial\tilde{u}(r,z)}{\partial r} = \tilde{\varepsilon}_r(r,z) = \frac{1}{E_k}[\overline{\sigma}_r(r,z) - \nu_k\{\tilde{\sigma}_\theta(r,z) + \overline{\sigma}_z(r,z)\}] \tag{10b}$$

The tilda quantities represent the "improved" stresses, strains and displacements, obtained using the equations of compatibility, equilibrium, etc., but may not satisfy the prescribed z=constant boundary conditions. Eliminating $\tilde{u}^{(k)}(r,z)$ from eqns (10) and solving for $\tilde{\sigma}_\theta^{(k)}(r,z)$ will yield

$$\tilde{\sigma}_\theta^{(k)}(r,z) = r^{-(1+\nu_k)}\left[\int\left\{(1+\nu_k)r^{\nu_k}\overline{\sigma}_r(r,z) + \nu_k r^{(1+\nu_k)}\frac{\partial\overline{\sigma}_r(r,z)}{\partial r} + \nu_k r^{(1+\nu_k)}\frac{\partial\overline{\sigma}_z(r,z)}{\partial r}\right\}dr + F_1^{(k)}(z)\right] \tag{11}$$

where $F_1^{(k)}(z)$, k = 1, 2,... is a "constant" of integration (with respect to r) and an arbitrary function of z. The hoop stress in an annular layer ($k \neq 0$, may be fiber or matrix) is obtained by substitution of $\overline{\sigma}_z^{(k)}(r,z)$ and $\overline{\sigma}_r^{(k)}(r,z)$ from eqns (9a,c) into the right side of eqns (11)

$$\tilde{\sigma}_\theta^{(k)}(r,z) = \sigma_\theta^{(k)*}(r,z) + \frac{1}{r^{1+\nu_k}}F_1^{(k)}(z);\quad k = 1, 2, 3,... \tag{12a}$$

where

$$\sigma_\theta^{(k)*}(r,z) = (1+\nu_k)[p_{31}^{(k)}(z)\frac{1}{(r_2-r_1)}(\frac{r_2}{\nu_k+1} - \frac{r}{\nu_k+2}) + p_{32}^{(k)}(z)\frac{1}{(r_2-r_1)}(\frac{r}{\nu_k+2} - \frac{r_1}{\nu_k+1})$$

$$+ p_{33}^{(k)}(z)\{\frac{r^3}{\nu_k+4} - (r_1^2+r_1r_2+r_2^2)\frac{r}{\nu_k+2} + r_1r_2(r_1+r_2)\frac{1}{\nu_k+1}\} + p_{34}^{(k)}(z)\{\frac{r^2}{\nu_k+3}$$

$$-(r_1+r_2)\frac{r}{\nu_k+2} + \frac{r_1r_2}{\nu_k+1}\} + p_{34}^{(k)}(z)\frac{1}{r_1r_2}\{\frac{r}{\nu_k+2} - (r_1+r_2)\frac{1}{\nu_k+1} + \frac{1}{r\nu_k}r_1r_2\}]$$

$$+ \nu_k[-p_{31}^{(k)}(z)\frac{r}{(r_2-r_1)(\nu_k+2)} + p_{32}^{(k)}(z)\frac{r}{(r_2-r_1)(\nu_k+2)} + p_{33}^{(k)}(z)\{\frac{3r^3}{(\nu_k+4)}$$

$$-(r_1^2+r_1r_2+r_2^2)\frac{r}{(\nu_k+2)}\} + p_{34}^{(k)}(z)\{\frac{2r^2}{(\nu_k+3)} - (r_1+r_2)\frac{r}{(\nu_k+2)}\}$$

$$+ p_{35}^{(k)}(z)\{\frac{r}{r_1 r_2(\nu_k+2)} - \frac{1}{r\nu_k}\} - p_{11}^{(k)}(z)\frac{r}{(r_2-r_1)(\nu_k+2)} + p_{12}^{(k)}(z)\frac{r}{(r_2-r_1)(\nu_k+2)}] \quad (12b)$$

The radial displacement component, $\tilde{u}^{(k)}(r,z)$, in the kth layer can now be obtained by substituting eqns (12) into eqn (10a) as follows:

$$\tilde{u}^{(k)}(r,z) = u^{(k)*}(r,z) + \frac{r^{-\nu_k}}{E_k} F_1^{(k)}(z); k = 1,2,... \quad (13a)$$

where

$$u^{(k)*}(r,z) = \frac{r}{E_k}\{\sigma_\theta^{(k)*} - \nu_k(\bar{\sigma}_r^{(k)} + \bar{\sigma}_z^{(k)})\}; k = 1, 2, ... \quad (13b)$$

whence $\tilde{\varepsilon}_r^{(k)}(r,z)$ can be obtained by direct differentiation with respect to r. Interlaminar shear strain, $\bar{\gamma}_{rz}^{(k)}(r,z)$, can now be obtained using Hooke's law:

$$\bar{\gamma}_{rz}^{(k)}(r,z) = \frac{1}{G_k}\bar{\tau}_{rz}^{(k)}(r,z) \quad (14)$$

It may be noted that the above two strains, $\tilde{\varepsilon}_r^{(k)}(r,z)$(or u) and $\bar{\gamma}_{rz}^{(k)}(r,z)$, are not compatible with the axial strain,$\bar{\varepsilon}_z^{(k)}(r,z)$, computed using Hooke's law

$$\bar{\varepsilon}_z^{(k)} = \frac{1}{E_k}[\bar{\sigma}_z^{(k)} - \nu_k(\tilde{\sigma}_\theta^{(k)} + \bar{\sigma}_r^{(k)})] \quad (15)$$

because the remaining compatibility equation

$$\frac{\partial^2 \varepsilon_z}{\partial r^2} + \frac{\partial^2 \varepsilon_r}{\partial z^2} = \frac{\partial^2 \gamma_{rz}}{\partial r \partial z} \quad (16a)$$

or the ensuing identity

$$\frac{\partial\, \varepsilon_z}{\partial r} + \frac{\partial^2 u}{\partial z^2} = \frac{\partial\, \gamma_{rz}}{\partial z} \quad (16b)$$

is not satisfied. It may be noted here that since the radial displacement $\tilde{u}^{(k)}(r,z)$ has already been derived using the appropriate kinematic and constitutive relations of axisymmetric elasticity theory, given by eqns (10), the radial strain, $\tilde{\varepsilon}_r^{(k)}(r,z)$, will automatically be compatible with the other two strains of eqn (16a), if $\tilde{u}^{(k)}(r,z)$ is compatible. Hence, the appropriate compatibility equation to be satisfied is eqn (16b) instead of eqn (16a). It is noteworthy that the compatibility equation (16a) is an Euler equation of the principle of minimum complementary energy, while eqn (16b) corresponds to an Euler equation of a yet to be derived variational principle intermediate between the principle of minimum potential energy and Reissner's variational principle.

The axial strain, $\varepsilon_z^{(k)}(r,z)$, is therefore, obtained by substituting $\tilde{u}^{(k)}(r,z)$ and $\bar{\gamma}_{rz}^{(k)}(r,z)$ given by eqns (13) and (14) into the integrated (with respect to r) version of the compatibility equation (16b) as follows:

$$\tilde{\varepsilon}_z^{(k)}(r,z) = \varepsilon_z^{(k)*}(r,z) + \frac{r^{1-\nu_k}}{E_k} F_1^{(k)}(z)'' + F_2^{(k)}(z);\ k = 1, 2,... \tag{17a}$$

where

$$\varepsilon_z^{(k)*}(r,z) = -\int \frac{\partial^2 u^{(k)*}(r,z)}{\partial z^2} dr + \int \frac{\partial \bar{\gamma}_{rz}^{(k)}(r,z)}{\partial z} dr;\ k = 0, 1, 2,... \tag{17b}$$

where $u^{(k)*}(r,z)$ and $\bar{\gamma}_{rz}^{(k)}(r,z)$ are given by eqns (13b) and (14), respectively. The corresponding normal stress, $\tilde{\sigma}_z^{(k)}(r,z)$, can be obtained by using Hooke's law as follows:

$$\tilde{\sigma}_z^{(k)}(r,z) = \sigma_z^{(k)*}(r,z) + r^{1-\nu_k} F_1^{(k)}(z)'' + E_k F_2^{(k)}(z) + \frac{\nu_k}{r^{1+\nu_k}} F_1^{(k)}(z);\ k = 1, 2, ... \tag{18a}$$

where

$$\sigma_z^{(k)*}(r,z) = E_k \varepsilon_z^{(k)*}(r,z) + \nu_k \{\bar{\sigma}_r^{(k)}(r,z) + \sigma_\theta^{(k)*}(r,z)\} = 0;\ k = 1, 2, ... \tag{18b}$$

It is noteworthy that although the axial strain, $\tilde{\varepsilon}_z^{(k)}(r,z)$, or its stress counterpart, $\tilde{\sigma}_z^{(k)}(r,z)$, satisfies the compatibility eqn (16), it is no longer in equilibrium (in the pointwise sense) with the stresses, $\bar{\tau}_{rz}^{(k)}(r,z)$and $\bar{\sigma}_r^{(k)}(r,z)$. These stresses are, therefore, rederived from the following two equilibrium equations:

$$\frac{\partial \tau_{rz}}{\partial r} + \frac{\tau_{rz}}{r} + \frac{\partial \sigma_z}{\partial z} = 0 \tag{19a}$$

$$\frac{\partial \sigma_r}{\partial r} + \frac{\sigma_r - \sigma_\theta}{r} + \frac{\partial \tau_{rz}}{\partial z} = 0 \tag{19b}$$

$\tilde{\tau}_{rz}^{(k)}(r,z)$ and $\tilde{\sigma}_r^{(k)}(r,z)$ can now be obtained from eqns (19) upon integration as follows:

$$\tilde{\tau}_{rz}^{(k)}(r,z) = \frac{1}{r}\left[-\int r \frac{\partial \tilde{\sigma}_z^{(k)}(r,z)}{\partial z} dr + F_3^{(k)}(z)\right] \tag{20a}$$

$$\tilde{\sigma}_r^{(k)}(r,z) = \frac{1}{r}\left[\int \tilde{\sigma}_\theta^{(k)} dr - \int r \frac{\partial \tilde{\tau}_{rz}^{(k)}(r,z)}{\partial z} dr + F_4^{(k)}(z)\right] \tag{20b}$$

Substitution of $\tilde{\sigma}_z^{(k)}(r,z)$, given by eqn (18) into the integrated equilibrium equation (20a) yields

$$\tilde{\tau}_{rz}^{(k)}(r,z) = \tau_{rz}^{(k)*}(r,z) - \frac{r^{2-\nu_k}}{3-\nu_k} F_1^{(k)}(z)''' - \frac{\nu_k r^{-\nu_k}}{1-\nu_k} F_1^{(k)}(z)' - E_k \frac{r}{2} F_2^{(k)}(z)' + \frac{F_3^{(k)}(z)}{r} \tag{21a}$$

where

$$\tau_{rz}^{(k)*}(r,z) = -\frac{1}{r}[\int r \frac{\partial \sigma_z^{(k)*}(r,z)}{\partial z} dr] \tag{21b}$$

Substitution of $\tilde{\sigma}_\theta^{(k)}(r,z)$and $\tilde{\tau}_{rz}^{(k)}(r,z)$ given by eqns (12) and (21), respectively, into the second integrated equilibrium equation (20b) yields

$$\tilde{\sigma}_r^{(k)}(r,z) = \sigma_r^{(k)*}(r,z) + \frac{r^{3-\nu_k}}{(3-\nu_k)(4-\nu_k)} F_1^{(k)}(z)'''' + \frac{\nu_k r^{1-\nu_k}}{(1-\nu_k)(2-\nu_k)} F_1^{(k)}(z)''$$

$$-\frac{r^{-1-\nu_k}}{\nu_k} F_1^{(k)}(z) + E_k \frac{r^2}{6} F_2^{(k)}(z)'' - F_3(z)' + \frac{F_4(z)}{r} \tag{22a}$$

where

$$\sigma_r^{(k)*}(r,z) = \frac{1}{r}[\int \sigma_\theta^{(k)*}(r,z)dr - \int r \frac{\partial \tau_{rz}^{(k)*}(r,z)}{\partial z} dr] \tag{22b}$$

The above procedure loses the end boundary conditions in the axial direction (i.e., at surfaces z = constant), a kind of mathematical Alzheimer's, thus rendering the boundary-value formulation ill-posed. This is due to the fact that the above boundary-value problem has now been transformed into an initial value problem, which is analogous to Hadamard's treatment of the Cauchy problem (see Tikhonov and Arsenin, 1979). This ill-posedness will be removed by deriving appropriate boundary constraint terms, which are functions of r, that help satisfy the end boundary conditions at surfaces z = constant as follows:

$$\tilde{\varepsilon}_z^{(k)}(r,z) = w_{,z}^{(k)}(r,z) = \frac{1}{E_k}\{\tilde{\sigma}_z^{(k)} - \nu_k(\tilde{\sigma}_r^{(k)} + \tilde{\sigma}_\theta^{(k)})\} \tag{23}$$

whence

$$w^{(k)}(r,z) = \frac{1}{E_k}\int\{\tilde{\sigma}_z^{(k)} - \nu_k(\tilde{\sigma}_r^{(k)} + \tilde{\sigma}_\theta^{(k)})\}dz + H_1^{(k)}(r) \tag{24}$$

Substitution of $\tilde{\sigma}_z^{(k)}(r,z)$, $\tilde{\sigma}_r^{(k)}(r,z)$ and $\tilde{\sigma}_\theta^{(k)}(r,z)$, given by eqns (18), (22) and (12), respectively into eqn. (24) gives

$$w^{(k)}(r,z) = \int \varepsilon_z^{(k)*}(r,z)dz + \frac{1}{E_k}\int\left\{r^{1-\nu_k}F_1^{(k)}(z)'' + E_k F_2^{(k)}(z) + \frac{\nu_k}{r^{1+\nu_k}}F_1^{(k)}(z)\right\}dz$$

$$-\frac{\nu_k}{E_k}\int\{\frac{r^{3-\nu_k}}{(3-\nu_k)(4-\nu_k)}F_1^{(k)}(z)'''' + \frac{\nu_k r^{1-\nu_k}}{(1-\nu_k)(2-\nu_k)}F_1^{(k)}(z)'' - \frac{r^{-1-\nu_k}}{\nu_k}F_1^{(k)}(z)$$

$$+E_k\frac{r^2}{6}F_2^{(k)}(z)'' - F_3(z)' + \frac{F_4(z)}{r} + \frac{1}{r^{1+\nu_k}}F_1^{(k)}(z)\}dz + H_1^{(k)}(r) \tag{25a}$$

where

$$\varepsilon_z^{(k)*}(r,z) = \frac{1}{E_k}[\sigma_z^{(k)*}(r,z) - \nu_k\{\sigma_r^{(k)*}(r,z) + \sigma_\theta^{(k)*}(r,z)\}] \tag{25b}$$

Next, using the relationship

$$u_{,z}^{(k)}(r,z) = \frac{1}{G_k}\tilde{\tau}_{rz}^{(k)}(r,z) - w_{,r}^{(k)}(r,z) \tag{26}$$

whence $u^{(k)}(r,z)$ can be obtained as follows:

$$u^{(k)}(r,z) = \frac{1}{G_k}\int[\tilde{\tau}_{rz}^{(k)}(r,z) - w_{,r}^{(k)}(r,z)]dz + H_2^{(k)}(r) \tag{27}$$

The stress, $\sigma_r^{(k)}(r,z)$, can now be obtained using Hooke's law

$$\sigma_r^{(k)}(r,z) = E_k\varepsilon_r^{(k)}(r,z) + \nu_k[\tilde{\sigma}_z^{(k)}(r,z) + \tilde{\sigma}_\theta^{(k)}(r,z)] \tag{28}$$

where

$$\varepsilon_r^{(k)}(r,z) = u_{,r}^{(k)}(r,z) \tag{29}$$

Finally, the stresses $\tau_{rz}(r,z)$ and $\sigma_z(r,z)$can be derived using the equations of equilibrium (19) as follows:

$$\tau_{rz}(r,z) = \frac{1}{r}\int[-\frac{\partial}{\partial r}\{r\sigma_r(r,z)\} + \tilde{\sigma}_\theta(r,z)]dz + H_3^{(k)}(r) \tag{30}$$

$$\sigma_z(r,z) = -\frac{1}{r}\int[\frac{\partial}{\partial r}\{r\tau_{rz}(r,z)]dz + H_4^{(k)}(r) \tag{31}$$

These expressions are derived using the symbolic language, MAPLE.

Core Region:

Pagano's (1991) stress field for the core layer (core, k = 0, always fiber) is as shown below:

$$\bar{\sigma}_z^{(0)}(r,z) = p_{11}^{(0)}(z)\frac{r_2 - r}{r_2} + p_{12}^{(0)}(z)\frac{r}{r_2} \tag{32a}$$

$$\bar{\tau}_{rz}^{(0)}(r,z) = p_{52}^{(0)}(z)\frac{r}{r_2} + p_{53}^{(0)}(z)(r - r_2)r \tag{32b}$$

$$\bar{\sigma}_r^{(0)}(r,z) = p_{31}^{(0)}(z)(\frac{r_2 - r}{r_2}) + p_{32}^{(0)}(z)\frac{r}{r_2} + p_{33}^{(0)}(z)(r^2 - r_2^2)r + p_{34}^{(0)}(z)(r - r_2)r \tag{32c}$$

The hoop stress, $\tilde{\sigma}_\theta^{(0)}(r,z)$, is derived using the same procedure as above:

$$\tilde{\sigma}_\theta^{(0)}(r,z) = r^{-(1+\nu_0)}\left[\int\left\{(1+\nu_0)r^{\nu_0}\bar{\sigma}_r(r,z) + \nu_k r^{(1+\nu_0)}\frac{\partial\bar{\sigma}_r(r,z)}{\partial r} + \nu_0 r^{(1+\nu_0)}\frac{\partial\bar{\sigma}_z(r,z)}{\partial r}\right\}dr + F_0^{(0)}(z)\right] \tag{33}$$

where $F_0^{(0)}(z)$, is a "constant" of integration (with respect to r) and an arbitrary function of z. The hoop stress in the core is obtained by substitution of $\bar{\sigma}_z^{(0)}(r,z)$ and $\bar{\sigma}_r^{(0)}(r,z)$ from eqns (32a,c) into the right side of eqn (33)

$$\tilde{\sigma}_\theta^{(0)}(r,z) = \sigma_\theta^{(0)*}(r,z) + \frac{1}{r^{1+\nu_f}}F_0^{(0)}(z) \tag{34a}$$

where

$$\sigma_\theta^{(0)*}(r,z) = (1+\nu_0)[p_{31}^{(0)}(z)(\frac{1}{\nu_0+1} - \frac{r}{r_2(\nu_0+2)}) + p_{32}^{(0)}(z)\frac{r}{r_2(\nu_0+2)} + p_{33}^{(0)}(z)(\frac{r^3}{\nu_0+4} - \frac{rr_2^2}{\nu_0+2})$$

$$+p_{34}^{(0)}(z)(\frac{r^2}{\nu_0+3} - \frac{r_2 r}{\nu_0+2})] + \nu_0[-p_{31}^{(0)}(z)\frac{r}{(\nu_0+2)r_2} + p_{32}^{(0)}(z)\frac{r}{r_2(\nu_0+2)} + p_{33}^{(0)}(z)(\frac{3r^3}{\nu_0+4}$$

$$-\frac{rr_2^2}{\nu_0+2}) + p_{34}^{(0)}(z)(\frac{2r^2}{\nu_0+3} - \frac{rr_2}{\nu_0+2}) - p_{11}^{(0)}(z)\frac{r}{r_2(\nu_0+2)} + p_{12}^{(0)}(z)\frac{r}{r_2(\nu_0+2)}] \tag{34b}$$

An examination of eqns (34) reveals that $\tilde{\sigma}_\theta^{(0)}(r,z)$ becomes unbounded at the core centerline, r = 0. Enforcement of the boundedness of $\tilde{\sigma}_\theta^{(0)}(0,z)$ reduces $F_0^{(0)}(z)$ to 0. However, it may be noted that unlike in the case of an annular region discussed earlier, the radial displacement component, $\tilde{u}^{(0)}(r,z)$, will not be obtained using the constitutive relation, given by eqn (10a), because this will result in an over-determined system by assigning one extra condition at r = constant boundary/interface.

Interlaminar shear strain, $\bar{\gamma}_{rz}^{(0)}(r,z)$ and the radial normal strain, $\bar{\varepsilon}_r^{(0)}(r,z)$, can now be obtained by substitution of the above stresses into Hooke's law:

$$\bar{\gamma}_{rz}^{(0)}(r,z) = \frac{1}{G_0}\bar{\tau}_{rz}^{(0)}(r,z) \tag{35a}$$

$$\bar{\varepsilon}_r^{(o)}(r,z) = \frac{1}{E_0}[\bar{\sigma}_r^{(o)} - \nu_0(\tilde{\sigma}_\theta^{(0)} + \bar{\sigma}_z^{(o)})] \tag{35b}$$

It may be noted that the above two strains, $\tilde{\varepsilon}_r^{(k)}(r,z)$ and $\bar{\gamma}_{rz}^{(k)}(r,z)$, are not compatible with the axial strain, $\bar{\varepsilon}_z^{(k)}(r,z)$, computed using Hooke's law

$$\bar{\varepsilon}_z^{(0)}(r,z) = \frac{1}{E_0}[\bar{\sigma}_z^{(0)} - \nu_0(\tilde{\sigma}_\theta^{(k)} + \bar{\sigma}_r^{(0)})] \tag{35c}$$

because the remaining compatibility equation, given by eqn (16a) is not satisfied. It may be remarked that in the absence of an exact solution to axisymmetric elasticity boundary-value problem, the appropriate variational principle will, as mentioned earlier, demand a different compatibility equation, such as eqn (16a) or (16b).

The axial strain is, therefore, obtained by substituting $\bar{\gamma}_{rz}^{(0)}(r,z)$ and $\bar{\varepsilon}_r^{(o)}(r,z)$ given by eqns (35 a,b) into the compatibility equation (16a) as follows:

$$\frac{\partial^2 \tilde{\varepsilon}_z^{(0)}}{\partial r^2} = -\frac{1}{E_0}.\frac{\partial^2}{\partial z^2}\{\bar{\sigma}_r^{(0)} - \nu_0(\tilde{\sigma}_\theta^{(0)} + \bar{\sigma}_z^{(0)})\} + \frac{1}{G_0}.\frac{\partial^2 \bar{\tau}_{rz}^{(0)}}{\partial r \partial z} \tag{36}$$

Further substitution of eqns 32(a,c) and (34) into eqn (36) and integration with respect to r twice lead to

$$\tilde{\varepsilon}_z^{(0)}(r,z) = \varepsilon_z^*(r,z) + rF_1^{(0)}(z) + F_2^{(0)}(z) \tag{37}$$

The corresponding stress, $\tilde{\sigma}_z^{(0)}(r,z)$, from

$$\tilde{\sigma}_z^{(0)}(r,z) = E_0\tilde{\varepsilon}_z^{(0)}(r,z) + \nu_0\{\bar{\sigma}_r^{(0)}(r,z) + \tilde{\sigma}_\theta^{(0)}(r,z)\} \tag{38}$$

can be expressed in the form:

$$\tilde{\sigma}_z^{(0)}(r,z) = \sigma_z^{(0)*}(r,z) + E_0\{rF_1^{(0)} + F_2^{(0)}\} \tag{39a}$$

with

$$\sigma_z^{(0)*}(r,z) = E_0\varepsilon_z^{(0)*}(r,z) + \nu_0\{\bar{\sigma}_r^{(0)}(r,z) + \tilde{\sigma}_\theta^{(0)}(r,z)\} \tag{39b}$$

It may be noted that the stresses $\tilde{\sigma}_\theta^{(0)}(r,z)$ and $\tilde{\sigma}_z^{(0)}(r,z)$, although compatible, are no longer in equilibrium with the remaining stresses, $\bar{\tau}_{rz}^{(0)}(r,z)$ and $\bar{\sigma}_r^{(0)}(r,z)$, given by eqns (32b,c). These stresses are obtained using the equilibrium equations of axisymmetric elasticity in a manner similar to an annular region, mentioned earlier:

$$\tilde{\tau}_{rz}^{(0)}(r,z) = \tau_{rz}^{(0)*}(r,z) - E_0\{\frac{r^2}{3}F_1^{(0)'}(z) + \frac{r}{2}F_2^{(0)'}(z)\} + \frac{F_3^{(0)}}{r} \tag{40a}$$

where

$$\tau_{rz}^{(0)*}(r,z) = -\frac{1}{r}[\int r\frac{\partial \sigma_z^{(0)*}(r,z)}{\partial z}dr] \tag{40b}$$

and

$$\tilde{\sigma}_r^{(0)}(r,z) = \sigma_r^{(0)*}(r,z) + E_0\{\frac{r^3}{12}F_1^{(0)''}(z) + \frac{r^2}{6}F_2^{(0)''}\} - F_3^{(0)'} + \frac{F_4^{(0)}}{r} \tag{41a}$$

where

$$\sigma_r^{(0)*}(r,z) = \frac{1}{r}[\int \sigma_\theta^{(0)*}(r,z)dr - \int r\frac{\partial \tau_{rz}^{(0)*}(r,z)}{\partial z}dr] \qquad (41b)$$

Finally, the displacement component, $w^{(0)}(r,z)$ is obtained by utilizing the following constitutive relations:

$$\tilde{\varepsilon}_z^{(0)}(r,z) = w_{,z}^{(0)}(r,z) = \frac{1}{E_0}\{\tilde{\sigma}_z^{(0)} - \nu_0(\tilde{\sigma}_r^{(0)} + \tilde{\sigma}_\theta^{(0)})\} \qquad (42)$$

The new $\tilde{\varepsilon}_z^{(0)}(r,z)$ is now given by

$$\tilde{\varepsilon}_z^{(0)}(r,z) = \varepsilon_z^{(0)*}(r,z) + rF_1^{(0)}(z) + F_2^{(0)}(z) - \nu_0\{\frac{r^3}{12}F_1^{(0)''}(z) + \frac{r^2}{6}F_2^{(0)''}\} + \frac{\nu_0}{E_0}\{F_3^{(0)'} - \frac{F_4^{(0)}}{r}\} \qquad (43a)$$

with

$$\varepsilon_z^{(0)*}(r,z) = \frac{1}{E_0}[\sigma_z^{(0)*}(r,z) - \nu_0\{\sigma_r^{(0)*}(r,z) + \sigma_\theta^{(0)*}(r,z)\}] \qquad (43b)$$

$w^{(0)}(r,z)$ can now be obtained by integrating $\tilde{\varepsilon}_z^{(0)}(r,z)$ as folows:

$$w^{(0)}(r,z) = w^{(0)*}(r,z) + r\int F_1^{(0)}(z)dz + \int F_2^{(0)}(z)dz - \nu_0\{\frac{r^3}{12}F_1^{(0)'}(z) + \frac{r^2}{6}F_2^{(0)'}\}$$
$$+ \frac{\nu_0}{E_0}\{F_3^{(0)}(z) - \frac{1}{r}\int F_4^{(0)}(z)dz\} + H_0^{(0)}(r) \qquad (44a)$$

with

$$w^{(0)*}(r,z) = \int \varepsilon_z^{(0)*}(r,z)dz \qquad (44b)$$

The remaining steps for derivation of $u^{(0)}(r,z)$, the stresses $\tau_{rz}(r,z)$ and $\sigma_z(r,z)$ are identical to those for an annular region, and will not be repeated here. These expressions are derived using the symbolic language, MAPLE. Finally, the unknown functions of z and r are currently being determined using MAPLE from boundary/interface conditions in the r- and z-directions, respectively.

This ends the first iteration or cycle. The stresses $\sigma_z^{(k)}(r,z)$ and $\sigma_r^{(k)}(r,z)$ computed above will now be substituted back into eqn (12) to compute the hoop stress $\tilde{\sigma}_\theta^{(k)}(r,z)$ in the second cycle, and the process can be continued for either a prescribed number of cycles, n, or until convergence is reached within certain pre-determined tolerance.

The above procedure introduces 4M "constants" of integration, $F_i^{(k)}(z)$, i = 1,...,4; k = 0, 1, 2,..., M-1 — 4 per iteration— which are functions of z, for each iteration. These are determined by using 4M appropriate boundary/interface conditions per iteration including those at the axis of symmetry.

Interfacing or Patching of Local Asymptotic Singular and Global Axisymmetric Micromechanical Stress Fields

Zak (1964) was the first to demonstrate interfacing a finite difference based global axisymmetric nonsingular stress field with the singular stress field obtained using a two-dimensional (plane strain) asymptotic solution due to Williams (1952). This approach was illustrated by Zak (1964) by matching the solution of his asymptotic analysis with the numerical finite difference analysis for the free-clamped boundary condition. Zak (1964) arbitrarily chose the point of matching at a distance of 2% of the thickness of the annular region for one of the three stresses, σ_z, σ_r and τ_{rz}, which would yield the stress intensity factor. The remaining two stresses computed using the asymptotic analysis, when multiplied with this factor, would match very closely their global axisymmetric counterparts computed using the finite difference analysis. Since the stress, σ_θ, cannot be derived using the plane strain asymptotic analysis, this stress was left out. Zak (1964) further remarked that if the finite difference based numerical results were available closer to the point of stress singularity, further improvement could be expected.

Pochiraju (1993) in continuing Zak's (1964) research computed the axisymmetric stress field using a highly refined finite element mesh near the point of stress singularity with the help of a commercially available package, ABACUS. He used the standard 8-noded quadrilateral elements with no special singularity formulation. Like Zak (1964), Pochiraju (1993 matched one stress component computed using the FEM (finite elements method) at a point on the fiber matrix interface at a very close distance of the order of 10^{-4} - 10^{-7} times the fiber radius with that obtained from the asymptotic analysis. The scaling or stress intensity factor thus computed when multiplied with the other stress components from the asymptotic analysis or with the same stress component at other angles led to matching of all the stress components from the two analyses at all angles at the same radius. Pochiraju (1993) also established the region of dominance for each term of the asymptotic analysis with an error function.

Although Zak's (1964) and Pochiraju's (1993) approach is effective in the context of two-dimensional finite difference and FEM models, such a brute force model cannot be used in situations such as the present variational axisymmetric model because of the afore-mentioned eigenvalue blow-up or in any three-dimensional modeling environment, where the cost of this degree of refinement will be prohibitive. The proposed research represents a novel attempt to recover a nearly "exact" elasticity solution by successively satisfying the equations of elasticity in the pointwise sense, as opposed to mean-square sense of the finite difference or FEM. The improved stress field can easily be matched with its asymptotic analysis counterpart without having to resort to highly refined layering approach of the existing numerical techniques.

3. Example Problem — Expected Results

As an illustration of the present approach and to examine the fidelity of its predictive capability, the body depicted in Figure 2 will be considered, and the improved results computed using the present compatibility/equilibrium based approach will be compared to the corresponding variational model solution due to Pagano (1991). All the external boundaries are assumed to be traction-free and that the body is subjected to a 1°C temperature rise. As a first step, the fiber is represented as a single solid cylinder or the core, while the matrix is represented as a single annular ring (Pagano's N=1). Both materials are assumed to be isotropic with the following properties

$$E_f = 413\text{GPa} \qquad \nu_f = 0.2 \qquad \alpha_f = 3.25\mu/°C$$
$$E_m = 63\text{GPa} \qquad \nu_m = 0.2 \qquad \alpha_m = 3.50\mu/°C$$

where the subscripts f and m stand for fiber and matrix, respectively, and the geometric parameters are taken as

$$\frac{l}{a} = 10 \qquad \frac{b}{a} = 2$$

The computed results (M = 2) will be compared with Pagano's (1991) multi-ring (e. g., M = 10, 20, etc.) solution in order to assess the degree of improvement of the computed stress field. Next, the effects of layering and number of iterations (n >1) on the convergence of the present solution will be assessed by way of comparison with the exact nonsingular series solution due to Kurtz and Pagano (1991). Finally, the improved "converged" stress field will be matched with the asymptotic singular stress field at a close distance of the order of 10^{-4}-10^{-7} times the fiber radius.

4. Closure

A semi-analytical iterative approach for enhancing the existing two-dimensional quasi-continuous axisymmetric stress field for a brittle matrix micro-composite (i. e., a single fiber surrounded by a concentric matrix cylinder), is proposed. In the present approach, the stress distribution in the radial direction obtained from the afore-cited variational model due to Pagano (1991) is improved *a posteriori* through an iterative approach that involves successive substitution of the previously computed strains (or stresses) into the equations of compatibility and equilibrium. A semi-analytical iterative approach for enhancing the existing two-dimensional quasi-continuous axisymmetric stress field for a brittle matrix micro-composite (i. e., a single fiber surrounded by a concentric matrix cylinder), is presented. The existing solution employs Reissner's variational theorem in conjunction with an equilibrium stress field in which the radial (r-) dependence is assumed *a priori.*

In the present approach, the stress distribution in the radial direction obtained from the afore-cited variational model is improved *a posteriori* through an iterative approach that

involves successive substitution of the previously computed strains (or stresses) into the equations of compatibility and equilibrium. The equations of compatibility are selected such that they form Euler equations corresponding to appropriate variational principle, such as the principle of minimum complementary potential energy, etc. The boundary/interface conditions at r = constant and z = constant surfaces/interfaces are satisfied in the pointwise sense. The expressions for the improved axisymmetric displacement and stress fields are derived using the symbolic language, MAPLE.

An illustrative thermal stress problem is currently being solved, and will be used to compare with the existing variational solution. When completed, this research will represent a novel semi-analytical post-processing tool to improve solution accuracy and numerical efficiency of existing variational solutions for micro-composites. The final results will be reported in Chaudhuri et al. (to be published).

References

C. Atkinson, J. Avila, E. Betz, and R. E. Smelser, "The Rod Pull Out Problem, Theory and Experiment," *J. Mech. Phys. Solids* , **30** (1982).

J. Aveston, G. A. Cooper, and A. Kelly, "Single and Multiple Fracture," *The Properties of Fibre Composites Conference Proceedings*, National Physical Laboratory, pp. 15-26. IPC Science and Technology Press Ltd, Guildford, UK (1971).

H. W. Brown III, "Analysis of Axisymmetric Micromechanical Concentric Cylinder Model," *Seventeenth Annual Mechanics of Composites Review, Air Force 86145/09-10-92-150,* Dayton, OH (1992).

B. Budiansky, J. W. Hutchinson, and A. G. Evans, "Matrix Fracture in Fiber-Reinforced Ceramics," *J. Mech. Phys. Solids*, **34**, 167 - 189 (1986).

R. A. Chaudhuri and P. Seide, "An Approximate Method for Prediction of Transverse Shear Stresses in a Laminated Shell," *Int. J. Solids and Structures*, **23**, 1145 - 1161 (1987).

R. A. Chaudhuri, "A Semi-Analytical Approach for Prediction of Interlaminar Shear Stresses in Laminated General Shells," *Int. J. Solids and Structures,* **26**, 499 - 510 (1990).

R. A. Chaudhuri, "A Novel Compatibility/Equilibrium Based Iterative Post-Processing Approach for Axisymmetric Brittle Matrix Composites," Summer Research Report submitted to RDL, October (1996).

R. A. Chaudhuri, N. J. Pagano, G. P. Tandon and A. H. Khan, "Interfacing of Local Asymptotic Singular and Global Axisymmetric Micromechanical Stress Fields in Brittle Matrix Composites," to be published.

H. L. Cox, "The Elasticity and Strength of Paper and Other Fibrous Materials," *British J. Appl. Phys.* , **3** (1952).

Z. Hashin and B. W. Rosen, "The Elastic Moduli of Fiber Reinforced Materials," *J. Appl. Mech.* **31** (1964).

R. D. Kurtz and N. J. Pagano, "Analysis of the Deformation of a Symmetrically Loaded Fiber Embedded in a Matrix Material," *Engr Composites* , **1** (1991).

V. K. Luk and L. M. Keer, "Stress Analysis for an Elastic Half Space Containing an Axially-Loaded Rigid Cylindrical Rod," *Int. J. Solids Structures* , **15** (1979).

L. N. McCartney, "New Theoretical Model of Stress Transfer Between Fibre and Matrix in a Uniaxially Fibre-Reinforced Composite," *Proc. R. Soc. London* **A 425**, 215 - 244 (1990).

N. J. Pagano, "Exact Solution for Composite Laminates in Cylindrical Bending", *J. Compos. Mater.* , **3**, 398 - 411 (1969).

N. J. Pagano, "Stress Fields in Composite Laminates," *Int. J. Solids Structures*, **14** (1978).

N. J. Pagano, "Axisymmetric micromechanical stress fields in composites," *Proceedings 1991 IUTAM Symposium on Local Mechanics Concepts for Composite Materials Systems*, p 1. Springer Verlag (1991).

N. J. Pagano and G. P. Tandon, "Elastic Response of Multi-directional Coated-Fiber Composites," *Comp. Sci. Tech..*, **31** (1988).

G. Pickett and M. W. Johnson, "Analytical Procedures for Predicting the Mechanical Properties of Fiber Reinforced Composites," *Technical Report AFML-TR-65-220* (1967).

E. Reissner, "On a Variational Theorem in Elasticity," *J. Math. Phys.*, **29** (1950).

G. E. Smith and A. J. M. Spencer, "Interfacial Tractions in a Fibre-Reinforced Elastic Composite Material," *J. Mech. Phys. Solids* , **18** (1970).

E. Sternberg, "Load-Transfer and Load-Diffusion in Elastostatics," *Proceedings of the Sixth U.S. National Congress of Applied Mechanics*, The American Society of Mechanical Engineers, New York (1970).

A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, John Wiley & Sons, New York (1977).

S. P. Timoshenko and J. N. Goodier, *Theory of Elasticity*, 2nd edn., McGraw-Hill, New York (1959).

A. R. Zak, "Stresses in the Vicinity of Boundary Discontinuities in Bodies of Revolution," *J. Appl. Mech.*, **31**, 150 - 152 (1964).

DETECTION TECHNIQUES USE IN FORWARD-LOOKING RADAR SIGNAL PROCESSING SYSTEM: A LITERATURE REVIEW

Mohammed Chouikha
Research Associate
Department of Electrical Engineering

Howard University
2600 6th Street NW
Washington, DC 20059

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

And

Wright Laboratory

September 1997

Detection techniques use in forward-looking radar signal processing system: A Literature Review

Mohammed Chouikha
Adedokun W Sule-Koiki, Ph.D Student
Howard University

Abstract

(FLAR) Forward-looking airborne radar system, as oppose to (SLAR) side looking airborne radar, allows high target to background contrast, accurate azimuth estimates, day and night operation, and can, to a limited degree, penetrate fog, haze, and dust. On the down side, forward-looking IR radar has range uncertainty, generate false alarm from background clutter, has difficult with occlusion of targets by vegetation and terrain, and is aspect angle dependent. The concept of detection and identification of targets, in nonstationary environment and obscure by inteferences such as clutter, jammer, and noise entails, not only suppressing the inteferences, but also classifying them into different clutters, jammer, and noise, with effective signal processing scheme.

In this report, we present a comprehensive review of the different STAP algorithms and computational learning-based methods, without analytical justification of these algorithms and methods, use for detection and classification of target obscured by interference such as clutter, jammer, and noise, applicable to forward-looking airborne radar system.

TABLE OF CONTENTS

I. Introduction....2-3

II Terminology and Signal Model....12-4

III Adaptive beamforming approach....12-6

A. Beamspace Processing....12-7

B. SMI Algorithm....12-9

C. Generalized Likelihood Ratio Test....12-11

D. Statistical Hypothesis....12-13

IV. Computational Learning Approach....12-14

A. Back-Propagation Algorithm....12-14

B. Self-Adaptive Recurrent Algorithm....12-16

V. Conclusion....12-16

VI. Future Work....12-17

References....12-17

Detection techniques use in forward-looking radar signal processing system: A Literature Review

Mohammed Chouikha
Adedokun W Sule-Koiki

I. INTRODUCTION

The general approach to detection of target(s) in nonstationary clutter and noise interference attempts to eliminate clutter and noise interference prior to detection by exploiting known differences in the statistical characteristics of target(s) relative to clutter and noise interference. One then performs detection and classification based on statistical models for the remaining residual clutter and noise interference. These models are typically based on experimental data.

It has been demonstrated, in SLAR (side looking airborne radar) applications, that space-time adaptive processing can result in weight solution equivalent to those required to perform DPCA. An effective and standard approach to clutter reduction makes use of the differential radial Doppler shift between target and clutter to discriminate against clutter, i.e. perform adaptive MTI processing. A standard MTI approach is only successful in achieve sufficient rejection over full clutter bandwidth at the expense of attenuation of the returns from slow moving targets. In SLAR applications, the DPCA (displaced phased center antenna) technique provide a mechanism for rejecting both mainlobe and sidelobe clutter by compensating directly for the motion of the antenna platform. These two methods mentioned are fully described in [1,2]

Adaptive beamforming is one of the many technique uses in a radar signal processing system. It involves forming multiple beams through applying appropriate delay and weighting elements to signal received by the sensors. The purpose is to suppress unwanted jamming inteferences and to produce the optimal beamformer response which contains minimal contributions due to noise. The most commonly employed technique for deriving the adaptive weights uses a closed loop gradient descent algorithm where the weight updates are derived from estimates of the correlation between the signal in each element and summed output of the array. This processing can be implemented in analog fashion using correlation loops or digitally in the form of the Widrow least mean square (LMS) algorithm [6]. The fundamental limitation for this technique is one of poor convergence for a broad dynamic range signal environment. The limitation was over come through the application of linear constraints to the weights. The basic concept of linearly constrained minimum variance (LCMV) beamforming is to constrain the response of the beamformer such that the desired signals are passed with specified gain and phase. The weights are chosen to minimize output power subject to the response constraint. When the beamformer has unity response in the look direction, the LCMV problem would become the minimum variance distortionless response (MVDR) beamformer problem, which is very general approach employed to control beamformer response. The weights of the beamformer should be updated in real-time in order to respond to rapid time-varying environment.

Meanwhile, the evaluation of weights is computational intensive and can hardly meet the real-time requirement. Systolic implementations of optimum beamformers have been studied to improve the computational speed by a number of investigators. McWhirter and Shepherd [7] showed how a triangular systolic array of the type proposed by Gentleman and Kung [8] can be applied to the problem of linearly constrained minimum variance problem, subject to one or more simultaneously linear equality constraints.

Tank and Hopfield and few other researchers [29and subref.] have shown how a class of neural networks with symmetric connections between neurons presets a dynamics that leads to the optimization of a quadratic functional. Chua et al in [30] and Kennedy et al in [31,32] extend the design of Hopfield network and introduce a canonical nonlinear programming circuit, which is able to handle more general optimization problems. Beck et al [34] presented a neural network approach to segmentation of forward-looking infrared and synthetic aperture radar imagery. Chang et al [35] presented a Hopfield-type neural network approach, which is similar to an analog circuit for implementing the real-time adaptive antenna, array. Davis et al [36] in applied a higher-order neural net work to the problem of 2-D target detection in noisy scene. Clark et al [39] presented the use of Gabor representations to generate feature vectors that are robust to variation in rotation, scaling, and translation. They described a prototype system for recognizing target in images by extracting image features, compressing the data, and classifying the targets using a neural network. Hara et al [37] in presented a terrain classification technique to determine terrain classes in polarimetric SAR images, utilizing unsupervised neural networks to provide automatic classification, and employing an iterative algorithm where the SAR image is reclassified using a Maximum likelihood (ML) classifier to improve the performance.

In Section II, a signal model useful for radar signal processing is presented. Section III describes adaptive beamforming. Section IV presents the computational learning methods and it importance to airborne radar. Section V concludes this report.

II Terminology and Signal model

Suppose it is desired to detect only the target signal in cases where the target Doppler is immersed in clutter/interference and noise and reject the other entire signal referred to as interference. Denote the sample data matrix $\mathbf{X}$, $N_t \times N_s$ is defined by

$$\mathbf{X} = \left[x_{t1} x_{t2} \cdots\cdots n_{tN_s}\right] = \begin{bmatrix} x_{s1}^T \\ x_{s2}^T \\ \vdots \\ \vdots \\ x_{sN_t}^T \end{bmatrix} \qquad 1$$

T denotes the transpose, and the row vectors of **X**, $x_{sn_t}^T$, $n_t = 1,2,\cdots,N_t$, are the snapshot obtained along The spatial channels. Under the signal-absence hypothesis H_o, the data matrix **X** consists of

Clutter/interference and noise components only, i.e.,

$$\mathbf{X = C + N} \qquad 2$$

C and N represent the clutter/interference and noise, respectively, and are assumed to be independent. Under the signal-presence hypothesis H_1, a target signal component also appears in the data matrix, i.e.,

$$\mathbf{X = AS + C + N} \qquad 3$$

A is an unknown complex constant representing the amplitude of the signal and S the signal matrix of unknown form. Under the assumptions, the $n_t n_s$ the entry of the signal matrix S has the following form:

$$s(n_t, n_s) = \exp\left[i2\pi(n_t - 1)\frac{2v}{\lambda PRF} + i2\pi(n_s - 1)\frac{d\sin\theta}{\lambda}\right], \qquad 4$$

v is the radial velocity of the target. θ is the direction of arrival of the target-return planewave with respect to the broadside of the array and λ the radar wavelength. Denoting

$$f_{st} = \frac{2v}{\lambda PRF} \qquad 5$$

and

$$f_{ss} = \frac{d\sin\theta}{\lambda} \qquad 6$$

f_{st} is the normalized Doppler frequency of the target signal and f_{ss} is the spatial frequency. S can be expressed by

$$S = s_s^T \otimes s_t \qquad 7$$

Where $\otimes$ is the Kronecker product, and

$$s_t = \left[1 \quad \exp(i2\pi f_{st}) \quad \ldots \quad \exp\left(i2\pi(N_t - 1)f_{st}\right)\right]^T \qquad 8$$

$$s_t = \left[1 \quad \exp(i2\pi f_{ss}) \quad \ldots \quad \exp\left(i2\pi(N_s - 1)f_{ss}\right)\right]^T \qquad 9$$

represent the signals in time and space. To detect signal S in received vector **X** one puts the received vector through a *filter* with weight

$$W = [w_1, w_2, \ldots\ldots w_n]^T. \qquad 10$$

The output of the filter W is the scalar

$$X = \sum_{k=1}^{n} w_k Z_k = W_T Z \qquad 11$$

T denotes the matrix transpose. For the case of signal noise alone, the expected values as

$$E[X] = \sum_{k=1}^{n} w_k E[z_k] = \sum_{k=1}^{n} w_k s_k = W_T S . \qquad 12$$

Similarly, the noise power or variance of X is

$$\begin{aligned} \sigma^2 &= E\left[|X|^2\right] - |E[X]|^2 \\ &= W_T^* E\left[N^*\right] E\left[N_T\right] W \\ &= W_T^* M W \end{aligned} \qquad 13$$

Where the asterisk denotes complex conjugate and M is the covariance matrix of the noise process i.e.

$$M = E\left[N^* N_T\right] = E\left[n_j^* n_k\right] \qquad 14$$

Based on the above concept, determining the weights with linear constraints to the weight vector called the linearly constraint minimum variance beamforming problem, which is usually formulated as

$$\begin{aligned} \min_{w} \quad & \phi(W) = W^H M W \\ \text{subject to} \quad & W^H S_o = r \end{aligned} \qquad 15$$

where H is the complex conjugate transpose of the vector, r is the a complex constant, S_o is the steering vector associated with the look direction and is given by [34]:

$$S_o = \left[1, \exp\left(j\frac{2\pi d}{\lambda_o}\cos\theta_o\right), \ldots, \exp\left(j\frac{2\pi d}{\lambda_o}(L-1)\cos\theta_o\right)\right]^T \qquad 16$$

where d is the element spacing, λ_o is the wave length of the plane wave in free space, and θ_o is the look direction angle (the angle between the axis of the linear array and the direction of the desired signal source). A solution of equation ( ), after using method of Lagrange multiplier, is given by:

$$\hat{W} = r\frac{M^{-1}S_o}{S_o^H M^{-1} S_o} \qquad 17$$

The problem in airborne radar signal processing is therefore to detect target while eliminate the unwanted signal returns commonly referred to as clutter, interference and noise.

III. ADAPTIVE BEAMFORMING [22]

The first adaptive system for cancellation of clutter with an unknown Doppler frequency is TACCAR (Time Averaged Clutter Coherent Airborne Radar) [20]. Phase differences between subsequent echoes are used

to adjust the COHO (coherent oscillator) frequency so that the mid-Doppler frequency becomes zeros and the clutter spectrum falls into the notch of the two-pulse canceler. In this section we will present some algorithm for adaptive beamforming.

A. BeamSpace

Some results on space-time adaptive processing have been presented in [1], [2], [11], [21], [25], [28.]. Earlier work of Reed, Mallet and Brennan (RMB) [3] described an adaptive array procedure for the detection of a signal of known form in the presence of noise (interference) which is assumed to be Gaussian, but whose covariance matrix is totally unknown. They discussed how adaptive array processing is used to achieve optimum detection as applicable to airborne radar signal processing problem The optimum decision criterion for detecting the signal S in the presence of noise begin by formulating the likelihood ratio test given by

$$L = \exp\left[\frac{1}{2\sigma^2}\left(2|S\|X|\cos(\theta-\delta)-|S_1|^2\right)\right] \qquad 18$$

Where $S_1 = W_T S$, $\theta = \arg(X)$; $\delta = \arg(S_1)$

The test is formed by averaging L with respect to θ and by comparing the result with a threshold. That is if

$$\int L(X/\theta)P(\theta)d\theta \geq c > 0 \qquad 19$$

One can then say a signal is detected: if this quantity is less than c, X is said to result from noise alone. L is therefore given by

$$L = \exp\left(\frac{-|S_1|^2}{2\sigma^2}\right) I_0\left(|S_1\|X|\right) \geq c \qquad 20$$

The probability of a false alarm P_F and the probability of detection P_D are given by

$$\begin{aligned} P_F &= prob\{|X| > c|Z = N\} \\ &= \int_c^\infty \int_0^{2\pi} \frac{1}{2\pi\sigma^2}\exp\left(\frac{-r^2}{2\sigma^2}\right) r dr d\theta \\ &= \exp\left(\frac{-c^2}{2\sigma^2}\right) \end{aligned} \qquad 21$$

$$P_D = prob\{|X| > c|Z = S+N\}$$

$$= \int_c^{\infty} \int_0^{2\pi} \exp\left(\frac{-1}{2\sigma^2}\left|re^{j\theta} - ae^{i\delta}\right|\right) \frac{rdrd\theta}{2\pi\sigma^2}$$

$$= \frac{1}{\sigma^2} \int_c^{\infty} I_0\left(\frac{ra}{\sigma^2}\right) \exp\left(\frac{-r^2 + a^2}{2\sigma^2}\right) rdr \qquad 22$$

$$a = |S_1| = |W_T S|; \text{ and } \sigma^2 = W_T^* M W$$

From this, an expression for integrated signal-to-noise ratio α^2 is given by

$$\alpha^2 = \left(\frac{S}{N}\right) = \frac{|W_T S|^2}{W_T^* M W} \qquad 23$$

In Q function, the probability of detection is then expressed as

$$P_D(\alpha) = Q\left(\alpha, \sqrt{2\log\frac{1}{P_F}}\right). \qquad 24$$

The bound on signal-to-noise is obtained to be

$$\max_w \alpha^2 = S_T M^{-1} S^*, \qquad 25$$

If we let

$$W = kM^{-1}S^* \qquad 26$$

Therefore, the maximum value of probability of detection P_D is given by

$$\max_w P_D = Q\left(\sqrt{S_T^* M^{-1} S^*}, \sqrt{2\log\frac{1}{P_F}}\right). \qquad 27$$

From the above expression, it was explicitly realized that matrix inversions of estimates of M often are not practical in real time. Instead, a recursive relaxation algorithm that simultaneously estimates M and recursively compute W was developed using the method of steepest ascent. The value of the weight W (j+1) is determined by the relationship given by:

$$W(j+1) = W(j) + \frac{1}{2}\mu(j)\tilde{\nabla}F[W(j)] \qquad 28$$

For j=1,2,3... $\tilde{\nabla}F[W(j)]$ is the "complex gradient" and is given by

$$\tilde{\nabla}F = 2\left(\frac{W_T S}{W_T^* M W}\right)\left[S^* - \left(\frac{W_T^* S^*}{W_T^* M W}\right) M W\right] \qquad 29$$

The best choice of $\mu(j)$ is the value of a real variable μ, which maximizes the function

$$F\left[W(j-1)+\mu\tilde{\nabla}F(W(j-1))\right] \qquad 30$$

For j = 1,2,3...

In some adaptive systems, linearization supplies the reduction in computational complexity needed to make the system practical. Therefore, a linearized version of the algorithm is given by

$$W(j-1)=W(j)+\mu a\left[S^{*}-a^{*}M(j)W(j)\right] \qquad 31$$

a is a complex scalar and $M(j)$ is the statistical estimate of the covariance matrix. To determine $M(j)$, the input data is assumed to be the sequence of independent vector $Z(j)$ for j = 0,1,2,

In order to apply this algorithm to airborne radar, the algorithm developed is equivalently expressed as a vector different equation and the form is given by:

$$\frac{d}{dt}W(t)+\frac{\mu}{\Delta}|a|^{2}M(t)W(t)=\frac{\mu}{\Delta}aS^{*} \qquad 32$$

Δ is the radar pulse width for the airborne radar problem.

B. SMI (Sample Matrix Inversion) Algorithm

Reed, Brennan and Mallet [4] described procedure in which the convergence rate, which limits the practical usefulness of adaptive arrays systems, is most severe with a large number of degrees of adaptivity and in situations where the eigenvalues of the noise covariance matrix are widely different. A direct theory of adaptive weight computation based on a sample covariance matrix of the noise field was determined to provide very rapid convergence in all cases i.e. independent of the eigenvalue distribution. The adaptive array filter was achieved by first estimating the covariance matrix M using K samples. Next, this estimate $\hat{M}$ of M is inverted to form finally the filter weight given by:

$$\hat{W}=K\hat{M}^{-1}S. \qquad 33$$

In order to obtain the estimate $\hat{M}$ to substitute for M *above of the optimum filter* W_o, one uses the maximum likelihood principle. The maximum likelihood estimate of M is given by:

$$\hat{M}=1/K\sum_{j=1}^{K}X^{(j)}X^{(j)^{*}}$$

$$=1/K\sum_{j=1}^{K}\left(x_r^{(j)}\bar{x}_s^{(j)}\right) \qquad 34$$

where $x_r^{(j)}\bar{x}_s^{(j)}$ denotes the elements in the *r*th row and *s*th column of matrix $X^{(j)}X^{(j)^{*}}$. This result is obtained by maximizing the algorithm of the likelihood function

$$L=P\left(X^{(1)},X^{(2)},\ldots\ldots X^{(K)}\right). \qquad 35$$

$\hat{M}$ is called the sample covariance matrix. The output signal-to-noise ratio, conditioned on $\hat{W}$, is given by:

$$\left(S/N|\hat{W}\right)_0 = \left[E\left(\hat{y}|\hat{W}\right)\right]^2 / Var\left(\hat{y}|\hat{W}\right)$$

$$= \left(S^*\hat{M}^{-1}S\right)^2 \Big/ \left[\left(S^*\hat{M}^{-1}M\hat{M}^{-1}S\right)\right] \qquad 36$$

and the signal-to-clutter ratio [the signal-to-noise ratio] is then given by:

$$S/C = \left|\hat{W}^*S\right|^2 / \hat{W}^*M\hat{W} \qquad 37$$

Formation of the sample covariance matrix requires $SN(N+1)/2$ complex multiplication, where S is the number of samples required for convergence and N is the total number of weights, i.e. antenna elements times pulses for a space time processor. To invert the matrix requires $(N^3/2 + N^2)$ complex multiplication, and to form each set of weights requires another N^2 multiplication. If only one set of weights is required, i.e., no filter banks or multiple beams are needed, then the weights can be found by solving the set of linear equations $W = \hat{M}^{-1}S$, which requires about $N^3/6$ complex computations. Forming a sample covariance matrix and solving for the weights provides a very fast convergence. This rate is dependent only on the number of weights and is independent of the noise and interference environment.

In another paper [12], Brennan *et al* discussed the ability of airborne moving target indication (AMTI) radar to reject clutter is often seriously degraded by the motion of the radar. The AMTI technique adapts the element(s) weights to compensate for the near-field scattering. Array excitation errors due to phase or amplitude differences between channels are sensed and compensated automatically in the adaptive AMTI system. The system has the capability of nulling discrete active interference sources without significantly degrading the AMTI performance.

Three well-known methods of approximating M^{-1} (sample covariance matrix, updated inverse, and adaptive loops) are compared. It was shown that adaptive loops (or gradient techniques) show poor convergence. The updated inverse algorithm is given by

$$\hat{M}^{-1} = \frac{1}{1-\alpha}M^{-1} - \frac{\alpha}{1-\alpha}\frac{\hat{M}_{t-1}^{-1}X_tX_t^H\hat{M}_{t-1}^{-1}}{1-\alpha+\alpha X_t^H\hat{M}_{t-1}^{-1}X_t} \qquad 38$$

The inverse sample covariance matrix is

$$\hat{M}^{-1} = \left(\frac{1}{t_{max}}\sum_{t=1}^{t_{max}}X_tX_t^H\right)^{-1} \qquad 40$$

It was shown that the inverse sample covariance matrix is slightly better than the updated inverse algorithm.

However, the gain maximum is reached in both cases after the same number of iterations (or data vectors X_t). The updated inverse algorithm needs about $(NM)^2$ operations per iteration step, whereas the number of operation required for inverse sample covariance matrix is greater than $(NM)^3$. The important step in using the updated inverse algorithm is the product $Y = M^{-1}X_t$, which is used for updating M^{-1} and is the filter operation at the same time. Y has to be multiplied by a combined beamformer/DOPPLER filter matrix, which can easily be realized by a double FFT (space and time) if a linear array with equidistant sensors and equidistant pulses are used.

C. *Generalized likelihood Processing*

The work of Wang and Cai in [24.26] pointed out that for the cases of limited training-data set the use of localized adaptive processing is almost mandatory, and they showed that localized adaptive processing can actually outperformed fully adapuve processing in nonstationary environments. They studied the problem of achieving the optimum moving target indicator (MTI) detection performance in strong clutter of unknown spectrum when the set of data available to the estimation of clutter statistics is small due to a severely non-homogeneous environment. They proposed an adaptive implementation called Doppler domain localized generalized likelihood processor and its detection performance was studied. They presented a processor refereed to as joint-domain optimum processor and which such as the adaptive algorithms such as the sample-matrix-inversion (SMI) can approach. the generalized likelihood (GLR) and the modified SMI. They picked the GLR because it offers the desirable embedded CFAR feature as well as robustness in non-Gaussian clutter/interference. The $N_l th$ order GLR processor performs adaptive filtering and threshold detection on the N_l bins of the *l*th group (i.e. angle- Doppler bins around the look direction) with the test statistics given by:

$$\eta_{nm}^{(l)} = \frac{\left|Vec\left(S_{nm}^{(l)}\right)^H \hat{R}_l^{-1} Vec(\chi_l)\right|^2}{Vec\left(S_{nm}^{(l)}\right)^H \hat{R}_l^{-1} Vec\left(S_{nm}^{(l)}\right)\left[1 + Vec\left(\chi_l^H\right)\hat{R}_l^{-1} Vec(\chi_l)\right]} \qquad 41$$

$$\hat{R}_l = \sum_{k=1}^{K} Vec(\gamma_{lk}) Vec(\gamma_{lk})^H \qquad 41a$$

and $S_{nm}^{(l)}$, $N_{ll} x N_{so}$, is the signal-steering matrix in the angle-Doppler domain for the *nm*-th bin of the *l*th GLR. The probability of detection at the *nm*th bin of the *l*th is found to be

$$P_d^{(l)}(n,m) = \int_1^0 P_{d/\rho}^{(l)}(n,m) f_{nm}^{(l)}(\rho) d\rho \qquad 42$$

The probability of false alarm for all bins in the lth GLR is given by

$$P_f^{(l)} = \left(1 - \eta_0^{(l)}\right)^{K-N_l+1} \qquad 43$$

Three facts can be concluded from the GLR approach: 1) The ability to de-couple the degrees of freedom necessary for handling suppression from data dimension was possible through the transformation of data from space-time to angle-Doppler domain, 2) The degrees of freedom necessary for handling clutter suppression at each angle-Doppler bin is much smaller and 3) The clutter components on all angle-Doppler bins can now be correlated, with closer bins having higher correlation in general. Klemm [14,15] in his paper described an adaptive airborne MTI (AMTI) which is based on the principle of two-dimensional radar signal processing. He suggested that the back-scattered echo field have to be sampled in space and time. Space-Time sampling is necessary because ground clutter echoes are Doppler-colored and depend on two parameters (azimuth and clutter velocity). His method is a generalization of the well-known sidelobe canceled technique [5] to two-dimensional (space-time) sampled field. Before clutter suppression the received echo samples of all sensor outputs are transformed into a vector space of much lower than NM. This is done by use of auxiliary channels matched in space and velocity to the clutter returns. Clutter rejection is carried out in the reduced vector space [15,18] described a technique of adaptive clutter suppression for airborne phased array radar. A statistical model of clutter returns as received by airborne phased array antenna is given. The model consists simply of the space-time covariance matrix samples is given in the literature.

The spatial information is contained in the submatrices $M(\tau)$, whereas the temporal information lies between them. The individual elements of M are given by integrals of the form

$$m_{ikr} = m(\tau) \int_{\phi=0}^{\pi} H^2(\phi) F(\phi,\tau) C_D(\phi,\nu,\tau) C_s(\phi,\tau) d\phi \qquad 44$$

where C_D is the phase change due to the Doppler effect in the interval τ and C_s is the phase difference due to the relative geometric displacement of target t at time t and element k at time t + τ. Once the clutter covariance matrix is obtained, one may get a first impression of the nature of the clutter echoes by calculating a power spectrum given by:

$$P(\nu,\phi) = W^H(\nu,\phi) M W(\nu,\phi) \qquad 45$$

with $W(\nu,\phi)$ being a space-time steering vector with elements given by:

$$W_{ikn}(v,\phi) = \exp(j2\pi\phi_D(v)nT)\exp\left[j\frac{2\pi}{\lambda}(x_i\cos\phi\sin\theta + y_k\sin\phi\sin\phi)\right] \quad 46$$

where T is the pulse-to-pulse interval. The above expression for power spectrum gives a poor resolution. For a high resolution, the power spectrum is given by:

$$P(v,\phi) = \frac{1}{\left|W^H(v,\phi)ME\right|^2} \quad 47$$

with E being a unity vector.

In [22], Short presented a digital realization of an adaptive clutter-locking MTI canceler. The technique proved to be effective in tracking and canceling a unimodal clutter spectrum. The probability of detection was

derived analytically for the adaptive canceler. Averaging this detection probability over all possible target Doppler frequencies, the average probability of detection was found as a function of the input target-to-clutter ratio.

D. Statistical Hypothesis Testing [17,19]

By using the techniques of statistical hypothesis testing, it was shown that the test exhibits the desirable property that its probability of false alarm (PFA) is independent of the covariance matrix of the actual noise encountered. It was shown that the effect of signal present depends only on the dimensional parameters of the problem and a parameter, which is the same as the *SNR* of a colored noise match filter. It was emphasized that the output of the likelihood ratio algorithm is a decision on signal presence and not a sequence of processed data samples from which the interference component has been nulled and in which actual signal detection remains to be accomplished. For this reason, the direct application of the algorithm to a real radar problem would require the storage of data from an array of inputs (such as adaptive array system), perhaps also sampled to form range-gated outputs for each pulse and collected for a sequence of pulses (such as those forming a coherent processing interval). Under two data sets: 1) primary data which contains the signal and noise interference and 2) secondary data which contain only noise interference, he obtained the likelihood ratio test in the form:

$$\max_b \ell(b) = \frac{\|T_0\|}{\max_b \|T_1\|} > \ell_0 \quad 48$$

where $T_0 and T_1$ are given in [12] and b is an unknown complex scalar amplitude.

The probability of detection for their test is given by:

$$P_D = 1 - \frac{1}{\ell_0^L} \sum_{k=1}^{L} \binom{L}{k} (\ell_0 - 1)^k H_k\left(\frac{a}{\ell_0}\right). \qquad 49$$

In this formula, the H functions are the expected values of the Gs:

$$H_k(y) = \int_0^1 G_k(ry) f(r) dr, \qquad 50$$

where $G_k and f(r)$ are given in the reference. The performance of the likelihood ratio test depends only on the dimensional integers N and K and the SNR parameter a. The latter is a function of the true signal strength and the intensity and character of the actual noise and interference.

IV. COMPUTATIONAL LEARNING-BASED APPROACH

In real-time, the presence of changing environment such as clutter and uncorrelated noise ensures that W may be invertible. Moreso, the beamforming problem defined in (15) and (17) is indeed a complex-value constraint quadratic problem, which cannot be solved by neural network directly. In order to meet the requirement of neural network-based optimizer, one should convert it into a real-value constrained quadratic programming formulation. A detailed analysis and illustration of this formulation is given in [34]. To allow a beamformer to respond to a rapid time-varying environment, the weights should be adaptively controlled to satisfy (17) in real time. Meanwhile, the evaluation of these weights is computationally intensive and can hardly meet the real-time requirement.

Artificial Neural Networks lend themselves nicely to problem of segmentation of radar images based on the classification of measurements of FLAR as their 'model freedom' and capacity to learn from labeled training data can help overcome the lack of sufficiently accurate statistical models for real world radar measurements. Hence, there has been a number of studies in which ANNs successfully have been applied to the classification radar signals. Backpropagation networks have already been applied successfully to the classification of radar signal, e.g. in [44], [43]. Learning vector quantization (LVQ) networks have also been applied to target classification from radar backscatter measurements [46]. Adaptive neural network have been applied to for radar detection and classification [40].

A. Back-Propagation Algorithm

Clark et al described [37] a prototype system for recognizing targets in radar images by extracting target features, compressing the data, and classifying the target using neural network. This requires a training phase and a recognition phase. Two key issues in ATR are the following :1) Making the recognition process robust with respect to image and target variations, including orientation/position, scale/size, perspective, occlusion, contrast, background, and noise. 2) Minimizing the computational complexity of the process via

data compression, so that real-time operation can be approached. The Gabor transform was used to create feature vectors for the neural network. A two-dimensional Gabor filter is the product of a Gaussian-shaped window and a complex exponential term:

$$\psi(x,y,\sigma,k_x,k_y)=\left\{\exp\frac{-\left(k_x^2+k_y^2\right)}{2\sigma^2}\left(x^2+y^2\right)\right\}\exp\left(j\left(k_x x+k_y y\right)\right). \qquad 51$$

where *(x,y)* are the variables representing position in spatial domain, (k_x,k_y) are the wavenumber, corresponding to spatial frequencies, and σ is the Gaussian window parameter. The Gabor transform $G(x,y)$ of the radar image $\mathrm{I}(x,y)$ is then defined as the convolution of a Gabor filter $\psi(x,y)$ with $\mathrm{I}(x,y)$:

$$G(x,y)=\psi(x,y)**\mathrm{I}(x,y). \qquad 51$$

where ** denotes two-dimensional linear convolution. The resultant image is a complex image. However, only the magnitude of the transform was used because any extra information in the phase- appears to be not worth the storage costs required to save it. Next, a data block was formed by using an ensemble of Gabor transforms. Selecting a column from the data block then created feature vectors. In this application, Artificial Neural Network of backpropagation was used. The feature vectors were then used as inputs.

A neural network approach [33] to segmentation of forward-looking infrared (FLIR) and SAR imagery is reported. This approach integrates three stages of processing. First, a wavelet transform of the image is performed by projection of the image onto a set of 2-D Gabor functions. This results in multiple-resolution decomposition of the image into oriented, spatial frequency channels. Second, a neural network optimization procedure, which is based on gradient descent, is used to estimate the wavelet transform coefficients. The third stage involves a segmentation technique which is accomplished by first projecting the image onto a set of vectors, which decompose the images based on the localized orientation and spatial frequency properties of the projection vectors. The resulting coefficients are sufficient for image reconstruction and region characterization. Finally, the amplitude of the vector responses in the multiple resolution reconstruction and the density of those amplitudes are use to segment the radar image using neural network.

Haykin et al in [39,40] constructed a neural network classifier to successfully distinguish between major classes of radar returns including weather, birds, and aircraft. This classifier incorporates both preprocessing and postprocessing approach. They presented result of an experimental studies aimed at the classification of radar clutter as experienced in air traffic control environment, described a neural network clutter classifier simulated on a Warp systolic computer. The neural network is trained with a modified version of the backpropagation algorithm. In the modified back-propagation algorithm, the weight and learning rate

updating rules are summarized as follows:

$$w_{ij}(n+1) = w_{ij}(n) + \eta_{ij}(n+1)\sum_{b=1}^{P} \delta_{bj}(n) y_{bi}(n) + \alpha \Delta w_{ij}(n-1) \quad 53$$

$$\eta_{ij}(n+1) = \eta_{ij}(n) + \Delta \eta_{ij}(n) \quad 53a$$

$$\Delta \eta_{ij}(n) = \begin{cases} \beta \ldots\ldots\ldots\ldots\ldots\ldots S(n-1)D(n) > 0 \\ -\phi \eta_{ij}(n) \ldots\ldots\ldots\ldots\ldots S(n-1)D(n) > 0 \\ 0 \ldots\ldots\ldots\ldots\ldots\ldots otherwise \end{cases} \quad 53b$$

$$D(n) = \sum_{b=1}^{P} \frac{\partial E_b(n)}{\partial v_{ij}(n)} \quad 53c$$

$$S(n) = (1-\theta)D(n) + \theta S(n-1). \quad 53d$$

The detailed description and implementation of this algorithm is in [41].

B. Self-Adaptive Recurrent Algorithm [38]

Ziemke presented an approach to segmentation and integration of radar images using a second-order recurrent artificial neural network architecture, which consisted of two subnetwork: a function that classifies radar measurement into different categories of objects in sea environment, and a context network that dynamically computes the function network's input weights. For the purpose of target classification basically three major values/features, reflecting the characteristic of the illuminated object or area, can be extracted from theses radar spectra [43]. These are radial velocity, intensity and spectrum width.

Training of SARN takes place in each time step as follows:

1) Propagate input forward through the function network to calculate output vector.
2) Compare output vector to target vector.
3) Backpropagate the error (difference between actual and desired output) through the function network to update W and input weights.
4) Use new input weights as target output for the context network, i.e. backpropagate error (difference between update and previous weights) through the context network to update.
5) Propagate state unit values forward through the context to calculate next time step's input weights.

V. CONCLUSION

The objective of this report has been to perform literature review and put into perspective the existing methods of detection and identification of target(s) signal in a nonstationary environment with interference background use in forward-looking airborne radar signal processing. In such an environment, the origin of the measurements can be uncertain: they could have come from the target(s) of interest or clutter or false alarm or be due to the background. Particular attention is paid to the specified existing method(s) and it's use in and

applicable to FLIR airborne radar signal processing problems.

VI. FUTURE WORK

Analysis of candidate methods using synthetic data will be performed by computer simulation. Understanding will follow this; preprocessing and reformatting of real data which will take us into an extensive simulation of learning methods using real data.

REFERENCES

1. Skolnik. M., "Radar Handbook"*(McGraw-Hill, 1970), pp. 18.1-18.16*
2. W. Tam and D. Faubert, "Displaced phased center antenna clutter suppression in space based radar applications," *Proceedings of Radar '87, IEE conference publication 281, pp. 385-389*
3. Brennan and I S. Reed, "Theory of adaptive radar" *IEEE Trans. on Aerospace and Electronic Systems. vol. AES-9, March, 1973*
4. Reed, J. D Mallett, I. E. Brennan . " Rapid Convergence rate in adaptive arrays" *IEEE Trans. on Aerospace and Electronic Systems, vol. AES-10. No 6, March, 1974*
5. S. P. Applebaum, "Adaptive array," *IEEE Trans. Antennas Propagate. Vol. AP-24, pp. 585-598, Sept. 1976.*
6. B. Widrow and S. Stearns, *Adaptive Signal processing. Englewood Cliff, NJ: Prentice-Hall; 1985*
7. J. G. McWhirter and T. J. Shepherd, "Systolic array processor for MVDR beamforming," *Proc. Inst. Elec. Eng., vol. 136, pt. F, no. 2, Apr: 1989*
8. W. M. Gentleman and H. T. Kung, "Matrix triangularisation by systolic array," *Proc. SPIE, Real time signal processing IV, 1981, p.298*
9. Andrews, G. A., "Radar pattern design for platform motion compensation", *IEEE Trans., 1978, AP-26, pp. 566-571*
10. Zeger, A. E, and Burgess, L. R., "A adaptive AMTI radar antenna array", *NAECON 74 record, pp. 126-133*
11. Chapman, "Adaptive Arrays and Side-lobe Cancellers: A Perspective," Microwave J., Aug. 1977
12. Brennan, J. D. Mallet, I. S. Reed, "Adaptive Arrays in Airborne MTI Radar", *IEEE Trans. On Antenna and Propagation, vol. AP-24, No. 5 September 1976*
13. Boroson, "Sample Size Consideration for Adaptive Arrays", *IEEE Trans. on Aerospace and Electronic Systems, vol. AES-16, No. 4, July, 1980*
14. R. Klemn, Dr. Ing. "Adaptive clutter suppression for airborne phased array radar", *IEE Proc., vol.*

130, Pts. F and H, no. 1, February *1983*

15. R. Klemn, Dr. Ing, "Adaptive Airborne MTI: an auxiliary channel approach", *IEE Proc., vol. 134, Pts. F, no. 3,* June *1987*
16. J. Ender and R. Klemn, Dr. Ing. "Airborne MTI via digital filtering", *IEE Proc., vol. 136, Pts. F, no. 1,* February *1988*
17. Kelly, "An Adaptive Detection Algorithm", *IEEE Trans. on Aerospace and Electronic Systems, vol. AES-22,No. 1, March, 1986*
18. Short, R. D., "An Adaptive MTI for Weather Clutter suppression", *IEEE Trans. on Aerospace and Electronic Systems, vol. AES-18, No 5, September, 1982*
19. Kelly, "Performance of an Adaptive Detection Algorithm; Rejection of Unwanted Signals", *IEEE Trans. on Aerospace and Electronic Systems, vol. AES-25, No 2, March, 1989*
20. Dickey, M Labitt, F. M. Staudaher, "Development of Airborne Moving Target Radar for Long Range Surveillance", *IEEE Trans. on Aerospace and Electronic Systems, vol. AES-27, No 6, November 1991.*
21. Kalson, "An Adaptive Array detector with Mismatched Signal Rejection" *IEEE Trans. on Aerospace and Electronic Systems, vol. AES-28, No 6, No. 1 Januarv. 1992.*
22. Feldman, L J. Griffiths, " A Projection Approach for Robust Adaptive Beamforming", *IEEE Trans. On Signal Processing, vol. 42, No. 4 April 1994*
23. Addio, M. Di Bisceglie, S. Bottalico, "Detection of Moving Objects with airborne SAR", *Signal Processing 36 (1994) 149-162*
24. Wang H., L. CAI, "On Adaptive Spatial-Temporal Processing for Airborne Surveillance Radar Systems", *IEEE Trans. on Aerospace and Electronic Systems, vol. AES-30, No. 3 July. 1994*
25. Barile, R.L. Fante, J. A. Torres, "Some Limitation on the Effectiveness of Airborne Adaptive Radar" *IEEE Trans. on Aerospace and Electronic Systems, vol. AES-28, No 28, October. 1992*
26. Wang, L. CAI, "On Adaptive Spatial-Temporal Processing for Airborne Surveillance Radar System", *IEEE Trans. on Aerospace and Electronic Systems, vol. AES-28, No 30, July. 1994*
27. P G Richardson "Analysis of the Adaptive Space Time Processing Technique for airborne radar", *IEE Proc.-Radar, Sonar, Navig, vol. 141, No. 4, August 1994*
28. J. Ward, "Space-Time Adaptive Processing for airborne radar," MIT, Technical Report 1015, DEC 1994
29. D. W. Tank and J. J. Hopfield, "*Simple 'neural' optimization network: An A/D converter, signal decision circuit, and a linear programming circuit,*" *IEEE Trans. Circuit Syst., vol. CAS-33, pp. 533-541, May 1986*
30. L. O. Chua and G. N. Lin, "Nonlinear programming without computation," *IEEE Trans. Circuit Syst., vol. CAS-31, PP 182-188, Feb. 1984.*
31. M. P. Kennedy and L. O. Chua, "Unifying the Tank and Hopfield linear programming network and the canonical nonlinear programming circuit of Chua and Lin," *IEEE Trans. Circuit Syst., vol. CAS-34, pp.*

210-214, Feb. 1987.

32. ---, "Neural network for nonlinear programming," *IEEE Trans. Circuit Syst., vol. 35, pp. 554-562, May 1988.*
33. H. Beck, D. Bergondy, J. Brown and H. Sari-Sarraf, " Multiresolution Segmentation of Forward Looking IR and SAR Imagery Using Neural Networks," *NASA/CCDS AV SPI, WNN-AIND 91.*
34. Po-Rong Chang, Wen-Hao, and Kuan-Kin Chan, "A Neural Network Approach to MVDR Beamforming Problem," *IEEE Trans. on Ant. and Propagate., vol. 40 no. 30, March 1992*
35. Jon. P. Davis and W. A Schmidt, " Use of Partial Templates with Higher-Order Neural Nets for 2-D Target Detection., *NASA 'CCDS AV SPI, WNN-AIND 91*
36. Y. Hara, R. G. Atkins, S. H. Yueh, R. T. Shin, and J. A. Kong. "Application of Neural Networks to Radar Image Classification," *NASA 'CCDS AV SPI, WNN-AIND 91.*
37. G. A. Clark, J. E Hermandez, B. S Lawver and R. J. Sherwood, " Gabor Transform and Neural Network for Automatic Target Recognition," NASA/CCDS AV SPI WNN-AIND 91
38. T. Ziemke, "Radar Image Segmentation using Self-Adapting Recurrent Network," University of-Skovde, Skovde, Sweden
39. S. Haykin and T. K Bhattacharya, 1992 "Adaptive radar detection using supervised learning networks," *Computational Neuroscience Symposium, pp. 35-51, Indiana University-Purdue University at Indianapolis*
40. S. Haykin and C. Deng, 1991. "Classification of radar clutter using neural network," *IEEE Trans. on Neural Networks 2, 589-600.*
41. J. A Anderson, M. T. Gately, P. P. Andrew, and D. R. Collin, "Radar Signal Categorization Using a Neural Network," *Proceeding of IEEE, 78(10), 1646-1657*
42. B. Kosko (1992) "Neural Network for signal processing," *Prentice Hall, Englewood Cliffs*
43. M. Madrid, Juan. Casar Corredera, Jose. R., and de Miguel Vela, G (1992), "A Neural Network approach to Doppler-based target classification," *IEEE International Radar Conference 'RADAR 92', 450-453*
44. S. C. Ahalt, F. D. Garner, I. Jouny, A. K. Krishnamurthy "Performance of Synthetic Neural Network Classification of Noisy Radar Signals," *D. S. Touretzky (Ed) Advances in neutral information processing systems I, Morgan Kauffmann, San Mateo, 281-288, 1989.*
45. O. R. Hager, "Object detection in clutter with learning maps," *SPIE Synthetic Aperture Radar, vol. 1620, 170-186, 1992*

SCHEDULING IN THE DYNAMIC SYSTEM SIMULATION TESTBED

Milton L. Cone
Associate Professor
Department of Computer Science/Electrical Engineering

Embry-Riddle Aeronautical University
3200 Willow Creek Road
Prescott, AZ 86301

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

And

Wright Laboratory

August 1997

SCHEDULING in the
DYNAMIC SYSTEM SIMULATION TESTBED

Milton L. Cone
Associate Professor
Department of Computer Science/Electrical Engineering
Embry-Riddle Aeronautical University

Abstract

The task of a sensor manager is to improve the performance of the individual avionics sensors by coordinating their activities based on the sensor manager's best estimate of the future. The Dynamic System Simulation Testbed (DSST) was developed by Data Fusion Corporation to test different sensor manager concepts. This report covers the use of the DSST to evaluate several genetic search schedulers. The goal is to find schedulers that can improve the performance of the baseline greedy scheduler in the DSST. Several performance measures are used to assess the performance of the schedulers. Results show that for the scenario considered here it is difficult to improve on the performance of the greedy scheduler. While the genetic scheduler improves some measures the improvement is not dramatic. The evaluation function used to drive the genetic search tries to execute higher priority jobs earlier in the scheduling window. While the evaluation function was effective at scheduling higher priority jobs in the scheduling window, it did not seem to schedule them earlier in the cycle. Rather the tasks seemed to be more randomly distributed over the planning cycle. It is suggested that a different scenario and a modified evaluation function might show more differences in the performance of the greedy and genetic based schedulers.

SCHEDULING in the

DYNAMIC SYSTEM SIMULATION TESTBED

by

Milton L. Cone

Introduction

The Dynamic System Simulation Testbed, DSST, [Kober, et. al., 1995], is a simulation developed in the MATLAB® language to aid development of sensor management systems. A more detailed description is contained in Kober but the main modules of the simulation consist of:

a request generator that selects the possible sensing tasks,

a prioritizer to establish the importance of each task to the mission,

a task scheduler to determine when each task is to be executed,

a task executor to determine which jobs are ready to run,

a sensor module to report the results of the sensing actions,

a fusion module that associates the collected data with existing data to develop target tracks.

The sensor manager coordinates the activities of the first three modules.

This report focuses on the performance of the task scheduler. The DSST comes with a greedy scheduler that assigns jobs based on the weighted shortest processing time first (WSPT) algorithm. In this scheduling approach the jobs with the largest ratio of weight (priority) to processing time are scheduled first. This favors high priority, short duration jobs over low priority, long jobs. WSPT turns out to be optimal for one machine trying to minimize the sum of weighted completion times. By minimizing the sum of weighted completion times, more high priority jobs get processed. High priority jobs that take a long time to process are less likely to be scheduled than high priority, short jobs since the value of the objective (sum of weighted completion times) is improved by processing more jobs.

The WSPT algorithm would be optimal for the scenario Kober considered except that each job in the DSST has a due date and the jobs are not all released at the same time. Since the problem of minimizing the sum of completion times with different release times for one machine is strongly NP-hard (see [Pinedo, 1995] for a discussion) this problem must be at least as difficult. Add in the complexity of several sensors, some of which may be functionally equivalent (can collect the same information), and the problem becomes even harder. An approach to scheduling that is more comprehensive than greedy should improve the performance of the sensor manager. In order to evaluate the performance differences between various scheduling techniques, the DSST was modified to increase the realism of the simulation. These modifications include:

adding a multiple sensor capability to the simulation,

providing setup times for each sensor task that depend on the previous task, and

adding a genetic algorithm based scheduler.

In addition, a module was developed to collect several measures of performance on the scheduler. With these changes in place, several simulation runs with the greedy scheduler and four variations of a genetic scheduler were compared to determine performance differences.

The rest of the report is divided into seven sections. The first section describes the rates important to scheduling in the DSST. In the next section, modifications made to enhance the realism of the DSST are documented as well as how to run the modified model. The following section discusses the scenario that is used to exercise the DSST. The next section describes the measures of performance (MOP) developed to characterize the scheduler. This section is followed by the results as characterized by the MOP's. This section also includes a discussion of the value of the MOP's to capture the performance of the scheduler as well as what the MOP's say about the performance of the various schedulers. The last section gives the conclusions that can be drawn from the results and describes the important future work that needs to be completed.

Discussion of DSST Rates

There are three periods important to understanding scheduling in the DSST. The first is determined by the planning rate. This is the frequency at which schedules are prepared. For all scenarios considered here the sensor manager reschedules every second. The next important period is determined by the horizon. This is the time over which a schedule runs. Every second the scheduler prepares a schedule covering the next number of "horizon" seconds. The "horizon" is fixed for three seconds for the data reported here. The last period is the execution period and is the time over which scheduled jobs will be executed. It is called the task execution rate in the simulation, is one second and is the finest rate at which jobs can begin. In summary, the sensor manager develops a schedule every second for three seconds that executes tasks for one second before rescheduling unexecuted tasks and new requirements.

Discussion of the Modifications Made to the DSST

A central file, SMgr, is read by the DSST to control the operation of the simulation. A simulation is characterized by the values assigned to the simulation parameters in the SMgr file. The first set of changes are the additions to the SMgr file necessary to characterize additional sensors in the tactical aircraft. The constants listed below are for a suite of three sensors, each having three modes. The set up in SMgr is:

```
%---initialize some variables
        cycle_rate(1) = 1;          %---object prioritizer rate (Hz)
        cycle_rate(2) = 1;          %---issue prioritizer (Hz)
        cycle_rate(3) = 1;          %---request generator rate (Hz)
        cycle_rate(4) = 1;          %---scheduler rate (Hz)
        cycle_rate(5) = 10;         %---task executor rate (Hz)
```

```
horizon = 3;                        %---scheduling horizon (secs)
num_eff_rows = 9;                   %---number of rows in eff matrix
num_modes =3;                       %---number of sensor modes
num_regions = 6;            %---number of search regions
num_snsrs = 3;                      %---number of sensors
run_title = 'left~SMgr';            %---title of run for plots
stopT = 65;                         %---duration of simulation (secs)
t_end = zeros(1, num_snsrs);        %---sensor availability
trck_obj = [];                      %---track objects

%---Initialize matrices
tasks = [];
jobs = [];
t_post = [];
rel_post = [];
mu_post = [];

%---initilize some computational matrixes
id_mode = [1 1 1                    %---pilot "designated" object
           0 1 0                    %---search object
           0 1 0                    %---unknown "un-detected track object
           1 1 1];                  %---unknown "detected" track object
mu_Td = [];
mu_T0 = [];
service = [.1 .2  .3                %---duration of task in this mode for sensor 1
           .1 .2 .3          %---sensor 2
           .1 .2 .3];               %---sensor 3
snsr_angl = [0                      %---mounting angle of sensor to centerline of aircraft
             0                      %---sensor 2
             0];                    %---sensor 3
snsr_eff =[1 1 0.95 0.95 0.95 0.95 0.95     %---effectiveness of sensor for this task
           1 2 0.1 0.1 0.1 0.1 0.1
           1 3 0.5 0.5 0.5 0.5 0.5
           2 1 0.95 0.95 0.95 0.95 0.95
           2 2 0.1 0.1 0.1 0.1 0.1
           2 3 0.5 0.5 0.5 0.5 0.5
           3 1 0.95 0.95 0.95 0.95 0.95
           3 2 0.1 0.1 0.1 0.1 0.1
           3 3 0.5 0.5 0.5 0.5 0.5];
snsr_FOR = [pi/3 pi/3 pi/3                  %---sensor field of regard
            pi/3 pi/3 pi/3
            pi/3 pi/3 pi/3];
snsr_FOV = [pi/180 pi/6 pi/12               %---sensor field of view
            pi/180 pi/6 pi/12
            pi/180 pi/6 pi/12];
snsr_rng_max = [60000 60000 60000           %---sensor max range
                60000 60000 60000
                60000 60000 60000];
snsr_status = [1.0 1.0 1.0                  %---availability of sensor in each mode
               1.0 1.0 1.0
               1.0 1.0 1.0];
snsrs = [1 0 0                              %---what modes each sensor has
         0 1 0
         0 0 1];
```

The changes to SMgr that are necessary to run a multiple sensor scenario include:

1. num_eff_rows has to agree with the number of rows in the snsr_eff matrix. There is one row for each mode on every sensor.
2. num_modes is the total number of different modes on the aircraft. If one sensor has one mode and another sensor has two different modes then there are three modes on the aircraft. Sensors with a subset of the mode set can use either the snsr_status or snsrs matrices to indicate what modes they possess.
3. num_snsrs is the number of sensors on the aircraft.
4. t_end keeps track of when a sensor is available to schedule the next job. It is a vector that has num_snsrs entries. It normally is the ending time of the current sensing task.
5. The id_mode has four rows, one for each type of object. Each column represents the suitability of that particular mode to provide information on that class of object. For example, any of the three modes is suitable for collecting information on a pilot designated target. Only mode two is suitable for search and modes one, two and three can be used for tracking. In the simulation only rows two and four are used. All objects are classified as either track or search. Each new mode adds a new column to this matrix.
6. service is the length of time it takes a sensor in this mode to process a request. There is an entry for each sensor and each mode. The size of this matrix is (num_snsrs, num_modes). Fake entries for the service time can be entered if those modes aren't on a particular sensor.
7. snsr_angl has an entry for each sensor. The entries are the direction that each sensor is mounted on the aircraft. Normally, this is 0°.
8. snsr_eff is a matrix with num_snsrs*num_modes rows and number of issues plus 2 columns. The plus 2 is for the first two columns that contain the sensor id number and mode. The remaining 5 columns are for each of the five issues that it is possible to collect information on a target. These five issues include range, range rate, angle, rate of change of angle and id. The last issue is carried in the matrix but the effectiveness is not used.
9. snsr_FOR is the sensor field of regard. It is the full angle over which the sensor can look. The entry is the half-angle. The sensors in this example can look ±60° centered on the snsr_angl for this sensor. There is one entry for each sensor and mode.
10. snsr_FOV is the instantaneous field of view of the sensor. It is given in half-angle. One entry for each sensor and mode.
11. snr_rng_max is the maximum range for each sensor and mode. It is the maximum range that a sensor can detect a target. At the present time all targets within the snsr_FOV and snr_rng_max

are detected without error. The probability of detection is one; the probability of false alarm is zero.

12. snr_status is the availability of the sensor mode. Since there are num_modes columns in this matrix there very likely will be some modes not implemented on all sensors. A zero in an element location indicates that this mode is either not present or has been deactivated.

13. snsrs is like snr_status. It is used as a flag matrix indicating which modes are present on each sensor. The difference between snsrs and snsr_status is that snsr_status could be used with fractional entries to indicate degraded performance for a particular mode.

The second group of changes are changes to the program. These changes, although not extensive, did involve several modules. While the exact changes are available by obtaining a copy of the final program, the nature of the changes are summarized here.

The first change is to the DSST file aggregator_kk. The agg_modes matrix needs to be expanded to size (num_snsrs*num_modes). The agg_modes matrix identifies which modes can be aggregated with which other modes. For example, a request for mode 1 can't be aggregated under modes 2 or 3 using the present agg_modes matrix. The references to agg_modes in aggregator_kk now has to carry both indices as agg_modes is a matrix not a vector.

With multiple sensors, it is possible for the eigMax function to not find an inverse in function ahp_threat. This is usually associated with the call eigMax(id) towards the end of the ahp_threat function. Code was put in to test for that occurrence and when it happens to replace vect2, the eigen vector associated with the maximum eigenvalue, with all zeros.

A new function was added called velocity, that parallels the function distance. Its purpose is to identify those targets whose velocity vectors are substantially different as set by a criteria in velocity. Velocity returns a flag, delta_vel, that is zero when the velocities are the same. Velocity is used in function fusion to distinguish objects.

Function fusion has the most changes. The basic algorithm takes all of the measured objects and tries to match them to track objects(things like planes or tanks) and search objects(areas of space and velocity that the sensor is to search for things like planes and tanks). Most of the changes result from multiple sensors collecting information on the same target in one sensing increment. This means that the loop over search and track objects has to handle multiple entries for the same object in meas_obj (measured objects matrix). Without these changes the track objects matrix (trck_obj) adds false targets, either duplicate track objects or search objects improperly labeled as track objects.

The loop over search objects now searches the entire meas_obj matrix looking for all of the matches not dropping out of the loop at the first match. It also looks to match both distance and velocity (using the velocity function) not just distance to the search object. The mu for a search object is properly updated for each search object and all entries in the meas_obj matrix are deleted. Before, when only the first meas_obj that matched a search object was detected, the duplicate entries ended up as track objects. The fusion function then converted these search objects into track objects. These changes prevent that from happening.

The loop over track objects is handled similarly. The meas_obj matrix is searched for all matches with track objects in case a track object is reported by more than one sensor in a single time increment. The track mu is properly updated for each detection of a target.

After the loop over track objects, those objects left in the meas_obj matrix are new detections. They are added to the trck_obj matrix as new track objects. There is a possibility that more than one sensor reports the same object as a new track. To prevent multiple entries on the same object a search of the new track objects is performed and multiple entries are removed. A note is provided on the screen warning of this. So far there have not been any instances of multiple entries at this point. Finally, there is a test of the track objects matrix to look for any remaining duplicate entries. Thus far, none have been found.

The third group of changes are new functions added to the code. The additions are a genetic scheduler and a module to collect measures of performance.

The genetic search algorithm is based on the survival of the fittest. A population of chromosomes is defined to start the search. In this case the population is made up of potential schedules. Each chromosome (schedule) is composed of genes. Each job to be scheduled is thus a gene. The genetic search manipulates the chromosomes using the operators of crossover, mutation and selection towards either maximizing or minimizing a performance function. Crossover occurs between two chromosomes and is equivalent to breeding. In crossover, two chromosomes exchange genes to produce two new chromosomes. In mutation a single chromosome has two genes swapped. The two genes that exchange places are randomly chosen. Crossover and mutation combine to define a new population. Selection decides which chromosomes will survive into the next generation. The probability of survival depends on the suitability of the chromosome as determined by a fitness function. The more fit a chromosome is, the more likely it will survive and the more copies of it will constitute the next generation. Not all chromosomes make it through. The fitter chromosomes do not always survive and the weaker are not always eliminated, although the fittest chromosome is always retained.

The genetic module is managed by the function scheduler_ga. Function scheduler_ga is an exact replacement for scheduler_greedy. To change between the two it is only necessary to comment out the undesired scheduler. The parameters that control the genetic search are set in scheduler_ga. Those parameters are shown below with typical values.

```
%---set constants that define the genetic search
    Pop_size = 50;          %---number of chromosomes in a population. Even numbers only.
    C_rate = 1.0;           %---crossover rate between 0 - 1.0
    M_rate = 0.1;           %---mutation rate between 0 - 1.0
    Max_gen = 50;           %---max generations before termination
    Rank_min = 0.75;        %---used in select to calculate probability of selection
    Gapsize = 1.0;          %---percentage of the old population replaced each gen
    ELITE = 1;              %---add best_guy to each pop before mutation and xover
    STATS_FLG = 1;          %---save max, min, mean, std of each gen
    PR_GEN_FLG = 0;         %---prints first and last gen +fitness when 1
    MAX_I = 8;              %--- DONE=1 after max generations if best_fit unchanged
    PR_CHRO = 0;            %---when 1 prints fitness and associated chromosome to prnt_chro.ga
    FAST_TRCK = 1;          %---when 1 uses approx evaluation to speed ga
    MATCH_MAKER = 1;        %---when 1 uses match maker to speed convergence of ga
```

Pop_size defines the number of schedules that are maintained in one generation of the genetic search. These schedules are combined through the genetic operations of crossover and mutation. C_rate defines the crossover rate and is the fraction of the population that are pairwise combined to produce offspring that may survive into the next generation. Crossover uses the PMX algorithm in which a portion of each chromosome is exchanged and then woven into the chromosome so that the resultant schedule is still valid. M_rate is the probability that a chromosome undergoes mutation. For M_rate = 0.1, on average, one in every ten chromosomes is mutated. Max_gen and MAX_I define termination conditions on the search. Max_gen is the number of generations that the search is allowed to run before stopping. MAX_I is the number of generations that execute without an improvement in the best schedule before termination. Rank_min is a parameter used in the selection process. It is generally not changed. Gapsize allows for some members of the old population to survive into the next generation. (1 - Gapsize)*Pop_size is the number of chromosomes from the old population that come through. Members from the old population that survive are randomly chosen. The parameters in all capitals are generally flags that select that function or not. ELITE adds the best schedule back into the surviving generation. STATS_FLG creates a file called statistics.ga that contains the maximum, minimum, mean and standard deviation of each generation. PR_GEN_FLG prints the first and last generation of chromosomes to the screen. It is generally used for diagnostics. PRT_CHRO creates a file called prnt_chro.ga that stores generation, fitness and chromosome for each chromosome passed to it. This function is also mainly for diagnostic purposes.

FAST_TRK and MATCH_MAKER are two variations on the genetic search that help speed convergence. FAST_TRK uses an approximate evaluation of the schedule rather than a complete schedule build. The principle is that it is not necessary for the genetic search to have complete knowledge of the evaluation function to find a good result. MATCH_MAKER tries to accelerate convergence by matching chromosomes with similar fitness evaluations. The idea is that high fitness evaluations must have good partial sequences of schedules and that by combining good partial sequences the genetic search is more likely to produce better schedules faster.

The last major addition is the measures of performance function, mop. This function collects data on the performance of the scheduler. Its function is twofold: collect data the and evaluate the suitability of different measures. A more detailed discussion of the measures test appears in the section titled MOP's.

Scenarios

The scenario was the same one used in the study by [Kober, et. al., 1995]. It has 24 targets arranged inside a 120° arc centered on the centerline of ownship. Twelve targets are in the upper 60° arc and 12 targets in the lower 60°. The twelve targets below the centerline of ownship mirror the targets above. Both have the same x and y positions and x velocity and oppositely directed y velocities. The targets all move at a constant velocity. Figure 1 shows the location of the twelve targets in the upper 60° arc. Of the

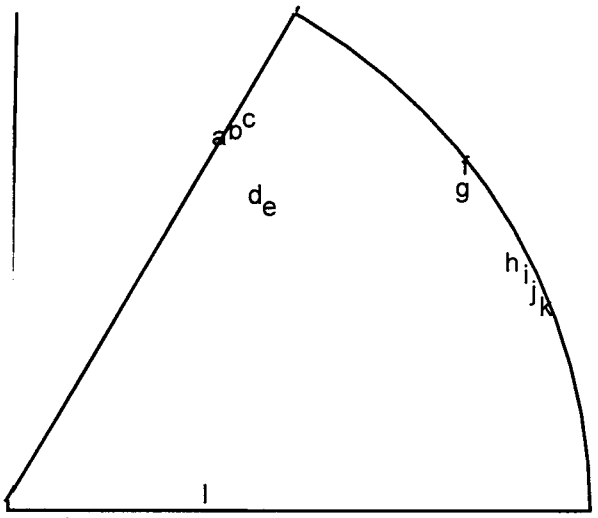


Figure 1. Location of twelve upper arc targets.

twelve targets in the upper arc, three targets (a, b, c) are on the outer edge of the field of regard slowly moving across the path of ownship. Two targets (d, e) are moving perpendicular to the flight path of ownship. There are two targets (f, g) on the outer boundary of the search sensor range moving diagonally

towards the ownship. The four targets (h, i, j, k) at the outer boundary are moving towards ownship, parallel to ownship's centerline. Target l is moving away from ownship.

MOP's

The measures of performance are chosen to evaluate some aspect of scheduler performance. There are 28 data items currently being recorded at every scheduling cycle from which eleven primary and four derived MOP's are computed. Most of the data involve tasks and requests. The difference between the two is that tasks are requests for sensing that get scheduled. Tasks fall into two categories: tasks that are scheduled and scheduled tasks that execute. Requests are composed of new requests from the request generator and scheduled tasks that did not execute but still are within their window of opportunity. Not all requests fit into the planning horizon. Comparing measures on tasks to measures on requests show how well the sensor manager is managing the workload.

A list of the items on which data was collected follows with a short explanation of what the item is.

1. number lost. These are tasks from the request generator that did not get executed.
2. tasks to horizon. The next three measures assess all of the tasks to the planning horizon. Most of the time this is 3 seconds. If the scheduling cycle is every second then these measures would be collected over the next three seconds every second. The three measures collected are:
    * Σpriority - The sum of the priorities of all the tasks to the planning horizon.
    * Σpriority*duration - The sum of the priority of a task times its duration.
    * Σpriority*completion time - The sum of the priority times the time the task completes normalized to the start of the planning cycle. This keeps later tasks from dominating this measure. It is basically the fitness criteria for the genetic algorithm.
3. requests to horizon. These two measures are similar to the first two tasks to horizon. The third measure, Σpriority*completion time, cannot be collected since the completion time of requests is not known.
    * Σpriority - The sum of the priorities of all the requests generated in this request generator cycle, generally every second.
    * Σpriority*duration - The sum of the priority of a request times its duration. Collected every request generator cycle.
4. lost jobs. The same two measures as requests to horizon but on jobs that did not execute. These may have been scheduled or not but they never executed because their priority was so low they kept getting pushed off by higher priority new requests until their window of opportunity closed.
5. idle time to horizon by sensor. There are two horizons in the sensor manager. One is determined by the execution cycle and is the interval over which tasks are actually executed. The

second horizon is the planning horizon over which jobs are scheduled. The purpose of the planning horizon is to allow the sensor manager to look into the future to better manage its workload. The execution horizon is set at one second for all of the suns in this report and the planning horizon is three seconds. To keep them separate, the execution horizon will generally be referred to as the execution cycle and the planning horizon will be simply called the horizon. For each sensor in the sensor suite "idle time to horizon" collects the total time each sensor does not have tasks scheduled during the current planning horizon. Measure 10 captures the same information for the execution horizon.

6. requests generated by mode. The way the simulation is set up every different sensor selects which modes out of the entire set of modes it will possess. For example, if there are three modes a sensor may have one, two, or all three modes. Sensors can have different modes or the same modes. This measure counts the number of requests generated for each mode.

7. tasks assigned by mode to horizon. Similar to requests generated by mode but counts only requests that are scheduled. Some of these scheduled requests may not execute.

8. tasks assigned by mode that execute. Similar to 6 and 7 but only covers jobs that will execute over the next execution cycle. These tasks are given to the sensors and they will execute. Usually these are the tasks executing over the next second of simulation time.

9. tasks assigned over execution cycle. This measure is similar to "2. tasks to horizon" except that the data is only collected over the execution cycle.

    * Σpriority - The sum of the priorities of the executing tasks.
    * Σpriority*duration - The sum of the priority of an executing task times its duration.
    * Σpriority*completion time - The sum of the priority times the time the task completes normalized to the start of the planning cycle. This keeps later tasks from dominating this measure.

10. idle time during execution cycle by sensor. This measure shows the time available to assign more tasks during the execution cycle. Measure 7 collects similar information but to the horizon.

11. sum mu. This measure is the sum of the average mu's calculated for each target that is being tracked. Mu is a qualitative measure of the amount of knowledge about a target. The closer mu is to one the more certain is the information about the target. While search requirements also use mu to determine if a search is required, that mu does not enter into this measure.

There are several derived measures that can be calculated from the above data. These include:

1. number. The total numbers of requirements, tasks to horizon and tasks that execute are derived from the individual mode results.

2. fraction. Records the (tasks)/(requests). Both tasks to horizon and tasks that execute are calculated.

3. normalized tasks to horizon, tasks to execute, and requirements. MOP's 2,3 and 9 calculate Σpriority, Σpriority*duration, and Σpriority*completion time. Since different simulation runs may assign different priorities based on the number of tasks that have to be scheduled, these MOP's try to normalize the previous results by dividing by the appropriate number of total tasks to horizon, tasks that execute or requests. This normalization should allow different schedulers to be more accurately compared.

4. fraction of jobs lost. This measure is the (number of jobs lost)/(total tasks to horizon). This measure is sensitive to the ability of the scheduler to push as many tasks as possible through the system. One minus this measure reflects the ability of the scheduler to execute all of the scheduled tasks.

Results

Results are given from DSST runs for two sensor configurations. The first is one sensor with three modes. Since this is the same scenario as reported in [Kober, et. al., 1995], the results shown here duplicate theirs. It roughly corresponds to a radar system with single target track (STT), search, and track while scan (TWS) identified as modes 1, 2, and 3, respectively. This configuration has the three modes competing with each other for time on the one sensor. The second configuration is three sensors with one mode each. This configuration exercises the multisensor capability of the simulation modifications. In the second configuration, STT, TWS and search only compete against each other since there is one sensor for each mode. The number of tasks in each mode measures how important this mode is relative to the other modes as determined by the request generator process.

The scheduling algorithms evaluated include greedy, which assigns tasks based on the highest ratio of priority to task duration and four genetic searches. The first genetic search, called full ga, is a full genetic search with 50 generations, 50 chromosomes, and a maximum of 8 generations without change before termination of the search. Fast track uses an approximate evaluation of the fitness function to accelerate the search process. Match maker uses a process described in [Cone, 1996] which combines chromosomes together with like fitness evaluations instead of randomly. The purpose is to also speed the convergence of the genetic algorithm. The last genetic search is a combination of fast track and match maker. Shadow is the greedy algorithm working on the same requests as full ga. Its purpose is to give a direct comparison of the greedy and genetic algorithms.

"Table 1. Measures of Performance" contains measures that report counts of requests or tasks over the 1 second execution cycle or 3 second planning cycle.

job	number lost	number requests			number tasks to horizon			number tasks over execution cycle		
		mode 1	mode 2	mode 3	mode 1	mode 2	mode 3	mode 1	mode 2	mode 3
one sensor										
greedy	452	1166	233	284	908	173	150	500	64	5
full ga	452	1199	234	286	939	174	154	496	64	6
shadow	489	1199	234	286	907	170	153	497	66	4
fast track	449	1175	227	284	918	167	152	499	64	5
mtch mkr	461	1204	243	283	936	182	151	490	65	8
ft_mm	446	1153	243	279	903	178	148	506	62	4
three snsrs										
greedy	123	1014	89	83	891	89	83	635	89	83
full ga	121	1209	85	97	1088	85	97	635	85	97
shadow	201	1209	85	97	1008	85	97	635	85	97
fast track	121	1151	89	96	1030	89	96	635	89	96
mtch mkr	120	1229	88	103	1109	88	103	635	88	103
ft_mm	123	1140	76	105	1017	76	105	635	76	105

Table 1. Measures of Performance

"Table 2. More Measures of Performance" lists values that are related to the priority of the job. The Σpr column is the sum of the priorities of all of the appropriate jobs. Larger values mean more, higher priority jobs are completed. The Σpr*dur multiplies the priority of each job times the duration of the job. This measure gives more weight to the longer jobs. The last measure is Σpr*ct. This measure multiplies the priority by the completion time of each job normalized to the start of the execution cycle. For example, if a job completes at 3.2 seconds and the execution cycle starts at 3.0 seconds, then ct is 0.2 seconds. This measure is minimized by completing higher priority jobs first. Requests does not have a Σpr*ct column since requests never carry a completion time.

job	tasks to horizon			requests		lost jobs		tasks to execution		
	Σpr	Σpr*dur	Σpr*ct	Σpr	Σpr*dur	Σpr	Σpr*dur	Σpr	Σpr*dur	Σpr*ct
one sensor										
greedy	443.	64.4	429.	539.	82.8	96.	18.4	279.	35.4	143.
full ga	474.	69.9	484.	572.	88.9	98.	19.0	284.	36.3	145.
shadow	463.	68.0	449.	572	88.9	109.	20.9	289.	36.6	147.
fast track	441.	64.2	430.	536.	82.3	95.	18.1	277.	35.1	142.
mtch mkr	481.	71.2	493.	580.	90.1	100.	19.0	287.	36.8	146.
ft_mm	449.	65.5	436.	547.	84.6	98.	19.1	280.	35.4	142.
three snsrs										
greedy	213.	26.8	103.	221.	27.6	8.1	0.81	194.	24.8	78.7
full ga	236.	29.0	141.	243.	29.7	6.9	0.69	192.	24.6	79.6
shadow	228.	28.2	121.	243.	29.7	14.6	1.46	199.	25.3	82.0
fast track	225.	27.8	125.	233.	28.5	7.1	0.72	190.	24.2	78.0
mtch mkr	230.	28.3	136.	236.	28.9	6.3	0.63	187.	24.0	77.3
ft_mm	213.	26.1	115.	220.	26.8	7.1	0.71	182.	23.0	74.0

Table 2. More Measures of Performance

The next table, "Table 3. Even More Measures of Performance," contains data on idle time for each sensor over the execution cycle and to the planning horizon. The execution cycle is just the time until replanning takes place which is one second for the scenarios considered here. Time to horizon is the time to the planning horizon. The total time is the horizon*(simulation duration) which is 195 seconds; three seconds for the planning horizon and 65 seconds for the simulation run. See the entries for the single sensor scenario. Similarly, the idle times over the one second execution cycle listed for sensors 2 and 3 in the single sensor scenario are just the length of the simulation. The lower idle time, the better the utilization of the sensor. It is unlikely that any of the numbers will get to 0 since the scenario starts without any knowledge of the environment. The first task is a search of the area and unless the search time divides evenly into the planning cycle there will be dead time before other tasks can be scheduled.

The sum mu column is calculated from the mu factor computed in the DSST simulation. Mu is a qualitative measure of the certainty of knowledge about the target or search object. The more information that has been collected on a target, the larger mu is. Sum mu is the sum of the mean mu's computed over

the simulation run for each track object. While there is a mu calculated for search objects, it is not factored into this measure.

job	idle time over execution cycle			idle time to horizon			sum mu
	Snsr 1	snsr 2	snsr 3	snsr 1	snsr2	snsr3	
one sensor							
greedy	0.8	65	65	24.6	195	195	5.9501
full ga	0.8	65	65	20.1	195	195	5.9094
shadow	0.8	65	65	24.4	195	195	--------
fast track	0.8	65	65	24.2	195	195	5.9512
mtch mkr	0.8	65	65	19.7	195	195	5.8861
ft_mm	0.8	65	65	24.7	195	195	5.9294
three snsrs							
greedy	1.5	47.2	40.1	105.9	177.2	170.1	7.8714
full ga	1.5	48.0	35.9	86.2	178.0	165.9	7.8959
shadow	1.5	48.0	35.9	94.2	178.0	165.9	--------
fast track	1.5	47.2	36.2	92.0	177.2	166.2	7.9212
mtch mkr	1.5	47.4	34.1	84.1	177.4	164.1	8.0182
ft_mm	1.5	49.8	33.5	93.3	179.8	163.5	7.9285

Table 3. Even More Measures of Performance

Table 4 collects several more measures, derived from the original measures, that might be useful to explain the performance of the schedulers. The first three columns are the total number of requests from the request generator, the total number of tasks scheduled to the planning horizon (3 seconds for all runs considered here), and the total number of tasks actually executed. The next column is the fraction of jobs lost determined by dividing the number of jobs lost by the number of tasks to horizon. This column is followed by the next two columns which record the fraction of the requests that were scheduled and the fraction of the requests that were executed. The next eight columns show various priority measures normalized by the number of tasks appropriate to that measure.

Discussion of Results

Consider first the simpler three sensor scenario. Each of the three modes has a sensor dedicated to its function. The modes do not conflict with each other for air time and can run simultaneously. Table 3 shows the amount of idle time over the 1 second execution cycle and the 3 second planning cycle (idle time

job	total number			fraction			tasks to horizon			tasks to execution			requests	
	req	tsks hor	tsks exc	jobs lost t_hor	tsks hor req	tsks exc req	Σpr t_hor	Σpr*dur t_hor	Σpr*ct t_hor	Σpr t_exc	Σpr*dur t_exc	Σpr*ct t_exc	Σpr req	Σpr*dur req
one sensor														
greedy	1683	1231	569	0.3672	0.7314	0.3381	0.3599	0.0523	0.3486	0.4900	0.0622	0.2506	0.3201	0.0492
full ga	1719	1267	566	0.3567	0.7371	0.3293	0.3742	0.0552	0.3823	0.5021	0.0641	0.2561	0.3330	0.0517
shadow	1719	1230	567	0.3976	0.7155	0.3298	0.3767	0.0553	0.3653	0.5098	0.0646	0.2587	0.3330	0.0517
fast track	1686	1227	568	0.3630	0.7337	0.3369	0.3565	0.0519	0.3478	0.4868	0.0618	0.2493	0.3178	0.0488
mtch mkr	1730	1269	563	0.3633	0.7335	0.3254	0.3788	0.0561	0.3891	0.5090	0.0654	0.2585	0.3355	0.0521
ft_mm	1675	1229	572	0.3629	0.7337	0.3415	0.3653	0.0533	0.3548	0.4890	0.0618	0.2475	0.3267	0.0505
three snsrs														
greedy	1186	1063	807	0.1157	0.8963	0.6804	0.2007	0.0252	0.0967	0.2401	0.0308	0.0976	0.1867	0.0233
full ga	1391	1270	817	0.0870	0.9130	0.5873	0.1857	0.0228	0.1108	0.2349	0.0301	0.0974	0.1744	0.0213
shadow	1391	1190	817	0.1689	0.8555	0.5873	0.1916	0.0237	0.1015	0.2432	0.0309	0.1004	0.1744	0.0213
fast track	1336	1215	820	0.0996	0.9094	0.6138	0.1855	0.0229	0.1028	0.2317	0.0296	0.0951	0.1741	0.0213
mtch mkr	1420	1300	826	0.0923	0.9155	0.5817	0.1768	0.0217	0.1048	0.2267	0.0290	0.0935	0.1662	0.0203
ft_mm	1321	1198	816	0.1027	0.9069	0.6177	0.1777	0.0218	0.0961	0.2231	0.0282	0.0908	0.1665	0.0203

Table 4. Derived Measures of Performance

to horizon). Sensors 2 and 3 are clearly not heavily loaded over either the execution or planning cycles. For example, of the 65 seconds available for execution, sensor 2 has 47.2 seconds left available when using the greedy scheduler. For the planning cycle, sensor 2 has 177.2 seconds left of the 195 seconds of planning (65*3). The only loaded sensor is sensor 1 over the execution cycle. For this sensor there are only 1.5 seconds left of the original 65 and those seconds occur at the start of the scenario before sensor 1 has any tasking. After the first 1.5 seconds, sensor 1 is fully occupied executing mode 1 tasks.

Table 1. tells the same story. All mode 2 and 3 tasks scheduled during the planning horizon are actually executed. None are lost. Only mode 1 tasks aren't executed. The genetic schedulers see more mode 1 tasks than greedy (Table 1). Modes 2 and 3 are about the same across all schedulers. Since requests are comprised of new requests created by the request generator as well as old tasks that do not get executed, it may be that the genetic schedulers keep tasks in the planning cycle longer than the greedy. Consider the greedy shadow scheduler. The greedy shadow scheduler sees the same number of tasks as the full genetic scheduler yet it schedules 80 less mode 1 tasks than the full ga. Jobs are lost when their priority is not high enough to get executed over the 3 seconds they have between creation and the closure of the window of opportunity. When the sensor is not always fully utilized it may be beneficial to execute lower priority jobs whose window is about ready to expire before higher priority jobs that have a later closure for the their window. That apparently happens here. Table 4 verifies that greedy has the highest average $\Sigma pr/t_exc$. This means that the genetic schedulers execute some jobs with a lower priority than greedy. Table 2 also shows that. Shadow executes the same number of tasks in each mode as the full ga yet has a higher Σpr. The full ga obviously chooses some lower priority jobs. By doing so the full ga was able to schedule but not execute 80 more mode 1 jobs than shadow. This result is not surprising since the fitness function that the genetic algorithm is trying to optimize is based on the planning cycle not the execution cycle. From an efficiency standpoint this may be desirable, from an operational perspective it may or may not be advisable.

Still considering the three sensor scenario, in Table 3, all genetic techniques have a higher sum mu. Generally this means that more data has been collected on the targets and that there is a higher degree of confidence in the knowledge of those targets. Only tasks that get executed increase mu. Since all schedulers execute the same number of mode 1 tasks, modes 2 and 3 make the difference in mu. All mode 2 and 3 tasks are executed, therefore the difference in the results has to be from the request generator creating more mode 2 and 3 requests for the genetic schedulers. Why that would happen is unknown.

Next consider the one sensor scenario. This scenario is more challenging since all modes compete for time on the same sensor. It also makes analysis harder since there are more competing themes to consider. A cursory look at all of the measures shows that there is not much difference in most of the values for any of the five schedulers. For example, consider fraction of jobs lost for the one sensor, Table 4. All of the genetic approaches lose a smaller fraction of jobs than greedy. Shadow, the greedy scheduler that schedules the same tasks at each planning epoch as the full ga, loses the highest fraction of jobs. The improvement ranges from 25 to 48%. The number of tasks executed is virtually the same for each of the schedulers varying from 563 to 572. All schedulers fully occupy the available execution time. The 0.8 seconds of idle time over the execution cycle occurs at the start of the simulation before the request generator can get started. There is some idle capacity during the three second planning horizon. Idle time happens when there is not enough jobs to fill the time available, usually at the beginning of the simulation before all of the targets have been found, and during the run when higher priority jobs keep lower priority jobs from executing before their window of opportunity closes.

Comparing full ga to shadow, both have the same number of requests in each mode but the full ga scheduler was able to schedule 37 more tasks in the planning cycle. This points out the ability of the genetic scheduler to improve on the performance of the greedy approach when working the same problem. Although percentage improvement is slight, about 3%, it does demonstrate the ability of the genetic algorithm to improve the performance in an area where the fitness function is supposed to improve. On the other hand, the number of tasks executed is almost identical, varying from 563 to 572. Since all schedulers fill all the available time the difference in number is due to the mix of jobs. Most tasks are mode 1, track. The next largest number of tasks is mode 2, broad search, with just a few mode 3 tasks.

Sum mu is in Table 3. Fast track has the highest rating while greedy has the second highest . There are two factors that affect the value of mu. The first is the number of mode 1 jobs that are executed. Mode 1 tasks cause mu to grow more than mode 2 or 3. The more mode 1 updates, the higher the certainty of a target's position and the higher mu. Greedy and fast track have the second and third highest number of mode 1 tasks. The second factor is the amount of information obtained on the targets in the half of the field of regard that the sensor was told not to search. Only searches collect information in this area. The more searches, the higher the mu's are and the higher the overall average is. Greedy and fast track found the right mix, match maker didn't.

Now consider the one sensor data in Tables 2 and 4 that deal with priority. These measures factor in the importance of priority in the evaluation of the quality of the schedules. The "Σpr for tasks to horizon" measures both the number of jobs and the importance of each of the jobs as measured by its priority.

Either more jobs or jobs with higher priority or both cause this measure to be higher. Match maker and full ga have the highest Σpr and they also have the highest number of tasks scheduled. Table 4 tries to remove the dependency on total number of tasks by normalizing by it. The tasks scheduled by matchmaker and full ga also have a higher average Σpr. Shadow has an even higher average Σpr. This is due to the reduced number of tasks shadow schedules and its ability to shed lower priority jobs.

The genetic schedulers tries to minimize the fitness function Σpr*ct. This is the sum of the products of the task priority times its completion time. Both match maker and full ga are high here when they should be low. Comparing full ga with shadow in Table 4, the normalized Σpr*ct is lower. Since shadow has a higher Σpr it might be concluded that greedy, in the form of shadow, does better. That's not true. What is happening is that the tasks in shadow complete sooner since there are less of them. Thus on average ct is smaller for shadow and the Σpr*ct is smaller. Looking at the same normalized values under tasks to execution for full ga and shadow where the number of tasks are almost identical the Σpr/t_exc and Σpr*ct/t_exc are both larger for shadow. This discussion points out the problem of using Σpr*ct as a measure of performance to compare different schedulers. While the measure is sensitive to higher priority jobs being scheduled earlier it also gives credit to schedulers that reduce the number of jobs scheduled.

Conclusions

The genetic schedulers generally perform better than the greedy scheduler for the scenario considered here, but only slightly. Improvement was not dramatic, generally less than 10% for all but the fraction of jobs lost. The evaluation function used for the genetic algorithm did not seem to put enough emphasis on scheduling high priority jobs early. Higher priority jobs were generally scheduled more often using the genetic schedulers but they were not necessarily scheduled earlier. The data suggests that the timing of the tasks in the scheduling window was more random than calculated. A modified evaluation function such as suggested in last year's report might help. This evaluation function consists of two complementary measures instead of one measure as used in this report. A different scenario may also show more significant differences. The scenario used here is one in which greedy is almost the optimal solution leaving very little room for improvement.

References

Cone, Milton (1996). "Of Match Maker and Metrics." In *RDL Final Report-1996*. RDL, Culver City, CA.

Kober, Woody, D. Meyer, J. Thomas, K. Krumvieda (1996). "Formal Mathematical Models for Sensor Management." WL-TR-96-1104, Jan.

Pinedo, Michael (1995). *Scheduling – Theory, Algorithms, and Systems*. Prentice Hall, Englewood Cliffs, NJ.

FEATURE BASED COST MODELING

Robert C. Creese
Professor
Department of Industrial and Management Systems Engineering

West Virginia University
P.O. Box 6107
Morgantown, WV 26506

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

And

Wright Laboratory

August 1997

Feature Based Cost Modeling

Robert C. Creese
Professor
Industrial and Management Systems Engineering Department
West Virginia University

Abstract

Feature based cost modeling is a parametric cost approach using product features that are available to the designer including shape features such as volume, external surface area, internal surface area, and projected area; production features such as production quantity; and material features such as density, strength, and modulus. A literature review revealed little data on manufacturing costs of parts, and most of the USA data referred to a study by Boeing in the 1950's. The feature that has been primarily utilized in the past has been part volume, but projected area and dimensional ratios have also been utilized. A form was developed to collect manufacturing cost data to develop cost expressions to be utilized by designers. Future work plans are to collect and analyze data to develop feature based cost expressions considering new feature variables.

Feature Based Cost Modeling

Robert C. Creese

1. Introduction

Feature based cost modeling is a parametric cost approach using product features such as shape, quantity, and material. The goal of the process is to utilize features easily available to the designer to estimate the product costs so designers estimate the effects of design alternatives upon the part cost in the conceptual design stage.

Parametric cost estimating is required as the process details cannot be determined until the material specifications, production quantities, and detailed shape requirements are finished. This tool is to be used to give an estimate of what a part shape should cost in the design stage. The designer can also determine whether it is better to produce two or more parts to be assembled or to design a single part.

The cost model will be designed and calibrated to predict the production costs of a unit rather than the life-cycle costs of the component. The features to be investigated are those available to the designer, such as total part volume, part box volume, surface area(external surface area), void volume, void surface area, projected area(maximum), projected perimeter, maximum height and depth above the projected area, number of critical dimensions, and etc. Initially the model will use only a few of the parameters because of difficulties in obtaining the necessary data.

2. Background

Feature based cost modeling is a relatively new concept, but it is similar in concept to the cost-capacity factors which have been used for over 50 years in the chemical processing industry. The general form of the relationship is:

$$C(1) = C(R)[Q_1 / Q_R]^n \quad (1)$$

Where

$C(R)$ = reference cost

$C(1)$ = cost at level or size 1

Q_1 = capacity at level or size 1

Q_R = reference capacity

n = cost capacity exponent

When n=0.6, the rule is called the six-tenths rule. The exponent can vary between 0.2 and 1.0, but most applications have exponents between 0.5 and 0.8. The exponent should be less than 1(there are a few exceptions) to indicate the economies of scale; for example, one ten ton truck is usually less costly than two five ton trucks. The log-log relationship has also been extended to predicting costs for castings and forgings(1).

2.1 Boeing Data Analysis

A study by Boeing Aircraft Engineers(1) was published by ASM to illustrate the relative tooling and part costs for five casting processes and also for close-tolerance forgings. Figure 1 indicates the tooling and unit cost data for aircraft aluminum and magnesium sand castings . The tooling and unit cost data for magnesium alloy, aluminum alloy, and steel investment castings is illustrated in Figure 2. The set-up, costs, die costs, and unit part cost data for close-tolerance forgings is in Figure 3. Figure 4 is a summary figure for the tooling and unit part costs for the five selected casting processes - sand casting, close-tolerance sand casting, die casting, investment casting, and permanent mold casting. These figures did not include the machining costs.

The summary figures for the casting processes were reproduced in the book Design to Cost by Michaels and Wood in 1989. These figures, in the book, were listed as being reprinted by permission of the Martin Marietta Corporation rather than either ASM or Boeing and this indicates why investigators thought more cost research was being performed by different companies.

The summary diagrams were also used by David Zenger in his Ph.D. dissertation at the University of Rhode Island in 1989. He obtained the data from the ASM Castings Design Handbook which did not include the forging data. Dr. Zenger developed equations from the data in the figures by subtracting an average metal cost from the unit cost to obtain the process cost. He also increased the costs by a value of 3 to reflect increases from inflation over the 30 plus years since the data was originally developed. He also included expressions for the material cost, machining cost, and core cost. The core costs were a function of the casting cost and the machining cost focused on the non-productive labor costs; these expressions have not been used as the original costs should have included the core costs and machining costs are separate. The expression recommended was:

$$C_u = C_m + C_p + C_t \qquad (2)$$

where C_u = total unit cost, C_m = material cost, C_p = process cost, C_t = tooling cost

$C_m = V\ \rho\ \ MCOST$ (3)

where

V = part volume(in^3)

ρ = material density(lb/in^3)

MCOST = Material Cost($/lb)

Zenger developed the following expressions for processing and tooling costs, including the adjustment factor of 3, for each of the casting processes:

Sand Casting(SC):

$C(SC)_p = 1.7\ V^{0.87} + 0.36$ (4)

$C(SC)_t = [272\ V^{0.57} + 47] / N$ (5)

Close Tolerance Sand Casting(Precision Sand Casting(PSC))

$C(PSC)_p = 8.2\ V^{0.72}$ (6)

$C(PSC)_t = [1030\ V^{0.51}\ \} / N$ (7)

Permanent Mold Castings(PM)

$C(PM)_p = 2V^{0.7} + 0.3$ (8)

$C(PM)_t = [517 + 101\ V\ ]/N$ (9)

Die Casting(DC)

$C(DC)_p = 0.3\ V + 0.56$ (10)

$C(DC)_t = [3660\ V^{0.27}] / N$ (11)

Investment Casting(IC)

$C(IC)_p = 15.3\ V^{0.57}$ (12)

$C(IC)_t = [1700\ V^{0.3}\ ] /N$ (13)

where

V = Casting Volume(in^3)

N = Number of Parts for Tooling

Similar expressions can be developed from the close tolerance forging data:

$C_u = C_m + C_p + C_{su} + C_{die}$ (14)

where

C_u = total unit cost($/unit)

C_m = material cost

C_p = process cost

C_{su} = set up cost

C_{die} = die cost

Expressions for the part and process cost, set-up cost and die cost were developed from the Boeing data and they are:

$$C_p = 0.55\ PA^{0.94} \quad (15)$$

$$C_{su} = [10\ PA^{1.0}\ ] / N_{su} \quad (16)$$

$$C_{die} = [200\ PA^{0.92}] / N \quad (17)$$

where

PA = Projected Area (in^2)

N = Number of Parts for Tooling Life

N_{su} = Number of Parts for a Setup

In summary the Boeing Data indicates that the part volume and the projected area are two key parameters for estimating costs. The data had considerable variation and other variables may be able to reduce that variation.

A regression model for the unit costs of gray iron castings was developed by Pacyna(5) and the equation was of the form:

$$C_u = K_p K_c\ n^{-0.0782}\ V^{0.8179}\ r_e^{-0.1124}\ r_v^{0.1786}\ r_l^{0.1655}\ n_c^{0.0387}\ S_u^{0.2301} \quad (18)$$

The important factor is that the casting volume is the dominant factor as it has the highest exponent. The factors r_e, r_v, and r_l are dimensional ratios of actual dimensions to the box dimensions of the shape. The second highest exponent was on the tensile strength of the material. The shape factors thus appear to be rather important in predicting the costs of products.

2.2 Cost Functions

Cost functions(6,7) are used to predict costs of products that are similar in form and manufacturing methods. These functions have been used in the prediction of costs for machining operations. The method utilizes a basic design with associated cost and characteristic design parameters and uses these to calculate the costs of new designs as a

function of the characteristic design parameter ratio. Some typical characteristic design parameters are: diameter(gears) and length(box structures). An example of a cost function is:

$$C_n = C_m R^{Em} + C_p R^{Ep} + C_{su} R^{Esu} /N + C_c/N \quad (19)$$

where

C_n = Unit Cost of New Design

C_b = Basic Design Unit Cost = Cost to make a lot of One Part

$$C_b = C_m + C_p + C_{su} + C_c \quad (20)$$

C_m = unit material cost of base design

C_p = unit process cost of base design

C_{su} = set-up cost for base design product

C_c = constant costs which are independent of characteristic design parameter

N = units in production run

E_m = exponent for material (usually 3, may be slightly lower)

E_p = exponent for processing (usually 2, typically 1.8-2.2)

E_{su} = exponent for set-up(usually 0.5, but may vary between .14 and 1.8)

R = characteristic design parameter ratio

The material costs may not only be the material costs, but may also include other costs which may vary with the cube of the characteristic design parameter ratio. For example, annealing and sandblasting costs tend to vary with the volume of the material. On the other hand, some processing costs may vary linearly with the characteristic design parameter ratio and an additional term could be added to Equations 19 and 20.

The unit costs for machines parts were plotted as a function of the part volume by Radovanovic(8) in his MS Thesis work. These machined parts were classified into seven shape categories, such as primary rotational, primary planar, primary rotational with secondary, and etc. These plots were then generalized into two groups, small (less than 20 cubic inches) and large. Equations were then developed(9) to predict the unit cost as a function of the part volume and the results are presented in Table 1. The exponents on the volume term were low for small parts, from 0.33 to 0.40, and much higher for large parts(0.70 to 0.89). This indicated that the material costs are dominant for large parts, where as the handling costs dominate for small parts. Part volume, however, was the best parameter for estimating the cost of the machined parts.

2.3 Process Selection Charts

M.F. Ashby(11) at the University of Cambridge in England has done considerable work on materials selection and has extended this to process selection. The materials selection has been developed as a software package called CMS, Cambridge Materials Selector. He has developed process selection charts which utilize features such as: surface area, section thickness, complexity(information content), size(weight), precision(tolerance range), and surface roughness. In the process selection the properties of melting point and material hardness are also considered.

An interesting approach to the development of a complexity factor has been undertaken by Ashby which considers the number of dimensions and the precision of the dimensions. The expression developed is:

$$C = n \log_2 (l_{gmd} / \Delta l_{gmp}) \quad (19)$$

where

C = information bits

n = number of dimensions to describe part

l_{gmd} = geometric mean dimension

Δl_{gmp} = geometric mean precision

where

$$l_{gmd} = (l_1 l_2 l_3 \ldots l_n)^{1/n} \quad (20)$$

and

$$l_{gmp} = (\Delta l_1 \Delta l_2 \Delta_{l3} \ldots \Delta l_n) \quad (21)$$

For example, suppose a 1 kg part has 5 specified dimensions and the mean dimension is 64 mm and the mean precision is +/- 0.5 mm, then

$$C = 5 \log_2(64 / 0.5) = 5 \log_2(128) = 5 * 7 = 35$$

From Figure 5 the processes to be considered would be: sand casting, deformation processing, machining, and permanent mold(gravity) casting. If C increased to 1000, the processes would be: sand casting, composite fabrication, machining, die casting , and conventional fabrication(joining).

Two relative cost expressions for cost have been proposed by Ashby, one directly related to the chart on tolerances and surface roughness, Figure 6, and the second for unit component costs. The relative cost expression for the effect of tolerances and surface finish can be expressed as:

$$C_R = 2^{(2 \cdot \log RMS \cdot \log T)} \quad (22)$$

where

C_R = relative cost factor

RMS = surface roughness in μm

T = tolerance in mm

For example, if a part has a roughness of 1 μm and a tolerance range of 0.1 mm, then the relative cost would be:

$$C_R = 2^{(2 \cdot \log 1 \cdot \log 0.1)} = 2^{(2 \cdot 0 \cdot (-1))} = 2^3 = 8$$

If the tolerance range is further decreased from 0.1mm to 0.01 mm, the relative cost would increase to a value of 16.

The second relative cost expression is for unit costs and is:

$$C_u = C_m + C_c / n + C_l / n^* \quad (23)$$

where

C_u = relative unit cost($/unit)

C_m = material cost($/unit)

C_c = capital cost($)

C_l = labor cost($/hr)

n = production quantity(pcs)

n^* = production rate(pcs/hr)

For process evaluation of a given material the material cost is set to 1.0 and all other costs are normalized to the material cost. If different materials are utilized, the lowest cost material would be the normalized reference material and all other costs would be adjusted appropriately. Equation 23 is similar to Equation 2, but it indicates more clearly the impact of the production quantity and the production rate.

3. Cost Modeling Overview

Cost modeling is involved at various levels in the design of a system or the production of a product. There are four activities in the product realization process(10) which are: Product Planning, Product Design, Process Planning, and Manufacture. These activities are illustrated in Table 2 with some of the specific action items along with a

general time line. During the process realization process there are three different cost estimating models that can be utilized at the different stages of design. In addition , there is a cost accounting model which will be utilized when the product is in production.

The conceptual cost model is the model at the earliest stage ands its purpose is to determine what the product should cost. It is based upon part features and is a base to be set for the target costing. After the preliminary processes have been selected and the preliminary design made, the Process Cost Model can be used. It is to be used to more accurately estimate costs and to evaluate design changes to meet the target cost. After the tooling design, detailed process plans and detailed design have been completed can the Detailed Cost Model be used to estimate the final product costs.

After production starts, the Cost Accounting Model(typically a ABC costing model) can be used to determine what are the cost drivers and activities . The results from the Cost Accounting Model will be used to fine tune the Detailed Cost Model. The Cost Accounting Model determines what the costs were whereas the cost estimating models are to predict the costs.

The modeling emphasis for this report is the conceptual cost model where only the product features are known and some of the material performance requirements relating to weight, strength, and stiffness. In the conceptual stage a few materials can be evaluated to meet the shape features of the product. The conceptual model will focus on material selection as after the material is selected, only then can the design and process selection phases be done for the Process Cost Model. The Process Cost Model can be used to evaluate alternative processes and design changes upon the processes selected. After the processes have been specified and the final shape features determined, the tool designs can be completed and then the Detailed Cost Model can be used to predict the unit costs. The Conceptual Cost Model will evaluate the approximate material, tooling, and processing costs for the selected material, product features, and production quantity. The Conceptual Cost Model can be used to determine if the target costs are realistic.

4. Feature and Cost Data

The feature parameters which have been utilized with some degree of success for conceptual cost models are the volume for casting and the projected area for forgings. The formulas to be developed must be independent of the process; that is a knowledge of the specific process will not be required by the design engineer. A data base must be constructed to obtain the data required to develop the design feature based cost relationships. Table 3 has been developed to obtain the shape and material feature data

for the manufactured components. The data includes features that should be easily obtained by the designer from the part data files. Table 4 contains the part production and cost feature data which would be obtained to develop the conceptual cost relationships with respect to the feature data in Table 3. Tables 3 and 4 would be developed into data sheets to obtain feature and cost data from various part manufacturers.

Data will be sought from part producers such as metalcasters and forgers to develop the Conceptual Cost Model. Most manufacturers are reluctant to give cost data, but then may be receptive to giving some cost data(tooling costs) and the selling price. Since the selling price is the cost to the customer, this data would be useful for the model development. The minimum cost data required would be the selling price and quantity. From the previous Boeing Data, the tooling costs tend to have smaller exponents than the processing costs, so perhaps the surface area and perimeter may better reflect these costs. From the previous benchmarking study(11), the lead time values tend to correlate with the tooling lead times. There may also be a correlation between tooling cost and tooling lead time, but this will not be investigated at this time. Cooper(13) reported one instance where suppliers in Japan were required to give a formal cost estimation document with the bid, which for a foundry, consisted of the following eight categories: 1) material cost ; 2) mold cost; 3) facility fees; 4) labor costs; 5) heat treatment costs; 6) shot blast costs; 7) management fee; and 8) profit. Most US companies would be reluctant to give out such information.

5. Proposed Model and Analysis

Two basic approaches to the development of the conceptual cost model are proposed where the parameters are expected to be in exponential form. The first model is:

$$C_u = a_0 + a_1 V_1^{e1} + a_2 V_2^{e2} + \ldots\ldots + \ldots\ldots \quad (24)$$

where

C_u = unit cost($/unit)

a_0 , a_1, a_2, = constants

e_1, e_2, e_3, = exponents

V_1, V_2 , V_3, = feature parameters

The model would need to determine the appropriate exponents and constants for the equation. This is difficult as the standard regression analysis approach could not be used. The second form of the equation would be similar to the factor method where the equation form would be:

$$C_u = a_0 V_1^{e1} V_2^{e2} V_3^{e3} \ldots \quad (25)$$

where

C_u = unit cost

a_0= equation constant

e_1, e_2, e_3, = equation exponents

V_1, V_2, V_3, = feature parameters

The second model has only one constant rather than one constant for each term. This can be evaluated using linear regression on the logarithmic form of Equation 25.

$$\log C_u = \log a_0 + e_1 \log V_1 + e_2 \log V_2 + e_3 \log V_3 + \ldots. \quad (26)$$

Equation 26 is the equation which will be investigated by regression analysis for the conceptual model.

Data will be sought from various metalcasters and forgers and separate models will be developed and then an attempt will be to develop a single generalized model. If a single model cannot be developed, then variants may be developed for the factors that are selected first in the design process, such as the material.

6. Conclusions

A literature review has been undertaken for the various feature based cost models that have been developed for casting, forging, and machining. These models have considered only a few variables, such as volume or projected area, and the estimate accuracy is not good. Most of the data used in the models was from the 1950's and new data must be analyzed to develop improved relationships. With the use of CAD files, more information can be utilized to better estimate the product costs. A model has been proposed and data must be collected to obtain the appropriate cost model coefficients and to validate the feature based model after development..

Bibliography

1) Taylor Lyman, Editor Metals Handbook, Vol. 1, Properties and Selection of Metals, 8th Edition, 1961, American Society for Metals, Metals Park, Ohio, pp 901-915.

2) J. V. Michaels and W.P. Wood Design to Cost, John Wiley & Sons, 1989 , New York, pp. 177-179.

3) David C. Zenger, Methodology for Early Materials/Process Cost Estimating for Product Design, Ph.D. Dissertation, University of Rhode Island, 1989, pp 113-125.

14) Casting Design Handbook, American Society for Metals, Metals Park , Ohio 1962, p.326.

5) Pacyna, H. Hillebrand, A. and Rutz, A. "Early Cast Estimation for Castings", VDI Berichte, Nr. 457, Designers Lower Manufacturing Costs, Dusseldorf, pp 103-114.

6) M. S. Hundal, "Designing to Cost", Concurrent Engineering, edited by H.R. Parsaei and W. G. Sullivan, Chapman & Hall, , 1993, pp. 329-351.

7) Leo S. Wierda, "Product Cost-Estimation by the Designer", Engineering Costs and Production Economics, Vol. 13, 1988, Elsevier Science Publishers, pp. 189-198.

8) P. Radovanovic, Approximate Cost Estimating for Machined Components, University of Rhode Island, MS Thesis, Department of Industrial and Manufacturing Engineering, 1989, p 140.

9) R.C. Creese, M. Adithan, and B.S. Pabla, Estimating and Costing for the Metal Manufacturing Industries, Marcel Dekker, Inc., 1992, pp. 110-113, 139-142.

10) M.S. Hundal, "Time and Cost Driven Design", Design for Manufacturability - 1995, DE-Vol 81, ASME, pp. 9-20.

11a) M.F. Ashby Materials Selection in Mechanical Design, Pergamon Press, Oxford, 1992, p. 311.

11-b) M.F. Ashby, Materials Selection in Mechanical Design - Materials and Process Selection Charts, Pergamon Press, Oxford, 1992 p 51, 57 (note: - this book permits reproduction of the materials, whereas 11a does not.)

12) R. C. Creese, R. Atluri, and N. Rammohan, "Sandcasting Lead-Time Prediction Model", 1997 AACE International Transactions, AACE International, Morgantown, WV B&MP /A .01.1 - .01-6.

13) R. Cooper, When Lean Enterprises Collide, Harvard Business School Press, Boston, Mass., 1995, p. 201.

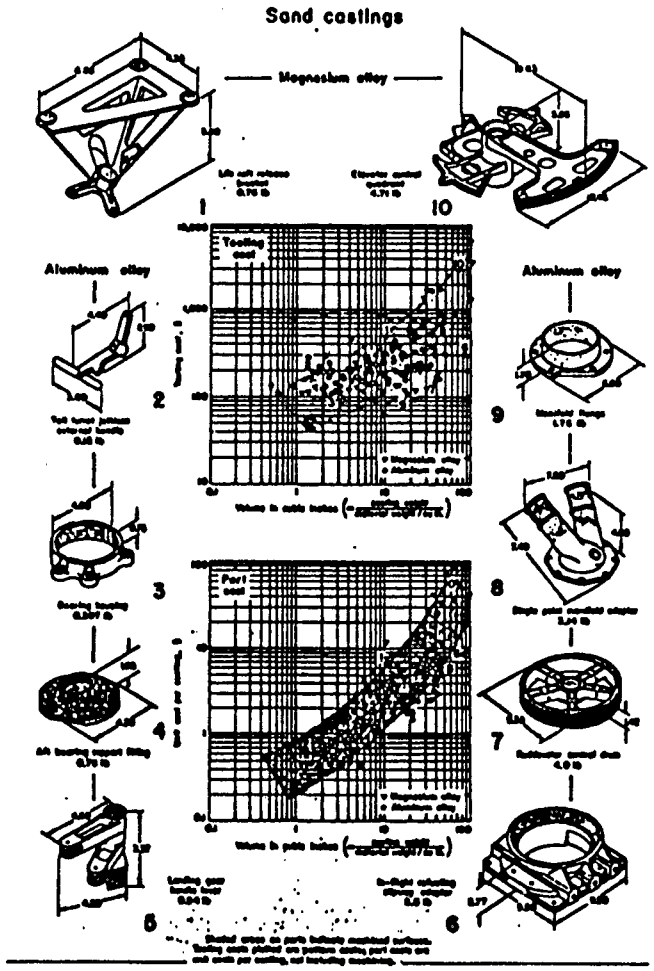


Figure 1 Recurring Unit Part Costs and Non-recurring Tooling Costs for Sand Castings.(Ref. 1)

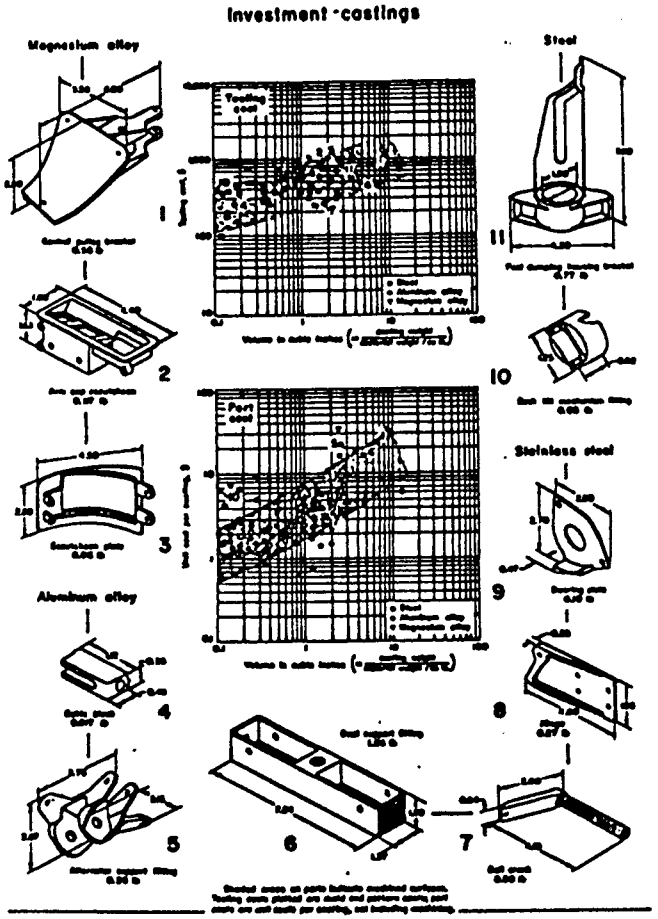


Figure 2. Recurring Unit Part Costs and Non-recurring Tooling Costs for Investment Castings.(Ref. 1)

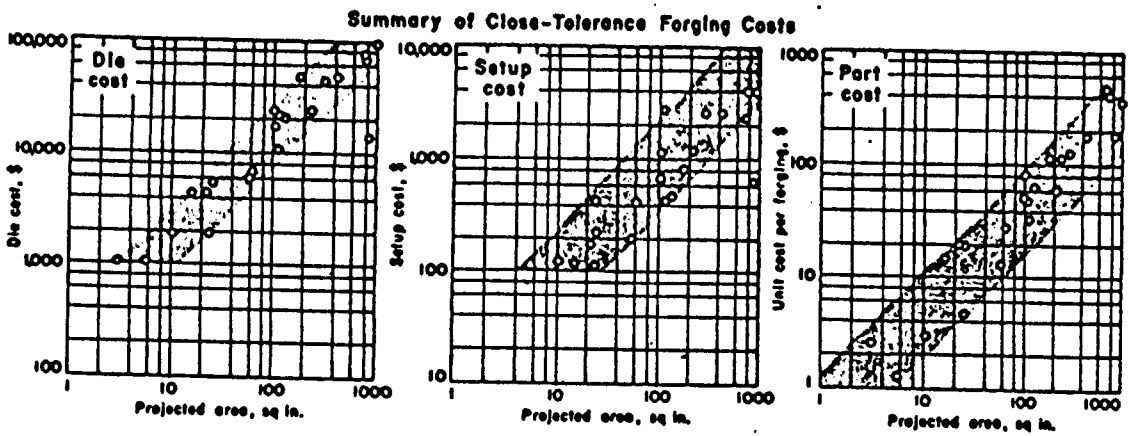


Figure 3. Recurring Unit Part Costs and Non-recurring Die Costs and Set-up Costs for Close-Tolerance Forgings.(Ref. 1)

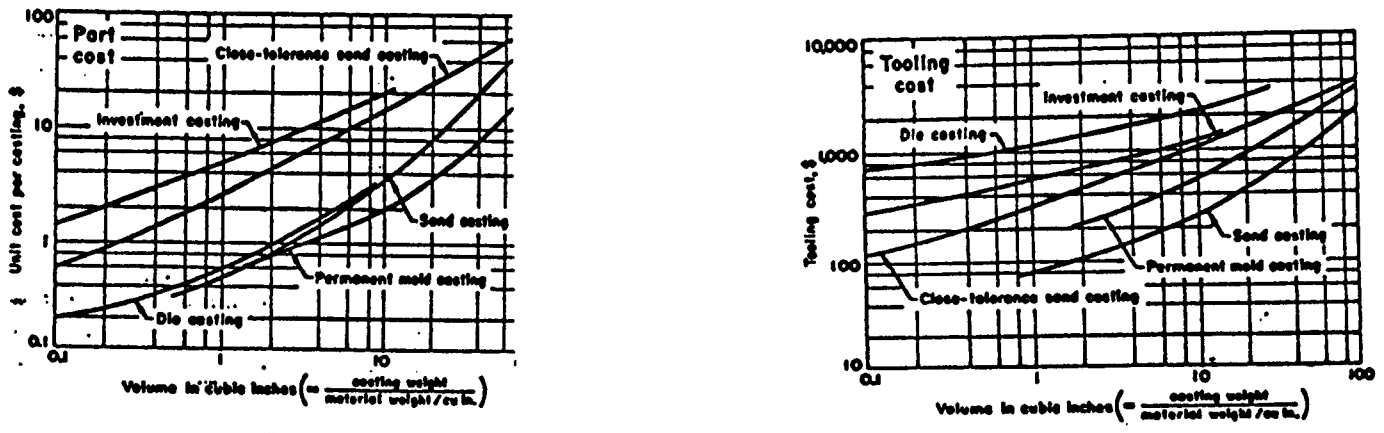


Figure 4. Recurring Unit Part Costs and Non-recurring Tooling Costs for Five Casting Processes - Sand Casting, Close-Tolerance Sand Casting(Composite Mold Casting), Investment Casting, Die Casting, and Permanent Mold Casting.(Ref. 1)

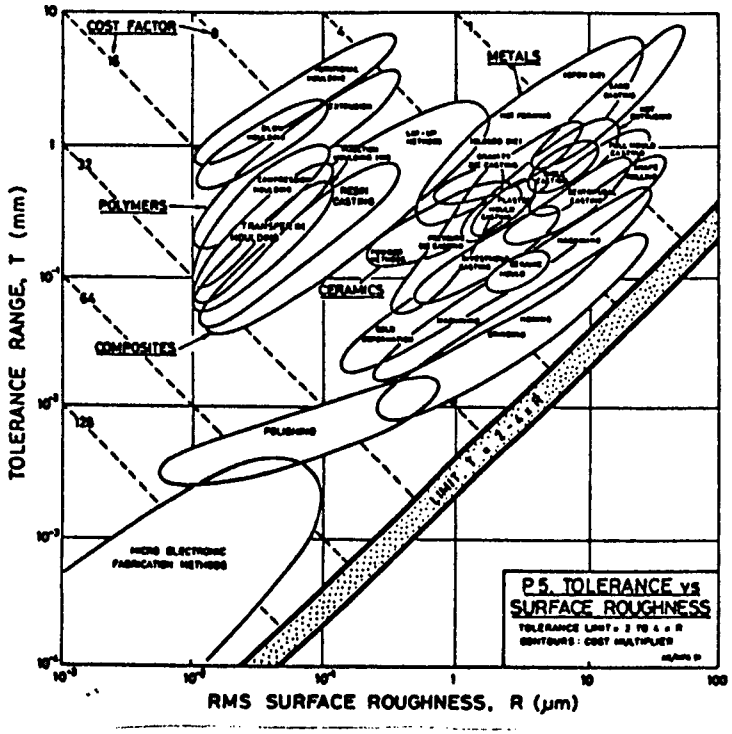


Figure 5. Complexity(Dimensions and Dimensional Precision) versus Part Weight.(Ref. 11-b, Ashby)

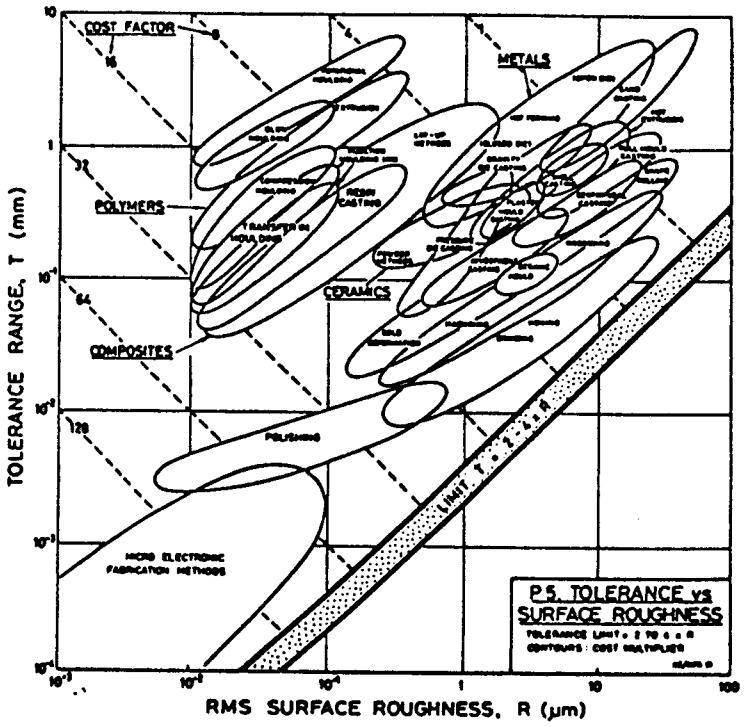


Figure 6. Tolerance Range and Surface Roughness with Relative Cost Factors.(Ref. 11-b, Ashby)

Table 1. Feature Based Estimating Expressions for Machined Parts

	Part Volume Range(cu.in.)	
Shape Category	**0.1 - 20 cu.in.**	**20 - 400 cu.in.**
1. Primary Rotational	$C = 4V^{0.33}$	$C = 0.7V^{0.89}$
2. Primary Rotational with Secondary	$C = 5V^{0.33}$	$C = 0.9V^{0.86}$
3. Primary Planar; Primary Planar and Rotational; and Primary Planar and Rotational with Secondary	$C = 8V^{0.33}$	$C = 2V^{0.81}$
4. Primary Planar with Secondary	$C = 9V^{0.40}$	$C = 4V^{0.70}$

Table 2 CONCURRENT ENGINEERING ACTIVITIES FOR MANUFACTURING

ACTIVITY	PRODUCT PLANNING	DESIGN	PROCESS PLANNING	MANUFACTURE
Project Time Scale	Product Goals Product Benefits Production Estimate (Maximum Quantity)			
		Specifications Technology Survey		
	Market Survey Product Features Product Performance			
		Conceptual Design **Conceptual Cost Model**		
	Target Cost Sales Volume(+/- 30%)		Primary Process Selection	
		Preliminary Design Prototype Design **Process Cost Model** Preliminary Tooling Design		Mfg. Capability Analysis Minimum Production Requirements
			Preliminary Process Design	
	Projected Sales(+/- 15%)			Prototype Manufacture Prototype Testing Preliminary Set-up Design --Prod.Tooling
		Detailed Design **Detailed Cost Model** Detailed tooling design	Detailed Process Design Detailed Bill of Materials	
		Operating Instructions Service Instructions	Product Scheduling	Tooling Order Part Procurement Materials Ordering
Start Production			Assembly Instructions	Production Runs **Accounting Model**
Production Ends				Packaging & Shipping

Table 3. Part Shape and Material Feature Data

A. Shape Feature Data

Part Volume
Box Volume

Projected Area(Maximum)
Projected Perimeter

Surface Area - Total
Surface Area - External
Surface Area - Internal
Number of Internal Surfaces

Surface Area - Roughness > 250 μ inch
Surface Area - Roughness 32-250 μ inch
Surface Area - Roughness < 32 μ inch

Tolerances/Dimensions
- Number of Dimensions
- Geometric Mean Dimension
- Geometric Mean Precision
- Minimum Precision Value

B. Material Feature Data

Density
Strength(Yield Strength)
Modulus of Elasticity
Number of Materials
Number of Layers

Table 4. Production and Cost Feature Data

A. Production Feature Data

Total Quantity
Lot Quantity
Maximum Production Rate Req'd(pcs/yr)
Maximum Production Time(Months)

B. Cost Feature Data

Production Cost Base Year
Total Cost
Unit Cost

Cost Components
Material Cost Components
- Raw Material Cost($/lb)

Product Material Cost($/unit)
- Material Yield

Tooling Cost
Total Tooling Cost
- Unit Tooling Cost($/unit)

Processing Cost
- Total Unit Processing Cost
- Unit Processing Cost - Internal
- Unit Processing Cost - External

Specification Assurance Costs
- Unit Testing Costs
- Unit Inspection Costs
- Other Unit Assurance Costs

Customer Costs
- Customer Change Costs

OBJECTS AND METHODS FOR AIRCRAFT CONCEPTUAL DESIGN AND OPTIMIZATION IN A KNOWLEDGE-BASED ENVIRONMENT

William A. Crossley
Assistant Professor
School of Aeronautics and Astronautics

Purdue University
1282 Grissom Hall
West Lafayette, IN 47907-1282

Final Report for:
Summer Faculty Research Program
Wright Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling AFB, DC
and
Wright Laboratory
Wright-Patterson AFB, OH

August 1997

OBJECTS AND METHODS FOR AIRCRAFT CONCEPTUAL DESIGN AND OPTIMIZATION IN A KNOWLEDGE-BASED ENVIRONMENT

William A. Crossley
Assistant Professor
School of Aeronautics and Astronautics
Purdue University

Abstract

This report discusses the development of object models and methods in the "Adaptive Modeling Language" (AML) environment to assist in the conceptual design of aircraft. Two related research efforts were made. The first emphasized the development of aircraft design objects that include properties and geometry necessary for aircraft conceptual design. The second focused on the creation of optimization methods for use in the AML environment. Several advantageous features of AML were exploited in both of these efforts. Recommendations resulting from this work offer promising advances for faster, more accurate aircraft conceptual design leading towards improved affordability of aircraft.

OBJECTS AND METHODS FOR AIRCRAFT CONCEPTUAL DESIGN AND OPTIMIZATION IN A KNOWLEDGE-BASED ENVIRONMENT

William A. Crossley

Introduction

The process of aircraft design receives much attention in the aerospace industry, because aircraft, both military and commercial, are very expensive systems. The cost of these systems is largely determined early in the design process, well before actual vehicles are constructed. Sophisticated analysis techniques are being used during aircraft design, and several sizing codes are found across the industry to assist in the conceptual design of aircraft. Numerical optimization techniques have also been introduced to improve solutions of several aerospace design problems. In spite of these efforts, there has been no concerted effort to enhance the capability of a design engineer during conceptual design to search through a vast, ill-defined design space to find an aircraft configuration well suited to perform a particular mission or task. Much of this search through the design space is still guided by qualitative design decisions influenced by the designer's experience, personal preferences and external pressures.

For "conventional" aircraft, today's designers make use of their accumulated knowledge to design new aircraft. As a result, most commercial jet transports have essentially similar configurations, a generally cylindrical fuselage with two high bypass turbofans mounted beneath a low-wing and a conventional tail arrangement. Aspiring to design and build newer aircraft, aerospace design engineers can rely only little on the knowledge and experience gained from designing current aircraft. New aircraft like Uninhabited Combat Air Vehicles (UCAVs) have virtually no experience base from which a design team can begin. Assistance in searching for the best concepts and configurations should be provided to the design team in these situations.

Conceptual design is the first phase of an aircraft design project. During this phase, little is known about the aircraft design, as a designer (or team of designers) begins to develop a design that will meet a given set of requirements. It is generally accepted that between 70 - 80% of an aircraft's cost is determined during the conceptual design phase,[1] when only a small portion of the knowledge required to complete the design actually is known. To decrease the risk involved with using this limited amount of knowledge, more detailed information about an aircraft design should be provided to designers early in the conceptual design process. Currently, this information relies on several historical databases that provide information about

the weight and size of an aircraft concept, but these databases are constructed with information about existing aircraft. These estimates work acceptably for aircraft concepts similar to those that comprise the database. However, if a design team attempts to design an aircraft significantly different than those in the database, the predictions and information will often be unreliable.

Much recent work toward optimal aerospace design has yielded promising results, but has focused more on preliminary design of aircraft components. This assumes that the design team has already made many decisions about the aircraft during the conceptual design phase. Providing the ability to perform optimization on the "ideas" generated during conceptual design can assist the creativity of a design team. When asking "what-if" types of questions while developing a conceptual design, decisions about the merit of various options are currently made using the aforementioned database and designers' experience as a guide. Optimization would allow the designer to "see" the best versions of the idea he/she had posed; in turn, unconventional concepts and ideas may lead to better designs than more traditional designs. To adequately perform optimization, reasonably accurate analyses are required; this is the major reason why optimization has not been fully employed in conceptual design of aircraft.

To address this issue, an ongoing effort at Wright Laboratory has been customizing a design modeling architecture for aircraft conceptual design. This architecture, called Adaptive Modeling Language (AML), is a knowledge-based environment that makes use of object-oriented programming features. Notable features for this work are *dependency tracking* and *demand-driven calculation*. Dependency tracking allows models to be developed with a set of properties and / or design variables; a property knows which other model properties influence its value and which other properties are influenced by it. Demand-driven calculation allows for properties to be evaluated only when needed. If one property is changed, other properties affected by this change are not updated until their values are needed. This can significantly reduce computational effort. Properties of a model can be design variables, and other properties may be constraint or objective function values. An optimization scheme could change design variables during its search, yet only require computation of those properties affected by those changes rather than all of the other properties.

OBJECTS AND METHODS FOR AIRCRAFT CONCEPTUAL DESIGN AND OPTIMIZATION IN A KNOWLEDGE-BASED ENVIRONMENT

William A. Crossley

It is intended to use this knowledge-based environment for conceptual design of aircraft. This report discusses two aspects required for this implementation: 1) the development of objects with properties relevant to aircraft conceptual design, and 2) the development of methods to perform numerical optimization in this environment.

Discussion of the Problem

As part of the current design process, aircraft designers generally employ "sizing codes" to assist during conceptual design. These computer programs are used because of their ability to estimate weight, required propulsive power and physical dimensions of given aircraft configurations to meet specified requirements over a defined mission. These include codes like ACSYNT[2] and FLOPS[3] for fixed-wing aircraft. This capability is also required in a knowledge-based environment for conceptual design.

At a minimum, a sizing code must incorporate primary design parameters, such as wing loading, aspect ratio, and taper ratio. These properties need to be included in a model representing an aircraft. Further, much of the aircraft sizing requires information about the aircraft's components to predict drag and empty weight of the vehicle. Because of this need for component information, the aircraft model object must be an assembly or group of component subobjects, which each has their own properties.

While conventional sizing codes have worked admirably, they have a significant shortcoming in that these codes cannot "draw" the aircraft that was "sized". The physical layout of an aircraft is crucial to the feasibility of a design, so a link to the aircraft geometry is important. Using AML as the design environment provides solid and surface modeling capabilities. Objects can be given a geometric description as well as parametric properties. This will allow the aircraft objects and component subobjects to have a geometrical description in addition to the solely numerical description in traditional sizing codes. Appropriate geometric descriptions need to be developed for this.

Equations to predict aircraft weight that incorporate only aircraft and component parameters, based on a database of existing aircraft designs, are available. With these, the weights of the aircraft's components are predicted, and the sum provides the empty weight of the aircraft. However, these parametric weight equations may not be appropriate for all possible concepts; again, especially those that vary significantly

from designs comprising the database. Similar predictions are made for the cost of an aircraft design via statistically based equations. Additionally, these are generally at the entire vehicle level rather than at the component level further reducing fidelity. Aerodynamic predictions for lift and drag using first-order predictions are slightly higher fidelity. Existing sizing codes make use of these computations out of necessity; higher fidelity analyses would generally require far too much computational overhead to be included.

The AML-based environment allows for aircraft and component properties, like the weight of a component or drag of a component, to be calculated with a "method". Because of the dependency tracking and demand-driven calculations, these methods can be changed during the design process. For example, a designer could begin formulating a design using empirical weight equations in a method, then replace these with a more sophisticated calculation method accounting for the amount of material used in each component. This *hybrid interpretive-compiled* feature of AML can provide significant advantage to a designer who is attempting to explore various concepts that may not fit one of the aforementioned databases. Therefore, properties and associated methods should be developed as part of the component objects.

Also, optimization of various components of the aircraft design can also be accomplished via methods in the design environment. These methods may either be developed in the AML construct (an adapted version of Common Lisp) or as external modules in C, C++, Fortran or Lisp. To examine these methods, development in AML should provide the most straightforward integration.

Methodology and Results

As a first step in this effort, geometric models for aircraft components were developed. A top-down view of the aircraft was adopted, so that an aircraft object is first created. Major components were then considered to be subobjects of the aircraft. The component subobject models developed were fuselage, wing, horizontal tail, vertical tail, and engine. Basic properties for this example aircraft were based on a cargo or transport aircraft, although this is not a limitation of the approach.

Aircraft Object

The aircraft object was developed using the define-class form in AML. It inherits from the group-object, which creates one aircraft geometry instance, yet allows the various components to retain their original colors, line types, etc., when presented graphically. Properties assigned to the aircraft object were wing loading, thrust-to-weight ratio, number of engines, and gross weight. These properties help to describe the aircraft in an overall sense. Each component is then constructed with its own properties and, in most cases, subobjects. Figure 1 presents the aircraft model in the graphical representation of AML.

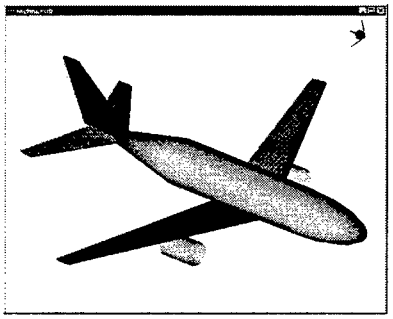


Figure 1 Aircraft model in AML.

Fuselage Object

The fuselage of an aircraft serves as an envelope in which the payload, equipment, etc. are enclosed. Therefore, the size of the aircraft must be such that the desired contents will fit inside. The external shape then affects the aerodynamic performance and structural weight of the aircraft. Properties assigned to the fuselage included length, diameter, and fineness ratio; currently length and diameter are entered and fineness ratio calculated, although these property definitions can easily be modified. For this example, the fuselage object was constructed as an assembly of three subobjects, the nose, the body and the tail. The body and tail subobjects inherit from AML's open-cylinder-object and open-cone-object, respectively. The body and tail also have parametric properties, which are "driven" by the fuselage's properties. The diameter of the body is equal to the diameter of the fuselage. The tail's base diameter also is equal to the fuselage diameter, while its length is taken as 2.5 times the diameter; this is based on guidelines in Ref. 4. An intersection-object of an ellipsoid and a half-plane provides the nose object. The base diameter again equals the fuselage diameter, while the length is specified 1.75 times the diameter.[4] Finally, the length of

the body is determined as the fuselage length less the nose and tail lengths. These "rules" to describe the dimensions are easily modified, so the shape of the aircraft is not limited to these relationships.

Wing Object

The wing object model provides an additional level of complexity. Here, the wing is defined as a geom-object with subobjects which describe the wing planform, the airfoil shapes at the root at the tips, and the surface (or skin) of the wing. The wing is assigned properties that parametrically describe the wing as discussed in Refs. 4 and 5; these are: area (here reference or planform area), aspect ratio, dihedral, incidence, taper, twist, sweep (of both quarter-chord and leading-edge), root, tip and mean aerodynamic chords, span, and aerodynamic center location. As with the fuselage, several of these drive the other properties. However, some of the aircraft object properties drive these fuselage properties. The aircraft property of wing loading determines the area of the wing, which, along with other properties, defines the span and chords of the wing.

For the wing, a configuration-related property was added. This property, "vertical-position" is used with several conditional statements to place the wing relative to the aircraft and to select appropriate dihedral and incidence angles. Currently, the "vertical-position" property recognizes the strings "low", "mid" and "high"; these correspond to the general categories discussed in Raymer.[5] When modifying this property, all of the affected properties are recalculated and a new geometry drawn. Because this is implemented in AML, only the affected components are redrawn, rather than the entire aircraft. Figure 2 shows the effect of the three property values. The effect on wing dihedral and vertical position are evident. Only changing the vertical-position property makes these changes in the wing object; affected values are recomputed and the wing redrawn when demanded.

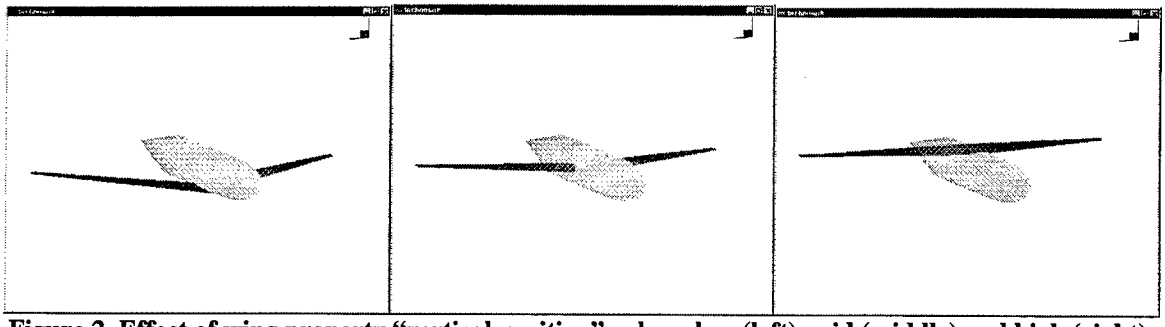


Figure 2 Effect of wing property "vertical-position" values, low (left), mid (middle) and high (right).

The planform object inherits from the AML polygon-object and provides the two-dimensional representation of the wing. In its current form, this object's vertices are driven by the wing properties, like taper, sweep, span, etc. because of the straight-tapered shape assumption. If desired, these vertices could be used to drive the wing properties, like area and aspect ratio.

The airfoil object provides the ability to move to higher levels of fidelity in aerodynamic modeling of the aircraft. Most sizing codes only require the planform dimensions and a measure of the lift coefficient to estimate much of the aerodynamic performance of the wing. An airfoil object allows various shapes to be used for the wing section. This also provides a natural means for incorporating more complex aerodynamic analyses ranging from inviscid panel codes to Navier-Stokes codes. To develop the airfoil object, the common approach of describing airfoil shapes with station and ordinate locations was followed. The airfoil object is a geom-object with the subobjects "upper-surface" and "lower-surface". These surfaces inherit from the AML NURB-object, so that a third-order B-spline is fit to the station and ordinate points describing upper and lower surfaces. Example airfoil objects are presented in Fig. 3.

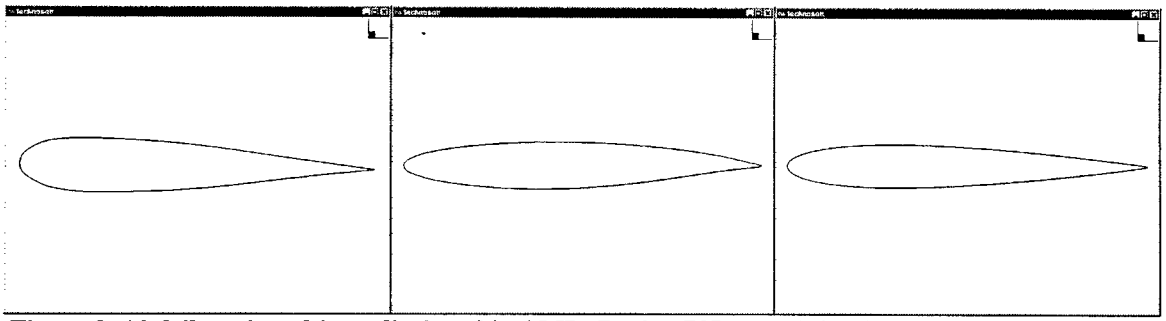


Figure 3 Airfoil section objects displayed in AML; Boeing 737 root airfoil (left), Lockheed C5A root airfoil (middle), NACA 0012 (right).

The properties of the airfoil object include the "airfoil-name" property. This text string property describes the name of a file containing the non-dimensional station and ordinate points of the airfoil; a second property, "airfoil-data" then uses the "formatted-list-from-file" function to read the data. The input file is named *airfoil-name*.dat so that changing the airfoil-name property will change the input file allowing a wide variety of airfoil shapes. In the airfoil-name.dat file, any lines preceded by a ";" will be ignored, so descriptions can be included in the body of the file. The first line in the file contains the number of points in the file describing the upper and lower surfaces, respectively. Then, the next lines are the x/c and y/c for the upper surface, starting from the leading edge (x/c = 0.0) to the trailing edge (x/c = 1.0). Finally, the remaining lines are the x/c, y/c values for the lower surface points, again from leading to trailing edge. The non-dimensional values are converted to dimensional values using the appropriate wing chord properties from the wing object. These are then further converted to the (x, y, z) coordinates of the aircraft, based on wing properties.

Finally, the surfaces of the wings are also treated as subobjects. These are based on the AML "surface-skin-object" and use the upper and lower surface curves of the airfoil subobjects to describe the shape of the wing surface. In the current implementation, there is a separate surface for the upper right wing, lower right wing, upper left wing and lower left wing.

Tail Objects

Because tails are essentially small wings, the basic wing object was copied and slightly modified to create the horizontal and vertical tail objects. The tail objects have similar dimensional properties, including aspect ratio, chords, taper and sweep angles, to the wing object. These tail objects also have airfoil and skin subobjects as described for the wing. Differences between the tail objects and the wing object include the use of tail volume coefficients and moment arm lengths to define areas for the tails. Additionally, the horizontal tail has symmetric left and right sides, while the vertical tail is non-symmetric and, in this implementation, extends only above the fuselage.

A configuration variable has also been included in the tail objects. The property "tail-configuration" has a string value that is used with several conditional statements to determine the relative placement of the

vertical and horizontal tail. To date, the values "conventional", "cruciform" and "T-tail" are recognized. As with the vertical-position property of the wing, appropriate changes are made to the geometry based on the change in the tail-configuration property. Figure 4 shows the effect of the three available choices.

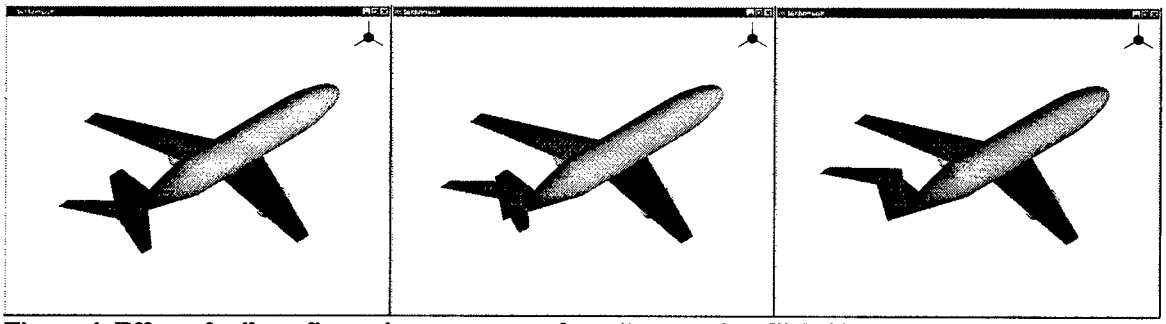


Figure 4 Effect of tail-configuration property values, "conventional" (left), "cruciform" (middle) and "T-tail" (right).

Engine Object

The engine object is the simplest of the component objects developed for this effort. Currently, it inherits from the open-cylinder-object of AML. The properties of length and diameter are based on a "rubber" engine approach. From the aircraft properties of thrust-to-weight and number of engines, the thrust of each engine is determined as a property. A baseline engine has been defined, so the length and diameter of the engine object are based on the ratio of thrust of the current engine to the baseline engine.

A configuration property, "engine-placement" was also investigated for the engine object. This has been given two valid string values: "under-wing" and "aft-fuselage". These choices provide a Boeing 737-like arrangement, or a McDonnell Douglas MD-80 arrangement. Further, the "under-wing" value allows the engines to move with the wing; for example, if the wing vertical-position is changed from "low" to "high", the engines move with the wing. These are displayed in Fig. 5.

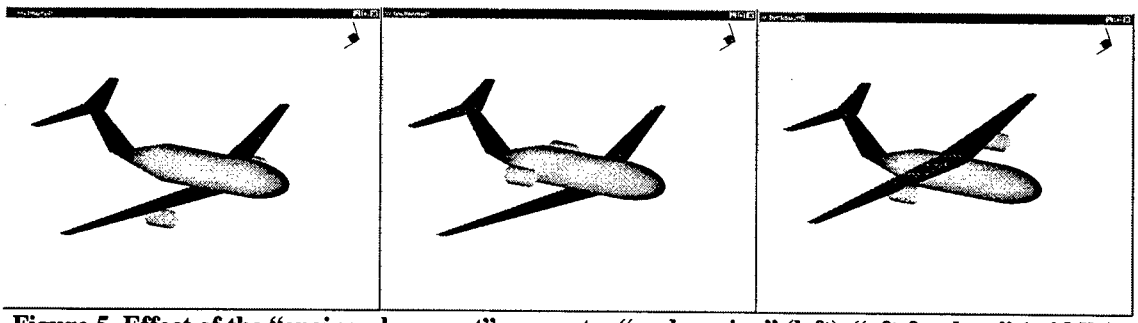


Figure 5 Effect of the "engine-placement" property; "under-wing" (left), "aft-fuselage" (middle), and "under-wing" with high-wing (right).

Aircraft Methods

In addition to the properties described with the aircraft models above, several properties of the aircraft and its components will require more than simple formulae to compute their values. These properties, like lift and drag coefficients or structural weight, can be computed through the use of AML methods. The method can be defined for objects of a given class; these methods can then be used to perform various calculations and analyses to provide values for a property. To examine this capability, an approach to calculate the weights of the various aircraft components was employed. These weight calculations were based on the empirically derived equations given in Raymer's text; these historically based equations have several drawbacks for aircraft design as discussed previously. For this effort, these were used to demonstrate the utility of the AML model construct, rather than the actual equations themselves.

These methods were created for each of the aircraft components. The method to compute the wing weight, for example, was named "get-wing-weight" and was defined for the wing-object. Then, in the wing object definition, the property "weight" is assigned by calling the get-wing-weight method. An advantage of this approach are that the calculations used in these methods could easily be replace by more sophisticated analysis techniques which would allow a designer to have a better description of the current design. Because more sophisticated analysis implies a longer calculation time, this approach gains further benefit in the AML environment. These analyses will only be conducted when the weight value is demanded, and then due to dependency tracking, this will be calculated only if properties which affect the weight have been changed.

William A. Crossley

Optimization Methods and Objects

In addition to the basic objects created for the conceptual design of aircraft, methods to include numerical optimization techniques into the design environment were also desired. These methods are intended to be combined with AML objects so that a designer could include an "optimization-problem" subobject in the object of interest for optimization. Properties of this optimization-problem subobject would then refer to properties of the parent object. In this manner, the design variables, constraint functions and objective functions for the optimization can be changed easily by modifying the properties of the optimization subobject. Methods and functions are then used to perform the numerical optimization on the optimization subobject.

The basic idea of a numerical optimization search is to move from a given design point to a new design point at which the objective function is reduced. This continues until no further reduction is possible. To carry out this search, two major tasks are performed. The first is determining an appropriate search direction, $\mathbf{s}$; the second is determining the appropriate step size, α^*, to move in the search direction. The next design point can be expressed as

$$\mathbf{x}^i = \mathbf{x}^{i-1} + \alpha^* \mathbf{s}^i \tag{1}$$

Step-length Method

The step-length determination requires a one-dimensional search (or line-search) to find the value of α^* which, for a given search direction minimizes the function

$$f\left(\mathbf{x}^{i-1} + \alpha^* \mathbf{s}^i\right) \tag{2}$$

Various approaches exist for this type of effort. For the method defined in AML, a polynomial interpolation scheme provides the line search mechanism. This version of the polynomial interpolation uses a cubic fit using both the function and the slope value at two points that are evaluated in the following manner

$$f_1 = f\left(\mathbf{x}^{i-1} + \alpha_1 \mathbf{s}^i\right) \tag{3}$$

$$f_2 = f\left(\mathbf{x}^{i-1} + \alpha_2 \mathbf{s}^i\right) \tag{4}$$

$$g_1 = \frac{df(\alpha_1)}{d\alpha} = \left[\nabla f\left(\mathbf{x}^{i-1} + \alpha_1 \mathbf{s}^i\right)\right]^T \mathbf{s}^i \tag{5}$$

$$g_2 = \frac{df(\alpha_2)}{d\alpha} = \left[\nabla f\left(\mathbf{x}^{i-1} + \alpha_2 \mathbf{s}^i\right)\right]^T \mathbf{s}^i \tag{6}$$

These points α_1 and α_2 are assumed to be bounds of the minimum.

From these function evaluations, the cubic polynomial approximation is constructed with the form

$$p(\alpha) = a(\alpha - \alpha_1)^3 + b(\alpha - \alpha_1)^2 + c(\alpha - \alpha_1) + d \tag{7}$$

where the coefficients are found as

$$a = \frac{[-2(f_2 - f_1) + (g_1 + g_2)(\alpha_2 - \alpha_1)]}{(\alpha_2 - \alpha_1)^3} \tag{8}$$

$$b = \frac{[3(f_2 - f_1) - (2g_1 + g_2)(\alpha_2 - \alpha_1)]}{(\alpha_2 - \alpha_1)^2} \tag{9}$$

$$c = g_1 \tag{10}$$

$$d = f_1 \tag{11}$$

The α value that minimize the polynomial approximation is easily expressed as

$$\alpha^* = \alpha_1 + \frac{-b + \sqrt{b^2 - 3ac}}{3a} \tag{12}$$

This requires that $a \neq 0$ and b^2-$3ac \geq 0$ to solve. If either of these is not satisfied, through the use of conditional statements, the method fits a quadratic polynomial using the values of f_1, g_1, and f_2 in the following form

$$p(\alpha) = b(\alpha - \alpha_1)^2 + c(\alpha - \alpha_1) + d \tag{13}$$

where the coefficients are found as before. Now, the solution is

$$\alpha^* = \alpha_1 + \frac{c}{2b} \tag{14}$$

Because the polynomial interpolation is only approximate, an iterative scheme is required to actually find the value of α which minimizes the actual function. The value of α^* is calculated, and $f(\alpha^*)$ and $g(\alpha^*)$

are also computed. If $g(\alpha^*)$ is suitably close to zero, the iteration stops. If not, the value of α^* becomes one of the new points used to improve the polynomial approximation. Assuming a valid search direction has been found, the slope of the polynomial is negative for $\alpha < \alpha^*$ and positive for $\alpha > \alpha^*$. Using this information, if the non-zero value of $g(\alpha^*)$ is negative, α^* replaces α_1; otherwise it replaces α_2 and the process repeats until convergence is reached.

To develop a method in AML that will conduct the line-search, the "do*" construct from common lisp has been adopted, which allows iteration to be controlled via an "end-test statement". Using a starting set of α_1 and α_2 values, the above iteration procedure is conducted in the do* until the absolute value of $g(\alpha^*)$ is less than 1×10^{-5}.

Bounds Method

The line search mentioned above assumes that the values of α_1 and α_2 bound the minimum value. To ensure this, a bounding method was also developed in AML, following the suggested algorithm given in Vanderplaats.[6] This bounding method begins by assuming that $\alpha_L = 0$, since the slope here is known to be negative. Using $\alpha_U = 1$ as the other estimate for a bound, the values of $f(\alpha_L)$ and $f(\alpha_U)$ are calculated. If $f(\alpha_L) < f(\alpha_U)$, then it is assumed that α_U is an acceptable upper bound, as the slope of $f(\alpha)$ must change from positive to negative between the two points. If $f(\alpha_L) > f(\alpha_U)$, an interim value, α_1, is set to the previous value of α_U and a new value for α_U is chosen, as it cannot be guaranteed that the minimum lies between the two points α_L and α_U. Now the value of $f(\alpha_1)$ is computed, and if $f(\alpha_1) < f(\alpha_U)$, then the slope between α_1 and α_U is positive, so α_U is an appropriate upper bound. If $f(\alpha_1) > f(\alpha_U)$, then there is still no guarantee on bounds. In this case, α_L is set to the previous value of α_1, α_U is set to the previous value of α_1, and a new α_U is calculated. This continues until values of α_L and α_U are found which bracket a minimum point. Because this also requires an iterative process, the do* construct is used to control the iterations until acceptable bounds are found.

Search Direction Method

The last method required to perform optimization in AML is one that determines the search direction. An acceptable search direction is one for which a move in that direction in the design space will reduce (increase) the objective function for minimization (maximization). For this work, minimization is assumed.

The simplest gradient-based approach for minimization is the steepest descent method.[6] This approach chooses the search direction as the vector opposite to the gradient of the objective function.

$$\mathbf{s}^i = -\nabla f\left(\mathbf{x}^{i-1}\right) \tag{15}$$

Always selecting this as the search direction leads to rather poor convergence, as this approach does not use information from previous evaluations to speed up the search process. However, this is generally used to start more sophisticated methods, and its behavior is well understood. This was the first approach developed using AML; again the do* iteration scheme was employed.

Convergence of a gradient-based method is generally determined in a combination of three basic approaches. First, a maximum limit on iterations is generally imposed; the version developed for AML was given a limit of 100. Second, if the change in the objective function is essentially zero convergence is often assumed; this is usually a practical consideration and not mathematically rigorous. In the AML method, this is not currently included. Lastly, the Kuhn-Tucker necessary conditions are checked. This necessary condition is a mathematically rigorous condition for describing an optimal point. The version developed in AML checks that the magnitude of the gradient vector is less than 1×10^{-5}. The end-test statement in the do* construct stops iteration if either the first or the third condition exists.

Search Results

To begin this effort, an approach to solve a simple example was adopted. A simple two-dimensional, unconstrained function was chosen to provide the objective function.

$$f(\mathbf{x}) = x_1^{\,4} - x_1^{\,2}x_2 + e^{x_2} + e^{-x_2} \tag{16}$$

This function has a known minimum of $f(\mathbf{x}^*) = 2$ at $\mathbf{x}^* = [0\ 0]^T$. This provides a simple test for an optimization routine, as it requires more than a few iterations to solve via most methods.

An optimization-problem object was constructed in AML. This object inherited from the object-class and has the properties: objective-function, gradient-vector, x-vector, and best-x-vector. For now, the objective function is a user-defined function. The gradient is also a user-defined function that returns a list containing the values of the gradient-vector components. The x-vector property describes the initial design point from which the search begins; currently $\mathbf{x} = [4\ 4]^T$ is used. Finally, the best-x-vector property is computed via the optimize method which uses the steepest descent algorithm described previously.

An instance of the optimization-problem was then created as a model. The modify-properties command was used to open the interactive window for modifying properties. Because AML makes use of the demand-driven calculations, the best-x-vector is not computed until required. In the modify-properties window, choosing the best-x-vector property initiates the optimization. After several iterations, the best-x-vector property is assigned the list (-0.016581625648272 1.29068458986499E-4). At this point, the objective function value is assigned the value 2.00000005676893, which is essentially equal to 2. Also, the gradient vector property is assigned the list (-1.39561626522799E-5 -1.6813390449899E-5), which is very close to the zero-vector. This demonstrates that the optimization objects and methods developed in AML work as expected.

As with most optimization approaches, the use of several starting points is warranted. This was also done for the example problem. Results similar to those described above were generated.

Conclusions

While this work is preliminary in nature, several conclusions and recommendations can be made about the use of objects and methods to conduct aircraft conceptual design and optimization in a knowledge-based environment.

First, the aircraft design objects created displayed the power of dependency tracking, as a change in one aircraft property or component affected all other dependent properties. Changing the wing loading property of the aircraft not only re-sized the wings, but the tails also change shape to maintain the appropriate tail volume coefficient, for example.

The demand-driven calculation behavior in AML was also demonstrated for aircraft design. Making a change in wing placement, for example, would also change the location of wing-mounted engines. However, the values describing the new location of the engines and their graphical representation was not presented until asked for. This would save time if many changes needed to be made to one property or component, as the affected components would not require time for recalculation and re-drawing until needed.

It is possible to construct an object whose properties successfully describe the objective and design variables of an optimization problem. The simple steepest-descent method was easily coded and successfully employed to solve a mathematical test problem. As with the aircraft design objects, the demand-driven calculation behavior of AML required the optimization to be conducted only when the solution was demanded, not whenever an associated property was changed. This appears to be of significant potential to incorporate optimization methods in the conceptual design phase for aircraft. An optimization-problem object could be used as a subobject of other objects of interest. For example, the internal structure of a wing may be a subobject of an aircraft model. If the structure is to be optimized for minimum weight, the appropriate properties can be assigned in the optimization subobject. When the aircraft geometry is changed, the wing structure is affected, but the optimization is not conducted until the new optimized parameters are needed. This would allow conceptual design of the aircraft to continue, including large changes in wing layout, planform, size, etc., with the optimal wing structure always available upon demand.

Recommendations

Two major recommendations can be made as a result of this summer research project.

First, the geometry of the aircraft and component models is not sufficient to represent a wide range of aircraft; developing improved geometry, incorporating the properties discussed in this report, should continue.

Second, further investigation and development of the optimization subobjects and methods is required. The approach used in this work is too simple and computationally expensive for practical application; more

sophisticated methods should provide faster and more accurate results. Additionally, the method explored in this work used and unconstrained minimization technique. Most problems for aircraft design will have constraints and either direct or indirect constrained optimization approaches should be explored.

These recommendations suggest several enhancements to aircraft conceptual design. Significant among these are the ability to help designers avoid being "trapped" by design databases, and the opportunity to rapidly include advanced analysis techniques early in the design process, both of which support the potential for more affordable aircraft.

References

1. McMasters, J. H. and Matsch, L. A., "Desired Attributes of an Engineering Graduate - An Industry Perspective," AIAA Paper 96-2241, 19th AIAA Advanced Measurement and Ground Testing Technology Conference, New Orleans, LA, Jun. 17-20, 1996.
2. Myklebust, A., and Gelhausen, P. "Putting the ACSYNT on Aircraft Design," *Aerospace America*, Vol. 32, No. 9, 1994, pp. 26-30.
3. McCullers, A., *FLOPS User's Manual*, NASA Langley Research Center, 1995.
4. Torenbeek, E., *Synthesis of Subsonic Airplane Design*, Delft University Press and Kluwer Academic Publishers (jointly published edition), Norwell, MA, 1988.
5. Raymer, D. P., *Aircraft Design: A Conceptual Approach*, AIAA Educational Series, AIAA, Washington DC, 1989.
6. Vanderplaats, G. N., *Numerical Optimization Techniques for Engineering Design*, McGraw-Hill, New York, 1984.

VIBRATIONAL ANALYSIS OF SOME HIGH-ENERGY COMPOUNDS

Gene A. Crowder
Professor
Department of Chemistry

Louisiana Tech University
Ruston, LA 71272

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

And

Wright Laboratory

August 1997

VIBRATIONAL ANALYSIS OF SOME HIGH-ENERGY COMPOUNDS

Gene A. Crowder
Professor
Department of Chemistry
Louisiana Tech University

Abstract

Infrared spectra were obtained and vibrational (normal coordinate) calculations were made for 1,3,3-trinitroazetidine (TNAZ), 1-acetyl-3,3-dinitroazetidine (ADNAZ), and 1-nitroso-3,3-dinitroazetidine (NO-DNAZ) in an effort to learn more about the conformational behavior of these compounds and to make assignments of the observed infrared bands to the appropriate normal modes of vibration. Molecular mechanics and semi-empirical molecular orbital calculations (MNDO-AM1) were also made in order to obtain additional information about the molecular structures. Normal coordinate calculations were made first for the slightly less complex molecule 1,3-dinitro-3-bromoazetidine in order to obtain force constants to transfer to ADNAZ. Appropriate force constants obtained for ADNAZ were then used as starting values for TNAZ. The resulting force constants obtained for TNAZ were transferred successfully to NO-DNAZ. It was shown that the observed frequency shift of the C=O stretch band of ADNAZ and the (N)-NO_2 antisymmetric stretch and (N)-NO_2 out-of-plane wag bands of TNAZ can be explained by a change in the conformation that is initially present to another conformation during recrystallization after mixtures are melted. The normal coordinate calculations produced vibrational potential energy functions that resulted in calculated vibrational frequencies that were in excellent agreement with the observed frequencies for all four compounds.

VIBRATIONAL ANALYSIS OF SOME HIGH-ENERGY COMPOUNDS

Gene A. Crowder

Introduction

The phase diagrams of the 1,3,3-trinitroazetidine/1-acetyl-3,3-dinitroazetidine (TNAZ/ADNAZ) and the TNAZ/1-nitroso-3,3-dinitroazetidine (NO-DNAZ) systems show some unusual behavior [1], perhaps because one or both of these compounds exhibit polymorphism. The infrared vibrational spectra of mixtures of ADNAZ and TNAZ and of TNAZ and NO-DNAZ change slightly after melting and recrystallization, so the possibility arose that some information about the unusual behavior might be obtained from an in-depth study of the infrared spectra. The real question was whether the abnormal behavior was the result of polymorphism or some other type of behavior such as rotational isomerism. It is likely that each of these compounds exists as a dynamic mixture of more than one stable conformation in the liquid or solution states. These conformations would be interconverted by internal rotation of the NO_2 and $OCCH_3$ groups about the single bond that connects each group with the remainder of the molecule. Studies of the vibrational spectra of the compounds and of mixtures of the compounds may yield some information about rotational isomerism but probably not about polymorphism.

Methodology

The vibrational frequencies of a compound depend on the masses of the atoms, the structure of the molecule, and the force constants for bond stretching, angle bending, and interactions between these different coordinates (stretch-stretch, stretch-bend, and bend-bend). If the structure and vibrational frequencies of a molecule are known, force constant values can be determined that reproduce the frequencies satisfactorily. If the force constants and frequencies are known, information about the molecular structure can often be determined. That is the focus of the present research. It was thought that the changes in the infrared spectra mentioned in the introduction might be explained by determining force constant values of suitable vibrational potential energy functions for TNAZ, ADNAZ, and NO-DNAZ and using those force constants to calculate vibrational frequencies for each compound in different molecular conformations. In that way, some information about the molecular conformations that exist might be obtained.

Those conformations would be interchanged by internal rotation of the $O=C-CH_3$ and/or NO_2 groups in ADNAZ and of one or more of the NO_2 groups in TNAZ and NO-DNAZ about the single bonds that connect the groups to the ring. Calculations of this type are called normal coordinate calculations.

Results and Discussion

1,3,3-trinitroazetidine and 1-acetyl-3,3-dinitroazetidine

Infrared spectra that were obtained for pure TNAZ and ADNAZ are shown in Fig. 1 for the 500-2000 cm^{-1} region, and spectra for an initial 50:50 (mol percent) TNAZ/ADNAZ mixture and after melting and recrystallization are shown in Fig. 2. The spectra shown in Fig. 1 show quite a few similarities because of the similar structures of the two compounds, and differences in the two spectra shown in Fig. 2 can also be seen.

Normal coordinate calculations were started first for TNAZ. This means that the 42 fundamental vibrational frequencies for this molecule (excluding the torsions) are being calculated using the computer program MOLVIB (version 6.0), which was written for a PC by Dr. Thomas Sundius of the University of Helsinki and distributed by the Quantum Chemistry Program Exchange office of Indiana University [2]. Unfortunately, no force constant value data were available, so an educated guess at all the force constant values had to made. A large eighty-one parameter potential energy function was used, and most of the calculated frequencies were not very close to the observed values. After approximately 30 computer runs, the frequencies above 1500 cm^{-1} were fit satisfactorily. However, the molecule is so complex that the frequencies below 1500 cm^{-1}, where a considerable amount of mixing of normal modes is expected, were not fit very well, and it was difficult to determine which force constants needed to be adjusted. Therefore, normal coordinate calculations were made for the simpler molecule 1,3-dinitro-3-bromoazetidine, which has one of the 3-nitro groups of TNAZ replaced by a bromine atom. This should result in better starting force constant values to use for TNAZ.

Normal coordinate calculations were completed satisfactorily for 1,3-dinitro-3-bromoazetidine. The infrared spectrum for a crystalline sample of 1,3-dinitro-3-bromoazetidine in a KBr pellet was obtained from Los Alamos National Laboratory. The structural parameters (bond lengths, bond angles, torsional angles) were obtained by doing semi-empirical molecular

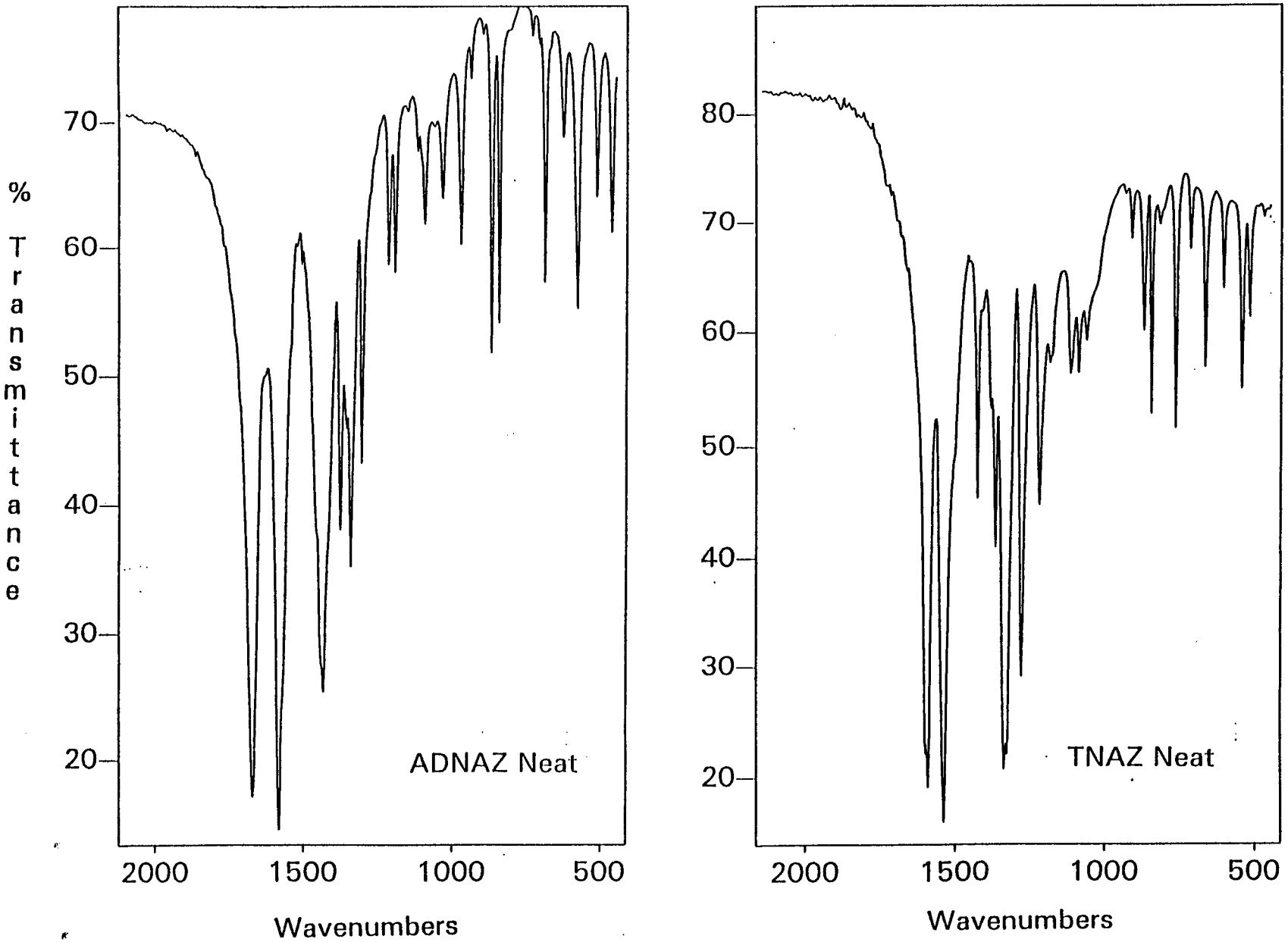


Fig. 1. Infrared spectra for ADNAZ and TNAZ in KBr pellets

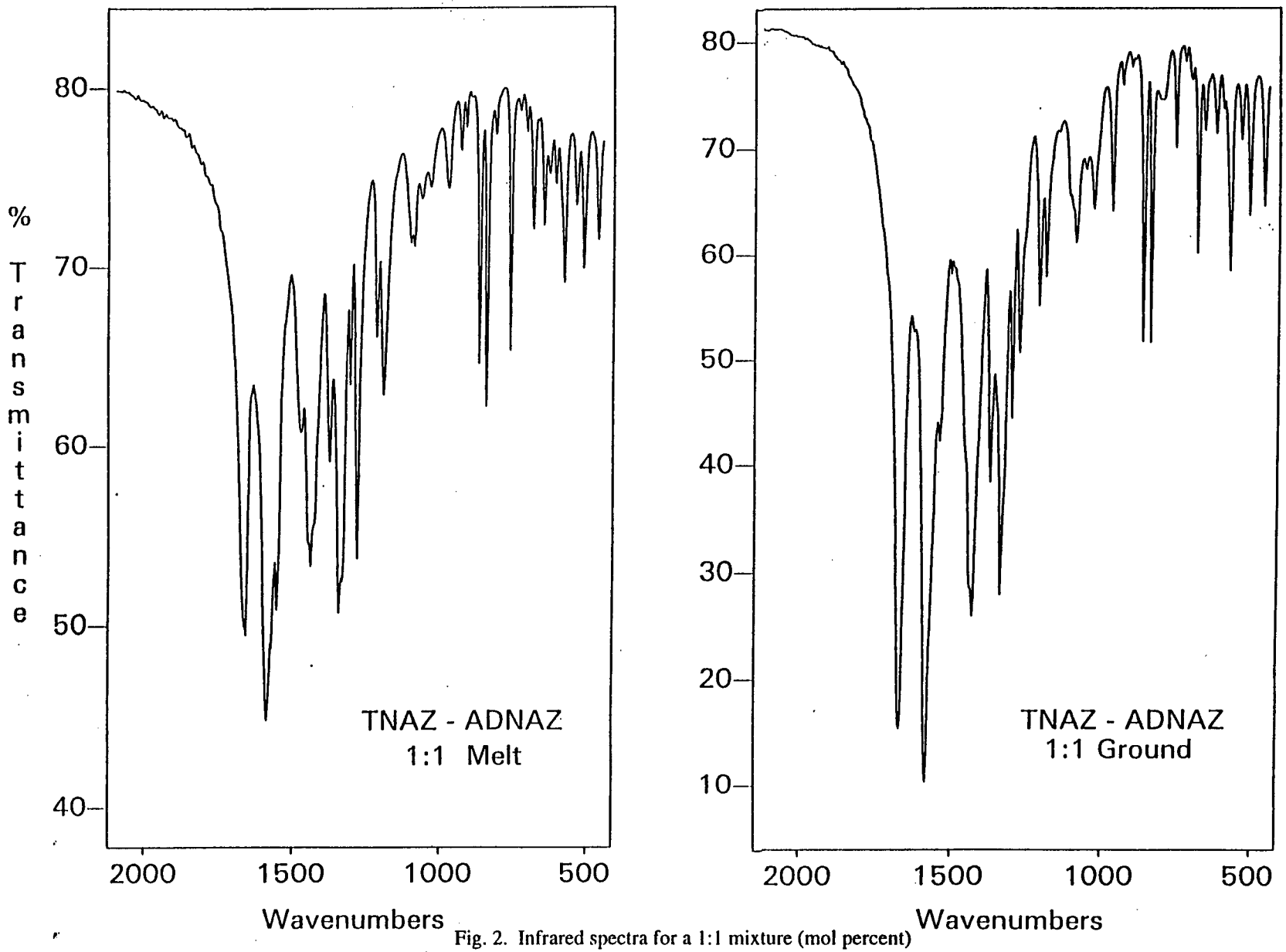


Fig. 2. Infrared spectra for a 1:1 mixture (mol percent) of TNAZ and ADNAZ. Right, initial mixture; Left, same mixture after melting and recrystallization

orbital calculations with a MNDO-AM1 force field. The Hyperchem™ commercial PC program was used for those calculations. A 71-parameter modified valence force field (vibrational potential energy function) was used for the vibrational calculations for the molecule in the configuration that has the two nitro groups on opposite sides of the cyclic plane. The NO_2 torsions and the ring puckering mode were neglected. Thirty-three computer runs were made, with different force constants being manually adjusted to improve the fit between observed and calculated vibrational wavenumbers. In the final run, the least-squares part of the computer program was allowed to adjust twenty force constants to fit twenty-eight observed wavenumbers. The average difference between observed and calculated values was 2.6 cm^{-1}. The observed and calculated wavenumbers and approximate potential energy distributions in terms of the normal modes are given in Table 1. An X-ray structure was not available for this compound, so it is not known if the two nitro groups actually are on opposite sides of the cyclic plane (hereafter called isomer 1 - see Fig. 3) or on the same side of that plane [the N-nitro group is on the side opposite the bromine atom] (called isomer 2 - see Fig. 3). Therefore, the structure of isomer 2 was calculated (AM1), and the force constants that were obtained for isomer 1 were used to calculate the vibrational frequencies of isomer 2. Comparison of the calculated frequencies of these two isomers shows that the infrared spectra of the two would be indistinguishable. This is apparently because the nitro group bonded to the ring nitrogen is too far from the bromine and nitro group bonded to the carbon for them to interact with each other. The heat of formation of isomer 1 was calculated to be 64 Kcal/mole with AM1 but only 22 Kcal/mole with PM3. The heat of formation of isomer 2 was calculated to be 64 (AM1) or 23 (PM3) Kcal/mole.

It was decided that normal coordinate calculations should be made next for 1-acetyl-3,3-dinitroazetidine (ADNAZ) in an attempt to interpret the infrared vibrational spectrum of this compound, because the spectrum should be easier to interpret than that of TNAZ. The reason is that there will be more overlapping of NO_2 bands for TNAZ since there are three nitro groups. The spectrum was obtained at 4 cm^{-1} resolution in this laboratory with a Mattson Cygnus 25 FTIR spectrometer. Appropriate initial force constant values were transferred from 1,3-dinitro-3-bromoazetidine, for which normal coordinate calculations had just been completed. The remaining force constants for ADNAZ were estimated or were transferred from an acetone force

TABLE 1. OBSERVED AND CALCULATED WAVENUMBERS AND POTENTIAL ENERGY DISTRIBUTIONS FOR 1,3-DINITRO-3-BROMOAZETIDINE

OBS. cm^{-1}	CALC. cm^{-1}	Main % contributions to the P.E.D. in sym. Coordinates (contributions less than 10% are omitted)
---	135.	CNBr twist(23), CNN bend(42), CNBr rock(19), NO2 rock(16)
---	170.	NCBr bend(47), NO2 out-of-plane bend (34)
---	218.	CNBr rock(50), CNBr twist(38)
---	252.	CNN bend(28), CN6 stretch(19), NO2 out-of-plane bend (9)
---	279.	NO2 rock(33), CNN bend(30), CNBr twist(26)
---	281.	CBr stretch(64)
---	364.	NO2 out-of-plane bend(35), CNN bend(24), CN stretch(15)
---	472.	NO2 rock(63), CC2 asym. stretch(11)
511	501.	NO2 out-of-plane bend(28), NO2 bend(20), CBr stretch(16), NN stretch (15)
536	538.	NO2 rock(43), CNBr rock(14), CNN bend(24)
611	613.	NO2 out-of-plane bend(41), NO2 bend(31)
642	644.	NO2 out-of-plane bend(52), CNN bend(28)
---	681.	NO2 bend(56)
761	764.	NO2 bend(25), NO2 out-of-plane bend(13)
809	814.	NN stretch (18), CC2 sym. stretch(15), NC2 sym. stretch(10)
847	846.	CC2 asym. stretch(44), CH2 wag(44)
904	903.	CN stretch(18), NO2 sym. stretch(11), CH2 rock(20), CH2 twist(16)
935	936.	CH2 wag(47), NN stretch (13), CC2 sym. stretch(13)
1025	1022.	CH2 twist(86), CH2 rock(12)
1055	1056.	CH2 rock(83), CH2 twist(12)
1098	1101.	CH2 twist(48), CH2 rock(18), NC2 sym. stretch(13)
---	1138.	NC2 asym. stretch(73)
1160	1158.	CH2 wag(22), NC2 sym. stretch(21), CC2 sym. stretch(17), CH2 twist(12)
1180	1181.	CH2 wag(43), CC2 asym. stretch(31)
1260	1258.	NO2 sym. stretch(50), CH2 rock(24), NO2 bend(11)
1276	1275.	CC2 sym. stretch(28), NC2 sym. stretch(21), NO2 sym. stretch(13)
1330	1331.	NO2 sym. stretch(48), NN stretch (26), NO2 bend(12)
1347	1347.	CN stretch(35), NO2 sym. stretch(25), CH2 rock(16), NO2 bend(11)
1443	1435.	CH2 bend(95)
1443	1452.	CH2 bend(83)
1529	1530.	NO2 asym. stretch(84), NO2 rock(12)
1561	1560.	NO2 asym. stretch(84), NO2 rock(10)
2967	2965.	CH2 sym. stretch(99)
2967	2968.	CH2 sym. stretch(99)
3026	3024.	CH2 asym. stretch(99)
3026	3026.	CH2 asym. stretch(99)

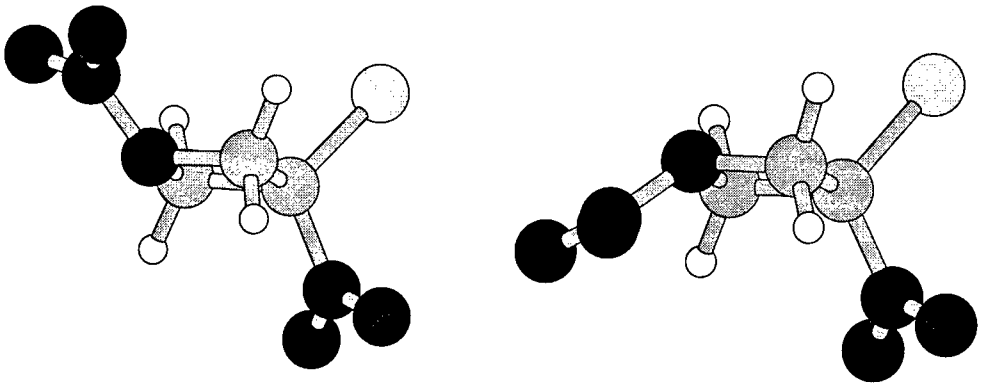


Fig. 3. Structures of 1,3-dinitro-3-bromoazetidine for which normal coordinate calculations were made. Left, isomer 1; right, isomer 2

field (for the CH_3-C=O group). An eighty-seven parameter modified valence force field was used, including thirty-three diagonal and fifty-four interaction force constants. Structural parameters for ADNAZ were obtained with the HyperChem™ MNDO-AM1 semi-empirical molecular orbital program.

The initial C-H stretching and bending frequencies that were calculated for ADNAZ were quite good, and the C=O and two NO_2 antisymmetric stretches were initially calculated to be 1640, 1584, and 1604 cm^{-1} (observed: 1669, 1580, 1580 cm^{-1}). The methyl bends and NO_2 symmetric stretches were calculated to be 1430, 1427, 1349, 1370 and 1334 cm^{-1} (obs., 1433, 1433, 1375, 1375, and 1341 cm^{-1}). Several force constants (C-H stretches, N-O stretches, H-C-H and C-C-H bends, and appropriate interaction constants) were refined to least-squares fit the calculated frequencies to the observed values just given. A total of eighteen computer runs were made for ADNAZ, with one, two, or three force constants being adjusted manually each time to fit one or more frequencies that had not been previously fit satisfactorily. All the force constants that had been adjusted manually a few at a time were then adjusted by the least-squares program in the final run to provide the best agreement between observed and calculated

frequencies. A total of thirteen force constants were adjusted to fit twenty-eight assigned frequencies, with the average difference between observed and calculated values being 4.6 cm^{-1}.

After completion of the normal coordinate calculations for ADNAZ, a structure that had been determined by X-ray diffraction became available [3]. The bond lengths were considerably different from the AM1 values that were used in the previous calculations. For example, the C-$N(O_2)$ bond lengths are 1.499 (X-ray) and 1.537 (AM1), and the ring C-C bond lengths are 1.539 (X-ray) and 1.583 (AM1). The N-C (ring nitrogen to acetyl group carbon) bond lengths are 1.351 (X-ray) and 1.406 (AM1). The different bond lengths and angles would significantly affect the calculated vibrational frequencies, so it was decided to repeat the normal coordinate calculations for ADNAZ with the experimentally-determined structure.

The methyl hydrogens were omitted in the new calculations in order to neglect any interaction of C-C-H bending with modes of the remainder of the molecule. The force constants obtained in this way should then be more transferable to TNAZ than those that included methyl interactions. The final force constant values that had been obtained for the AM1 structure were used for the zero-order calculation. The average difference between calculated and observed wavenumbers in the zero-order run was 11.2 cm^{-1}, as compared with a final average error of 4.6 cm^{-1} with the AM1 structure. However, seven wavenumbers in the zero-order run had an average error greater than 15 cm^{-1} each, and the maximum error was 36 cm^{-1}. Different sets of force constants were adjusted manually during several computer runs. Eleven runs were made in this way, and sixteen force constants were least-squares adjusted in the final run to fit twenty-seven assigned frequencies. The final average difference between observed and calculated wavenumbers was a very good 3.4 cm^{-1}, and the maximum error was an acceptable 11 cm^{-1}. The observed and calculated wavenumbers and approximate potential energy distributions are given in Table 2 for ADNAZ.

Table 3 lists the observed infrared bands for TNAZ, ADNAZ, and a 50:50 (mol percent) mixture of TNAZ and ADNAZ before and after melting and recrystallization. It can be seen that the carbonyl stretch band (1669 cm^{-1}) for pure ADNAZ shifts downward to 1657 cm^{-1} in the melt. It was thought that an explanation for this shift, and perhaps even for the unusual phase diagram of this mixture, might be a change in structure from the X-ray structure (Fig. 4, - conformation A) to some other structure, such as conformation B in Fig. 4. This difference in

TABLE 2. OBSERVED AND CALCULATED WAVENUMBERS AND POTENTIAL ENERGY DISTRIBUTIONS FOR ADNAZ (X-RAY STRUCTURE)

obs. cm^{-1}	calc. cm^{-1}	Main % contributions to the P.E.D. in terms of symmetry coordinates (contributions less than 10% are omitted)
---	148.	CNC out-of-plane bend(34), CN2 twist(19), CN2 wag(16), NO2 out-of-plane bend(13), CCN bend(12)
---	156.	NCN bend(60), out-of-plane bend (20)
---	231.	CNC in-plane bend(39), CO out-of-plane bend(19), CN2 rock(11)
---	254.	CN2 wag(43), NO2 rock(25), CN2 twist(23)
---	301.	CN2 twist(33), CCN bend(20), NO2 out-of-plane bend(19), CNC out-of-plane bend(18)
---	318.	CN2 rock(38), NO2 bend(22)
---	342.	CO out-of-plane bend(23), CN2 sym. stretch(14), CNC in-plane bend(10)
---	428.	NO2 rock(35), NO2 bend(15), CN2 sym. stretch(15), CN2 asym. stretch(10)
465	466.	NO2 bend(23), out-of-plane bend(19), CO rock (17), CN2 asym. stretch(10)
---	510.	CCN bend(32), CNC out-of-plane bend(22)
515	522.	NO2 rock(45), NO2 out-of-plane bend(23), CN2 twist(15)
---	552.	NO2 out-of-plane bend(32), CO rock(16), NCN bend(11)
579	579.	CO out-of-plane bend(50), CNC in-plane bend(36)
627	622.	ring deformation(41), NO2 out-of-plane bend(31), CN2 wag(17)
---	653.	NO2 bend(35), out-of-plane bend(12), CN2 rock(10)
689	685.	NO2 bend(17), CN2 rock (9), CC stretch(8), CN stretch(7)
---	752.	NO2 bend(30), NO2 sym. stretch(10)
855	853.	CC stretch(22), CH2 wag(27), CN stretch(14)
899	893.	CN2 asym. stretch(31), CH2 twist(16)
939	928.	ring deformation(30), NO2 sym. stretch(18)
972	975.	CH2 wag(39), ring deformation(31), CH2 wag(15)
1034	1034.	CH2 twist(99)
1093	1097.	CH2 rock(71)
1117	1111.	CH2 rock(23), CH2 twist(23), CC stretch(14)
1150	1143.	CH2 twist(34), ring deformation(27),
1194	1198.	ring deformation(61), CH2 wag(10)
1215	1221.	CN stretch(22), ring deformation(20), CC stretch(13)
---	1246.	CH2 rock(54), CN2 rock(12)
1304	1300.	ring deformation(41), CH2 wag(15)
1341	1335.	NO2 sym. stretch(51), CN2 asym. stretch(22), NO2 bend(14)
1375	1373.	NO2 sym. stretch(59), CN2 sym. stretch(21)
1375	1378.	ring deformation(38), CH2 wag (25)
1433	1430.	CH2 bend(95)
1433	1436.	CH2 bend(77)
1580	1583.	NO2 asym. stretch(62), NO2 rock(19)
1580	1583.	NO2 asym. stretch(78), NO2 rock(15)
1669	1669.	CO stretch(74)
2963	2962.	CH2 sym. stretch(99)
2963	2964.	CH2 sym. stretch(99)
3013	3013.	CH2 asym. stretch(99)
3013	3013.	CH2 asym. stretch(99)

TABLE 3. OBSERVED VIBRATIONAL WAVENUMBERS (CM^{-1}) FOR TNAZ (NEAT), ADNAZ (NEAT), AND A 1:1 MIXTURE OF TNAZ AND ADNAZ (INITIAL MIXTURE AND MELT)

TNAZ	ADNAZ	1:1 GROUND	1:1 MELT
n.b.	1669	1669	1657
1589	1580	1582	1586
1537	n.b.	1541	1553
1427	1433	1433	1439
1368	1375	1375	1373
1339	1341	1341	1343
n.b.	1304	1304	1304
1279	n.b.	1281	1281
1219	1215	1215	1215
1182	1194	1194	1190
1086	1094	1094	1090
n.b.	1034	1034	1036
n.b.	972	972	976
868	868	868	870
843	843	843	845
762	n.b.	762	762
n.b.	689	689	685
665	n.b.	666	650
n.b.	579	579	581
517	515	515	515
n.b.	465	465	465

n.b. = no band observed

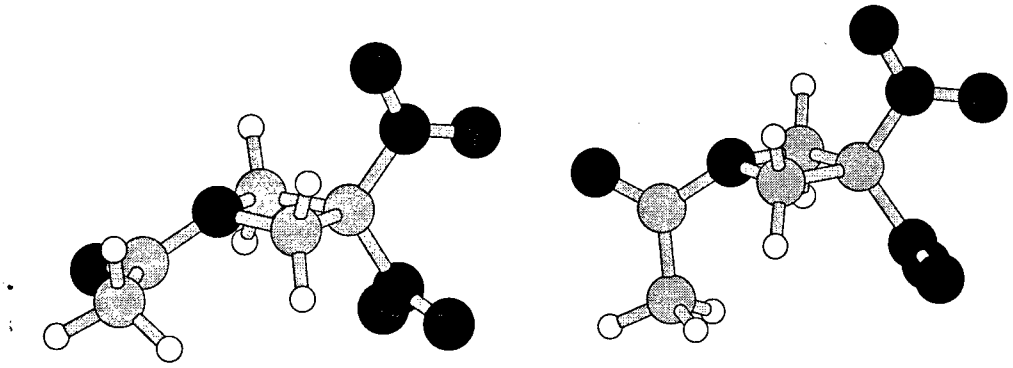


Fig. 4. Left, structure of ADNAZ as determined by X-ray diffraction (conformation A); Right, ADNAZ with the O=C-CH_3 group rotated 90° about the N-C bond (conformation B)

structure results from internal rotation about the N-C bond between the ring nitrogen and the acetyl carbon. In order to determine the dependence of C=O stretching frequency on structure, the vibrational frequencies of structure B were calculated with the force constants that were determined for structure A. The C=O stretching frequency for B was calculated to be 1646 cm^{-1}, which is a downward shift from the value of 1669 cm^{-1} in A, in agreement with the observed trend. None of the other bands for ADNAZ are expected to be more than 2 or 3 cm^{-1} different in structures A and B, and this is what is observed. The HyperChem™ program was used to do a conformational search on the ADNAZ molecule. It was found that both conformations shown in Fig. 4 are stable and therefore should exist in appreciable amounts.

All the force constants that had been determined for ADNAZ, except for those of the O=C-CH_3 group, were used in the TNAZ vibrational potential energy function to calculate the vibrational frequencies of that compound with the molecule in the configuration that had been determined by X-ray diffraction [4]. Initial force constant values for the N-NO_2 group were taken from the C NO_2 group. The X-ray structure is shown in Fig. 5 as conformation A. The bond lengths, bond angles, and torsional angles that were determined by X-ray diffraction were

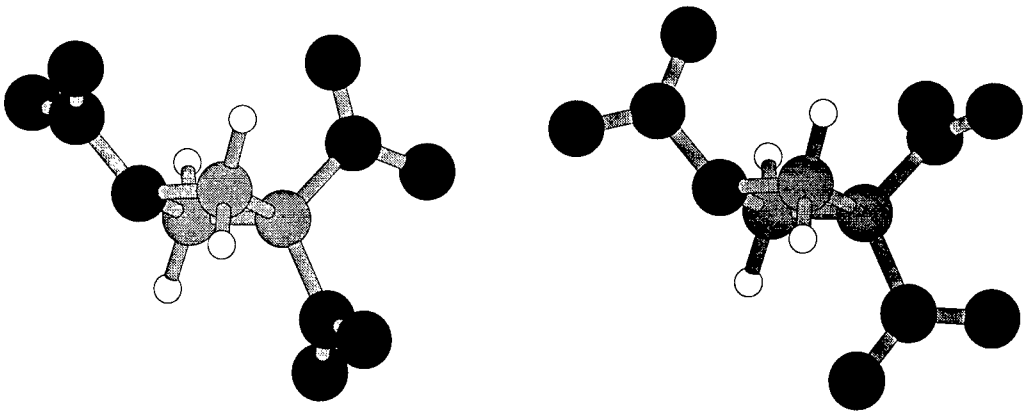


Fig. 5. Conformations of TNAZ for which calculations were done. Left, conformation as determined by X-ray diffraction (conformation A); Right, conformation with the three NO_2 groups rotated about the C-N or N-N bond by approximately 90° (conformation B).

used in the calculations. Several of the original force constants were eliminated because they had a negligible effect on the calculated frequencies, and the final potential energy function that was used consisted of seventy-three force constants, including twenty-seven diagonal and forty-six interaction constants. The average difference between calculated and observed differences in the zero-order run was 15 cm^{-1}, so changes had to be made in quite a few of the force constant values. In the second run, the C-H stretch, N-O stretch, H-C-H bend, C-H,C-H interaction, C-N,C-N interaction, and C-N,N-O interaction constants were adjusted to fit several observed frequencies. The Jacobian matrix elements were used in all force constant adjustments to determine which force constants to adjust. Over the next several runs, different force constants or sets of force constants were adjusted manually to better fit the frequencies above 900 cm^{-1}. Fourteen force constants were then least-squares adjusted to better fit those frequencies above 900 cm^{-1}, with the average error being 4.8 cm^{-1}. Several runs were then made, with different force constants being adjusted manually, in an effort to fit the frequencies below 900 cm^{-1}. In the final run, twenty force constants were least-squares adjusted to fit twenty-four assigned

frequencies. The average difference between calculated and observed values was 1.3 cm^{-1}. The observed and calculated wavenumbers and the band assignments in terms of approximate potential energy distributions are given in Table 4. There were no frequency data obtained for the region below 450 cm^{-1}.

Table 3 also shows that the 1537-cm^{-1} TNAZ band (NO_2 antisymmetric stretch of the N-NO_2 group) shifts upward to 1553 cm^{-1} in the 1:1 melt. This is the opposite direction of the shift of the C=O stretch band in ADNAZ. There is also a downward shift of the band observed at 665 cm^{-1} in neat TNAZ to 650 cm^{-1} in the melt. This band should be due to the NO_2 out-of-plane bend of the N-NO_2 group. The other bands listed in Table 3 that are due only to TNAZ (1279 and 762 cm^{-1}) do not shift in the melt. In an effort to explain the frequency shift of the two bands just mentioned, additional normal coordinate calculations were made for TNAZ, assuming that the molecule exists in the conformation shown on the right side of Fig. 5. A conformational search with the HyperChem™ program shows that both of these conformations, which are interconvertible, are stable.

The force constants that provided the fit shown in Table 4 for conformation A were used to calculate the vibrational frequencies of conformation B. The C=O stretch and out-of-plane bending wavenumbers were calculated to be 1525 and 694 cm^{-1}, respectively, both of which are shifted from the conformation A values in the direction opposite to that which is observed. Therefore, it was assumed that conformation B is the one that gave rise to the bands listed for TNAZ in Table 3, and the force constants were adjusted to fit the calculated wavenumbers of conformation B to the observed values. This means that the 1537 and 665 cm^{-1} bands were assigned to B rather than to A. The force constants that were determined in this way for conformation B were then used to calculate the frequencies of conformation A. The C=O stretch and C=O out-of-plane bend wavenumbers were calculated to be 1550 and 647 cm^{-1} for conformation A, which are in good agreement with the observed values.

1-nitroso-3,3-dinitroazetidine

The infrared spectrum was obtained for neat NO-DNAZ, but it will not be shown here in order to conserve space. The two NO_2 antisymmetric stretching frequencies obviously overlap, giving rise to the most intense band in the spectrum at 1573 cm^{-1}. The next most intense band, at 1333 cm^{-1}, must be due to overlapping NO_2 symmetric stretches. The normal region for the N=O

TABLE 4. OBSERVED AND CALCULATED WAVENUMBERS AND POTENTIAL ENERGY DISTRIBUTIONS FOR TNAZ (X-RAY STRUCTURE)

obs. cm^{-1}	calc. cm^{-1}	Main % contributions to the P.E.D. in terms of symmetry coordinates (contributions less than 10% are omitted)
---	148.	CNN bend (44), CN2 wag (19), CN2 twist (19), NO2 rock(16)
---	159.	CN2 bend (44), NO2 out-of-plane bend (27), NO2 rock(11)
---	240.	CNN bend(28), CN2 rock (20)
---	242.	CN2 wag (36), NO2 out-of-plane bend (27), CN2 twist (18)
---	303.	CN2 twist (28), NO2 rock (29), CNN bend(24)
---	339.	CN2 rock (42)
---	362.	CNN bend(22) NO2 out-of-plane bend (18), NCN sym. stretch (17),
---	416.	NO2 rock(41), ONO bend (13), NCN stretch (23)
---	507.	NO2 out-of-plane bend(23),NN stretch(21),NCN stretch(11),ONO bend(22)
517	517.	NO2 rock(64), NO2 out-of-plane bend (12)
544	544.	NO2 out-of-plane bend (25), NO2 rock(21), CNN bend(14) CN2 twist (13)
604	604.	NO2 out-of-plane bend (37), ONO bend(14)
---	654.	NO2 out-of-plane bend (31), CN2 wag (29), NO2 rock(14)
665	661.	NO2 out-of-plane bend (67), CNN bend(11), CNN bend(11)
714	713.	ONO bend (33), CN2 rock (20), NO2 sym. stretch(12)
762	762.	ONO bend (42), NO2 sym. stretch(20)
843	836.	ONO bend (27), NO2 sym. stretch(21), CH2 wag(12), ring deformation(16)
---	890.	ring deformation(29), NO2 sym. stretch(11)
908	911.	NCN stretch (25), NO2 sym. stretch(20), CH2 twist(15)
---	968.	CH2 wag(32), NN stretch (17), NO2 sym. stretch(14), ring deformation(23)
---	1004.	ring deformation(82)
1061	1061.	CH2 twist(88)
1086	1085.	CH2 wag(36), ring deformation(21), CH2 rock(10)
1113	1113.	CH2 rock(75)
---	1164.	CH2 twist(56), CH2 rock(12), ring deformation(10)
1182	1178.	ring deformation(32), CH2 twist(15), NO2 sym. stretch(12)
1219	1222.	CH2 wag(61), CN2 wag (11)
---	1242.	CH2 rock(29), ring deformation(14), CH2 wag(13)
---	1264.	NO2 sym. stretch(19), CH2 wag(18), ring deformation(16), CH2 rock(13)
1279	1281.	NO2 sym. stretch(33), NN stretch (21), ONO bend (21)
1339	1339.	NO2 sym. stretch(35), NCN asym. stretch (27), ONO bend (17)
1367	1367.	NO2 sym. stretch(45), ONO bend (21), NCN sym. stretch (20)
---	1397.	CH2 bend(85)
1428	1428.	CH2 bend(59), CH2 wag(13), ring deformation(11)
1538	1537.	NO2 asym. stretch(76), NO2 rock(16)
1589	1589.	NO2 asym. stretch(86), NO2 rock(11)
1589	1589.	NO2 asym. stretch(72), NO2 rock(16)
2969	2969.	CH2 sym. stretch(100)
2976	2976.	CH2 sym. stretch(100)
3022	3022.	CH2 asym. stretch(99)
3037	3037.	CH2 asym. stretch(99)

stretch of a nitrosoamine (R_2N-N=O) is ca. 1450 cm^{-1} for the compound in solution or 1490 cm^{-1} for the vapor [5]. The phase difference can certainly affect the frequency, but the nearest band to these regions for NO-DNAZ (KBr Pellet) is a medium intensity band that was observed at 1421 cm^{-1}, so this band must be due to that mode. Since one of the two CH_2 bending bands should also be observed in this region, it is assumed that a CH_2 bend overlaps the nitroso stretch at 1421 cm^{-1}. Most of the other bands will involve considerable mixing of normal modes, so calculations must be done to describe the motions responsible for those bands.

The structure of crystalline NO-DNAZ has been determined by X-ray diffraction [6]. The nitroso nitrogen is bent out of the CNC plane by only 11.4°, whereas the N-N bond in TNAZ was bent out of the plane by 39.6° [4]. The structural parameters for NO-DNAZ that were used in the normal coordinate calculations were taken from the X-ray structure.

A sixty-one parameter modified valence force field (vibrational potential energy function) was used for NO-DNAZ, which included twenty-four diagonal and thirty-seven interaction force constants. The torsions and ring puckering modes were omitted because their frequencies are unknown. Initial values of all force constants were taken from the TNAZ force field that was determined in this work. The N-N and nitroso N=O stretching constants were assumed to be larger than for TNAZ, and were therefore set a little higher than the TNAZ values. The N-N bond length in NO-DNAZ is shorter than in TNAZ (1.292 vs 1.351 Å), which indicates more double-bond character to this bond in NO-DNAZ and therefore a larger force constant. In addition, the N-O bonds in a nitro group are equivalent, and the resonance structure must therefore be intermediate between a double and single bond. There will be less resonance between N-N and N=O in NO-DNAZ than between the two N-O bonds in a nitro group, so the nitroso N=O will have more double-bond character than for nitro, and the nitroso N=O force constant must be larger than that of the nitro N=O.

The transferred force constants resulted in calculated wavenumbers in the zero-order run that were quite good, with the average difference between calculated and observed values being 10.6 cm^{-1} for twenty-one assigned wavenumbers. Only one calculated value (895 cm^{-1}) differed more than a reasonable amount from the observed value (855 cm^{-1}). This may indicate that the analogous band for TNAZ had been assigned incorrectly. The Jacobian matrix and potential energy distributions were used as indicators of the force constants that needed to be adjusted to

obtain a better fit of calculated to observed wavenumbers. Several computer runs were made, with one or more force constants being adjusted manually each time, until a better fit was obtained. In the sixth run, sixteen force constants were refined by the least-squares part of MOLVIB to fit twenty-three wavenumbers. Several more runs were made with manual adjustments being made, and then in the final run, thirteen force constants were least-squares refined to fit twenty-three assigned wavenumbers. The average difference between calculated and observed values was a very low 1.9 cm^{-1}. The observed and calculated wavenumbers, along with a description of the vibrations in terms of symmetry coordinates, are listed in Table 5.

The TNAZ/ADNAZ work showed a frequency shift of the (N)-NO_2 antisymmetric stretch band from 1537 cm^{-1} in neat TNAZ to 1553 cm^{-1} in a 1:1 mixture after melting and recrystallization. In addition, the (N)-NO_2 out-of-plane wagging band shifted from 665 cm^{-1} to 650 cm^{-1}. The proposed explanation was that either (1) internal rotation of the NO_2 groups results in a conformational change, as had been observed in DNNC [7], or (2) component interaction produces minor changes in a few force constants. The IR spectra for a 35:65 (mol percent) TNAZ/NO-DNAZ mixture show the same behavior for TNAZ, with the shifts [from neat (and mixture prior to melting and recrystallization) to recrystallized mixture melt] of the two bands just mentioned being the same as in the TNAZ/ADNAZ mixtures. It therefore seems that the conformational behavior of TNAZ is the same in a mixture with NO-DNAZ as it is in a mixture with ADNAZ. However, none of the bands due solely to NO-DNAZ show such a shift, so either (1) this compound does not undergo a conformational change, or (2) no bands above 500 cm^{-1} are dependent on conformation.

Conclusions

The calculations discussed in this report provide one possible explanation of the behavior of the TNAZ/ADNAZ system. Each of the two neat compounds could be in one conformation, which changes to another conformation during recrystallization after being melted together. This could be partially due to low barriers of internal rotation of the NO_2 and O=C-CH_3 groups, which could make many conformations possible. It has been shown that nitromethane [8] and nitroethane [9] have extremely low barriers to internal rotation of the NO_2 group. Another possible explanation of the frequency shifts described in this report is a change in the force constants of the C=O bond of ADNAZ and the N=O bond and angles that comprise the out-of-plane wag of the N-NO_2 group of TNAZ. These minor changes in force constant values could be

TABLE 5. OBSERVED AND CALCULATED WAVENUMBERS AND POTENTIAL ENERGY DISTRIBUTIONS FOR 1-NITROSO-3,3-DINITROAZETIDINE

obs.[a] cm^{-1}	calc. cm^{-1}	Main % contributions to the P.E.D. in symmetry coordinates (contributions less than 10% are omitted)
---	159	CN2 bend(59), NO2 out-of-plane wag(16)
---	173	CNN bend(71)
---	183	CN2 wag(39), CNN bend(24)
---	257	CN2 twist(62), NO2 out-of-plane wag(26)
---	285	NNO bend(18), CNN bend(15), CN2 wag(14), NO2 rock(12)
---	313	CN2 rock(52), NO2 out-of-plane wag(10)
---	369	CN stretch(28), ONO bend(16), CNN bend(12)
---	419	NO2 rock(43), CN stretch(15)
513	517	NO2 rock(23), NO2 out-of-plane wag(19), NNO bend(14)
---	535	NO2 out-of-plane wag(40), NO2 rock(14), CN2 twist(13)
594	590	NO2 out-of-plane wag(29), NNO bend(18)
663	664	CN2 wag(29), NO2 out-of-plane wag(27), NO2 rock(14)
708	699	ONO bend(27), CN2 rock(14), NO2 sym. stretch(12)
726	729	ONO bend(35), NO2 sym. stretch(12), NO2 out-of-plane wag(11)
812	810	NN stretch(20), NO2 sym. stretch(18), Ring deformation(13), CH2 wag(10)
855	853	Ring deformation(56)
913	914	CN stretch(36), CH2 rock(12)
---	1018	CH2 wag(37), NN stretch(34)
---	1060	Ring deformation(70), CN2 wag(11)
---	1088	CH2 twist(87), CH2 rock(11)
1101	1102	Ring deformation(32), CH2 wag(23), CNN bend(15)
---	1133	CH2 rock(79)
---	1150	CH2 twist(55), CH2 rock(25)
---	1165	Ring deformation(29), CH2 twist(12), NO2 sym. stretch(10)
1203	1203	CH2 wag(53), Ring deformation(26)
1263	1263	CH2 rock(34), NO2 sym. stretch(30)
1288	1288	CH2 wag(31), Ring deformation(30), CH2 bend(13),
1333	1332	NO2 sym. stretch(28), CN stretch(28), ONO bend(14)
1333	1342	NO2 sym. stretch(35), CN stretch(26), ONO bend(17)
1383	1383	CH2 bend(78)
1421	1415	CH2 bend(42), CH2 wag(16), Ring deformation(15)
1421	1422	NO stretch(67), CNN bend(11)
1573	1574	NO2 asym. stretch(66), NO2 rock(17)
1574	1576	NO2 asym. stretch(86), NO2 rock(11)
2962	2962	CH2 sym. stretch(99)
2980	2980	CH2 sym. stretch(99)
3014	3014	CH2 asym. stretch(99)
3037	3037	CH2 asym. stretch(99)

[a] no data were obtained below 500 cm^{-1}

caused by intermolecular interactions between the ADNAZ and TNAZ molecules. Appropriate force constants from the modified valence force field that had been determined for TNAZ were transferred successfully to NO-DNAZ. Manual and least-squares adjustments of selected force constants resulted in an exceptionally low average error for the calculated wavenumbers that were assigned to observed values. The modified valence fields developed should be useful for normal coordinate calculations for other substituted dinitroazetidines.

References

1. R. L. McKenney, Jr., unpublished work.
2. Indiana University, Quantum Chemistry Program Exchange, Program No. QCPE103.
3. Prof. H. L. Ammon, University of Maryland, private communication.
4. T. G. Archibald, R. Gilardi, K. Baum, and C. George, J. Org. Chem. **1990**, *55*, 2920.
5. N.B. Colthup, S.E. Wiberley, and L.H. Daly, "Introduction to Infrared and Raman Spectroscopy," Academic Press, New York, 1964, p. 290.
6. T.B. Brill private communication
7. Oyumi, T.B. Brill, A.L. Rheingold, and T.M. Haller, J. Phys. Chem. **1985**, *89*, 4317.
8. E. Tannenbaum, R. J. Myers, and W. D. Gwinn, J. Chem. Phys. **1956**, *25*, 42.
9. J. Ekkers, A. Bauder, and Hs. H. Gunthard, Chem. Phys. Lett. **1973**, *22*, 249.

GEOMETRICALLY INVARIANT NON-LINEAR RECURSIVE FILTERS, WITH APPLICATION TO TARGET TRACKING

R. W. R. Darling
Professor
Department of Mathematics

University of South Florida
4202 East Fowler Avenue PHY 114
Tampa, FL 33620

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

And

Wright Laboratory

September 1997

Geometrically Invariant Nonlinear Recursive Filters, with Application to Target Tracking

R. W. R. Darling
Professor
Department of Mathematics
University of South Florida

ABSTRACT

The Geometrically Invariant Nonlinear Recursive Filter, or GI Filter, is a coordinate-independent geometric generalization of the Kalman Filter for a continuous-time nonlinear state process subject to discrete-time observations. It is optimal in the sense that, if a linear system were subjected to a nonlinear transformation f of the state-space and analyzed using the GI Filter, the resulting state estimates and conditional variances would be the push-forward under f of the Kalman Filter estimates for the untransformed system – a property which is not shared by any of the variants of the Extended Kalman Filter.

The state process, which can be any Markov diffusion process, induces (through its covariance structure) a Riemannian metric on state space, and the observation covariance induces a Riemannian metric on observation space. Using the associated Levi-Civita connections, and gamma-martingale theory developed in a separate article, we are able to construct intrinsic location parameters (ILP) for the state and the observation, and thereby propagate the system dynamics in a coordinate-free way. An intrinsic generalization of the Kalman filter update formula allows recursive updating of state estimates and covariances.

As an example, the GI Filter is applied to the problem of tracking and intercepting a target, using sensors based on a moving missile, and explicit formulas are given. A software implementation using MATLAB is under development.

AMS (1991) SUBJECT CLASSIFICATION

Primary: 93E11. Secondary: 60H30, 65U05

KEY WORDS

GI filter, nonlinear filter, Kalman filter, stochastic differential equation, forward-backwards SDE, geometrically invariant

1 Background on Nonlinear Filtering

1.1 Example: A Nonlinear Filtering Problem in Target Tracking

D'Souza, McClure, and Cloutier (1994, 1997) consider the following tactical air-to-air missile intercept problem. The state of the target is represented by a position, velocity, and acceleration in space, making nine dimensions in all; the authors also model three time constants as state variables. Data consists of a sequence of observations of: range, angle from vertical, azimuth, and range-rate, all measured from a missile with known position, velocity, and acceleration. The goal of filtering in this case is to provide a sequence of "good" estimates of the state of the target, based on all measurements so far, so as to defeat the target's possible evasive maneuvers and intercept it.

D'Souza et al (1994) point out that, although the state dynamics can be modeled linearly (see Section 6.1 below), the observations are a highly nonlinear function of the state (see Section 6.4). Alternatively, if a spherical coordinate frame, based on the missile, is used, then observations are linear, but the state dynamics are highly nonlinear. Moreover the Extended Kalman Filter, or any of its variants, will give a different set of answers in the Cartesian coordinate frame than in the spherical one, because it is "non-intrinsic", i.e. lacking in absolute geometric meaning.

1.2 Drawbacks of Current Approaches

1.2.a The Infinite-Dimensional Approach

The standard mathematical presentation of the nonlinear filtering problem, as can be seen for example in Lipster and Shiryaev (1977), and Pardoux (1991), involves a measure-valued SDE called the Zakai equation (or the Fujisaki-Kallianpur-Kunita formula). This is virtually never used in real-time filtering applications because it is impossible to store enough data to update an infinite-dimensional SDE (although see Lototsky, Mikulevicius, and Rozovskii (1997) for a computational method using a Wiener chaos expansion).

1.2.b Finite-Dimensional Filters

Under certain circumstances, the conditional law can be described using a finite set of parameters. Although this topic is outside the scope of this article, an account of recent progress using geometric methods can be found in Cohen de Lara (1997). Apart from the Kalman filter, these methods are not widely used in practice, since the parameters may be difficult to determine in theory, large in number, and difficult to update computationally.

1.2.c The Extended Kalman Filter and Other Approximations

Linearizing the state and observation about the most recent state estimate, and then applying the Kalman Filter, gives the Extended Kalman Filter. The goal here is no longer to describe the full conditional distribution of X_t given $\{Y_s, 0 \le s \le t\}$, but merely to approximate the conditional expectation and the conditional covariance. This sometimes gives good results, and sometimes does not, and in any case the output is coordinate-dependent. A careful asymptotic analysis of this and other approximation schemes has been given by Picard (1991) - see also references therein.

1.3 A New Paradigm for Nonlinear Filtering

1.3.a State Evolves Continuously, Observations are Discrete

The state dynamics (for example, the dynamics of an aircraft) should be modeled by a stochastic process $\{X_t, t \geq 0\}$ in **continuous time**, on a differentiable manifold N. However since digital implementation of a filtering algorithm is carried out using discrete-time observations, the filter should involve observations $Y_1, Y_2, \ldots$ collected at **discrete times** $t_1 < t_2 < \ldots$ on another manifold M.

1.3.b State Estimates Should Not Be Coordinate-Dependent

Let $\{x_t^{(1)}, t \geq 0\}$ and $\{x_t^{(2)}, t \geq 0\}$ be representations of $\{X_t, t \geq 0\}$ in two coordinate systems, where $x_t^{(2)} = \phi(x_t^{(1)})$. Likewise let $y_1^{(1)}, y_2^{(1)}, \ldots$ and $y_1^{(2)}, y_2^{(2)}, \ldots$ be representations of $Y_1, Y_2, \ldots$ in two coordinate systems. We require that our state estimate of $x_{t_n}^{(2)}$, given $\{y_1^{(2)}, \ldots, y_n^{(2)}\}$, be the image under ϕ of our state estimate of $x_{t_n}^{(1)}$, given $\{y_1^{(1)}, \ldots, y_n^{(1)}\}$. Notice carefully that this criterion **excludes** use the conditional expectation $E\left[x_{t_n}^{(1)} \middle| y_1^{(1)}, \ldots, y_n^{(1)}\right]$ as the state estimate, because it does **not** have this kind of invariance. The replacement of conditional expectation by an "intrinsic location parameter" is the main theoretical contribution of this work.

1.3.c Must Coincide with the Kalman Filter in the Linear Case

When $\{X_t, t \geq 0\}$ is a continuous-time Gaussian process, and Y_n is a linear function of X_{t_n} with additive Gaussian noise, our filtering algorithm must give the Kalman filter state estimates (which fully describe the conditional distribution of the state, given the observations, in this context.)

2 A Simple Presentation of the Kalman Filter

The following presentation of the Kalman Filter will serve as a model for the Geometrically Invariant Nonlinear Recursive Filter. Suppose that U and V are random vectors in R^k and R^q respectively, whose joint distribution is multivariate Normal with

$$\begin{bmatrix} V \\ U \end{bmatrix} \sim N_{q+k}\left(\begin{bmatrix} \mu_V \\ \mu_U \end{bmatrix}, \begin{bmatrix} S & A \\ A^T & Q \end{bmatrix} \right). \tag{1}$$

We allow Q, but not S, to be degenerate. The Kalman Filter is based on the following Lemma.

2.1 Lemma

The conditional distribution of U given V = v is:

$$U|v \sim N_k(\mu_U + A^T S^{-1}(v - \mu_V), Q - A^T S^{-1} A). \tag{2}$$

Proof: Divide the joint density of U and V by the marginal density of V, and use elementary facts about inverses and determinants found, for example, in Rao (1973), pages 32 and 33. ◊

2.2 Prototype: the Gaussian Linear Model.

Suppose the following random variables are independent: $X_0 \sim N_k(\mu_0, \Sigma_0)$, and for $n = 1, 2, \ldots$, $\xi_n \sim N_k(0, K_{n-1})$ and $\eta_n \sim N_q(0, H_{n-1})$. Consider a discrete-time state process $\{X_1, X_2, \ldots\}$, evolving according to linear dynamics

$$X_n = F_{n-1}X_{n-1} + \xi_n, \; n = 1, 2, \ldots, \tag{3}$$

subject to linear observations

$$Y_n = G_{n-1}X_n + \eta_n, \; n = 1, 2, \ldots. \tag{4}$$

Introduce the filtration $\{\mathfrak{I}_n^Y, n \geq 0\}$, where

$$\mathfrak{I}_0^Y = \{\varnothing, \Omega\}\; ; \; \mathfrak{I}_n^Y \equiv \sigma\{Y_1, \ldots, Y_n\} \text{ for } n \geq 1.$$

We are assuming here that $F_{n-1}, K_{n-1} \in L(R^k; R^k)$, $G_{n-1} \in L(R^k; R^q)$, and (non-degenerate) $H_{n-1} \in L(R^k; R^q)$ are $\mathfrak{I}_{n-1}^Y$-measurable random matrices; this measurability is natural because we typically use observations $\{Y_1, \ldots, Y_{n-1}\}$ to control step n of the trajectory, using F_{n-1}. Define

$$\hat{\mu}_n \equiv E[X_n | \mathfrak{I}_n^Y], \; \hat{\Sigma}_n \equiv \mathrm{Var}(X_n | \mathfrak{I}_n^Y). \tag{5}$$

2.3 Proposition (the Kalman Filter)

The conditional distribution of X_n given $Y_1, \ldots, Y_n$ is $N_k(\hat{\mu}_n, \hat{\Sigma}_n)$, where $\hat{\mu}_n$ and $\hat{\Sigma}_n$ are computed recursively according to the formulas:

$$\hat{\mu}_0 = \mu_0; \; \hat{\mu}_n = x^\circ + \Phi_{n-1}[Y_n - y^\circ] \text{ for } n \geq 1; \tag{6}$$

$$\hat{\Sigma}_0 = \Sigma_0; \; \hat{\Sigma}_n = (I - \Phi_{n-1}G_{n-1})Q_{n-1} \text{ for } n \geq 1; \tag{7}$$

where $x^\circ \equiv F_{n-1}\hat{\mu}_{n-1}$, $y^\circ \equiv G_{n-1}F_{n-1}\hat{\mu}_{n-1}$ and $Q_{n-1}, \Phi_{n-1}, S_{n-1}$ are the $\mathfrak{I}_{n-1}^Y$-measurable random matrices given by:

$$Q_{n-1} \equiv K_{n-1} + F_{n-1}\hat{\Sigma}_{n-1}F_{n-1}^T, \tag{8}$$

$$S_{n-1} \equiv H_{n-1} + G_{n-1}Q_{n-1}G_{n-1}^T, \tag{9}$$

$$\Phi_{n-1} \equiv Q_{n-1}G_{n-1}^T S_{n-1}^{-1}. \tag{10}$$

Proof: A straightforward induction, using Lemma 2.1 to compute the posterior distribution of $U = X_n$ given $V = Y_n$, conditioned upon $\mathfrak{I}_{n-1}^Y$. ◊

2.4 Structure of the Recursive Estimation Procedure

We can divide the calculations (6) - (10) into five stages, which we will imitate in the differential geometric context. Here a "prior" distribution means one that is conditioned on $\mathcal{Y}_{n-1}^{Y}$, and a "posterior" means one that is conditioned on $\mathcal{Y}_{n}^{Y}$.

(a) Given $\hat{\mu}_{n-1}$, compute the prior mean x° of X_n; this is called "propagation".

(b) Compute a prior covariance Q_{n-1} of X_n.

(c) Compute a prior mean y° for Y_n.

(d) Compute a prior covariance S_{n-1} for Y_n.

(e) Combine these ingredients using Lemma 2.1 to find the posterior mean and covariance of X_n.

In the nonlinear context, steps (a) and (c) are the difficult ones, since they necessitate our theory of intrinsic location parameters for diffusion processes.

3 The Nonlinear Model and its Induced Geometry

3.1 Notational Convention

From here on, the state process $\{X_t, t \geq 0\}$ (upper case) will take values in a connected manifold N of dimension k, called the "state space". When we choose to compute in a specific chart, with chart map φ, we shall use lower case letters $\{x_t, t \geq 0\}$, where $\varphi(X_t) = x_t \equiv (x_t^1, \ldots, x_t^k) \in R^k$. The observations $Y_1, Y_2, \ldots$ (upper case), which may well be angles, will be assumed to lie on a manifold M of dimension q, called the "observation space". Likewise $\Phi(Y_n) = y_n \equiv (y_n^1, \ldots, y_n^q)$ will denote observation Y_n studied within the domain of a specific chart map Φ.

3.2 Basic Nonlinear Model, Expressed in Coordinates

3.2.a State Process

The state process will be represented in coordinates by a non-degenerate[1] Markov diffusion process

$$dx_t^i = b^i(x_t)\,dt + \sum_{j=1}^{k} \sigma_j^i(x_t)\,dW_t^j, \tag{11}$$

where $\sum b^i \frac{\partial}{\partial x_i}$ is a vector field on R^k, W is a Wiener process in R^k, and $\sigma(x) \equiv (\sigma_j^i(x))$ is a $k \times k$ matrix. We assume the coefficients[2] b^i, σ_j^i are C^2 and bounded.

---

1. We can also handle the degenerate case; we prefer to give an example in Section 6.2, rather than explain the general theory.

2. As in Section 2.2, we may also wish for the coefficients to depend on the observations prior to time t, but we omit this for notational simplicity; it will not affect the validity of our subsequent calculations.

3.2.b Observations

The observations $Y_1, Y_2, \ldots$ at times $t_1 < t_2 < \ldots$ are supposed to be of the form "$\psi(X_{t_n})$ corrupted by noise", for $n = 1, 2, \ldots$, where $\psi: N \to M$ is some C^2 function. Observation noise is assumed to be independent of the process X. We will see below how to express the "noise covariance" using a section of $TM \otimes TM$ (TM denotes the tangent bundle of M).

3.3 Geometry Induced by the Model

3.3.a The Diffusion Covariance Metric on State Space

Given a stochastic differential equation of the form (11) in each chart, define a C^2 Riemannian metric $\langle .|. \rangle$ on the cotangent bundle of N, which we call the **diffusion covariance metric**, by the formula

$$\langle dx^i | dx^l \rangle_x \equiv (\sigma(x) \cdot \sigma(x))^{il} \equiv \sum_{j=1}^{k} \sigma_j^i(x)\, \sigma_j^l(x) , \tag{12}$$

which is non-degenerate by assumption. This metric is actually intrinsic: changing coordinates for the diffusion will give a different matrix (σ_j^i), but the same metric. A corresponding metric g on the tangent bundle can be constructed as follows: if $\{\theta^1, \ldots, \theta^k\}$ is an orthonormal frame field of 1-forms for $\langle .|. \rangle$, the corresponding dual frame $\{e_1, \ldots, e_k\}$ is an orthonormal frame field of vector fields for g. The metric tensor (g_{ij}) is the inverse of the metric tensor $((\sigma \cdot \sigma)^{ij})$ for $\langle .|. \rangle$. For background information on bundles and frame fields, see Darling (1994), especially Chapters 7 and 9.

The **intrinsic** way to rewrite (11) is to postulate that X is a diffusion process on the Riemannian manifold N with generator

$$L \equiv \xi + \frac{1}{2}\Delta \tag{13}$$

where Δ is the Laplace-Beltrami operator associated with $\langle .|. \rangle$, and ξ is a vector field whose expression in the local coordinate system $\{x^1, \ldots, x^k\}$ is as follows:

$$\Delta = \sum_{i,j} (\sigma \cdot \sigma)^{ij} \{D_{ij} - \sum_m \Gamma_{ij}^m D_m\} , \quad \xi = \sum_m \{b^m + \frac{1}{2}\sum_{i,j} (\sigma \cdot \sigma)^{ij} \Gamma_{ij}^m\} D_m , \tag{14}$$

where $\{\Gamma_{ij}^m\}$ are the Christoffel symbols of the Levi-Civita connection ∇ obtained from $\langle .|. \rangle$, namely

$$\Gamma_{ij}^m \equiv \frac{1}{2}\sum_l \{D_i g_{jl} + D_j g_{li} - D_l g_{ij}\} (\sigma \cdot \sigma)^{lm} . \tag{15}$$

In situations where we want to suppress the indices, we will refer instead to a "local connector" $\Gamma: R^k \to L(R^k \otimes R^k; R^k)$. Thus $\Gamma(x)(\sigma \cdot \sigma)$ has m-th coordinate

$$\Gamma^m(x)(\sigma \cdot \sigma) = \sum_{i,j} \Gamma_{ij}^m (\sigma \cdot \sigma)^{ij} . \tag{16}$$

For details on the coordinate-free construction of the diffusion X, given a Wiener process W on R^k, see Elworthy (1982) p. 252. The main point to grasp here is:

Axiom A: *The appropriate metric for the state space is the diffusion covariance metric, not the Euclidean metric.*

3.3.b Covariance Tensor of a Random Variable in a Riemannian Manifold

We now introduce a local covariance concept, which we use for describing the uncertainty in the state estimates. Suppose η is a random variable in **any** Riemannian manifold N, for which a point $\mu \in N$ is some kind of location parameter. Let $T_\mu N$ denote the tangent space to N at μ. We shall say that a symmetric element

$$\Sigma \in T_\mu N \otimes T_\mu N \tag{17}$$

is the **covariance tensor** of η at μ if, for any cotangent vectors θ and λ at μ,

$$E\left[(\theta \cdot \exp_\mu^{-1}\eta)\ (\lambda \cdot \exp_\mu^{-1}\eta) \right] = \theta \otimes \lambda (\Sigma)\ . \tag{18}$$

Here $\exp_\mu$ is the exponential map from $T_\mu N$ to N, and we assume here (and henceforward) that η takes values in the image of the set on which $\exp_\mu$ is injective. In more concrete terms, if $\{e_1, \ldots, e_k\}$ is some basis of $T_\mu N$, and $\Sigma = \sum_{i,j} \Sigma^{ij} e_i \otimes e_j$, then Σ is the covariance matrix of the random vector $(\eta^1, \ldots, \eta^k)^T$ defined by

$$\exp_\mu^{-1}\eta = \sum_i \eta^i e_i .$$

Note that the components of $(\eta^1, \ldots, \eta^k)^T$ are uncorrelated, with unit variance, in the special case where $\{e_1, \ldots, e_k\}$ is an orthonormal basis, and $\Sigma = \sum_i e_i \otimes e_i$, i.e. the Riemannian metric itself.

3.3.c The Covariance Tensor of the State Estimate

In the linear, Gaussian case, we modelled the uncertainty about the initial state X_0 by giving it a $N_k(\mu_0, \Sigma_0)$ distribution. In the case where X_0 must take values on N, μ_0 will be a point in N, and the covariance matrix will be replaced by the covariance tensor of X_0 at μ_0, namely

$$\Sigma_0 \in T_{\mu_0} N \otimes T_{\mu_0} N .$$

Likewise X_{t_n} will be estimated by an $\mathfrak{I}_n^Y$-measurable random variable $\hat{\mu}_n$ with values in N; the uncertainty about X_{t_n}, given $\{Y_1, \ldots, Y_n\}$, will be modelled by an $\mathfrak{I}_n^Y$-measurable covariance tensor

$$\hat{\Sigma}_n \in T_{\hat{\mu}_n} N \otimes T_{\hat{\mu}_n} N ,$$

with the notational convention that $\hat{\mu}_0 = \mu_0$ and $\hat{\Sigma}_0 = \Sigma_0$.

3.3.d The Observation Covariance Metric

On the observation manifold M, we do not have a Riemannian metric *a priori*, but given a coordinate system $\{y^1, \ldots, y^q\}$ on $U \subseteq M$, given by a chart map φ, we have for each $y \in \varphi(U)$ the notion of an observation noise covariance in this coordinate system, i.e. a non-degenerate symmetric tensor H_y in $T_y R^q \otimes T_y R^q$. Provided these covariances are compatible in different coordinate charts cover-

ing M, we can interpret them as metric tensors for a metric $\langle .|. \rangle_o$ on the cotangent bundle of M, called the **observation covariance metric,** namely

$$\langle dy^i | dy^j \rangle_o = H^{ij}.$$

In short:

Axiom B: *The appropriate metric for the observation space is the observation covariance metric, not the Euclidean metric.*

We denote by $(\bar{g}_{ij})$ the metric tensor on TM, inverse to (H^{ij}).

If η is a random variable with values in M, for which a point $\mu \in M$ is some kind of location parameter, then we can use the construction of Section 3.3.b to describe the covariance tensor of η at μ with respect to the metric $\langle .|. \rangle_o$.

3.4 Summary: the Model in Intrinsic Terms

Following the discussion given in this section, we can rephrase the nonlinear filtering problem of paragraph 3.2 in an abstract way. The model consists of Riemannian manifolds N and M, called the state space and the observation space, respectively, a vector field ξ on N, and a C^2 function $\psi: N \to M$. Data consists of a point $\mu_0 \in N$ and a tensor $\Sigma_0 \in T_{\mu_0}N \otimes T_{\mu_0}N$, a sequence of times $0 < t_1 < t_2 < \dots$, and for each $n \geq 1$ a noisy observation[1] Y_n of $\psi(X_{t_n})$ (in the sense of paragraph 3.3.d), where the state process X is a diffusion process on N with generator $L \equiv \xi + \frac{1}{2}\Delta$.

3.4.a Goal

The goal is to construct a sequence of state and covariance estimates $(\hat{\mu}_n, \hat{\Sigma}_n)$ for $n = 1, 2, \dots$, in the sense of paragraph 3.3.c, with the following two properties:

- In the linear case (3) - (4), these estimates coincide with the Kalman filter estimates (6) - (7).
- The construction of $(\hat{\mu}_n, \hat{\Sigma}_n)$ is intrinsic: i.e. unaffected by choice of coordinates.

4 Intrinsic Location Parameter for a Diffusion Process

The key technical tool for setting up the Geometrically Invariant Nonlinear Recursive Filter algorithm, or GI Filter for short, is the following construction, discussed at more length in Darling (1997).

The second fundamental form ∇du of a C^2 mapping $u: N \to M$ is described in Eells and Lemaire (1978), as follows: ∇du is a section of the vector bundle $\mathrm{Hom}(TN \otimes TN; u^{-1}TM)$ and may be expressed in local coordinates (i.e. as a map $u \equiv (u^1, \dots, u^q)^T$ from R^k to R^q) by:

---

[1] If one wishes to be more rigorous, one could, for example, interpret Y_n as the value at time 1 of a Brownian motion on M, independent of X, started at $\Psi(X_{t_n})$ at time zero. This approach will not be pursued here.

$$(\nabla du)_{ij}^{m} = \frac{\partial^2 u^m}{\partial x_i \partial x_j} - \sum_h \Gamma_{ij}^h \frac{\partial u^m}{\partial x_h} + \sum_{p,l} \bar{\Gamma}_{pl}^m \frac{\partial u^p}{\partial x_i} \frac{\partial u^l}{\partial x_j} . \tag{19}$$

Let ξ be the vector field appearing in (14), and consider a family of mappings $\{u^\varepsilon(t,.) : N \to M\}$, for $t \in [0,\delta]$, $\varepsilon \in [0,1]$, such that, for each ε, u^ε solves the system of quasilinear parabolic PDE (a "heat equation with drift" for harmonic mappings):

$$\frac{\partial (u^\varepsilon)^m}{\partial t} - \xi (u^\varepsilon)^m - \frac{\varepsilon^2}{2} \sum_{i,j} (\sigma \cdot \sigma)^{ij} (\nabla du^\varepsilon)_{ij}^m = 0 , \; m = 1, \ldots, p , \tag{20}$$

$$u(0,p) = \psi(p) , \; p \in N . \tag{21}$$

From standard PDE theory (actually from a form of the Inverse Function Theorem), we have:

4.1 Local Existence and Uniqueness Theorem

Assume that $\|d\psi\|^2$ and ξ are bounded. There exists $\delta_1 > 0$ such that, for any $0 < \delta \le \delta_1$, there is a unique C^2 mapping $(\varepsilon, t, x) \to u^\varepsilon(t,x)$ from $[0,1] \times [0,\delta] \times N$ to M, such that the family $\{u^\varepsilon, 0 \le \varepsilon \le 1\}$ satisfies (20) and (21).

4.2 Definition of the Intrinsic Location Parameter

For the diffusion process X, started at $p \in N$, with generator $\xi + \frac{1}{2}\Delta$, the intrinsic location parameter of $\psi(X_\delta)$ is defined to be $u^1(\delta, p)$, provided $0 < \delta \le \delta_1$.

4.2.a Remarks

- In stochastic analytic terms, $u^1(\delta, p)$ is precisely the initial value of an $\{\mathfrak{I}_t^W\}$-adapted H^2 $\bar{\Gamma}$-martingale on M, with terminal value $V_\delta = \psi(X_\delta)$; see Darling (1996, 1997). This relationship uses systems of forward-backward SDE, as discussed in Pardoux and Peng (1992).
- Note in particular that, for all t and x, as $\varepsilon \to 0$, $u^\varepsilon(t,x) \to u^0(t,x) \equiv \psi(\phi_t(x))$, where $\{\phi_t, t \ge 0\}$ is the flow of the vector field ξ. In topological terms, the family $\{\{u^\varepsilon(t,.), 0 \le t \le \delta\}, 0 \le \varepsilon \le 1\}$ is assumed to be homotopic to $\{\psi \bullet \phi_t, 0 \le t \le \delta\}$.

It is not realistic to compute $u^1(\delta, p_0)$ in a filtering algorithm. However fortunately there is an easily computable approximation, described in the following theorem. Using the chart maps φ and $\bar{\varphi}$ on N and M, respectively, $\{\phi_t, t \ge 0\}$ denotes the flow of the vector field ξ, with derivative flow given locally by

$$\tau_s^t \equiv D(\phi_t \bullet \phi_s^{-1})(x_s^0) = \exp\{\int_s^t D\xi(x_u^0)\, du\} \in L(R^p; R^p) . \tag{22}$$

Let $x_s^0 = \phi_s(x)$, for $0 \le s \le \delta$, and let $G \equiv D\psi(x_\delta^0)$.

4.3 Theorem

In local coordinates, the derivative of $u^\varepsilon(\delta, x)$ *with respect to* ε *in the tangent space at* $y = \psi(x_\delta^0)$ *when* $\varepsilon = 0$, *may be expressed as*

$$\frac{\partial}{\partial \varepsilon} u^\varepsilon(\delta, x)\Big|_{\varepsilon=0} \in T_y M \tag{23}$$

$$= \frac{1}{2}\{G\int_0^\delta \tau_s^\delta [D^2\xi(x_s^0)(\chi_s) - \Gamma(x_s^0)(\sigma\cdot\sigma(x_s^0))]\,ds + D^2\psi(x_\delta^0)(\chi_\delta) + \bar{\Gamma}(y)(G\chi_\delta G^T)\}, \tag{24}$$

where $\sigma\cdot\sigma$ *is as in (2), and*

$$\chi_t \equiv \int_0^t \tau_s^t(\sigma\cdot\sigma)(x_s^0)(\tau_s^t)^T ds \in R^p \otimes R^p. \tag{25}$$

5 GI Filter Algorithm

Suppose that $Y_1, \ldots, Y_{n-1}$ have been observed, from which we have calculated the state estimate $\hat{\mu}_{n-1}$ and its associated covariance tensor $\hat{\Sigma}_{n-1}$. We shall now describe the GI Filter algorithm for intrinsic state estimation, which recapitulates the five steps delineated in Section 2.4.

5.0.a Location Parameter for the State

We are going to linearize in N (or in local coordinates, in R^k) about the point

$$x^\circ \equiv \phi_\delta(\hat{\mu}_{n-1}), \tag{26}$$

where $\delta \equiv t_n - t_{n-1}$, using the flow $\{\phi_t, t \geq 0\}$ of the vector field ξ on R^k, as given in (14). Computing x° amounts to solving a first-order ordinary differential equation.

We cannot work with the true state process $\{X_t, t_{n-1} \leq t \leq t_n\}$ for filtering purposes, because $X_{t_{n-1}}$ is unknown. Instead we use a diffusion process $\{\hat{X}_t, t_{n-1} \leq t \leq t_n\}$, with the same generator (13), but started at $\hat{\mu}_{n-1}$. We take an approximation to the intrinsic location parameter of $\hat{X}_{t_n}$, namely

$$\mu_x \equiv \frac{\partial}{\partial \varepsilon} u^\varepsilon(\delta, x)\Big|_{\varepsilon=0} \in T_{x^\circ}N \equiv T_{x^\circ}R^k, \tag{27}$$

which is calculated from (24), taking ψ = identity and $\bar{\Gamma} = \Gamma$. This "propagation" step is more complicated than in the Kalman filter, where μ_x would be zero.

5.0.b Compute a Prior Covariance Q_{n-1} for the State

The uncertainty about X_{t_n} will be affected by the derivative of the flow $\{\phi_t, t \geq 0\}$ during the time interval $[0, \delta]$. The derivative of $\phi_\delta: N \to N$ may be extended to a "push-forward" map

$$(\phi_\delta)_*: T_{\hat{\mu}_{n-1}}N \otimes T_{\hat{\mu}_{n-1}}N \to T_{x^\circ}N \otimes T_{x^\circ}N, \tag{28}$$

given locally by $(\phi_\delta)_*(\zeta_1 \otimes \zeta_2) = D\phi_\delta(\zeta_1) \otimes D\phi_\delta(\zeta_2)$. Define a covariance tensor at x° by

$$Q_{x^\circ} \equiv (\phi_\delta)_* \hat{\Sigma}_{n-1} + \int_0^\delta (\phi_{\delta-s})_* \langle .|. \rangle_{\phi_s(x)} ds \in T_{x^\circ}N \otimes T_{x^\circ}N \cong L(T^*_{x^\circ}N; T_{x^\circ}N) . \tag{29}$$

The second term on the right side of (29) is χ_δ as defined in (25). Here $\langle .|. \rangle_{\phi_s(x)}$ is the diffusion covariance metric, evaluated at $\phi_s(x)$. Formula (29) is precisely analogous to (8), in that uncertainty about $\hat{\mu}_{n-1}$ is being convolved with uncertainty resulting from running a diffusion with generator (13) during a time interval of length δ. Computing the first term in (29) requires solving the linear differential equation for the derivative of the flow $\{\phi_t, t \geq 0\}$, in other words the $k \times k$ matrix valued ODE $\dot{F}_t = D\xi(\phi_t(\hat{\mu}_{n-1}))F_t$, $F_0 = I$; then in matrix terms, $(\phi_\delta)_* \hat{\Sigma}_{n-1}$ is $(D\phi_\delta)\hat{\Sigma}_{n-1}(D\phi_\delta)^T$.

5.0.c Location Parameter for the Observation

We are going to linearize in M (or, in local coordinates, in R^q) about the point

$$y^\circ \equiv \psi(x^\circ) . \tag{30}$$

An approximation to the $\mathfrak{I}^Y_{n-1}$-measurable location parameter for $Y_n \equiv \psi(X_{t_n})$ is given by formula (24), in other words

$$\mu_y \equiv \frac{\partial}{\partial \varepsilon} u^\varepsilon(\delta, x)\Big|_{\varepsilon=0} \in T_{y^\circ}R^q , \tag{31}$$

taking ψ as in Theorem 4.3.

5.0.d Prior Covariance Tensor for the Observation

Following (9), the covariance tensor of Y_n at y° is given by

$$S_{y^\circ} \equiv H_{y^\circ} + \psi_* Q_{x^\circ} \in T_{y^\circ}M \otimes T_{y^\circ}M \cong L(T^*_{y^\circ}M; T_{y^\circ}M) , \tag{32}$$

where H_{y° is the observation covariance metric evaluated at y°, and

$$\psi_* : T_{x^\circ}N \otimes T_{x^\circ}N \to T_{y^\circ}M \otimes T_{y^\circ}M$$

follows the notation of (28). In matrix terms, $\psi_* Q_{x^\circ}$ becomes $(D\psi)Q_{x^\circ}(D\psi)^T$.

5.0.e Posterior Distribution of the State

We are now back in the situation treated in the proof of the Kalman Filter, Section 2.3. The only subtle point is that, in place of the conditional covariance of Y_n and X_{t_n}, given $\mathfrak{I}^Y_{n-1}$, we now have

$$A_{n-1} \in T_{x^\circ}N \otimes T_{y^\circ}M \cong L(T^*_{x^\circ}N; T_{y^\circ}M) ,$$

with adjoint

$$A^*_{n-1} \in L(T^*_{y^\circ}M; T_{x^\circ}N) ,$$

which is defined by its action on a cotangent vector $\eta \in T^*_{y^\circ}M$, namely

$$A^*_{n-1}(\eta) \equiv Q_{x^\circ}(\psi^*\eta) , \tag{33}$$

where $\psi^*: T^*_{y^\circ}M \to T^*_{x^\circ}N$ is the pullback, i.e. $\psi^*\eta(\zeta) = \eta(T\psi(\zeta))$, for $\zeta \in T_{x^\circ}N$ (see Darling (1994)). In matrix terms, A^*_{n-1} becomes $Q_{x^\circ}(D\psi)^T$.

The following key formula uses both the exponential map $\exp_{x^\circ}: T_{x^\circ}N \to N$ determined by the Levi-Civita connection Γ on N, and $\exp_{y^\circ}: T_{y^\circ}M \to M$ determined by the Levi-Civita connection $\bar{\Gamma}$ on M. Arguments are assumed to be within the injectivity radii of these mappings.

The **GI Filter update formula**, generalizing (6), is:

$$\hat{\mu}_n = \exp_{x^\circ}(\mu_x + A^*_{n-1} S^{-1}_{y^\circ}(\exp^{-1}_{y^\circ}(Y_n) - \mu_y)), \tag{34}$$

where the subtraction occurs in $T_{y^\circ}M$, and the addition in $T_{x^\circ}N$. Computation of $\exp_{x^\circ}(.)$ and $\exp^{-1}_{y^\circ}(.)$ involves solving second-order bilinear ordinary differential equations (i.e. geodesic equations) for the geodesic flows on N and M respectively, which we describe briefly in Section 5.1. However there is also a "quick and dirty" version described in Section 5.2. Formula (34) makes sense because the domains and ranges of the maps involved are as shown in the following diagram:

$$T_{y^\circ}M \xrightarrow{S^{-1}_{y^\circ}} T^*_{y^\circ}M \xrightarrow{A^*_{n-1}} T_{x^\circ}N \xrightarrow{\exp_{x^\circ}} N$$

To update the covariance tensor, we first compute $Q_{x^\circ} - A^*_{n-1} S^{-1}_{y^\circ} A_{n-1} \in T_{x^\circ}N \otimes T_{x^\circ}N$ (compare with (7)), and then push the covariance tensor forward to $T_{\hat{\mu}_n}N \otimes T_{\hat{\mu}_n}N$ using the geodesic flow $\{\pi_t, 0 \le t \le 1\}$ described in Section 5.1, namely

$$\hat{\Sigma}_n = (\pi^H_1)_*(Q_{x^\circ} - A^*_{n-1} S^{-1}_{y^\circ} A_{n-1}) \in T_{\hat{\mu}_n}N \otimes T_{\hat{\mu}_n}N. \tag{35}$$

5.1 Geodesic Flow

The geodesic equation on N can be represented as a first order ODE on the tangent bundle TN; under the chart map $\varphi \times d\varphi$, a solution is given by the geodesic flow

$$\begin{bmatrix}\gamma(s)\\ \zeta(s)\end{bmatrix} = \pi_s\left(\begin{bmatrix}\gamma(0)\\ \zeta(0)\end{bmatrix}\right) = \begin{bmatrix}\pi^H_s((\gamma(0), \zeta(0))^T)\\ \pi^V_s((\gamma(0), \zeta(0))^T)\end{bmatrix}, \quad 0 \le s \le 1, \tag{36}$$

in $R^k \oplus R^k$, satisfying the system of ODE

$$\begin{bmatrix}\gamma'\\ \zeta'\end{bmatrix} = h(\gamma, \zeta) \equiv \begin{bmatrix}\zeta\\ -\bar{\Gamma}(\gamma)(\zeta \otimes \zeta)\end{bmatrix}. \tag{37}$$

(See Sakai (1996), p. 56.) For example, to compute $\gamma(1) = \exp_{x^\circ}(v)$, the initial conditions will be $\gamma(0) = x^\circ$, $\zeta(0) = v$. To compute $(\pi^H_1)_*: T_{\gamma(0)}N \to T_{\gamma(1)}N$, we must compute the derivative flow $\{H(s), 0 \le s \le 1\}$ in $L(R^k \oplus R^k; R^k \oplus R^k)$ satisfying

$$H' = Dh(\gamma, \zeta)H, \quad H(0) = I. \tag{38}$$

We partition $H \equiv \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}$ into $k \times k$ matrices; then $(\pi_1^H)_* = H_{11}(1)$.

5.2 Single-step Discretization of the GI Filter Update Formula

If we want to save computation time, we can forget about the geodesic flow. In local coordinates, the single-step discretization of formula (34) goes as follows:

$$\exp_{y^\circ}^{-1}(Y_n) \approx v_n \equiv y_n - y^\circ + \frac{1}{2}\bar{\Gamma}(y^\circ)\left((y_n - y^\circ) \otimes (y_n - y^\circ)\right) \tag{39}$$

$$u_n \equiv \mu_x + Q_{x^\circ} G_{x^\circ}^T S_{y^\circ}^{-1}(v_n - \mu_y) \tag{40}$$

$$\hat{\mu}_n \approx x^\circ + u_n - \frac{1}{2}\Gamma(x^\circ)(u_n \otimes u_n)\,. \tag{41}$$

Then take $(\pi_1^2)_*$ to be the identity in formula (35).

5.3 Optimality Property of the GI Filter

Suppose $\{X_t, t \geq 0\}$ is a continuous Gaussian process in R^k such that

$$X_{t_n} = F_{n-1} X_{t_{n-1}} + \xi_n,\ n = 1, 2, \ldots, \tag{42}$$

subject to observations

$$Y_n = G_{n-1} X_{t_n} + \eta_n,\ n = 1, 2, \ldots, \tag{43}$$

in R^q, where other terms have the meanings assigned in Section 2.2. More specifically, we suppose that $\{X_t, t \geq 0\}$ is an Ornstein-Uhlenbeck process satisfying the SDE

$$dX_t = A_{n-1} X_t dt + \sigma_{n-1} dW_t,\ t_{n-1} \leq t \leq t_n; \tag{44}$$

here A_{n-1} and σ_{n-1} are $\mathfrak{I}_{n-1}^Y$-measurable $k \times k$ matrices, and $F_{n-1} = \exp\{(t_n - t_{n-1})A_{n-1}\}$.

5.3.a Theorem

Suppose $\phi: R^k \to R^k$ and $\theta: R^q \to R^q$ are any pair of C^2 diffeomorphisms. When the GI Filter is applied to the process $\{\phi(X_t), t \geq 0\}$, with observations $\theta(Y_1), \theta(Y_2), \ldots$, the state estimator of $\phi(X_{t_n})$, given $\theta(Y_1), \ldots, \theta(Y_n)$, is precisely $\phi(\hat{\mu}_n)$, with conditional covariance $D\phi(\hat{\mu}_n)\hat{\Sigma}_n(D\phi(\hat{\mu}_n))^T$, where $(\hat{\mu}_n, \hat{\Sigma}_n)$ are the Kalman Filter estimates given by (5) - (7).

The theorem says, in effect, that when a nonlinear system is a transformed version of a linear system, then the GI Filter estimates are similarly transformed versions of the Kalman Filter estimates, as we desired in Section 1.3.c.

Proof: Since every step in the GI Filter is coordinate-independent, it suffices to prove the theorem when ϕ and θ are both the identity. When $\{X_t, t \geq 0\}$ satisfies (44), then (11) holds with $\sigma(x)$ not depending on x, and $b(x)$ is of the form Ax, where A stands for A_{n-1} when $t \in [t_{n-1}, t_n]$. The connector Γ is zero on $N \equiv R^k$, so $\xi = \sum b^i \frac{\partial}{\partial x_i}$, and $D^2\xi(x)$ is zero.

Likewise since the covariance tensor is constant on observation space, the connector $\bar{\Gamma}$ is zero, and since $\psi(x) = G_{n-1}x$ is linear, we have $D^2\psi = 0$. Considering (24), we see that $\mu_x = 0$ in (27) and $\mu_y = 0$ in (31). In the constant-metric case, $\exp_x v = x + v$. Hence (34) becomes

$$\hat{\mu}_n = x^\circ + A^*_{n-1} S^{-1}_{y^\circ} (Y_n - y^\circ)$$

which is the same as (6), and (35) coincides with (7). ◊

6 Application to Target Tracking from a Moving Missile

6.1 State Dynamics

The state x consists of the location, velocity, and acceleration of the target in three-dimensional space, namely $x^T \equiv (p^T, v^T, a^T) \in R^3 \times R^3 \times R^3$. We model the acceleration as an Ornstein-Uhlenbeck process, with the constraint that acceleration must be perpendicular to velocity. Thus within a Cartesian frame, the equations of motion take the nonlinear form:

$$\begin{bmatrix} dp \\ dv \\ da \end{bmatrix} = \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & I \\ 0 & -\rho(x) I & -\lambda P(v) \end{bmatrix} \begin{bmatrix} p \\ v \\ a \end{bmatrix} dt + \begin{bmatrix} 0 \\ 0 \\ \sqrt{\lambda c_1} P(v)\, dW(t) \end{bmatrix}, \tag{45}$$

where the square matrix consists of nine 3×3 matrices, λ is a positive time constant,

$$\rho(x) \equiv \|a\|^2 / \|v\|^2, \tag{46}$$

$$P(v) \equiv I - \frac{vv^T}{\|v\|^2} \in L(R^3; R^3), \tag{47}$$

and W is a three-dimensional Wiener process. Note that $P(v)$ is precisely the projection onto the orthogonal complement of v in R^3, and $\rho(x)$ has been chosen so that $d(v \cdot a) = 0$. D'Souza et al (1997) describe a procedure for estimating λ, but in our simulations we assign to it a predetermined value.

6.2 Geometry of the State Space

The diffusion variance metric (12) is degenerate here. In order to compute the Christoffel symbols, we use the following trick. For a small $\varepsilon > 0$, we replace the diffusion variance metric by the following:

$$\sigma \cdot \sigma \equiv \begin{bmatrix} \varepsilon I & 0 & 0 \\ 0 & \varepsilon I & 0 \\ 0 & 0 & \lambda c_1 P(v) + \varepsilon Q(v) \end{bmatrix}, \tag{48}$$

where $Q \equiv I - P$, noting that $P^2 = P$. To compute the inverse matrix g, note that the bottom right 3×3 block is $T(v) \equiv c^{-1}[I + ((c - \varepsilon)/\varepsilon) Q(v)]$, with matrix derivative

$$DT(v)w = \left(\frac{1}{\varepsilon} - \frac{1}{c}\right)\left\{\frac{vw^T + wv^T}{\|v\|^2} - \frac{2(v \cdot w)}{\|v\|^4}vv^T\right\}.$$

Using linear algebra and (15), we calculate the local connector $\Gamma: R^9 \to L(R^9 \otimes R^9; R^9)$ as a limit (which exists) as $\varepsilon \to 0$. It is clear from (45) that any tangent vectors ζ and η to the state space may be broken down into three 3-dimensional components

$$\zeta \equiv \begin{bmatrix} \zeta_p \\ \zeta_v \\ \zeta_a \end{bmatrix}, \eta \equiv \begin{bmatrix} \eta_p \\ \eta_v \\ \eta_a \end{bmatrix}, \tag{49}$$

where ζ_p, η_p are both in direction v, and ζ_a, η_a are both in direction a. Thus

$$\Gamma(x)(\zeta, \eta) = \frac{(\zeta_a \cdot v)(\eta_a \cdot v)}{\|v\|^4}\begin{bmatrix} 0 \\ v \\ 0 \end{bmatrix} - \frac{1}{2\|v\|^2}\left\{(\zeta_a \cdot v)\begin{bmatrix} 0 \\ \eta_a \\ 0 \end{bmatrix} + (\eta_a \cdot v)\begin{bmatrix} 0 \\ \zeta_a \\ 0 \end{bmatrix}\right\}. \tag{50}$$

6.3 Deterministic Flow

When we evaluate the vector field ξ according to (14), we find that the $\sum (\sigma \cdot \sigma)^{ij}\Gamma_{ij}^m$ term is zero, because by (48) (with $\varepsilon = 0$) and (50), it involves terms such as $v^T P(v) v = 0$. Hence by (45),

$$\xi(x) = \begin{bmatrix} \xi_p \\ \xi_v \\ \xi_a \end{bmatrix} = \begin{bmatrix} v \\ a \\ -\rho(x)v - \lambda P(v)a \end{bmatrix}, \tag{51}$$

$$D\xi(x)\left(\begin{bmatrix} \zeta_p \\ \zeta_v \\ \zeta_a \end{bmatrix}\right) = \begin{bmatrix} \zeta_v \\ \zeta_a \\ -D\rho(x)(\zeta)v - \rho(x)\zeta_v - \lambda DP(v)(\zeta_v)a - \lambda P(v)\zeta_a \end{bmatrix}, \tag{52}$$

$$D\rho(x)(\zeta) = \frac{2}{\|v\|^2}\{a \cdot \zeta_a - (v \cdot \zeta_v)\rho(x)\}, \; DP(v)(\zeta_v) = \frac{2(v \cdot \zeta_v)}{\|v\|^4}vv^T - \frac{v\zeta_v^T + \zeta_v v^T}{\|v\|^2}. \tag{53}$$

One may compute $D^2\xi(x)$ similarly.

6.4 Observation Function

The observables are respectively: range, angle from vertical, azimuth, and range-rate (all measured from a missile with known state (p_M, v_M, a_M)) and a fictitious measurement; the latter is a zero-mean Gaussian random variable representing a fictitious observation of the inner product of velocity and acceleration of the target, which according to our model should be zero. Take $\psi: R^3 \times R^3 \times R^3 \to R_+ \times S^2 \times R^2$ to be:

$$\psi(p, v, a) \equiv (\Phi(p - p_M), \|v - v_M\|, a \cdot v) \tag{54}$$

where $\Phi \equiv h^{-1}$, and h is the spherical coordinate transformation

$$h(r, \theta, \phi) = (r\sin\theta\cos\phi, r\sin\theta\sin\phi, r\cos\theta) . \tag{55}$$

For the sake of brevity, we omit here the calculations of the first and second derivatives of ψ.

6.5 Geometry of the Observation Manifold

The covariance matrix for the five observed quantities is taken to be of the form

$$H_{\psi(x)} = \text{diag}\left(r^2 s_0, \frac{s_1}{r^2} + s_2, \frac{s_3}{r^2} + s_4, r^2 s_5, \sigma_F^2\right), \tag{56}$$

where r denotes range, and the other quantities are constants. The inverse of H is the metric tensor $(\bar{g}_{ij})$. Since the off-diagonal entries of R are zero, the formula for the Christoffel symbols in our coordinate system on the observation manifold becomes:

$$\bar{\Gamma}_{ij}^m = \frac{1}{2}\{D_i\bar{g}_{jm} + D_j\bar{g}_{mi} - D_m\bar{g}_{ij}\} R^{mm}, \; i, j, m \in \{1, 2, 3, 4, 5\},$$

$$= \frac{1}{2}\left\{[\delta_{jm}\delta_{i1} + \delta_{j1}\delta_{mi}]\frac{\partial\bar{g}_{mm}}{\partial r} - \delta_{m1}\delta_{ij}\frac{\partial\bar{g}_{jj}}{\partial r}\right\} R^{mm},$$

using the fact that the only differentiation which gives a non-zero result is with respect to the first coordinate, r. Taking $m = 1$, we obtain:

$$\bar{\Gamma}_{ij}^1 = \frac{r^2 s_0}{2}\left\{2\delta_{j1}\delta_{i1}\frac{\partial}{\partial r}\left(\frac{1}{r^2 s_0}\right) - \delta_{ij}\frac{\partial\bar{g}_{jj}}{\partial r}\right\};$$

$$(\bar{\Gamma}_{ij}^1) = \text{diag}\left(\frac{-1}{r}, \frac{-s_0 s_1}{r(s_2 + s_1/r^2)^2}, \frac{-s_0 s_3}{r(s_4 + s_3/r^2)^2}, \frac{s_0}{s_5 r}, 0\right). \tag{57}$$

The only other non-zero Christoffel symbols are the following:

$$\bar{\Gamma}_{12}^2 = \bar{\Gamma}_{21}^2 = \frac{s_1}{r(s_2 r^2 + s_1)}; \; \bar{\Gamma}_{13}^3 = \bar{\Gamma}_{31}^3 = \frac{s_3}{r(s_4 r^2 + s_3)}; \; \bar{\Gamma}_{14}^4 = \bar{\Gamma}_{41}^4 = \frac{-1}{r}. \tag{58}$$

6.6 Ingredients of the Update Formula

We assume that step $n-1$ of the algorithm has been performed, yielding a state estimate $\hat{\mu}_{n-1}$ at time t_{n-1}, with conditional covariance tensor $\hat{\Sigma}_{n-1}$. Let $\delta \equiv t_n - t_{n-1}$. We discretize (45) to obtain a numerical solution of the ODE $dx_t^0/dt = \xi(x_t^0)$ with $x_0^0 = \hat{\mu}_{n-1}$, to obtain (as in (26))

$$x^\circ \equiv \begin{bmatrix} p_n \\ v_n \\ a_n \end{bmatrix} = x_\delta^0 .$$

Working similarly with the gradient flow, using the vector field (52),

$$dF_t/dt = D\xi(x_t^0)F_t,\ F_0 = I, \tag{59}$$

we compute the matrix F_δ for $(\phi_\delta)_*$ as given in (28), τ_s^t as in (18), and χ_δ given by (25).

6.6.a Intrinsic Location Parameter for the State

We must now evaluate (24) (with $\psi = 0$) to compute $\mu_x \equiv T_{x^\circ}N \cong T_{x^\circ}R^9$. The $\Gamma(x_s^0)\,(\sigma \cdot \sigma(x_s^0))$ term is zero, as noted above, and G is the identity. This gives the formula

$$\mu_x = \frac{1}{2}\{\int_0^\delta \tau_s^\delta D^2\xi(x_s^0)\,(\chi_s)\,ds + \Gamma(x^\circ)\,(\chi_\delta)\}\,. \tag{60}$$

6.6.b Prior Covariance Tensor for the State

We compute from (59):

$$Q_{x^\circ} = F_\delta \hat{\Sigma}_{n-1} F_\delta^{\,T} + \chi_\delta . \tag{61}$$

6.6.c Intrinsic Location Parameter for the Observation

Take $y^\circ \equiv \psi(x^\circ)$. We calculate:

$$G_{x^\circ} \equiv D\psi(x^\circ)\,, \tag{62}$$

Our expression for μ_y is given by (24), in which part of the integral term is zero, leaving:

$$\mu_y = \frac{1}{2}\{G_{x^\circ}\int_0^\delta \tau_s^\delta D^2\xi(x_s^0)\,(\chi_s)\,ds + D^2\psi(x^\circ)\,(\chi_\delta) + \bar{\Gamma}(y^\circ)\left(G_{x^\circ}\chi_\delta G_{x^\circ}^{\,T}\right)\}\,. \tag{63}$$

6.6.d Prior Covariance Tensor for the Observation

We calculate from (56), (61), and (62):

$$S_{y^\circ} \equiv H_{y^\circ} + G_{x^\circ}Q_{x^\circ}G_{x^\circ}^{\,T} \tag{64}$$

6.6.e Posterior Distribution of the State

We now pull the observation Y_n back to the tangent space at $y^\circ \equiv \psi(x^\circ)$ using (39), and the formulas (57) and (58), to obtain v_n. Finally (40) and (41) give the new state estimate $\hat{\mu}_n$. According to (35), the new conditional covariance tensor is

$$\hat{\Sigma}_n = (I - Q_{x^\circ}G_{x^\circ}^T S_{y^\circ}^{-1} G_{x^\circ})\,Q_{x^\circ}\,. \tag{65}$$

This completes the updating procedure.

6.7 Extended Kalman Filter (for Comparison)

In the same coordinate systems as above, formula (65) is unchanged, but the state estimate becomes

$$\hat{\mu}_n^{EKF} = x^\circ + Q_{x^\circ} G_{x^\circ}^T S_{y^\circ}^{-1} (Y_n - y^\circ) \ . \tag{66}$$

In computational terms, the difference between the GI Filter and the EKF is that the GI filter, unlike the EKF, takes account of the following terms:

- The geometry of state space induced by the model dynamics, as expressed by the connector Γ as in (50).
- The geometry of observation space induced by the observation covariance structure, as expressed by the connector $\bar{\Gamma}$ as in (57) and (58).
- The second derivative $D^2\xi(x)$ of the deterministic flow.
- The second derivative $D^2\psi(x)$ of the observation function.

Finally the GI filter, unlike the EKF, will give the same results in any coordinate system.

6.8 Software Implementation

MATLAB codes for the GI filter and the Extended Kalman Filter (EKF) have been developed for the tracking problem above. Computational effort (in terms of number of flops) was about 40% more for the GI filter then for the EKF. The comparative results will be reported in a future publication.

7 References

[1] M. Cohen de Lara (1997). Finite-dimensional filters. Part I: The Wei-Norman technique. Part II: Invariance group techniques. SIAM J. Control Optim. 35, 980 - 1029..

[2] R. W. R. Darling (1994). *Differential Forms and Connections*. Cambridge University Press.

[3] R. W. R. Darling (1996). Martingales on noncompact manifolds: maximal inequalities and prescribed limits. Ann. de l'Institut H. Poincaré B, 32, No. 4, 1-24.

[4] R. W. R. Darling (1997). Intrinsic location parameter for a diffusion process. Preprint, Mathematical Sciences Research Institute, Berkeley.

[5] D'Souza, C. N., Mcclure, M. A., and Cloutier, J. R (1994). Spherical target state estimators. Proceedings of the American Control Conference, Baltimore MD, 1675-1679.

[6] D'Souza, C. N., Mcclure, M. A., and Cloutier, J. R (1997). Second-order adaptive spherical target state estimation. Wright Laboratory Preprint.

[7] J. Eells, L. Lemaire (1978). A report on harmonic maps. Bull. London Math. Soc. 10, 1–68.

[8] K. D. Elworthy (1982). *Stochastic Differential Equations on Manifolds*. Cambridge University Press.

[9] R. S. Lipster, A. N. Shiryaev (1977). *Statistics of Random Processes*. Springer, New York.

[10] Lototsky, S., Mikulevicius, R., Rozovskii, B. L. (1997). Nonlinear filtering revisited: a spectral approach. SIAM J. Control Optim. 35, 435--461.

[11] E. Pardoux (1991). Filtrage non linéaire et équations aux derivées partielles stochastiques associées. *Ecole d'Eté de Probabilités XIX*, Lecture Notes in Mathematics 1464, 67-163.

[12] E. Pardoux, S. Peng (1992). Backward stochastic differential equations and quasilinear parabolic partial differential equations, in *Stochastic Partial Differential Equations and Their Applications*, B.L. Rozowskii, R.B. Sowers (eds.), Lecture Notes in Control and Information Sciences 176, Springer-Verlag, Berlin.

[13] J. Picard (1991). Efficiency of the extended Kalman filter for nonlinear systems with small noise. SIAM J. Applied Math. 51, 843-885.

[14] C. R. Rao (1973). *Linear Statistical Inference and Its Applications*, 2nd Edition. Wiley, New York.

[15] T. Sakai (1996). *Riemannian Geometry*. American Mathematical Society, Providence.

QUANTITATIVE DESCRIPTION OF WIRE TEXTURES IN CUBIC METALS

Robert J. De Angelis
Professor
Department of Mechanical Engineering

University of Nebraska-Lincoln
255 Walter Scott Engineering Center
Lincoln, Nebraska 68588-0656

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

October 1997

QUANTITATIVE DESCRIPTION OF WIRE TEXTURES IN CUBIC METALS

by:
Robert J. De Angelis
Department of Mechanical Engineering
University of Nebraska-Lincoln
Lincoln, Nebraska 68588-0656

Abstract

The results of x-ray measurements of texture in metallic materials are usually represented graphically employing a method dictated by the measurement technique. Two of the most commonly used representation of texture are the crystallographic pole figure and the inverse pole figure. There exists two types of textures sheet and wire. The rotational symmetry about an axis makes the wire texture simpler to described than a sheet texture. Two x-ray techniques are presented which quantify wire textures. They are the $\Theta/2\Theta$ and the ODF methods. The $\Theta/2\Theta$ method has the advantage of being much quicker than the ODF method, however it is not as robust. This investigation reports the results obtained from employing the $\Theta/2\Theta$ and the ODF methods to quantify the texture in two investigations. The texture changes occurring during annealing of ten specimens of cold worked and annealed copper cut from a eight inch diameter 3/8 inch thick plate were determined in one investigation. The texture evolution in a series of copper compression specimens was monitored in a second investigation. In the two studies both x-ray methods generated almost identical texture descriptions.

QUANTITATIVE DESCRIPTION OF WIRE TEXTURES

IN

CUBIC METALS

Robert J. De Angelis

Introduction

The preferred orientation of grains in a polycrystalline material is refereed to a crystallographic texture. There are two broad categories of textures known as sheet and wire or fiber textures. To completely describe a sheet texture requires the determination of the crystallographic plane that tends to be aligned in the rolling plane and the direction in the crystals aligned in the rolling direction (or the transverse direction). A wire or fiber texture is completely described by the definition of the crystallographic direction aligned parallel to the wire axis.

Since the 1930's two types of x-ray determined pole figures have been used to describe the texture or crystal orientations in polycrystalline wires and sheets. In a crystallographic pole figures the orientation of a grain is plotted on the specimen processing axial system producing a crystal orientation distribution. An comparable method of plotting texture information is by employing the inverse pole figure. In this case the specimen orientation of each grains is plotted on the crystal axial system giving a sample orientation distribution. The inverse pole figure is a projection of the sample orientation distribution on the crystal axial system and the crystallographic pole figure is a projection of the crystal orientation distribution on the axial system of the sample.

In the crystallographic pole figures the density of normals (or poles) to a specific (hkl)

plane are plotted on a polar grid with coordinates based on the specimen processing coordinate system (e.g. rolling direction, transverse direction and normal to the sheet surface). In the case of a sheet texture two directions 90° apart on the perimeter of the polar grid are selected as the rolling and the transverse directions and the center of the pole figure is the direction normal to the plane of the sheet. In the case of a wire texture the center of the polar plot is normally selected to be wire axis direction and there exist a rotational symmetry about this direction.

Pole figures are usually represented on two dimensional polar coordinate plots where the center of the pole figure is the normal to the sheet surface and the direction to the right is parallel to the rolling direction. Thus the pole figure gives the position of the orientation of the specified {hkl} poles of each grain in the polycrystal relative in the axial system of the sample; e.g. normal to the rolling plane and rolling direction in the case of sheet textures or wire axis direction in the case of wire textures. The inverse pole figure gives the distribution of crystal orientations plotted on a section of a suitable net (e.g. Wulff net) on an axial system related to the crystallographic axis.

Problem Description:

There are a number of circumstances in materials application and design which depend on the existing texture being quantified in a condensed way. In some of these cases the fraction of the three major crystallographic poles present on a surface is a sufficient degree of texture characterization to satisfy the application. These three fractions are usually all that are required for a face centered cubic materials having wire textures. The case that is of interest here. The fraction of (100), (110) and (111) poles existing on the transverse plane to the wire axis provides a concise, but very practical, partial description of a wire texture. Two techniques to arrive at such

a partial description of a wire texture will be considered in this investigation. The two techniques, crystallographic pole figures (crystal orientation distribution) and inverse pole figures (sample orientation distributions) present descriptions of a texture utilizing different graphical methods. Initially a detailed development of the crystallographic pole figure (or $\Theta/2\Theta$) method will be described followed by a short development of the inverse pole figure (or ODF) method.

In the case of the (100), (110) and (111) crystallographic pole figures with the plane transverse to the wire axis being the projection surface the centers of the three pole figures contain the density of normals to grains with (hkl) parallel to the wire axis. The density of poles at the centers of the (100), (110) and (111) pole figures are evaluated from a $\Theta/ 2\Theta$ scan of the 2Θ region that contains the (100), (110) and (111) Bragg peaks. This routine scan is made quickly with an x-ray diffractometer.

The fractions of poles of the three orientations are computed from the intensity data of the $\Theta/ 2\Theta$ scan by assuming that the deviations of the ratios of the heights of the Bragg peaks from that observed from a random sample correspond to the strength of the texture. This method of wire texture characterization was suggested by Junginger and Elsner (1) and expanded on by Rhode et al. (3). This $\Theta/ 2\Theta$ method is based on the following considerations. X-Ray diffractometer $\Theta/ 2\Theta$ patterns are collected from random and texture samples. Examples of $\Theta/ 2\Theta$ patterns containing the (111), (200) and (220) peaks for textured and random polycrystalline copper are shown in Fig. 1. Notice the intensities are normalized such that the (111) peak intensity is 100 in both patterns. In cases where the random sample is not available the random peak intensities are taken to be those reported in the JCPDS-ICDD database; e.g. for copper, the reported intensity for the (111):(200):(220) are100:46:20 as contained in Fig. 1.

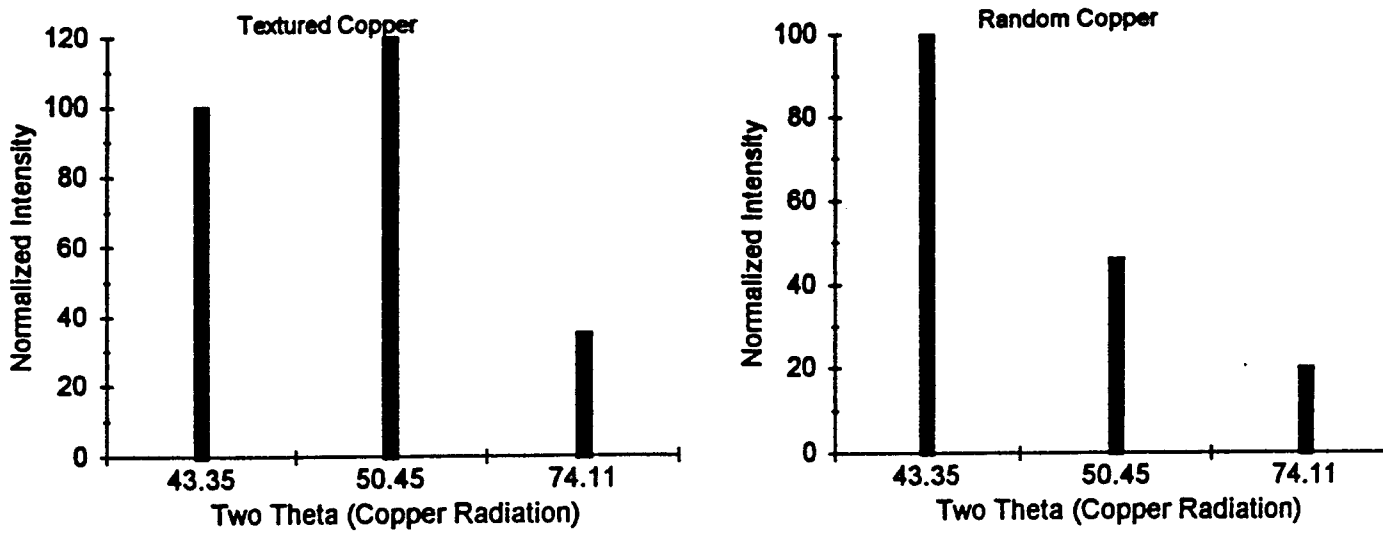


Fig. 1. Normalized X-Ray Diffraction Patterns from Textured and Random Copper.

The texture quantities, the fraction of grains in a specific orientation, are arrived at by calculating the ratios of the peak intensities of the textured material to the peak intensities of the random material. Here the $h_{(hkl)}$ are the textured peak intensities and the $\bar{h}_{(hkl)}$ are intensity values from the random material. The fraction of (200) grains, P_{200} is:

$$P_{200} = \frac{\frac{h_{200}}{\bar{h}_{200}}}{\frac{h_{200}}{\bar{h}_{200}} + \frac{h_{220}}{\bar{h}_{220}} + \frac{h_{111}}{\bar{h}_{111}}} \qquad (1)$$

and the fraction of (220) grains, P_{220}, is

$$P_{220} = \frac{\frac{h_{220}}{\bar{h}_{220}}}{\frac{h_{200}}{\bar{h}_{200}} + \frac{h_{220}}{\bar{h}_{220}} + \frac{h_{111}}{\bar{h}_{111}}} \qquad (2)$$

Equations (1) and (2) are reduced to the following expressions for the fraction of grains of (200) and (220) orientation.

$$P_{200} = \frac{h_{200} * \bar{h}_{220}}{(h_{200} * \bar{h}_{220}) + (h_{220} * \bar{h}_{200}) + (\bar{h}_{200} * \bar{h}_{220})} \qquad (3)$$

And:

$$P_{220} = \frac{h_{220} * \bar{h}_{200}}{(h_{200} * \bar{h}_{220}) + (h_{220} * \bar{h}_{200}) + (\bar{h}_{200} * \bar{h}_{220})} \qquad (4)$$

The fraction of (111) grains is obtained from the condition that the sum of the three $P_{(hkl)}$ are set to unity.

$$P_{111} = 1.0 - P_{200} - P_{220} \qquad (5)$$

In the case of the inverse pole figure of the transverse plane to the wire axis the fraction of the three low index orientations can be computed by integrating the pole densities existing in a fixed angular range around each of the (100), (110) and (111) poles. A numerical technique for accomplishing this integration which includes the required normalization and corrections has been developed by Hosford (3). The three numerically values give the fraction of the total population of grains that are (100), (110) and (111) and the sum of these three values gives the fraction of grains in these three orientations. To compare these values with the values obtained from the $\Theta/2\Theta$ method requires that the sum be normalized to equal unity (see Eq. (5)). The two data sets generated employing the Hosford technique were normalized to satisfy Eq.(5). Since the Hosford method requires the calculation of the ODF to form the inverse pole figure his method is referred to as the ODF method.

Experimental:

A 0.375 inch thick copper plate was produced from a one inch thick pancake by cold rolling, employing a clockwise rotation of 135° between passes. The copper pancake was made by upset forging a three inch diameter, three inch long bar cut from a hot rolled three inch thick slab. One half of the "as cold rolled" plate was provided by Mr. Joel W. House of Wright Laboratory (AWEF) at Eglin Air Force Base. Nine x-ray diffraction specimens were machined from the plate at 1/4, 3/4 and 4/4 radial positions. These specimens were located in the plate half section such that their radial center lines coincided with zero, forty five and ninety degrees to the cut surface. These nine specimens, plus the specimen from the center, were the ten locations in the plate where texture determinations were made. The specimen layout in the plate is shown in Fig. 2. These ten samples were split near to mid-plane and milled flat. The midplane surface was

prepared for x-ray investigation by metallographically polishing and etching. The (111), (200) and (220) pole figures were collected from the midplane surface of the ten specimens. The pole figure data was transformed to ODFs employing both popLA and Siemens software. The ODF data was projected onto an inverse pole figure and the Hosford (3) analysis was performed on these data.

The matching ten specimens were vacuum annealed at 300°C for one hour. The matching surfaces were prepared as described above and the (111), (200) and (220) pole figures were determined on the ten annealed specimens. The ODFs of the annealed specimens were calculated from the pole figure data inverse pole figures were formed and the Hosford analysis was done to quantify pole densities. $\Theta/2\Theta$ scans were collected from all twenty specimens using the Siemens diffractometer at WL/MNMW. Data from the $\Theta/2\Theta$ scans employed with Eqs.(3-5) to obtain the values of $P_{(hkl)}$.

Similar determinations of texture were made on compression specimens supplied by Mr. Joel House. Thirteen compression specimens 0.3 inch diameter by 0.3 inches long were prepared a half hard cold drawn copper bar. The original bar material contained a very strong [111] and [200] combination wire texture. The specimens were compressed to natural strains between -0.051 and -0.792. Specimens subjected deformations in the higher range of strains were remachined to a constant diameter after -0.275 and -0.541 strains. The textures of these thirteen and those of the cold drawn specimens were determined by both the ODF and the $\Theta/2\Theta$ methods.

Results:

The comparison of the ODF and the $\Theta/2\Theta$ methods of quantifying texture in the copper

plate specimens in shown in Figs. 3 to 8. As the data contained in these figures show texture quantification by the ODF and the $\Theta/2\Theta$ methods agree very well. The differences between the results obtained from the two methods are very small for the cold worked specimens. The differences are slightly greater in the case of the annealed specimens, however the overall agreement is excellent. Generally the (220) texture component of the cold worked sample increases from 0.4 to 0.8 at radial distances between 1.5 to 2.5 inches and remains relatively high out to a 4 inch radius. The fraction of (200) poles in the surface of the plate decreases from 0.5 in the center to less than 0.2 at radial distances of 2 to 3 inches and increase back to 0.4 at the outer radial positions. The (111) component is almost completely absents from the cold worked texture at all radial positions. The annealed samples show less variation in texture with radial position than the cold worked specimens. Annealing the cold worked plate has the effect of reducing the strengths of the (200) and (220) texture components and increases the strength of the (111) component substantially.

The texture data obtained on the copper compression employing the $\Theta/2\Theta$ and the ODF methods are shown in Figs.9 and 10.. The fractions of the (111), (110) and (100) poles are plotted versus the total amount of natural compressive strain experienced by the specimen. This data compared very favorably to the data obtains from the ODF method. The agreements between the two methods for the (111) and (200) poles are extremely good. In the case of the (220) the agreement at low and high strains is excellent, however the ODF method transition from low to higher values initiates at a lower compression strain than observed in the data obtained by the $\Theta/2\Theta$ method. Overall the agreement between the two methods for the compression data is considered to be very good.

Conclusions:

The $\Theta/2\Theta$ method provides a rapid, simple and concise method to quantify wire texture in cubic materials.

References:

1. R. Junginger and G. Elsner, "On the Texture of Electroless Copper Films", J. Electrochem. Soc., 135, (1988) 2304-2308.

2. S.L. Rhode, Y.K. Kim and R.J. De Angelis, "Processing-Texture Relationships in Dual-Unbalanced Magnetron Deposition of TiN Films", J. of Electronic Matls., 22 (1993) 1327-1330.

3. William F. Hosford, Final Report for Summer Faculty Research Program, AFOSR, Bolling Air Force Base, DC, 1997.

Figure Captions:

Fig. 1. Normalized X-Ray Diffraction Patterns from Textured and Random Copper.

Fig. 2. X-Ray Specimen Layout in the 0.375 inch Thick Copper Plate.

Fig. 3. Texture Coefficients for the Nebraska Plate in the Cold worked Condition at 0 Degrees.

Fig. 4. Texture Coefficients for the Nebraska Plate in the Cold worked Condition at 45 Degrees.

Fig. 5. Texture Coefficients for the Nebraska Plate in the Cold worked Condition at 90 Degrees.

Fig. 6. Texture Coefficients for the Nebraska Plate in the Annealed Condition at 0 Degrees.

Fig, 7. Texture Coefficients for the Nebraska Plate in the Annealed Condition at 45 Degrees.

Fig. 8. Texture Coefficients for the Nebraska Plate in the Annealed Condition at 90 Degrees.

Fig. 9. Texture Coefficients for Copper Compression Specimens by the $\Theta/2\Theta$ Method.

Fig. 10. Texture Coefficients for Copper Compression Specimens by the ODF Method.

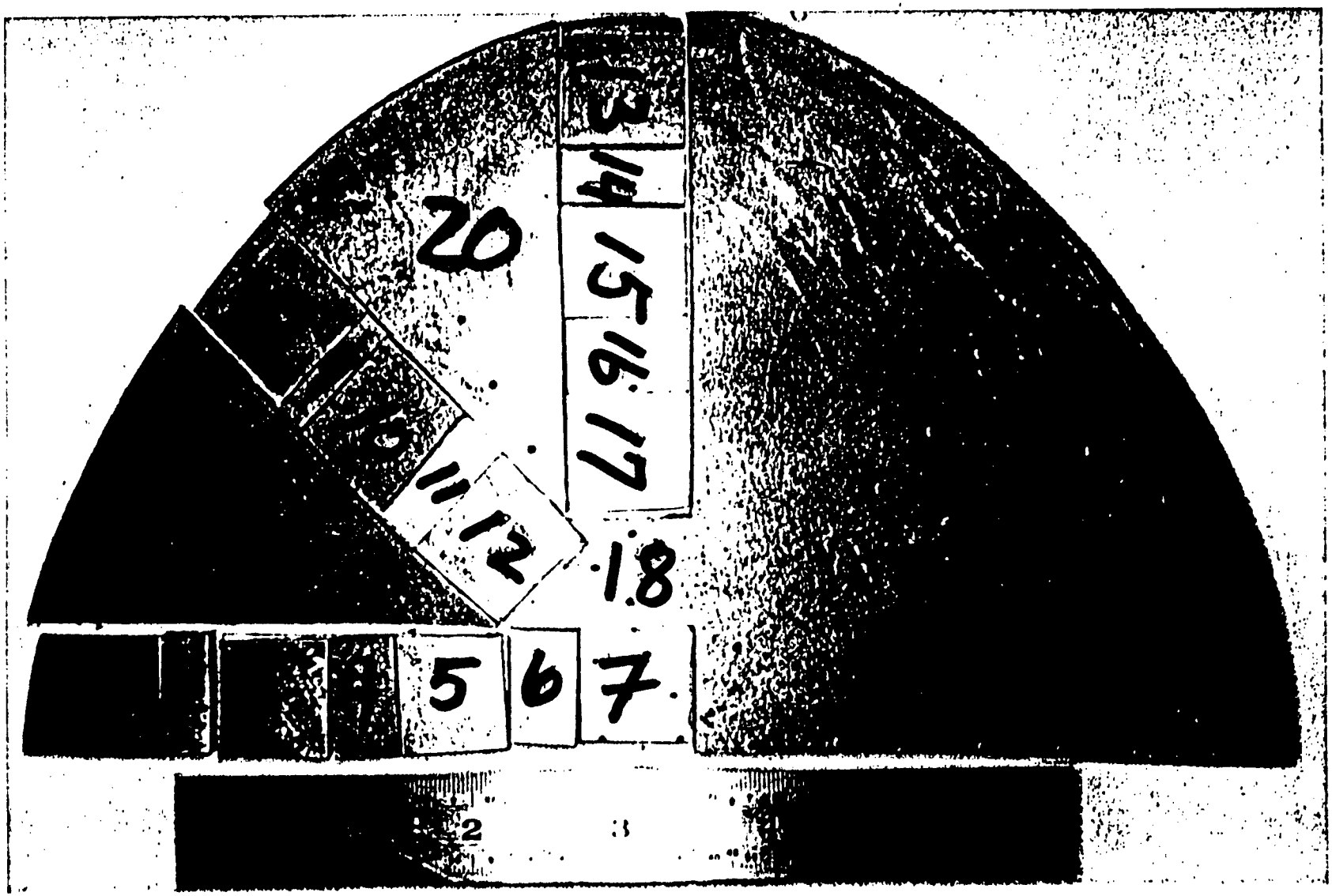


Fig. 2. X-Ray Specimens 1, 3, 5, 7, 8, 10, 12, 13, 15 and 17 in the 0.375 inch Thick Copper Plate.

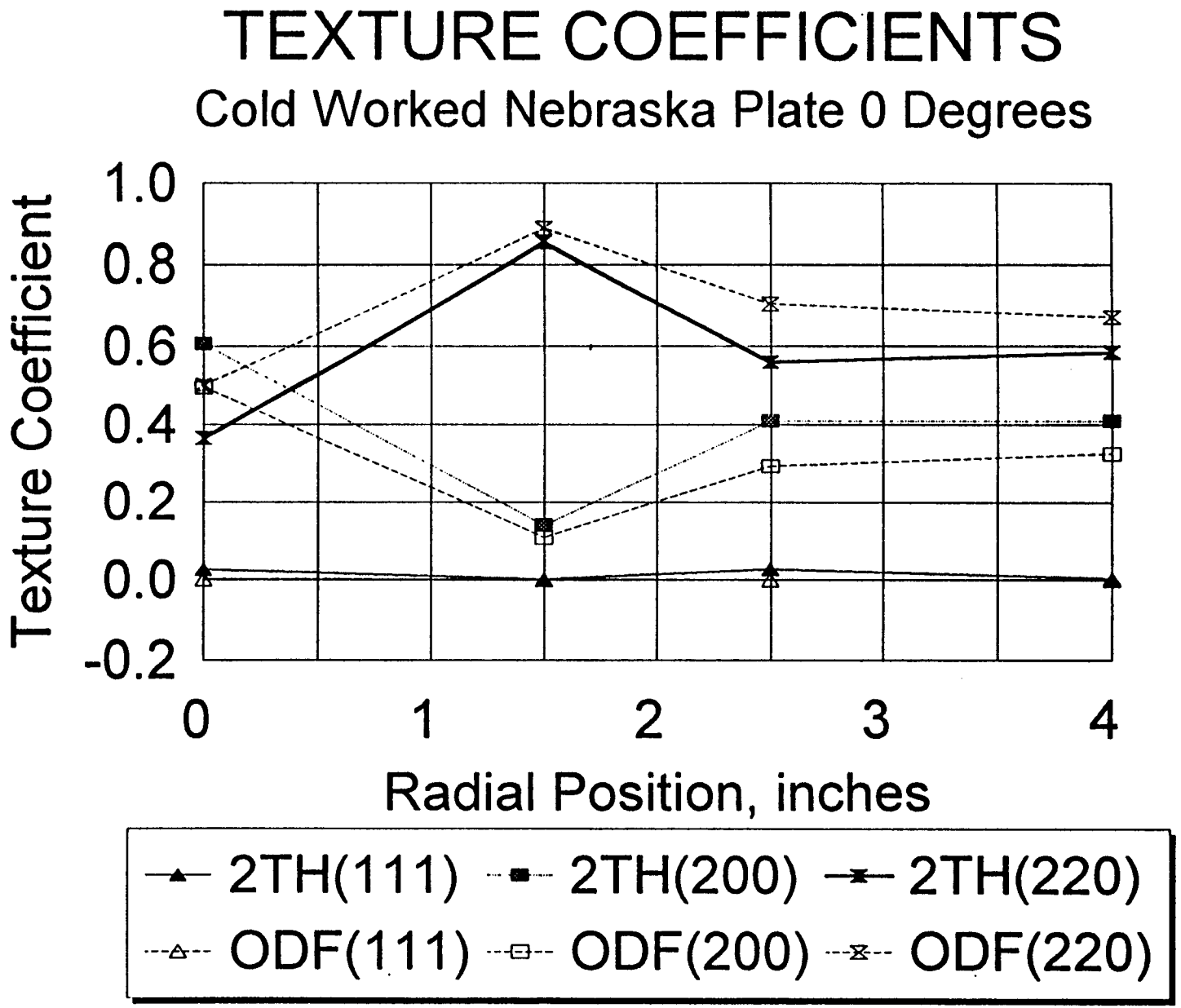


Fig. 3. Texture Coefficients for the Nebraska Plate in the Cold worked Condition at 0 Degrees.

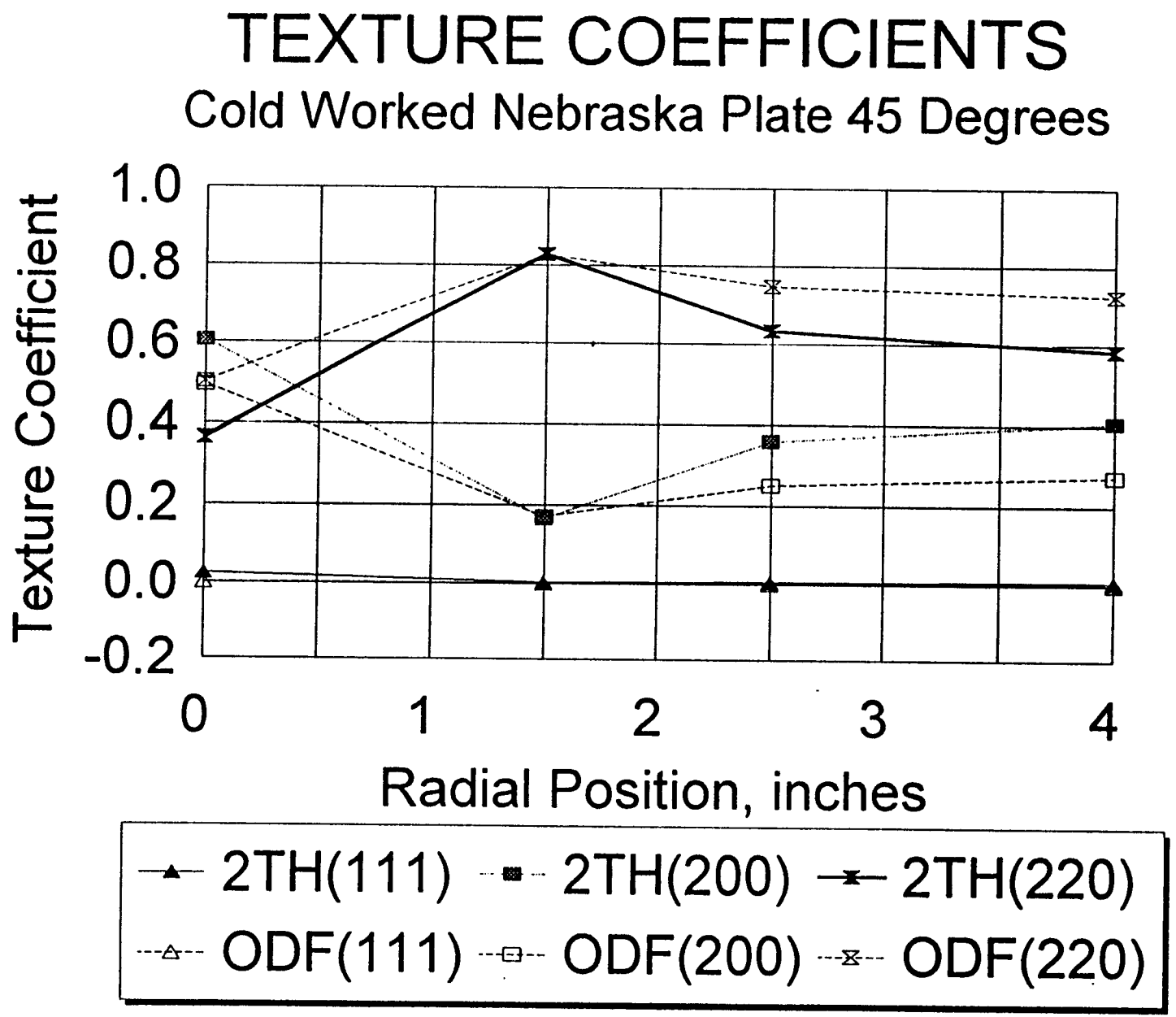


Fig. 4. Texture Coefficients for the Nebraska Plate in the Cold worked Condition at 45 Degrees.

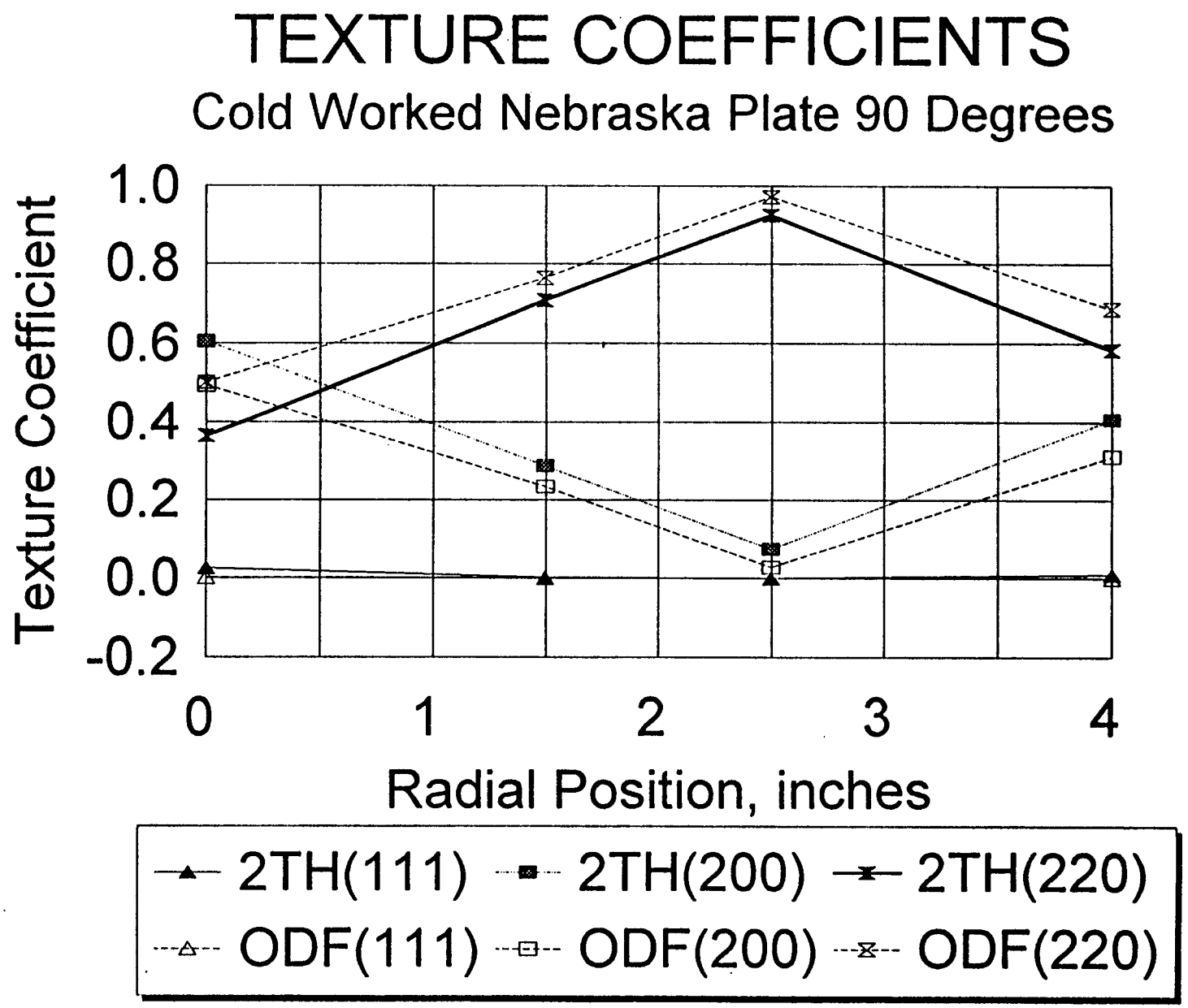


Fig. 5. Texture Coefficients for the Nebraska Plate in the Cold worked Condition at 90 Degrees.

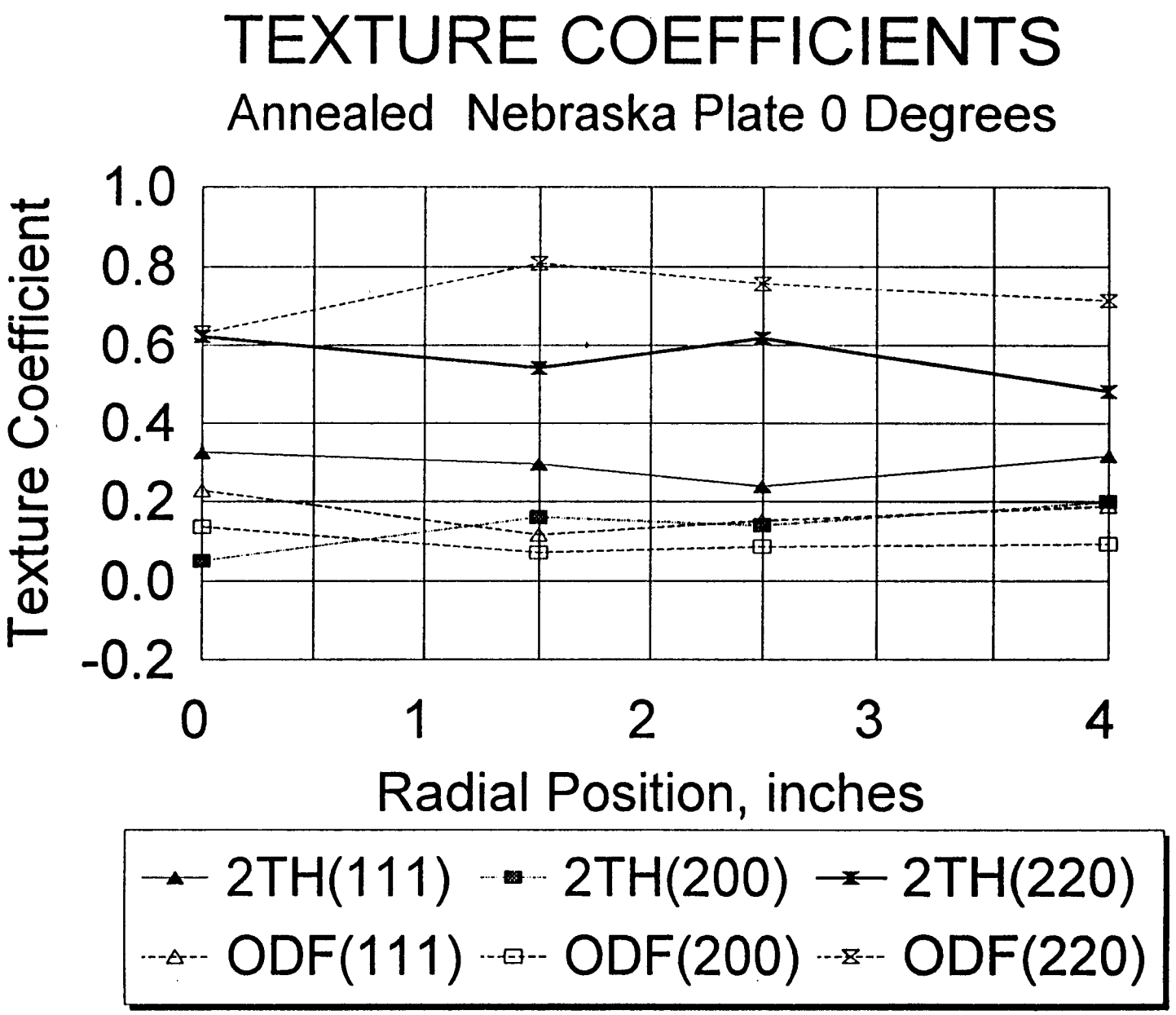
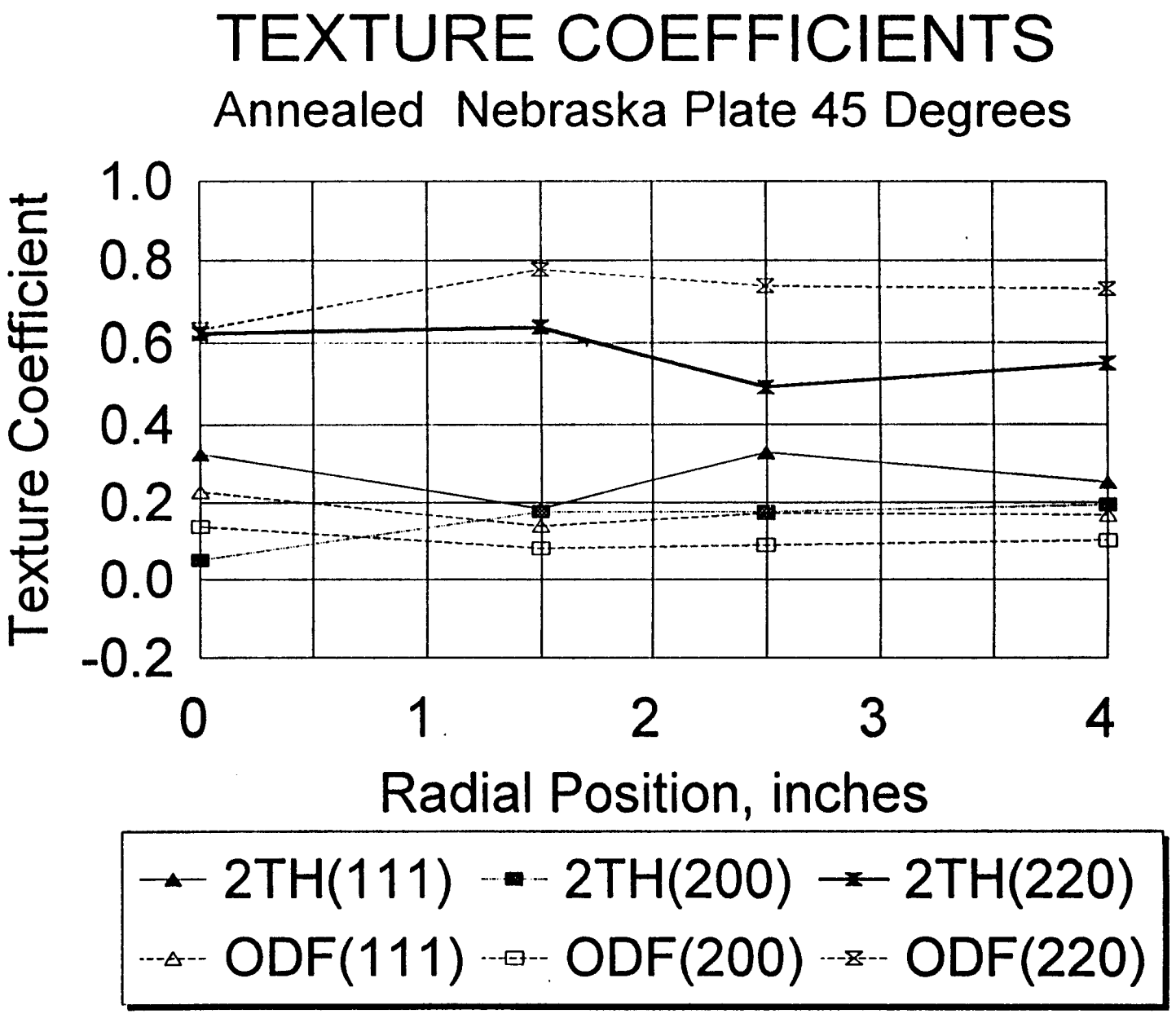


Fig. 6. Texture Coefficients for the Nebraska Plate in the Annealed Condition at 0 Degrees.



Fig, 7. Texture Coefficients for the Nebraska Plate in the Annealed Condition at 45 Degrees.

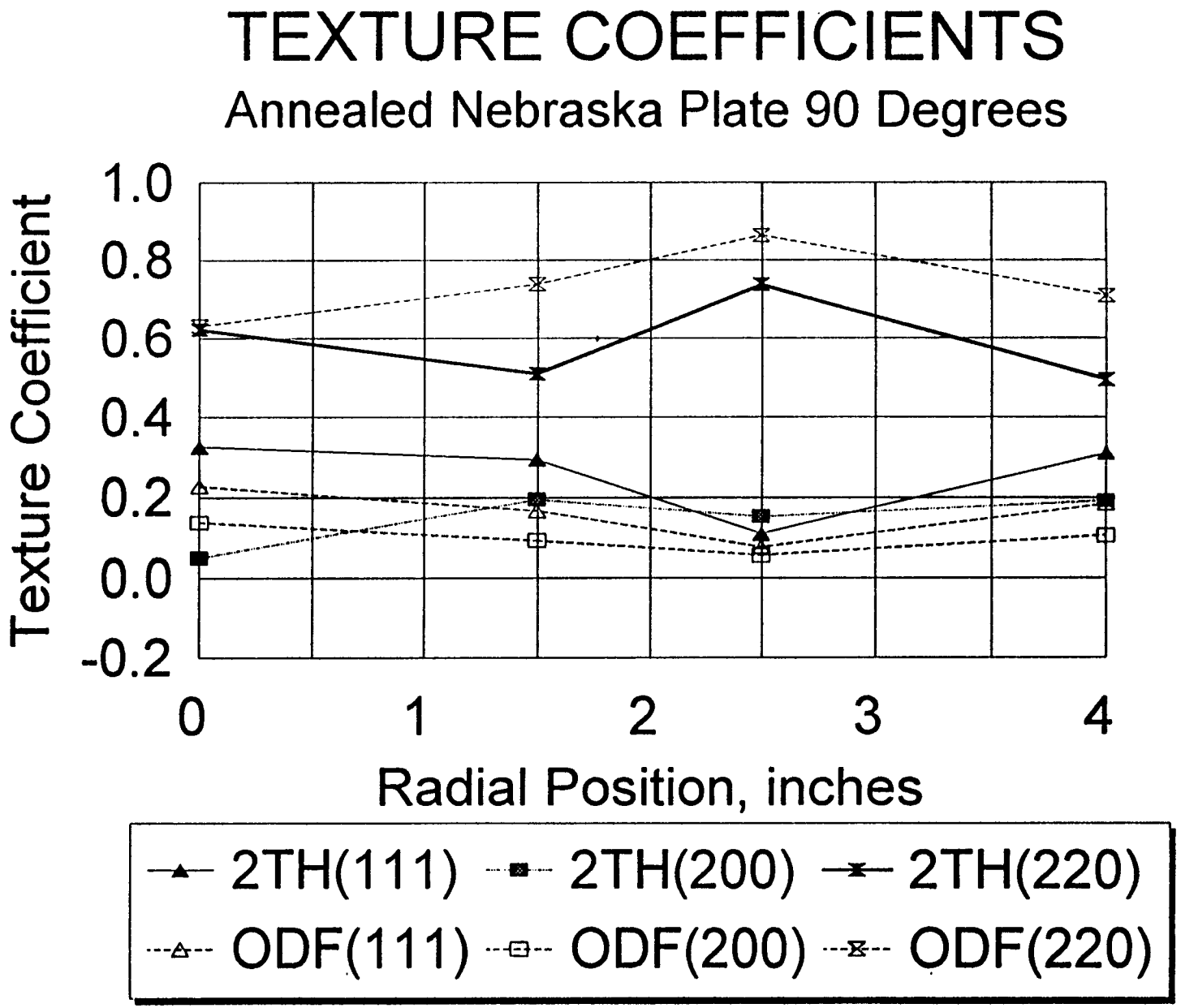


Fig. 8. Texture Coefficients for the Nebraska Plate in the Annealed Condition at 90 Degrees.

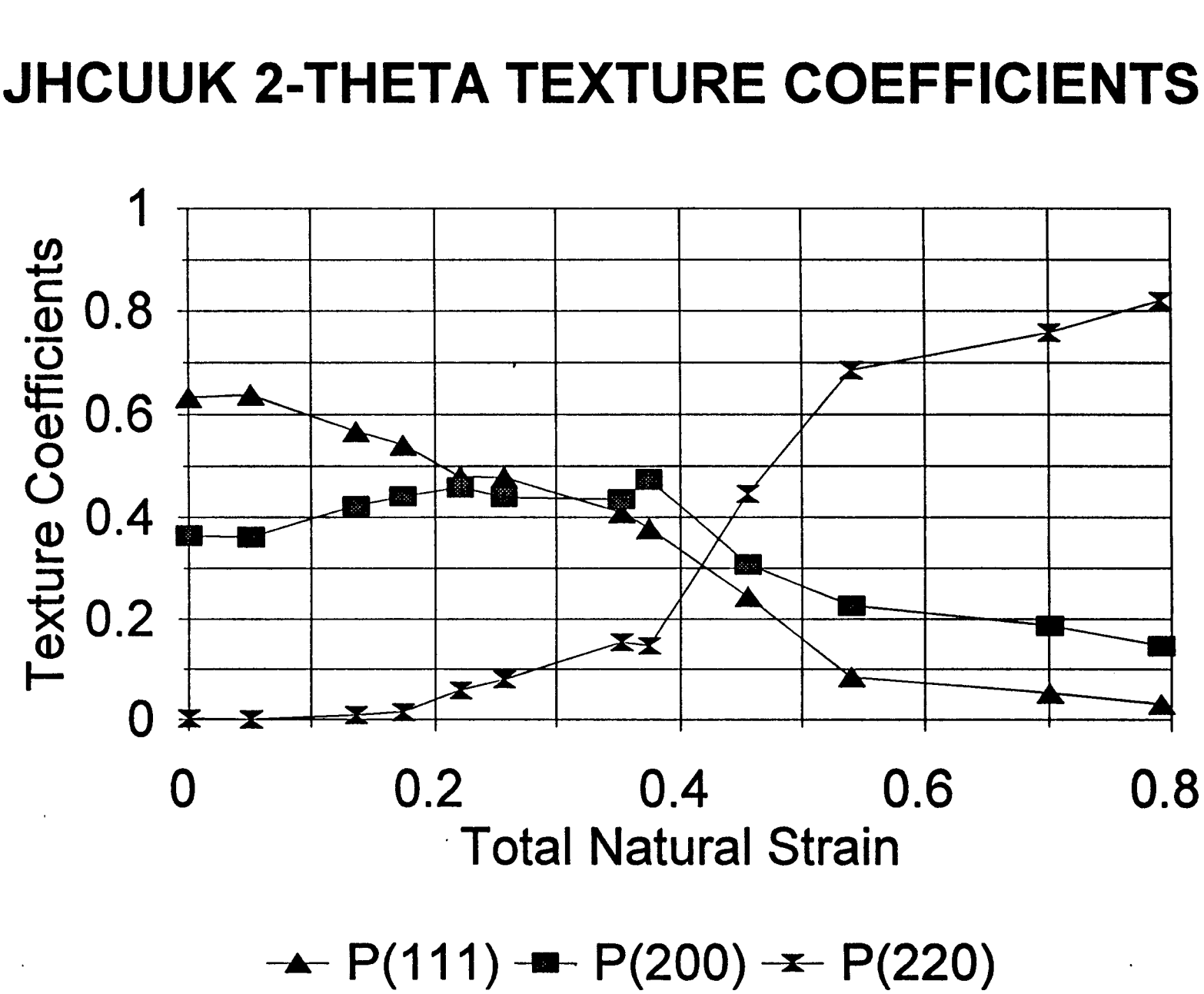


Fig. 9. Texture Coefficients for Copper Compression Specimens by the Θ/2Θ Method.

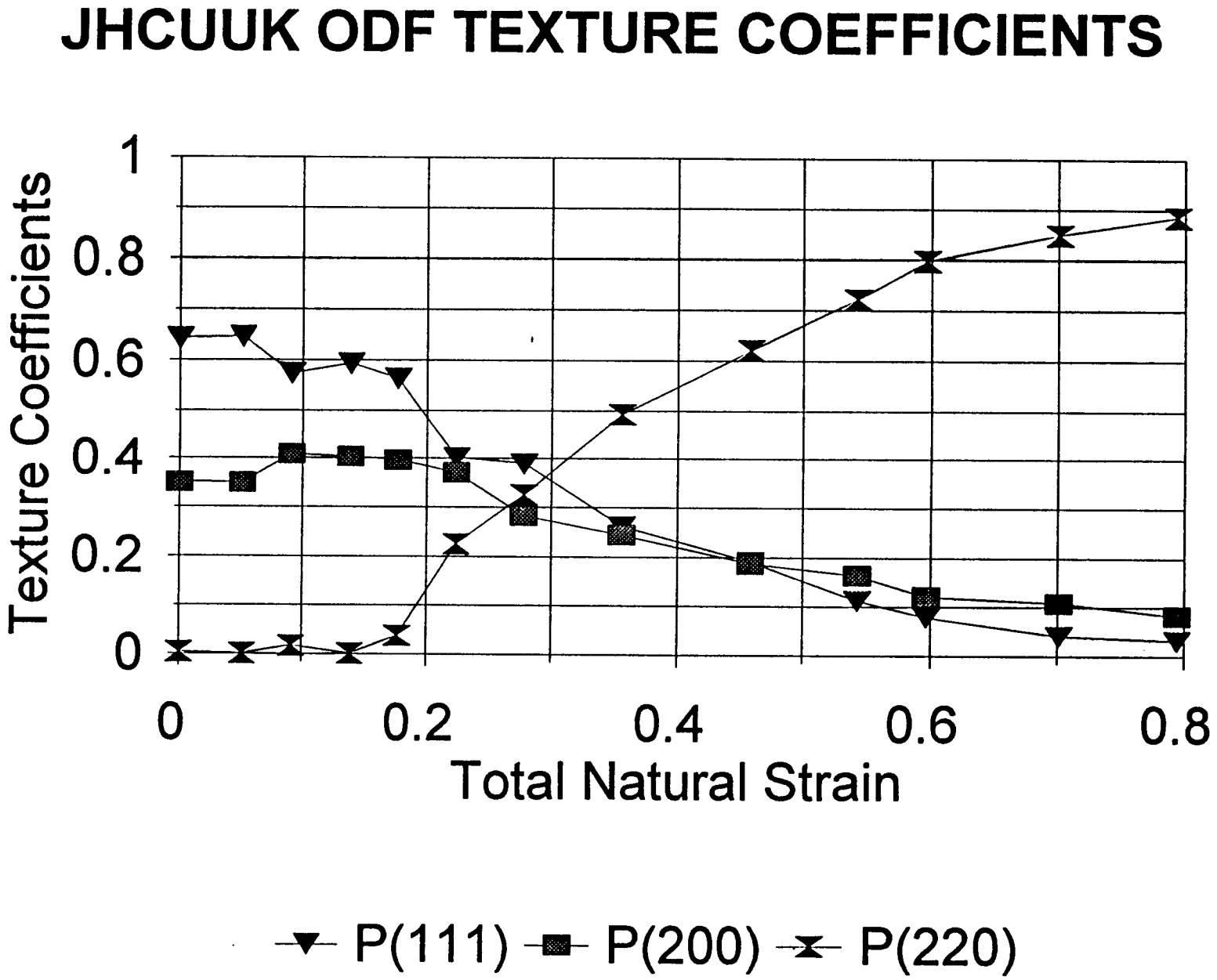


Fig. 10. Texture Coefficients for Copper Compression Specimens by the ODF Method.

ANALYSIS AND CONTROL DESIGN FOR A NOVEL RESONANT DC-DC CONVERTER

Bill M. Diong
Assistant Professor
Engineering Department

The University of Texas - Pan American
1201 W. University Drive
Edinburg, TX 78539-2999

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

August 1997

ANALYSIS AND CONTROL DESIGN FOR A NOVEL RESONANT DC-DC CONVERTER

Bill M. Diong
Assistant Professor
Engineering Department
The University of Texas - Pan American

Abstract

A novel topology for a resonant DC-DC converter was studied. Theoretical analysis, control design and computer simulations of this converter system were performed. Based on the results of this work, a patent will soon be applied for.

ANALYSIS AND CONTROL DESIGN FOR A NOVEL RESONANT DC-DC CONVERTER

Bill M. Diong

Introduction

Under the Air Force's More Electric Aircraft (MEA) initiative, future planes will emphasize the use of electrical power over the use of hydraulic, pneumatic and mechanical power. More and larger electrical loads than in present aircraft, with differing voltage requirements, will be connected to the main 270 Vdc power bus. High-voltage high-efficiency DC-DC converters are therefore of considerable interest to the Air Force. If such converters can also be easily 'programmed' as needed for loads with different requirements, then acquisition and maintenance costs will be reduced as commonality is increased. This report describes the theoretical analysis, control design and computer simulations performed on a novel DC-DC resonant converter topology aimed at satisfying the above requirements.

Methodology

A block diagram of the converter system that was studied is shown in Figure 1 below: we are unable to show the actual circuit diagram at this time due to patent considerations.

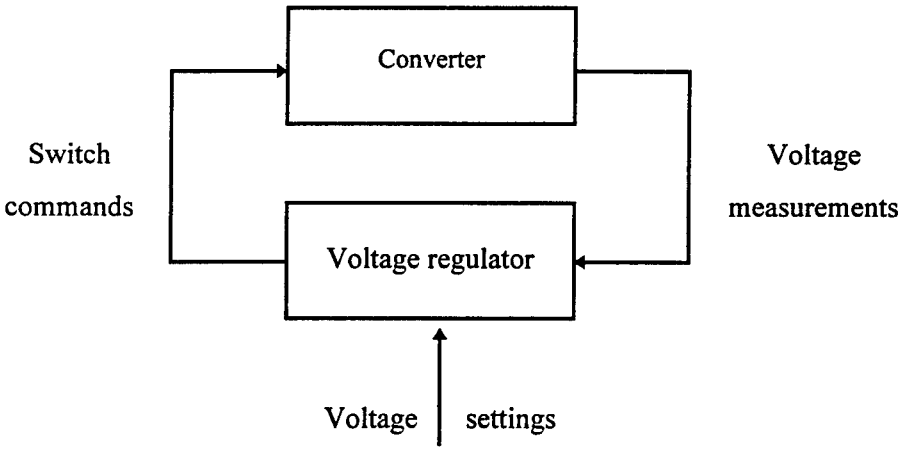


Figure 1 - Block diagram of converter system

The approach followed for this study was to

1. analyze the converter circuit to determine its properties for design purposes,
2. design the control circuit for output voltage regulation and
3. run computer simulations of the closed-loop system to verify its performance.

Results

The resonant converter topology studied differs from presently available topologies [1] in the resonant circuit. However, it is also similar enough so that the various analysis performed in [1] could be adapted to the analysis of this new converter circuit. Among the properties determined were

1. input impedance of the resonant circuit,
2. voltage transfer function,
3. maximum voltage and current stresses for each device,
4. efficiency of the converter and
5. short-circuit and open-circuit operation.

The feedback control circuit (see Figure 2 below) to regulate the output voltage of the converter was designed based on the previous analysis. The proportional-integral (PI) part of the control circuit was used because a simple, low-bandwidth controller was sufficient for this present application. The pulse-width-modulation (PWM) part of the circuit sends fixed-frequency gating signals to the converter circuit with duty-cycles as determined by the output of the PI circuit.

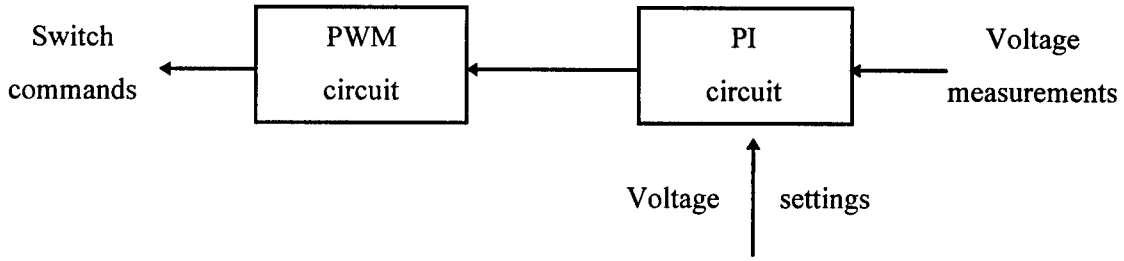


Figure 2 - Block diagram of feedback control system

Computer simulations were performed using MicroSim PSpice software to verify the analysis and validate the control system design. The simulated output voltage response of the converter, regulated at 240 V, during a 9 kW load application followed by a 6 kW load removal is shown as Figure 3 below.

Conclusion

Theoretical analysis, control design and computer simulations were performed on a novel DC-DC resonant converter topology. The results strongly suggest that it is able to fulfill all of the requirements associated with operation on a MEA.

References

[1] M. K. Kazimierczuk and D. Czarkowski, *Resonant power converters*, John Wiley & Sons, Inc., New York, 1995.

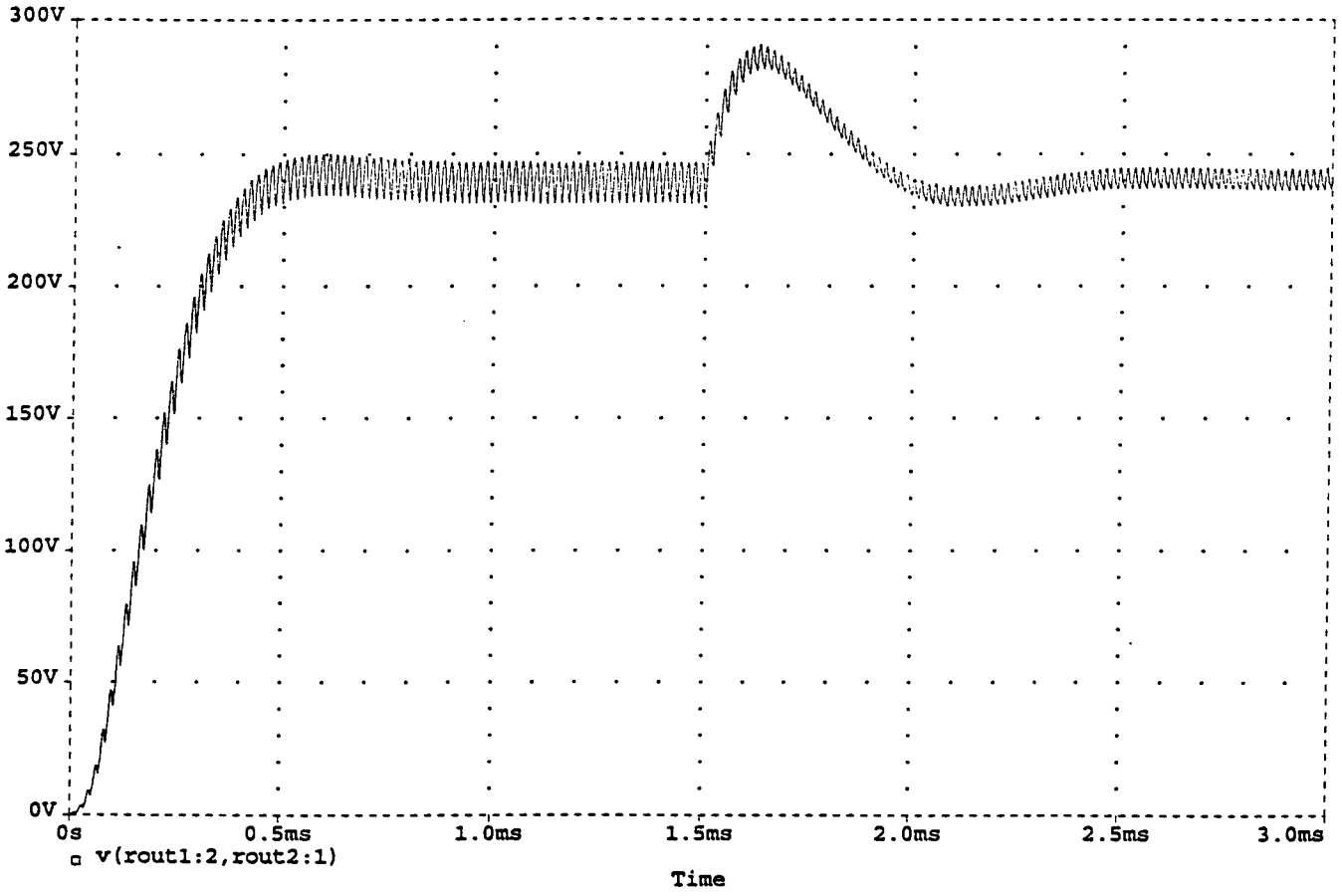


Figure 3 - Simulated output voltage response of the converter for a 9 kW load application followed by a 6 kW load removal

GIDDING MISSILES (ON THE FLY): APPLICATIONS OF NEUROBIOLOGICAL PRINCIPLES TO MACHINE VISION FOR ARMAMENTS

John K. Douglass
Professor
Division of Neurobiology

University of Arizona
611 Gould-Simpson Bldg.
Tucson, AZ 85721

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

And

Wright Laboratory

August 1997

GUIDING MISSILES "ON THE FLY:" APPLICATIONS OF NEUROBIOLOGICAL PRINCIPLES TO MACHINE VISION FOR ARMAMENTS

John K. Douglass
Arizona Research Laboratories Division of Neurobiology
University of Arizona

ABSTRACT

The goal of this project was to evaluate the potential for applying information processing strategies from the visual systems of flying insects to biomimetic designs for missile guidance technology. A review of various neurobiological publications was supplemented by additional readings on analog VLSI and other biomimetic technology, and by briefings provided by personnel at the Wright Laboratory, Eglin AFB, regarding current technology for guided munitions. Three major information processing paradigms were identified as essential components of the functional organization of visual processing in the fly brain: (1) multiple spatially mapped representations that are connected and processed in parallel, (2) "coarse coding" of input parameters, and (3) matched filters tailored to specific, often complex features of the raw sensory inputs. Together, these paradigms provide a conceptual basis for designing revolutionary new guidance systems, some proposed components of which are described in this report. The numerous advantages of incorporating these biological principles into man-made designs include the elimination of any need to solve complex equations, or even to perform explicit mathematical computations at all. Instead, in the insect brain and the emulations proposed here, all algorithms and computations are encoded implicitly in the overall physical architecture and the functional properties of its components.

GUIDING MISSILES "ON THE FLY:" APPLICATIONS OF NEUROBIOLOGICAL PRINCIPLES TO MACHINE VISION FOR ARMAMENTS

John K. Douglass

INTRODUCTION

The overall goal of this project is to evaluate the potential for applying information processing paradigms from insect visual systems to innovations in missile guidance technology. This paper provides an overview of some relevant features of the functional organization of visual processing in the fly brain, reviews certain fundamental information processing principles that have been identified in the insect visual system as well as in nervous systems of other animals, and identifies current issues, the resolution of which will provide additional insights for future designs of artificial visual systems. Although the emphasis is on vision in flies, illustrative examples from other sensory processing systems are included where appropriate.

The field of biomimetic design has produced impressive results in the form of VLSI chips that emulate certain basic image processing capabilities of natural visual systems (e.g. Mead 1989; Koch and Li 1995). Already some five years ago, a robot on wheels was built using fly-like visual inputs and elementary motion detectors (Franceschini et al., 1992). The field is nonetheless still in its infancy, and some of the simplest, most ubiquitous and powerful principles of neurobiological information processing are currently underutilized. These principles offer the potential to revolutionize the way that man-made devices acquire and process sensory information, for applications ranging from remote sensing and image processing, to automated guidance of unmanned aerial, terrestrial and submersible vehicles, as well as enhanced visual aids for manned vehicles.

Diverse aspects of natural visual systems are applicable to technology; this is not intended to be a comprehensive review. Due to space limitations, references are limited to review articles and papers selected from a large body of literature. Important areas that are not covered here include lateral inhibitory mechanisms (Tonkin and Pinter, 1996), feedback control loops, dynamical and nonlinear properties of neurons and neural networks, and the functional architecture of information processing in non-mapped neuropils.

Basic Air Force Requirements for guided munitions

Shrinking budgets and the unpredictability of future threats require that missile guidance systems of the future be smaller and more affordable, yet more accurate and more resistant to countermeasures. Ideally, the specific capabilities of unmanned aerial vehicles (UAVs) will include automated navigation and course correction systems, target detection, identification, tracking and interception, reliable methods for identifying friend or foe, minimal power requirements, maximal processing speeds, and modular designs that can be quickly and easily reconfigured to serve multiple mission objectives. The US Air Force envisions five basic conventional armament designs: the dual-range missile (DRM), small smart bomb, smart hard target munition, smart soft target munition, and anti-materiel munition (LOCAAS) (Anon., 1997). Each of these armaments requires fast, sophisticated sensory processing and robust performance characteristics under varied weather conditions. For example, in order to pursue either fore or

aft targets, the DRM must maintain an acceptable level of aerodynamic stability during a 180-degree about-face from the forward launch direction. The LOCAAS must autonomously detect, identify and pursue a target by visual means, preferably through simultaneous use of several processing mechanisms in order to increase reliability and decrease vulnerability to countermeasures.

Rationale for Developing Biomimetic Designs that Focus on Insects

What do insects have to offer to the world of machine vision? Flying insects such as the housefly have evolved extraordinarily effective mechanisms of sensory processing and locomotor control that meet many of the above-noted requirements for guided munitions. Flies depend heavily upon vision in order to survive, and a large proportion of their brain is devoted to processing visual information and generating appropriate commands to wing, neck and leg muscles in response to changes in visual inputs. The visual system is largely responsible for the ability of flies to detect and evade capture by predators, avoid obstacles as well as track, pursue and evaluate potential mates while in flight, and land safely on an unstable substrate such as a twig swaying in the wind (Collett and Land, 1978; Egelhaaf et al., 1988; Wagner, 1986). In addition, the neuroanatomy of the fly visual system shows a very precise and modular architecture, which facilitates the repeatability of experiments and the prospects for fully understanding the mechanisms of information processing in this system (c.f. Strausfeld 1976, 1989; Douglass and Strausfeld 1995, 1996). Finally, compared to vertebrates, flies are far more convenient and inexpensive to maintain as experimental animals. For all of these reasons, the visual system of the fly has been, and continues to be, an excellent model system for investigating neural principles of sensory information processing.

Despite their impressive behavioral sophistication, insects are not rocket scientists: they accomplish all that they do with a very small brain, a minuscule "power supply," and no explicit mathematical training. Yet, precisely for these reasons, rocket scientists should look to flying insects for efficient, yet elegant and robust solutions to problems in guidance and control. It will be argued below that much of the computational power of the insect visual system is actually built in to the architecture of the brain. The functional properties of individual neurons support powerful implicit information-processing capabilities that are compatible with this architecture. Therefore, in order to fully exploit neural information processing mechanisms, technology must be designed to imitate specific structural as well as functional features of neural systems.

This report describes four crucial principles of insect visual processing: spatially mapped architectures for information representation, coarse coding, matched filters, and the use of implicit algorithms. The full-blown application of these principles to components of automated guidance systems is expected to provide tremendous improvements in processing speed, design modularity, robustness, and cost-effectiveness, with no need to explicitly solve any equation or otherwise employ a digital interface.

METHODS

In order to evaluate applications of neural information processing systems to military technology, I began by familiarizing myself with current and planned Air Force missile capabilities, then considered what aspects of

natural sensory systems may be both practical and appropriate for new man-made designs. Readings included the FY97 Conventional Armament Technology Area Plan (Anon., 1997). In addition, various personnel at the Wright Laboratory, Eglin AFB, provided non-classified briefings regarding existing guided munitions technology, and I participated in numerous discussions at the Wright Laboratory on aspects of insect visual systems and neurophysiology. I reviewed various publications selected from the neurobiological literature which pertain to visual as well as other mechanisms of sensory processing, and read several papers pertaining to VLSI and machine vision technology. Many of these publications were obtained from the Interlibrary Loan system through the Technical Library at Eglin AFB, FL. A program written in Turbo Pascal (Borland International) was used to model the ability of alternative coarsely-coded sensory processing mechanisms to resolve differences in stimulus parameters.

RESULTS

There are many features of neuronal information processing that are unusual from the standpoint of current technology, and may therefore be useful to include in future designs. Several features are rather obvious, such as redundancy, small size, and low energy consumption. Other significant features of biological sensory processing systems have already been implemented with considerable success, including reconfigurable analog processing, multiplexing, lateral inhibition, and automatic local gain control (e.g. Massie et al., 1994; Tonkin and Pinter, 1996; Villasenor and Mangione-Smith, 1997).

The Functional Organization of Visual and other Sensory Systems

The visual systems of various organisms, despite their distinct evolutionary histories and functional specializations for different environments, share many fundamental features. In both insects and primates, for example, each of two eyes projects an image of the external world, with some binocular overlap, onto a retina that is characterized by a 2-dimensional array of photoreceptor neurons. When a single photon entering a photoreceptor is absorbed by a rhodopsin molecule, this initiates a biochemical cascade that alters the electrical potential across the receptor cell's membrane, initiating the process by which changes in light intensity at a particular location in space are signalled to the rest of the visual system. Some attributes of a visual system, such as imaging optics and photolabile pigments, are absolute requirements for it to function; other features, such as the number of eyes and the mechanism of image formation, can be viewed as alternative "designs" that accomplish the same basic tasks. In assessing the biomimetic potential of natural sensory systems, a major goal therefore should be to understand which features are essential, which are optional, and what the functional differences are among alternative designs.

Spatial Maps

A crucial architectural feature of natural visual systems, also prominent in other sensory systems, is that incoming signals are processed in parallel, mapped arrays at multiple, interconnected strata (Figs. 1&2). In visual systems, the primary spatial map of the visual world is formed by photoreceptors in the retina. Already at this level, visual information is actually represented by the analog voltage fluctuations within several superimposed arrays of

different photoreceptor types. In humans, these arrays are composed of rods and three types of cone photoreceptor. Similarly, insect retinas typically contain three or more distinct photoreceptor types, each endowed with its own form of the visual pigment rhodopsin and therefore sensitive to a somewhat different range of light wavelengths. Beyond the photoreceptor layer, subsequent processing levels receive massively parallel inputs from the retina that preserve the retinotopically mapped spatial relationships among individual elements, yet are specialized for processing distinct aspects of the visual scene.

At the first levels beyond the photoreceptors (e.g. in the fly's lamina; Fig. 1), very basic image processing functions such as contrast enhancement and gain control are carried out. (These functions are very important, but will not be elaborated upon due to time and space limitations.) Also at a very early level (beginning in the fly's medulla), the initial steps in motion detection take place. Already at this level, motion-processing pathways have diverged from separate pathways (also retinotopically organized) that are involved with form and spectral (color) processing (Ramachandran and Gregory, 1978). The achromatic, motion-sensitive pathways proceed to the lobula plate, while chromatic processing pathways proceed to the lobula (Strausfeld and Lee, 1991). From each retinotopic column, or "visual sampling unit" within the medulla, several types of neurons with small spatial receptive fields project to the lobula plate, several others to the lobula, and a third group comprised of "Y" cells (Fig. 2) having a bifurcated axon, projects to both the lobula and the lobula plate. To the extent that their neuronal activities are isolated from each other, each of these pathways to constitutes a distinct processing channel. On the other hand, surely there are synaptic interactions among these channels. Thus, the overall architecture of the fly's visual processing hardware can be

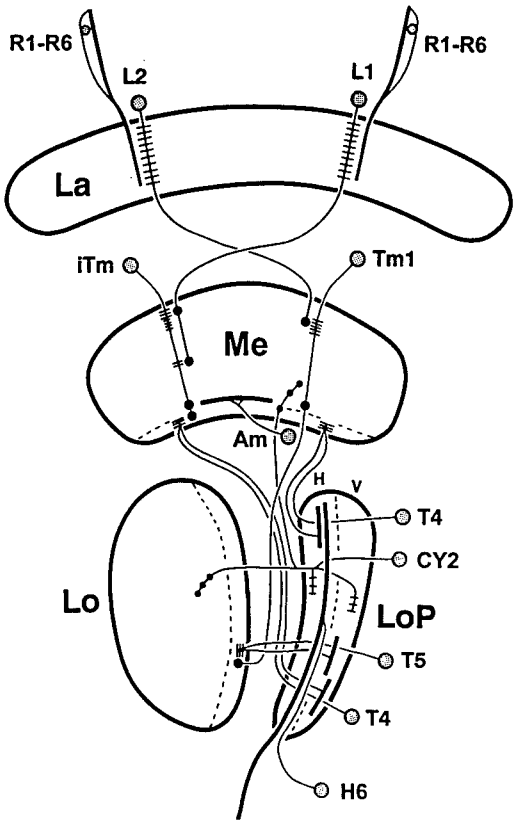


Figure 1. Schematic view of parallel pathways to the lobula plate in the calliphorid fly brain (after Douglass and Strausfeld, 1995; see also Buschbeck and Strausfeld, 1996), showing anatomical relationships among selected small-field retinotopic neurons, a medulla amacrine cell, and a wide-field lobula-plate efferent neuron (*stippled circles*, cell bodies). An ipsilateral optic lobe (*La*, lamina; *Me*, medulla; *Lo*, lobula; *LoP*, lobula plate) is shown in horizontal section, with anterior to the *left*. Dendrites (*thin bars*) of large monopolar cells L1 and L2 receive achromatic inputs from photoreceptors R1–R6 and terminate (*filled circles*) at characteristic levels that coincide with dendrites of the transmedullary cells iTm and Tm1. iTm terminates at T4 dendrites in the deep medulla. Tm1 terminates at the T5 dendritic layer in the outer lobula. iTm and Tm1 both have output zones (*filled circles*) just distal to the T4 dendritic layer in a stratum containing the deep medulla amacrine (*Am*). T4 and T5 terminate (*thick bars*) in lobula plate strata corresponding to horizontal (*H*) and vertical (*V*) motion sensitivity, in which they synapse with wide-field tangential cells, exemplified by the tangential cell H6.

Fig. 1. Schematic view of visual processing centers in the brain of a blowfly. **La**, lamina; **Me**, medulla; **Lo**, lobula; **LoP**, lobula plate. Reprinted from Figure 1 in Douglass and Strausfeld (1996), p. 4552.

viewed coarsely as a combination of multiple, superimposed spatial mapped pathways that diverge and merge their information in various ways.

Mapped arrays may also arise in the nervous system that are not explicitly represented in the raw sensory input, but have some systematic relationship to physical dimensions or distributions in the external world. Good examples of such "synthesized" maps are provided by the auditory system of owls. Barn owls are able to use passive auditory cues to locate prey (such as mice) in total darkness. In the barn owl's brain, the time delays between signals arriving at the two ears are spatially mapped to represent the azimuthal (horizontal) coordinate of the mouse. In addition, because one of the owl's ears is placed slightly higher than the other, slight interaural amplitude differences reveal the vertical position of the mouse; these differences are represented in the brain as an additional spatial map (Konishi 1986, 1993). The ordered spatial representation of input variables, creating a spatial "image" in the brain whether or not the variables have any real connection with external spatial coordinates, is thought to facilitate the analysis of the mapped parameters by local circuits (Heiligenberg, 1987). Beyond synthesized maps that organize neural activity along straightforward parametric dimensions, highly abstracted qualities of sensory inputs are also represented spatially. In the olfactory cortex of vertebrates, for example, much of the perceptual information about a specific odor appears to reside in spatial rather than temporal patterns of activity (Freeman, 1990).

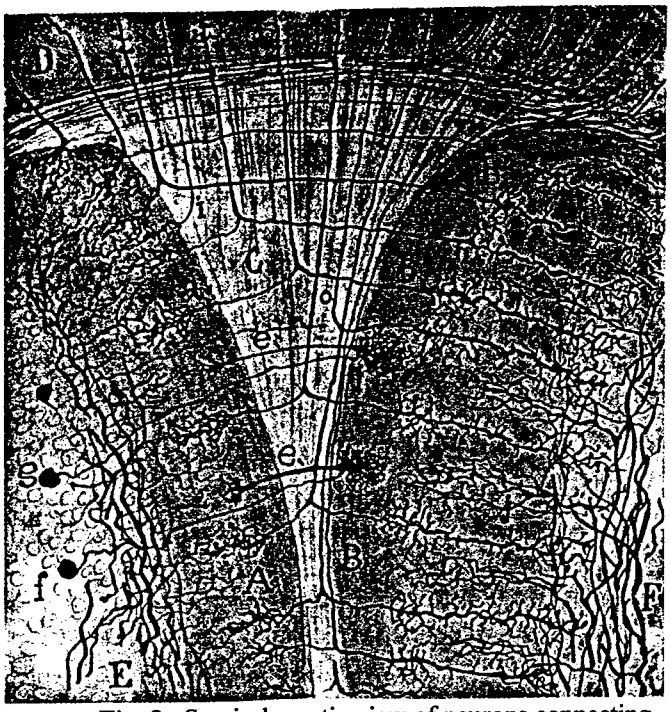


Fig. 2. Semischematic view of neurons connecting the medulla (at top), lobula plate (left) and lobula (right) of a tabanid fly (see Fig. 1). Reprinted from Figure 52 in Cajal and Sanchez (1915), p. 101.

Some (perhaps many) neural maps are three-dimensional. In the pancake-shaped lobula plate neuropil of the fly's brain, for example, the basic 2-D retinotopic (spatial) map is composed of distinct layers that encode the direction of motions within the monocular visual field (Buchner et al., 1984). The directional map within the lobula plate is a synthesized map in two senses, first in that it arises from temporal changes in the activity of neurons in 2-D spatial representations at lower processing levels, and second, because the relationship between "direction" as a mathematical coordinate and "direction" as a position within the neural map is not linear. Instead, two adjacent layers of the lobula plate encode opposite directions of horizontal motion (Fig. 1, stratum labelled, "H"), and the next two layers, opposing vertical directions (Fig. 1, "V"). This configuration is intriguing as a reminder that there is no requirement for the architecture of neural maps to meet human intuitive expectations. Moreover, because other

maps of direction and orientation in nervous systems (e.g. in vertebrate cortex) exhibit the "expected" order between anglular and mapped coordinates (Shepherd, 1994), this suggests that the alternative layout in the lobula plate may have important functional consequences.

Coarse Coding of Input Parameters

In addtion to mapped arrays, one of the most ubiquitous functional characteristics of sensory processing systems is that the response properties of individual neurons tend to be broadly tuned, or "coarsely coded" (Heiligenberg, 1987; Horridge, 1992). Visual pigments, for example, have broad absorption spectra, and individual photoreceptors therefore are sensitive to a broad range of light wavelengths. In various sensory modalities, receptors and interneurons that are sensitive to the direction of motion all exhibit directional response patterns that resemble a sinusoid or cardioid function. Examples are the saccular hair cells in vertebrate semicircular canals, hairlike sensors of wind and water motion on the body surfaces of invertebrates, and visual motion-sensitive neurons in the fly brain and the vertebrate retina. Sensory receptors that are part of a spatially mapped array can also be said to exhibit "coarse coding" to a limited degree; two examples are the overlapping angular acceptance functions of photoreceptors (Snyder, 1975) and the spatial overlap among pressure-sensitive receptors in human fingertips (Wheat et al., 1995).

At the level of primary receptor neurons, the broadness of these response functions is often a simple consequence of the physics of sensory transduction. Organic molecules, for example, generally have broad absorption spectra. Yet, it would appear that mechanoreceptive hairs could easily have evolved to be more directionally selective, and photoreceptors more wavelength-selective. Why, then, are neurons and their associated sensory transducers typically so coarsely tuned to the very variables they function to encode?

An obvious answer is that it would be too costly, in various ways, for a nervous system to have a separate neuron responding to each point along every parametric dimension. At the level of sensory transduction, such specificity would likely come at the metabolic expense of providing response-sharpening filters of some kind. The filters, in turn, would reduce the absolute sensitivity of individual sensory receptors. Moreover, at all levels of the system, the death of a single cell could leave an irreparable "blind spot" in the abstract sensory space being encoded. Finally, a brain full of such overspecialized neurons would have to be much bigger than its coarse-coded counterpart (but see Matched Filtering section, below).[1]

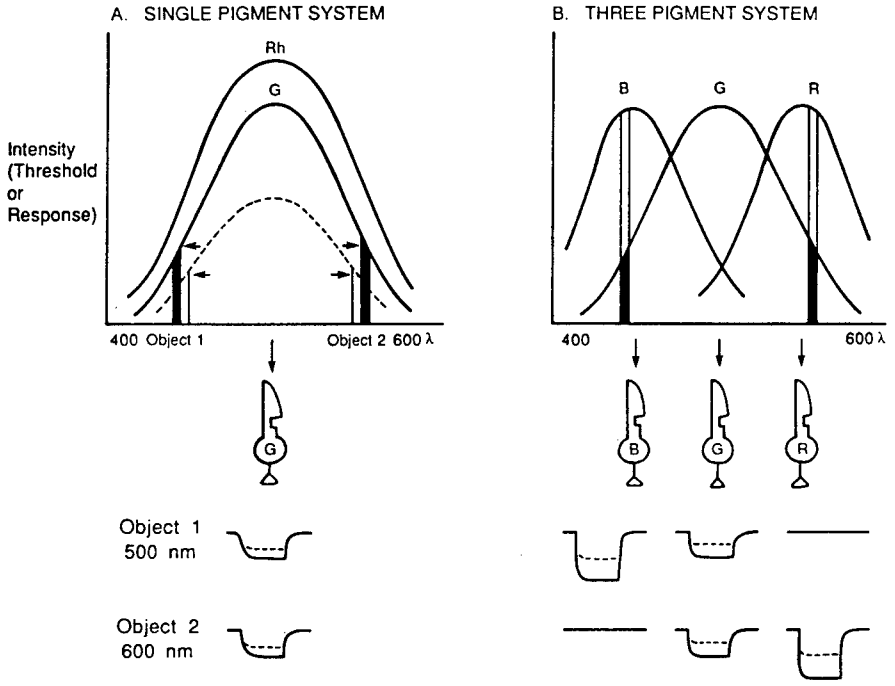
A second major benefit of coarse coding is that it actually works: behavioral experiments often show that animals can resolve much smaller differences in stimuli than would be suggested simply by the parametric "spacing" between neuronal response maxima. Thus, animals with color vision can discriminate a huge variety of hues, usually with only three or four kinds of visual pigments with widely separated light absorption maxima. Similarly, although

[1] The importance of coarse coding actually extends far beyond the concept of broad tuning curves along a single variable. Many neurons respond to several different parameters, contain multiplexed information, and serve in multiple circuits; thus they are effectively "coarse-coded" along several dimensions. In this broad sense, coarse coding goes a long way toward explaining how a brain as small as a fly's can do so much.

the density of pressure-sensitive mechanoreceptors under a human fingertip is only about 1 per mm^2, humans can resolve the location of a pinprick at a much finer resolution (Wheat et al. 1995).

Optimal Design of Coarsely Coded Receptors

Although the actual neuronal machinery for processing coarse-coded inputs is not well-understood, it must involve comparisons among the responses of different receptors or interneurons (e.g. Fig. 3). Theoretical studies (Heiligenberg 1987; Theunissen and Miller 1991), as well as common sense, suggest that given a population of coarsely-coded receptors, the amount of information they can provide about the value of a particular parameter depends on the number of receptors having distinct response functions, how broadly-tuned the individual receptors are to that parameter, the positions of their response maxima within the parametric range of interest, and the stochastic properties of individual receptors. The need to control costs, both in biological and technological systems, dictates that the number of



Color-coding mechanisms at the photoreceptor level, illustrating the necessity for more than one visual pigment. **A.** A single pigment (G, for maximum sensitivity in green) gives a receptor different sensitivities to different wavelengths (λ), but the receptor could not distinguish between objects reflecting wavelengths of 450 nm and 600 nm (which have identical sensitivities). If the luminosity is decreased (dashed line), the receptor could not distinguish between the change in luminosity and a change in wavelength (arrows). Sample recordings in the G receptor under these conditions are shown below. **B.** A three-pigment system can distinguish wavelength independently of intensity. The pigments must have overlapping spectra. The two objects stimulate the three photoreceptors (B, blue; R, red) in different amounts. Each object stimulates the receptors to different degrees, so that the color code for each object is unique, and maintained despite a reduction in luminosity (dashed lines in recordings). (Modified from Gouras, 1985)

Fig. 3. Reprinted from Shepherd (1994), p. 369.

receptors be minimized. For a given number of receptors or sensors, the response functions should be neither too broad nor too narrow.

Simulations based on information theoretic analyses of data from the cricket mechanoreceptive system (Theunissen and Miller 1991) suggest that maximal transfer of directional information from four hypothetical interneurons having cosine-shaped directional tuning curves corresponds to equal (90^0) spacing of the curves, with widths at half-maximum of 110^0. These results, which are based on responses of the receptor neurons to steady-state air motion, also suggest that this system of only four interneurons is capable of resolving directions that differ by as few as 5 to 8 degrees. Under more natural conditions, where stimulus velocity and direction are not constant, receptor responses may be much less variable (Steveninck et al., 1997), permitting even greater angular resolution. In a man-made circuit design, a few additional inexpensive, coarsely coding sensors could be included in order to yield still finer resolution of the parameter(s) of interest.

The analysis of Theunissen and Miller (1991) specifically avoided assumptions about the actual neuronal mechanisms that decode stimulus direction. An alternative approach is to consider specific decoding mechanisms that could be implemented in a biomimetic design, then evaluate the effects of different receptor and processor configurations on the overall quality of information transmission. One such "mechanism" which has a long history in color vision research (Levine, 1985) uses the response amplitudes of N photoreceptors containing different visual pigments to define an N-dimensional parameter space (in this case, a color space). In this example, the collective response (r_1, r_2,...r_n) of the receptors to any spectral distribution of light intensities within the visible range of wavelengths defines a point in space. Although an identified neural circuit that corresponds to such a "parameter space" mechanism is not known to this author, in principle each coordinate could be implicitly represented by the unique pattern of activity induced in a group of interneurons receiving inputs from the sensory receptors. A second biologically plausible mechanism for discriminating among coarse-coded sensory inputs involves an additional processing stage at which comparisons are made among the activity levels of the individual sensors (e.g. Masland, 1996), for example by computing the ratio of the responses of sensors having similar response functions. As above, the information encoded in a set of ratios can be represented mathematically as a point in n-space (ratio 1, ratio 2, ... , ratio n), or neurophysiologically as a pattern of activity in a group of neurons. (Both mechanisms may be affected by changes in the directional response functions due to fluctuations in input intensities, contrasts, and spatiotemporal frequencies. Gain control mechanisms may help to minimize such changes.) Note that in engineered applications, the parameter in question can be decoded either explicitly, i.e. using digital algorithms, or represented implicitly as the activity patterns in a biomimetic circuit.

The relative merits of the raw "parameter space" vs. "ratio space" type mechanisms for discriminating among distinct coarse-coded sensory inputs were investigated in a preliminary fashion using simulations written in Pascal. Individual receptor responses were defined as cosine-shaped functions, which closely approximate the response functions of many types of directional sensors (e.g. Heiligenberg, 1987; van Hateren, 1990; Douglass and Wilkens, 1997) and are roughly comparable to the broad wavelength response functions of many photoreceptor neurons. For the "raw parameter space" - based mechanism, the ability to discriminate two sensory inputs (ignoring noise) was defined as

the geometric distance between the two coordinates specified by the responses (ra_1, ra_2, ... ra_n) and (rb_1, rb_2, ... rb_n) of **n** individual receptors to two inputs **a** and **b**:

$$D = (\Sigma(ra_i - rb_i)^2)^{1/2}$$

Similarly, for the ratio-based mechanism, the discrimination index **D** was defined as the distance between points defined by the ratios of individual receptor responses.

Figure 4 illustrates an application of these methods to evaluate the ability of a sensory processor based on four hypothetical elementary visual motion detectors (emds) to discriminate the direction of motion. Since the detailed circuitry for elementary motion detection is still unknown, this simulation begins by assuming a simple configuration with inputs from only two neighboring visual sampling units (A visual sampling unit is a single photoreceptor, or its equivalent at higher, retinotopically mapped processing levels.). For constant-speed visual motion stimuli, this configuration is assumed to result in a cosine-shaped directional response function, with a maximum at the motion direction parallel to the line defined by the two inputs, and zero response to motion directions ± 90^0 from the preferred direction. Figure 4A shows the response functions of an array of four such emds with response maxima separated by 90°. The discrimination index for the raw parameter space - based mechanism (Fig. 4B, solid curve) indicates accurate discrimination of directions that lie in between the receptor maxima, but very poor discrimination near the receptor maxima. (This is because the slopes of the curves are smallest in these regions.). The discrimination index for the ratio - based mechanism (Fig. 4B, dotted curve) complements the former mechanism to a large extent, with maximal directional discrimination near, but not at, the receptor maxima. Both mechanisms, however, suffer from poor

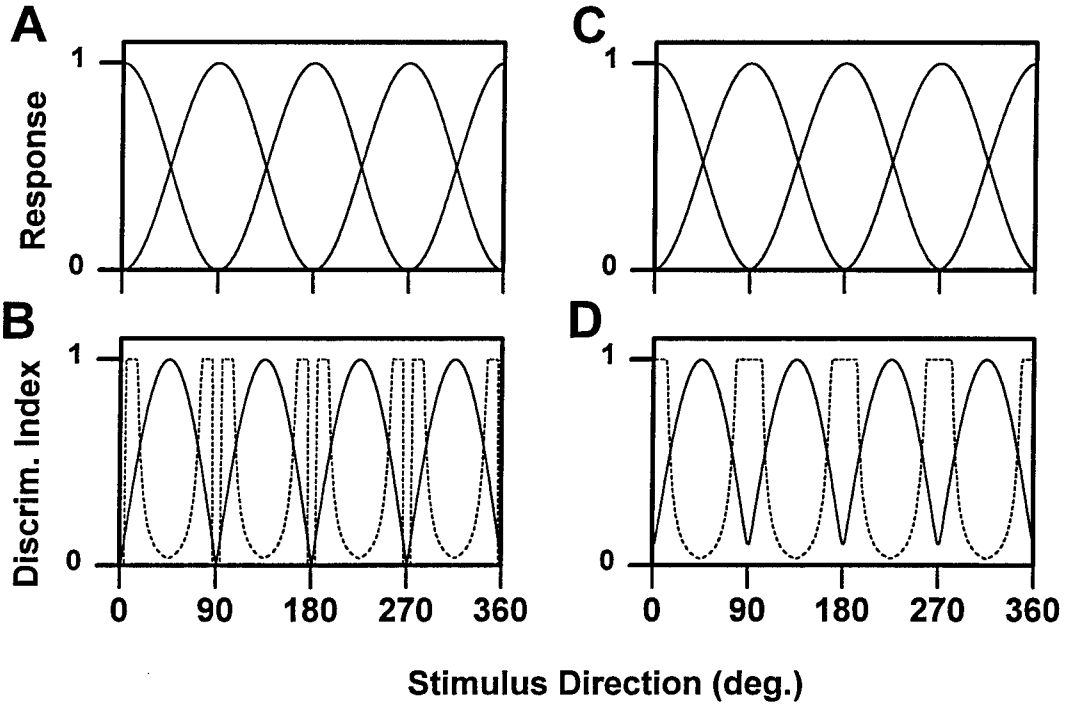


Fig. 4. Simulated responses and discrimination indices computed for four equally spaced, coarse-coded receptors having half-maximal bandwidths of 90^0 (A,B) and 93^0 (C,D). Discrimination indices (B and D) are based on differences between raw responses (solid curves) and ratios between the responses of adjacent receptors (dotted curves). See text for additional details.

discrimination at the receptor maxima. This shortcoming of the ratio mechanism is easy to eliminate, simply by broadening the receptor response functions slightly (Fig. 4C,D). In a real biological emd circuit, the actual response functions of the emds probably are at least this much wider than in Figure 4B, due to crosstalk among neighboring emds.

These results suggest two important conclusions. First, in order to optimize the parametric resolution of coarse-coded processors, it may be necessary to incorporate more than one decoding mechanism. In fact, there is evidence that color-coding in neurophysiological systems employs both of the types of mechanisms outlined above (Zeki 1978, 1980; Levine, 1985). A second, perhaps less intuitive conclusion is that in arrays of unit neuronal circuits or their biomimetic counterparts, a certain amount of "sloppiness" in the wiring can be beneficial.

Matched Filtering

Matched filtering represents a third important principle of information processing at work in the insect visual system, in addition to spatial mappings and coarse coding. The term, "matched filter," is borrowed from engineering and refers to processing mechanisms that are tailored to specific properties of a wide array of possible inputs (Wehner, 1987). Certain of the large, wide-receptive field neurons in the lobula plate of the fly brain provide excellent examples of matched filters. The identified neurons H1 and V1 are tuned to respond best to very specific panoramic flow fields. These flow fields correspond to natural motions of the fly during walking or flight, namely horizontal regressive motion (H1) and downward pitch rotation about a lateral axis (V1) (Krapp and Hengstenberg, 1996, 1997). Although the mechanisms responsible for the match to a particular flow field have not been established directly, the tuning almost certainly arises from individualized spatial mappings of synaptic inputs, from homogeneous arrays of small-field neurons that coarsely code motion direction (such as T5 cells, Douglass and Strausfeld, 1995), to each wide-field neuron (see Fig. 5, below).

Matched filters, where they exist, provide a very computationally efficient means of detecting and responding to a complex input. Thus, neurons tuned to specific flow fields associated with pitch, roll, yaw, and translatory motions can provide nearly instantaneous information on a fly's flight trajectory. Whereas engineers have struggled to devise relatively efficient algorithms to *compute* panoramic flow fields (rev. by Nelson and Aloimonos, 1988), flies are to a large extent simply pre-wired to analyze flow patterns "on the fly."

Relative Roles of Coarse and Matched Filtering

As basic paradigms for sensory processing, there appears to be a fundamental tradeoff between coarse coding and matched filtering. The concept of coarse coding (or, "coarse filtering") implies broad response functions that may span several dimensions of parameter space. In contrast, matched filters by definition are tuned to very specific and often complex features, and thus severely constrain what kind of information can be extracted from inputs. Although in some sense the distinction between these two types of filtering is simply a matter of degree, it will be crucial to understand the extent to which distinct aspects of visual information processing may be dominated by one or the other

extreme. The overall performance of the insect visual system may depend, in large part, on the successful integration of these strategies. Biomimetic designs, in turn, will benefit greatly from an understanding of the relative roles of coarse and matched filtering.

Some Proposed Applications of Spatial Maps, Matched Filters, and Coarse Coding to Missile Guidance

1. Biomimetic designs for directional motion processing

Figure 5 illustrates a fly brain - inspired design for a visual flow field processor which combines mapped arrays, coarse coding and matched filters so as to stabilize the trajectory of a low-flying missile. As in an insect brain, no explicit computation is required; all computations are implicitly coded in the filtering properties of individual channels and in the mapped circuitry. In this prototype design, the final outputs sum to control the operation of four equally-spaced side jets that compensate for yaw and pitch deviations from the intended trajectory. Additional, but similar circuitry would be needed to control rotatory motions.

In the first processing stage, standard optics form a moderately wide field of regard, forward-looking image on a photosensor array. As the main flow field information would come from below the horizon, the inputs from the upper visual field may be unnecessary. The outputs of the photosensors are mapped onto to four arrays of small-field, coarse-coded correlation-type elementary motion detectors (emds). The inputs to each emd are provided by two adjacent photosensors, wired so that all emds in an array are maximally sensitive to motion in the same direction. (In the fly brain, analogous, but unknown numbers of emd arrays are superimposed within a single mapped neuropil, the medulla.) Each emd is assumed to have a cosine-shaped directional response function. Some provision for gain control may be important at or before the emd stage, in order to prevent changes in the directional response functions of the emds due to variations in input amplitude, contrast, etc. The emd outputs are mapped onto five matched filters, each with the same field of regard as the photosensor array, but designed for maximal activation by a slightly different flow field. The center filter serves as a master controller, keeping all four side jets turned off with strong deactivation signals (open arrowheads) as long as the on-center flow field is maintained. During pitch or yaw, one to two of the four matched filters activates an appropriate side jet (filled arrowheads), and simultaneously provides contralateral deactivation signals (open arrowheads) to help the weakened signals from the central filter keep the opposite jet turned off. A completed design would require appropriate weightings of the activating and inactivating signals, as well as features designed to minimize oscillations.

2. Hyperacuity for long-range target localization and range-finding.

Hyperacuity is a phenomenon in which the ability of sensory system to discriminate small changes in the parametric value of an input is considerably greater than one would predict based solely on the parametric "spacing" between receptors. The example of precise localization of objects on the human fingertip has been noted above; the analogous spatial phenomenon in vision is known as vernier acuity. In the human fovea, the minimum separation between cone photoreceptors is equivalent to about 25 seconds of arc. Normal human observers, however, are capable of positioning the two halves of a vernier scale to an average precision of about 5 seconds (Woodhouse and Barlow,

1982). It should be noted that vernier acuity does not correspond to the ability to *resolve* two objects in a scene, but simply to judge an object's *position.* The very existence of hyperacuity implies coarsely-coded (overlapping) inputs. The neuronal mechanisms for hyperacuity presumably involve some form of interpolation among inputs (Barlow, 1979).

The potential of coarse coding for providing improved positioning accuracy appears to be far greater than the human vernier acuity example suggests. With only 3 or 4 visual pigment types to report inputs spanning over 300 nanometers, many organisms are able to distinguish spectral stimuli that differ by only a few nanometers. Long-range object-positioning devices should be designed to include a coarse-coded spatial sensor array optimized for fine positional discriminations. With such a design, it should be possible to determine the azimuth and elevation of high-contrast objects, though still far too distant to occupy more than a single pixel of a raw image, to an accuracy as high as 1/10 pixel or less. Even existing devices that employ interpolation may benefit from a design reevaluation based on optimized sensor spacing and response functions. The applications of this principle are not limited to the spatial domain; the only requirement is for comparisons between the responses of two or more coarsely-coded sensors having different peak sensitivities and overlapping response functions. Thus, for example, in a sampling frequency - based device designed to measure the times of occurrence of isolated spike-like events, a temporal resolution far superior to the sampling frequency could be obtained if the temporal window for each sample can be made to overlap with the previous and next samples. Such a coarsely-coded temporal sampling design may offer accuracy and/or cost benefits over conventional designs for echo delay - based rangefinding devices. In conclusion, the incorporation of coarse coding designs into both seeker and rangefinding equipment may significantly improve the cost/performance characteristics of various armaments, including LOCAAS and the DRM.

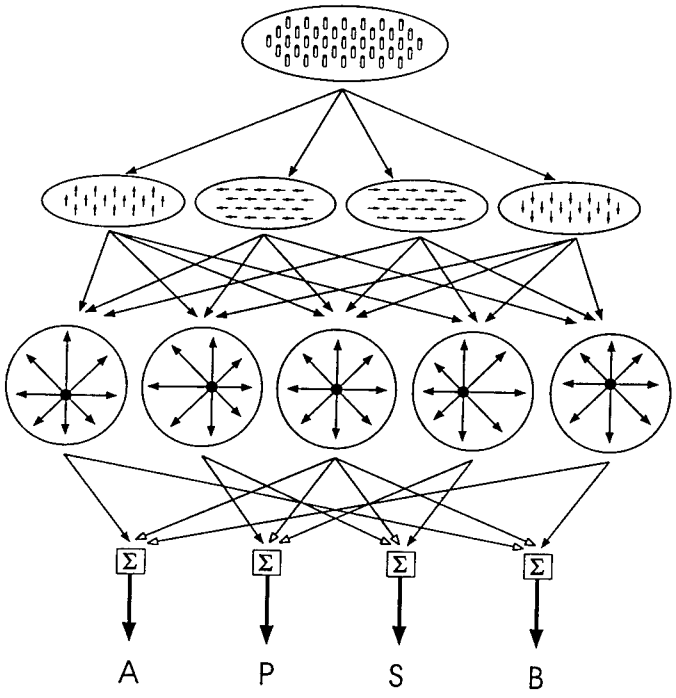


Fig. 5. Prototype design for a fly-inspired visual flow field processor, illustrating how retinotopic maps, coarse coding and matched filters can be combined in an analog design for stabilization of a low-altitude flight trajectory. Useful optic flow information is assumed to come mainly from below the horizon. Summed outputs of matched filters for flow fields activate (filled arrows) or deactivate (open arrows) one of four side jets (A, above; B, below; P, port; S, starbord) to compensate for pitch and yaw motions.

3. Polarization sensitivity

Sensitivity to linear polarization of light is well-developed in a variety of terrestrial and aquatic invertebrates, as well as some fishes, amphibians and birds. Polarization sensitivity is coarse-coded and spatially mapped at the level of the retina, suggesting the potential both for fine discriminations among polarization angles, and for discriminating spatial patterns of polarization by mechanisms similar to those responsible for color vision. What do animals actually do with polarized light? A variety of organisms use celestial or underwater polarization patterns (a result of Rayleigh scatter) as a type of sun compass for navigation or orientation to important features of the environment, with no need to directly view the location of the sun (Waterman, 1984). In addition, some flying insects can exploit the polarization inherent in specular reflections to help locate bodies of water (Schwind, 1991; Horvath & Varju, 1997). Recently, experiments on octopi have provided the first evidence that an animal can use patterns of polarized light for form vision (Shashar and Cronin, 1996).

There is much potential for exploiting polarized light for target identification and navigational purposes (i.e. as a component of backup systems in the event of Global Positioning System failure). Whereas polarization sensitivity in insects is generally restricted to the "dorsal rim" region of the compound eyes, man-made polarization imaging devices may benefit from a much wider field of regard for navigational purposes. It would also be beneficial to use more directional classes of polarization sensors than sometimes occur in nature. In the octopus retina, each photoreceptor apparently has maximal sensitivity to either vertical or horizontal polarization (see Shashar and Cronin, 1996), suggesting (Fig. 6A,B) that this animal should be unable to discriminate between polarization angles of 45^0 and 135^0 (the full range of polarization angles by any system is 180 degrees). To overcome this limitation and improve angular resolution, man-made polarization analyzers can be designed with three polarization channels, separated by 60^0 (Fig. 6C,D), instead of two channels separated by 90^0. Interestingly, despite its apparent limitation to two basic polarization channels, the octopus actually can discriminate behaviorally between **e** vectors at 45^0 and 135^0 (Moody and Parriss, 1961). The processing mechanisms that underly this ability are unknown, but presumably take advantage of irregularities in the orientations of individual receptors. To the extent that man-made sensory processing arrays can be equipped with analogous "imperfection-compatible" circuitry, relaxed manufacturing and calibration requirements could result in substantial cost savings.

As the experiments on *Octopus* have shown, polarization sensitivity can be useful for discriminating among different objects or scenes. In technological applications, polarization-based, false-colored images potentially can be used to augment the information in gray-scale or wavelength-based color images. Because man-made objects (particularly those with shiny surfaces) tend to reflect polarized light, polarization discrimination can be used to help detect and identify such objects. Do natural scenes exhibit sufficient amounts of linear polarization to be useful for imaging purposes? Several considerations suggest that they do:

Polarization is currently exploited for remote sensing (Egan and Sidran, 1994). This has obvious applications for military reconnaisance, but is sufficient polarization also present to be useful at close range in natural scenes? Studies of insect polarization sensitivity indirectly suggest an answer in the affirmative. The photoreceptors of many invertebrates are inherently polarization-sensitive, due in large part to the geometry of the

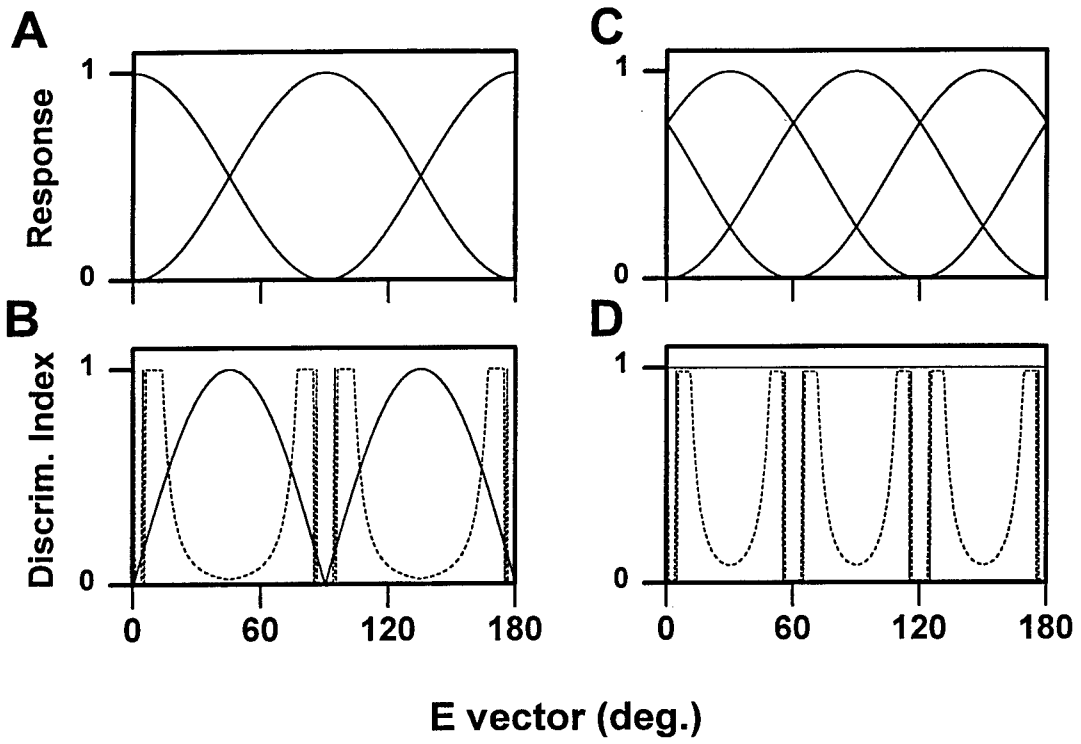


Fig. 6. Simulated responses and discrimination indices (as in Fig. 4) based on two (A and B) or three (C and D) Equally spaced, coarse-coded polarized light sensors. Discrimination indices (B,D) are derived from raw responses (solid curves) or ratios among responses (dotted lines), as in Figure 4.

cell membranes that contain the visual pigment, rhodopsin (Moody and Parriss, 1961). This presents a potential conflict with color processing channels, which could be contaminated by "false colors" in parts of a scene that are polarization-enriched. In the compound eyes of insects, however, the false color problem has been eliminated by an axial twist in the photoreceptors (with the exception of photoreceptors in the "dorsal rim" region of the compound eye, which is specialized for polarization sensitivity) (Wehner and Bernard, 1993). The presence of axial twist suggests that indeed, enough polarization is present in natural scenes to be useful in a system that is specifically designed for polarization discrimination.

The feasibility of constructing practical polarization imaging devices depends as much on the sensitivity of the imaging system as on the distribution of polarized light in the scenes of interest. Polarization discrimination is directly analogous to color discrimination, in that both involve sensitivity to variations in a physical parameter (wavelength or e-vector) of the light which emanates from a scene, and both require two or more receptor types having distinct, overlapping response profiles. Given the remarkable hue-discriminating capabilities of natural, "coarsely-coded" visual systems, it should be possible to design polarization imaging devices that can discriminate quite subtle variations in the polarization content of a scene. An image analyzing system that incorporates both color and polarization information, in turn, could provide an enhanced basis for discriminating among targets (friend from

foe, etc.), in ways that are superior to either a color- or polarization-based system alone. For example, three spectral and three e-vector imaging channels could be combined to generate a 6-dimensional representation of a scene. Pattern-recognition hardware could then be used to (a) display key attributes of the image on a conventional color monitor, or (b) provide direct inputs into an automated target-identification system.

CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

The goal of biomimetic design should not be faithfully to replicate an insect brain in silicon. Most individual technological applications will require only a subset of the complete capabilities of a brain, even a fly brain. Moreover, although natural sensory systems are optimized in many ways, they do not necessarily represent the optimal design for implementation in man-made devices. Yet, armed with a thorough understanding of biological solutions, humans will design new types of sensors or processors that have yet to evolve in nature (or at least to be discovered by scientists). Additional neurophysiological research will add to our understanding of the fly's brain, and will thereby augment its value as a paradigm for smart guided munitions. A major issue to be addressed in this research is the relative roles of matched filtering and coarse coding. To the extent that the interplay between these two strategies can be understood, they can be translated more appropriately into simple, yet computationally powerful chip designs that are both cost-effective and far superior to current technology.

ACKNOWLEDGEMENTS

This work would not have been possible without the assistance of Frances Chambers and Cheryl Mack at the Eglin AFB Technical Library. Stimulating discussions with Geoffrey Brooks, Pat Coffield, Chuck Conklin, Bill Eardley, Johnnie Evers, Dennis Goldstein, Paul McCarley, Major Jim Stright, and Ric Wehling were also essential.

REFERENCES

Anonymous (1997) FY 97 Conventional Armament Technology Area Plan Published by: Headquarters Air Force Materiel Command, Directorate of Science & Technology, Wright-Patterson AFB, Ohio.

Anonymous (no date) New World Vistas Air and space power for the 21st Century Human systems and biotechnology volume. USAF Scientific Advisory Board.

Barlow HB (1979) Vernier acuity and interpolation. Nature 279:189.

Buchner E, Buchner S and Bülthoff I (1984) Deoxyglucose mapping of nervous activity induced in *Drosophila* brain by visual movement. J. Comp. Physiol. A *155*:471-483.

Cajal, S. R. and D. Sanchez. (1915) Contribucion al conocimiento de los centros nerviosos de los insectos. Trab. Lab. Invest. Biol. *13*:1-167.

Collett TS, Land MF (1978) How hoverflies compute interception courses. J Comp Physiol 125:191-204.

Douglass JK and Strausfeld NJ (1995) Visual motion detection circuits in flies: Peripheral motion computation by identified small-field retinotopic neurons. J. Neurosci. 15:5596-5611.

Douglass JK and Strausfeld NJ (1996) Visual motion-detection circuits in flies: Parallel direction- and non-direction-sensitive pathways between the medulla and lobula plate. J. Neurosci. *16*:4551-4562.

Douglass JK and Wilkens LA (1997) Directional selectivities of near-field filiform hair mechanoreceptors on the crayfish tailfan (Crustacea: Decapoda). submitted to J of Comparative Physiology.

Egan WG, Sidran M (1994) Polarimetric detection of land sediment runoff into the ocean using space shuttle imagery. Applied Optics 33:8117-8119.

Egelhaaf M, Hausen K, Reichardt W, Wehrhahn C (1988) Visual course control in flies relies on neuronal computation of object and background motion . Trends Neurosci 11:351-358.

Franceschini N, Pichon JM and Blanes C (1992) From insect vision to robot vision. Philosophical Transactions of Royal Society of London B 337:283-294.

Freeman WJ (1990) Searching for signal and noise in the chaos of brain waves. In: S Krasner (ed) The Ubiquity of Chaos. AAAS, Washington, DC. pp. 47-55.

Hateren JH van (1990) Directional tuning curves, elementary movement detectors and the estimation of the direction of visual movement. Vision Research 30, 603-614.

Heiligenberg W (1987) Central processing of sensory information in electric fish. Journal of Comparative Physiology [A] 161: 621-631.

Horridge GA (1992) What can engineers learn from insect vision? Philosophical Transactions of Royal Society of London B 337: 271-282.

Horvath G and Varju D (1997) Polarization pattern of freshwater habitats recorded by video polarimetry in red, green and blue spectral ranges and its relevance for water detection by insects. Journal of Experimental Biology 200(7):1155-1163.

Koch C and Li H (1995) (eds) Vision Chips: Implementing vision algorithms with analog VLSI circuits. IEEE Computer Society Press, Los Alamitos, CA.

Konishi M (1986) Centrally synthesized maps of sensory space. Trends in Neuroscience (April) 163-168.

Konishi M (1993) Similar neural algorithms in owls and electric fish. Journal of Comparative Physiology [A] 173:700-702.

Krapp HG and Hengstenberg R (1996) Estimation of self-motion by optic flow processing in single visual interneurons. Nature *384*:463-466.

Krapp HG andHengstenberg R (1997) A fast stimulus procedure to determine local receptive field properties of motion-sensitive visual interneurons. Vision Res. *37*:225-234.

Levine (1985) Vision in man and machine. New York: McGraw Hill.

Masland, RH (1996) Unscrambling color vision. Science 271, 616-617.

Massie MA, Curzan JP and McCarley PL (1994) A neuromorphic sensor with retinal capabilities. Sensors, Sept 1994: 37-40.

Mead C (1989) Analog VLSI and Neural Systems. Addison-Wesley, Reading, MA.

Moody MF and Parriss JR (1961) Discrimination of polarized light by *Octopus*: a behavioral and morphological study. Z vergl Physiol 44:268-291.

Nelson, R.C. and J. Aloimonos (1988) Finding motion parameters from spherical motion fields (or the advantages of having eyes in the back of your head). Biological Cybernetics 58: 261-273.

Ramachandran VS, and RL Gregory (1978) Does colour provide an input to human motion perception? Nature 275:55-56.

Schwind R (1991) Polarization vision in water insects and insects living on moist substrate. J comp Physiol A 169: 531-540.

Shashar N and Cronin TW (1996) Polarization contrast vision in *Octopus*. Journal of Experimental Biology 199: 999-1004.

Shepherd, GM (1994) Neurobiology (Third Edition) New York, Oxford: Oxford University Press.

Snyder AW (1975) Photoreceptor optics - Theoretical principles. In: Snyder AW, Menzel R (eds), Photoreceptor Optics. Berlin, Heidelberg, New York: Springer. pp. 38-55.

Steveninck R de Ruyter van, Lewen GD, Strong SP, Koberle R, Bialek W (1997) Reproducibility and variability in neural spike trains. Science 275:1805-1808.

Strausfeld NJ (1976) Atlas of an insect brain. Berlin: Springer.

Strausfeld, N. J. (1989) Beneath the compound eye: Neuroanatomical analysis and physiological correlates in the study of insect vision. In: Facets of vision (Stavenga DG, Hardie RC, eds), pp 317-359. Heidelberg: Springer.

Strausfeld NJ and Lee J-K (1991) Neuronal basis for parallel visual processing in the fly. Visual Neurosci. 7:13-33.

Theunissen FE and Miller JP (1991) Representation of sensory information in the cricket cercal sensory system. II. Information theoretic calculation of system accuracy and optimal tuning-curve widths of four primary interneurons. Journal of Neurophysiology 66: 1690-1703.

Tonkin SP and RB Pinter (1996) Motion processing using asymmetric shunting lateral inhibitory networks. Network- Computation in Neural Systems 7:385-407.

Villasenor J and Mangione-Smith WH(1997) Configurable computing. Sci Am 276(6):66-71.

Wagner H (1986) Flight performance and visual control of flight of the free-flying housefly (*Musca domestica* L.) II. Pursuit of targets. Phil Trans R Soc Lond B 312:553-579.

Waterman TH (1984) Natural polarized light and vision. In: MA Ali (ed), Photoreception and vision in invertebrates. New York: Plenum Press, pp 63-114.

Wehner R (1987) "Matched Filters" - neural models of the external world. J Comp Physiol A 161:511-531.

Wehner R and Bernard GD (1993) Photoreceptor twist: A solution to the false color problem. Proceedings of the National Academy of Science (USA) 90:4132-4135.

Wheat NE, Goodwin AW, Browning AS (1995) Tactile resolution - peripheral neural mechanisms underlying the human capacity to determine positions of objects contacting the fingerpad. J. Neurosci 15:5582-5595.

Woodhouse JM, Barlow HB (1982) Spatial and temporal resolution and analysis. In: HB Barlow, JD Mollon (eds), The senses. Cambridge Texts in the Physiological Sciences, vol. 3. Cambridge, England: Cambridge University Press. pp. 133-164.

Zeki SM (1978) Functional specialization in the visual cortex of the rhesus monkey. Nature 274:423-428.

Zeki SM (1980) The representation of colours in the cerebral cortex. Nature 284:412-418.

MODELING THE CHARGE REDISTRIBUTION ASSOCIATED WITH DEFORMATION AND FRACTURE

M.E. Eberhart
Associate Professor
Department of Metallurgical and Materials Engineering

Colorado School of Mines
Golden, CO 80401

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

August 1997

MODELING THE CHARGE REDISTRIBUTION ASSOCIATED WITH DEFORMATION AND FRACTURE

M.E. Eberhart
Associate Professor
Departments of Metallurgical and Materials Engineering
Colorado School of Mines

ABSTRACT

Attempts to correlate the bonding in crystalline materials with mechanical properties often employ quantum mechanically determined charge density difference maps. The interpretations which result from the use of these maps lack rigor, are ambiguous, and often provide contradictory rationale as to the origins of specific properties. Based on these results, some have suggested that first principle calculations may have little to offer in the development of a more fundamental understanding of the mechanical properties of metals and alloys. Here, we build on the work of Bader and show that the geometric properties of the total charge density at its Morse points provide a concise and unambiguous description of crystalline bonding and the change in this bonding associated with deformation and fracture. The description which results is used to account for the previously unexplained trends in mechanical properties of the B2 aluminides of Fe, Co, and Ni.

MODELING THE CHARGE REDISTRIBUTION ASSOCIATED WITH DEFORMATION AND FRACTURE

M.E. Eberhart
Associate Professor
Departments of Metallurgical and Materials Engineering
Colorado School of Mines

Introduction

As part of an attempt to accelerate the pace at which high temperature intermetallic alloys are developed, increasing attention has been directed toward providing a more fundamental understanding of the relationships between mechanical properties and atomic scale structure. Among those properties which are believed to be controlled at the atomic scale are those influencing the competition between fracture and deformation. Both these processes are forms of materials failure but are distinguished: fracture, without prior deformation, is termed brittle; while fracture following deformation is termed ductile. As the consequences of ductile failure are often less catastrophic than those of brittle failure, there is a preference to design with ductile materials. Unfortunately, most of the candidate intermetallic alloys considered as replacements for the current generation of high temperature Ni-based superalloys, though potentially offering superior thermal properties, are intrinsically brittle. The obvious question arises as to whether or not desirable thermal properties are a consequence of undesirable failure properties. If so, then we must seek a design solution, learning how to better design with brittle materials. If not, then we may seek a materials solution, a ductile material with desirable thermal properties. Before this question can be answered we must first understand the origins of the intrinsic thermal and mechanical properties of an alloy.

Of the many intrinsic thermal and mechanical properties of metals for which more fundamental understandings are being sought, the competition between ductile and brittle

failure seems particularly tractable. This competition is thought to be determined by the potential surface on which the atoms at the tip of an atomically sharp crack move in response to a stress. Modern electronic structure techniques now permit one to calculate the total energy at specific points of this potential surface for a small number of crack tip geometries. Because it is not yet possible to calculate, strictly from first principles, the potential surface for the movement of atoms at a generalized crack tip, researchers have rationalized the competition between ductile and brittle behavior in terms of total energy differences between configurations and on features of the "bonding", inferred from the quantum mechanically determined charge density[1-3]. Unfortunately, investigators using the same electronic structure techniques to study the identical system, and presumably generating the same charge densities, are coming to contrary explanations as to the origins of brittle versus ductile behavior[1,2]. These opposing explanations have led some to conclude that quantum mechanical techniques may have nothing to offer toward a more fundamental understanding of mechanical behavior of metals.[3]

A case in point is provided by the transition metal aluminides with the B2 structure. Of these, NiAl is of most interest as a potential high temperature material. This alloy fails by undesirable brittle mechanisms and consequently is one of the systems used as a model to investigate the atomic scale interactions responsible for brittle failure. Of the many electronic structure calculations used to explore the mechanical properties of NiAl the systematic work of Schultz and Davenport[3] is of particular note. In this study they investigated the electronic structure and bonding of FeAl, CoAl, and NiAl. These intermetallics, though sharing the same structure (B2) and having similar elastic constants, melting points, and lattice constants have very different mechanical behavior. FeAl is more ductile than NiAl, which is more ductile than CoAl. Additionally, the cleavage and slip properties change through the series, with FeAl showing $\{110\}\langle 111\rangle$ slip and $\{100\}$ cleavage planes, while CoAl and NiAl show $\{110\}\langle 100\rangle$ slip and $\{110\}$ cleavage planes.

The calculations of Schultz et al. were motivated by the studies of other investigators who had come to strikingly contrary rationale as to the origins of the deformation and cleavage properties of NiAl. As examples, Hong and Freeman[1] using full-potential linear-APW methods concluded that directional bonding was responsible for NiAl's brittleness. Using the same technique, Fu[2] concluded that strong directional bonding was the source of FeAl's resistance to brittle failure and the brittle behavior of NiAl was due to reduced directional bonding. Schultz et al. used a full potential LASTO approach to compare charge densities through the series FeAl, CoAl, NiAl. It was reasoned that because the transition between ductile and brittle behavior occurs most dramatically between FeAl and CoAl, there should also be a corresponding change in the character of the charge density between these two alloys. As with the calculations of Hong and Freeman; as well as Fu, directional bonding was gauged in terms of qualitative and semi-quantitative features of the charge density and charge density differences. From these comparisons, Schultz et al. concluded that there was no discernible difference in the bonding which could account for the trends in mechanical properties. The bonding of CoAl was judged to be intermediate to that of FeAl and NiAl. They summarized by saying, "Taken together, these results paint a rather bleak picture of first principle theory to contribute to the alloy development process. Ideally the goal of theory in this enterprise is to obtain a more fundamental understanding of the microscopic properties that lead to the macroscopic behavior and yet none of the quantities accessible to theory appear to correlate with the observed behavior of technological interest."

Before we write an epitaph for twenty years of research using quantum mechanical methods to investigate the atomistic origins of mechanical behavior, perhaps we should take a closer look at how we interpret the results of our calculations. Of particular concern is the way in which we extract information regarding the "bonding". Invariably this information comes from charge density differences, where for example, the superimposed

charge density of isolated atoms occupying crystal sites is subtracted from the calculated charge density for that crystal. The resultant density-difference gives information has to how charge is redistributed when a crystal is formed from its elements. While this may be a useful heuristic devise, it is doubtful that these density difference maps provide any information about the properties of the crystal. The Kohn-Shamm theory tells us that the total energy is a functional of the charge density, and hence all molecular or solid state properties should in principle be derivable from the density alone. Charge density differences are not a functional of either the energy of the atomic system, the energy of the crystal, or the energy difference between the two, i.e. the heat of formation. In the spirit of the Kohn-Shamm theorem a description of the bonding should be based on the density, not density differences.

In what follows we briefly review a quantifiable description of the bonding based on the total charge density. This description is used to compare the bonding through the series FeAl, CoAl, and NiAl. It will be shown that this description provides for a consistent explanation for the observed changes in mechanical behaviors of these alloys. Our description of the bonding builds on the theory of Bader[4–9] describing molecular and solid state electronic structure. Bader's theory, in turn, draws part of its validity from Morse theory which allows for the assigning of a topology to every scalar field. As the charge density, $\rho(\vec{r})$, is a scalar field, it must have a topology.

The Topology of the Charge Density

The topology of a scalar field is determined by the position and type of its critical points, which are the zeroes of the gradient of this scalar field. There are four kinds of critical points in a three dimensional space: a local minimum, a local maximum, and two kinds of saddle points. These critical points (cps) are denoted by an index which is the number of positive curvatures minus the number of negative curvatures; for example, a

minimum cp has positive curvature in the three orthogonal directions, therefore it is called a (3, 3) cp, where the first number is simply the number of dimensions of the space, and the second number is the net number of positive curvatures. A maximum would be denoted by (3, -3), since all three curvatures are negative. A saddle point with two of the three curvatures negative is denoted (3, -1), while the other saddle point is a (3, 1) cp.

For every scalar field one can construct a system of space filling polyhedra such that there is a homeomorphism between the critical points of this field and the corners, edges, and faces of the packed polyhedra. Under this homeomorphism (3, -3) cps are mapped one-to-one onto polyhedral corners, the path connecting two (3, -3) cps which is a maximum with respect to every neighboring path must pass through a (3, -1) cp and maps one-to-one onto polyhedral edges, the smallest ring of such paths originating and returning to the same (3, -3) cp must necessarily define a surface which contains a (3, 1) cp. These critical points map one-to-one onto polyhedral faces. Finally the smallest volumes which can be constructed from the union of the (3, 1) defined surfaces must contain a (3, 3) cp. These can be placed in one-to-one correspondence with the polyhedra filling space. The scalar field and the system of packed polyhedra are topologically equivalent under this homeomorphism.

Bader realized that the topology of $\rho(\vec{r})$, describes much of what is consider chemical structure and bonding. A bond path is seen as the ridge of maximum charge density connecting two nuclei and passing through a (3, -1) cp. The charge density along such a path must be a maximum with respect to any neighboring path. Because a (3, -1) cp is both a necessary and sufficient condition for the existence of a bond path, this critical point is sometimes referred to as a bond critical point. Other types of critical points have been correlated with other features of molecular structure. A (3, 1) cp is required at the center of ring structures like benzene. Accordingly, this critical point has been designated as a ring critical point. Cage structures are always characterized by a single (3, 3) cp

somewhere within the cage and again have been given the descriptive name of cage critical points.

Both fracture and deformation are processes which involve changes in topological structure, as such, the model of Bader is ideally suited to describe the change in bonding accompanying these processes. For example, fracture must be accompanied by a topological change including the disappearance of polyhedral edges and the appearance of ring cps. We recognize this process as bond breaking and surface formation. In electronic terms, it must be associated with the disappearance of (3, -1) cps. However, topology only constrains the type of transformation allowed, it says nothing about how susceptible a particular bond is to breaking, how near a (3, -1) cp is to instability. This can only be ascertained by providing a metric which measures the "distance" between two charge densities.

We have constructed such a metric which is discussed in detail elsewhere.[10,11] Briefly, the quantity of charge at a cage critical point, i.e. the charge density at this point, must be less than that at a ring cp of this cage. In turn the charge density at this ring cp must be less than that at a bond cp of this ring, which must be less than that at an atom or pseudo atom cp at the ends of the bond path. Hence a topological transformation, which must involve changes in the character of critical points, will be accompanied by a predictable redistribution of charge. With some critical points gaining charge density and some losing density. The charge density metric allows one to estimate the distance between two charge densities by providing information regarding the amount of charge which must be lost or gained in order to change the character of a critical point. This distance then is related to the magnitude of the perturbation necessary to produce the specified change in the charge density.

The distance between two charge densities is given in terms of the quantities which determine the geometric properties of the charge density in the neighborhood of a critical point. Note that the character of a critical point is determined only by the curvatures

of the charge density at this point; that is, the Hessian of the charge density, $\mathcal{H}_{ij}\rho(\vec{r}) = \frac{\partial^2\rho(\vec{r})}{\partial x_i \partial x_j}$. This tensor has the same transformation properties as the coefficients of a quadratic polynomial, often referred to as the representation quadric. When expressed in a diagonal basis, the representation quadrics can be described by the equation

$$\rho_{11}x_1^2 + \rho_{22}x_2^2 + \rho_{33}x_3^2 = 1 \tag{1}$$

where x_1, x_2 and x_3 are the eigenvectors and ρ_{11}, ρ_{22}, and ρ_{33} are the curvatures of the charge density in these directions i.e. the eigenvalues or principal curvatures of the charge density.

The type of critical point described by equation (1) is determined only by the signs of the principal curvatures. At a cage critical point, for example, all of the eigenvalues of $\mathcal{H}_{ij}\rho(\vec{r})$ are positive. Therefore its quadric corresponds to an ellipsoid. The quadric of a ring critical point (two positive and one negative eigenvalue) is an hyperboloid of one sheet with the axis of the hyperboloid normal to the ring. The quadric of a bond critical point corresponds to an hyperboloid of two sheets. Here the axis of the quadric is parallel to the bond path. Finally the representation quadric for a maximum, a (3, 3) cp (an atom or pseudo-atom), corresponds to a negative ellipsoid, which has the same geometric properties as an ellipsoid.

The shape of the quadric can be specified to within a scale factor by the ratios of the principal curvatures. Thus some function of these ratios may serve as a charge density metric. We have found that the charge density at a critical point times the square root of the ratio of the appropriate principal curvatures correlates well with the amount of charge redistribution required to produce a specified change in the character of a critical point.[10] An intuitive understanding as to why this particular function should correlate with charge redistribution is seen in its geometrical significance. At a bond critical point, the square root of the ratio of one of the eigenvalues perpendicular to the bond path, to the eigenvalue

parallel to the bond path, is the tangent of the angle between the perpendicular eigenvector and the asymptotic surface (an elliptic cone) which bounds the representation quadric. The directions of zero curvature through the critical point lie in this cone. As charge is removed from this critical point, the cone becomes more obtuse, finally becoming a disk when the perpendicular eigenvalues vanish. The perpendicular eigenvector is then contained in the cone of zero curvature. Thus the inherent stability of a bond can be visualized in terms of the angles between the directions of principal curvature and asymptotic elliptic cone defining the directions of zero curvature.

Applications to the B2 Aluminides of Fe, Co, and Ni

With a knowledge of the crystalline charge density and a way to asses the distance between two densities, it now becomes possible to compare the amount of charge redistribution associated with deformation or fracture for different alloys with the same topology. Using a full potential Linear Augmented Slater-type Orbital (LASTO) electronic structure code,[12] the charge density of B2 FeAl, CoAl, and NiAl have been determined. The B2 structure is a simple cubic structure with basis atoms at the origin and at the body center. In this paper we will consider the transition metal as being located at the origin (cube corners) and the aluminum atom as being located at the body center. All of these aluminides share the same topology, with bond critical points located between each pair of first neighbor aluminum-transition metal atoms, not surprisingly indicating that these atoms are bound. However, there are also bond critical points centered on the cube edges, indicative of second neighbor transition metal-transition metal bonds. In the center of each cube face the charge density achieves a minimum, giving rise to cage critical points at these locations. The cage critical points characterize the polyhedra whose corners are represented by four transition metal atoms and two aluminum atoms. There are of course ring critical points in each of the eight faces of these polyhedra.

Figure 1 shows the network of bonds which characterize all of the aluminides discussed in this paper. Additionally, the representation quadrics of the aluminum-transition metal bond, and the cage critical point are shown with the angles needed to compare the charge redistribution of the three aluminides designated. We will consider the evolution of the charge density accompanying {110} ⟨111⟩ slip in an attempt to account for: i) the extreme brittle behavior of CoAl, ii) the existence of ⟨111⟩ slip in FeAl and it absence in CoAl and NiAl, and iii) the {110} cleavage plane of CoAl.

During ⟨111⟩ slip a semi-infinite slab containing the atoms designated 3, 4, 5, 6, 7, 8, and α will shift along ⟨111⟩ directions relative to the plane containing atoms 1, 2, γ, and β, so in the example, atom 3 moves to the position of atom α while α moves to the position of atom 5, etc., creating an antiphase boundary (APB). During ⟨111⟩ slip to form an APB, per formula unit, at most four aluminum-transition metal bonds will break. In the Figure these are shown as the 1-to-α and 2-to-α, 5-to-β, and 5-to-γ bonds. Also four new first neighbor bonds may form in the process, these would be the 3-to-1, 3-to-2, α-to-β, and α-to-γ bonds.

There are two outcomes to the slip process. If bond breaking occurs earlier in the slip than bond formation, the result will be cleavage along {110} planes. On the other hand, if bond formation occurs earlier, or at a comparable rate to bond breaking, slip, with the formation of an APB, will occur. The competition between bond breaking and bond formation controls the point along the slip of the Peierls barrier (the activation energy for slip). In turn, the greater the atomic displacements at the barrier, the larger the Peierls energy. We may assume that it is the making and breaking of first neighbor bonds which will dominate the position, and hence height, of the Peierls barrier. Of these first neighbor bonds, those in which the internuclear distance is increasing more rapidly, will be the first to break, while those which are closest throughout the slip, will be the first to form. In the Figure, the first bonds to break will be 2-to-α and 5-to-β bonds. The first to form will

be the 2-to-3 and α-to-β bonds.

The breaking of bonds can not proceed independently from the formation of bonds. The bond breaking must be accompanied by a loss of charge density from the bond cp, while bond formation is the result of charge accumulation at the forming bond cp. It is the redistribution of charge from breaking to forming bonds which drives the process. The qualitative differences in deformation properties of the three aluminides being studied can be explained in terms of this charge redistribution.

Consider first the bond breaking process. The tangent of the angle designated θ in Figure 1 is a measure of the amount of charge which must be transferred from the first neighbor bond critical point to cause bond breaking. As θ approaches zero there will be a flow of charge density from the critical point, causing the direction of zero curvature to coincide with the perpendicular curvatures of the Hessian of the charge density. At this point, by definition, the bond is broken. Our calculations reveal that the value of $\tan(\theta)$ for FeAl, CoAl, and NiAl are respectively 1.63, 1.37, and 1.43. Thus the first neighbor bonds in both CoAl and NiAl require smaller transfers of charge density from the bond critical point to induce the transition, with CoAl requiring the least.

The charge density lost from the breaking bonds will be accumulated in the critical points between the bonds being formed, in the example there will be a flow of electron density to the bond critical point between atoms 2 and 3 and the cage critical point between atoms α and β. The flow of charge to the bond critical point will not induce a topological transformation, as this bond already exists in all three aluminides. However, the flow of density to the cage critical point can be a component of a number of transformations. If a bond is to form between α and β, the curvature of the charge density perpendicular to the α-β axis must go from positive to negative and the angle designated as ϕ in Figure 1 must vanish at the transition point. On the other hand, a topological transition will result if the curvature along the α-β axis vanishes. In this case the angle designated ψ in Figure 1

will vanish. The loss of a cage critical point without the formation of a bond is topological allowed through the formation of a free surface, i.e. through fracture. The values of $\tan(\phi)$ for FeAl, CoAl, and NiAl are respectively 0.50, 0.68, and 0.51. While the values of $\tan(\psi)$ for FeAl, CoAl, and NiAl are respectively 1.97, 1.47 and 1.96.

It can be seen that FeAl and CoAl are extremes in terms of the charge transfer necessary to induce topological transformations. FeAl requires the greatest charge redistribution to break the Fe-Al bond while requiring the least to form an Al-Al bond. FeAl also requires the greatest charge redistribution to form a free surface. One can conclude that during $\langle 111 \rangle$ slip bonds will be formed early, resulting in an early Peierls barrier with a concomitant low energy. The competing process of free surface formation can not be realized, as the charge redistribution necessary for this process is accommodated in the forming bonds. CoAl, on the other hand, requires the least charge redistribution to break the Co-Al bond, the greatest to form Al-Al bonds, and the least to form a free surface. During $\langle 111 \rangle$ slip bond formation would happen very late in the slip, leading to a large Peierls energy. However before bond formation occurs, the charge lost from the bond critical points across the $\{110\}$ planes is accumulated in the cage critical point inducing a topological transformation resulting in free surface. In short, the amount of charge which must be redistributed in CoAl to produce free surface is less than required to produce an APB. In FeAl the opposite is the case. Consistent with all observations, NiAl has charge redistribution properties intermediate to FeAl and CoAl.

It would appear that if the bonding in a material is seen to be a consequence of the geometric features of the total charge density, the prospect for gaining a more fundamental understanding of the electronic origins of mechanical properties is quite bright.

References

1. T. Hong and A.J. Freeman, *Phys. Rev. B*, **43** 6446 (1991).

2. M.H. Yoo and C.L. Fu, *Scripta Metall.*, **39**, 669 (1991).

3. P.A. Schultz and J.W. Davenport, *J. of Alloys and Compounds*, **197**, 229 (1993).

4. R.F.W. Bader and H.J.T. Preston, *Int. J. Quantum Chemistry*,**3**, 327 (1969).

5. R.F.W. Bader, P.M. Beddall and J. Peslak, Jr., *J. Chem. Phys.*, **28**, 557 (1973).

6. G.R. Runtz, R.F.W. Bader and R.R. Messer, *Can. J. Chem.*, **55**, 3040 (1977).

7. R.F.W. Bader, T.T. Nguyen-Dang and Y. Tal, *Rep. Prog. Phys.*, **44**, 893 (1981).

8. R.F.W. Bader and P.J. MacDougall, *J. Am. Chem. Soc.*, **107**, 6788 (1985).

9. P.F Zou, and R.F.W. Bader, *Acta Crystallographica A*, **50**, (1994)

10. M.E. Eberhart, *Acta. Meter.*, **44**, 2495 (1996).

11. M.E. Eberhart, *Phil. Mag. A*, (1996)

12. J.W. Davenport, *Phys. Rev.* **B29**, 2896 (1984).

Figure 1

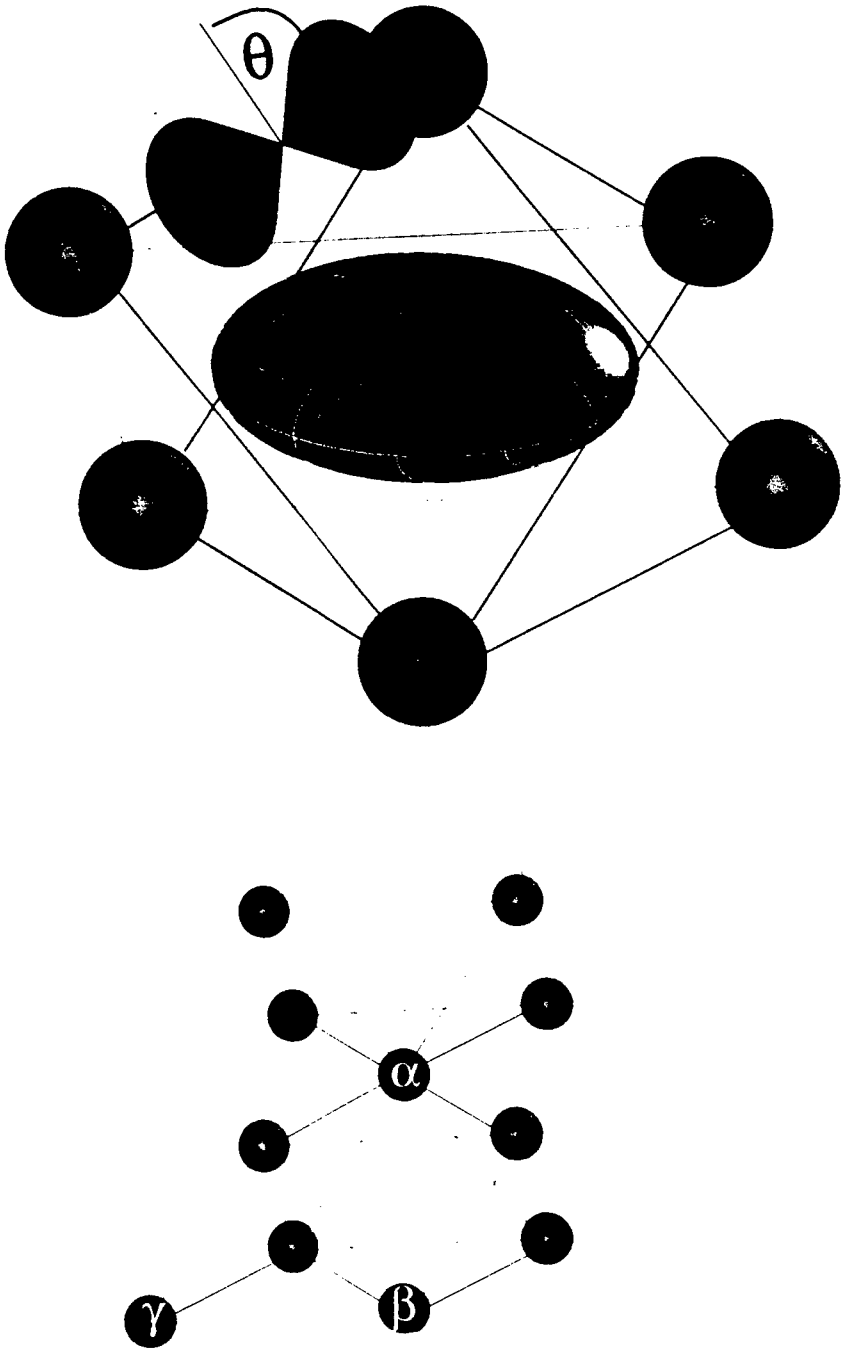


Figure Captions

Figure 1. Top, the bonding polyhedral for the B2 aluminides also showing the representation quadric of the cage critical point as well as one of the bond critical points between aluminum and transition metal atoms. The disappearance of the angles marked will result in various kinds of topological instabilities. Bottom, the designations given to the atoms to help explain the bond breaking and formation which will occur during $\langle 111 \rangle$ slip.

ON THE DEVELOPMENT OF PLANAR DOPPLER VELOCIMETRY

Gregory S. Elliott
Assistant Professor
Department of Mechanical and Aerospace Engineering

Rutgers University
Brett and Bowser Roads, P.O. Box 909
Piscataway, NJ 08855-0909

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

and

Wright Laboratory

September 1997

ON THE DEVELOPMENT OF PLANAR DOPPLER VELOCIMETRY

Gregory S. Elliott
Assistant Professor
Department of Mechanical and Aerospace Engineering
Rutgers University

ABSTRACT

A molecular filtered based laser diagnostic technique termed Planar Doppler Velocimetry (PDV) has been developed to measure the velocity field in a two dimensional image. Using an injection seeded Nd:YAG laser, the beam is expanded into a sheet to illuminate the flow under study. Two cameras are used to record the image. The filtered camera has an iodine cell in front of it which discriminates the intensity with respect to frequency. A second camera, which is not filtered, is used as a reference to eliminate variations due to seeding or laser fluctuations. From these two images the transmission ratio is calculated and used to obtain the velocity of the flow field. The PDV system and programs have been created to make the technique more user friendly and reduce the uncertainties in the measurement as much as possible. Several different flow fields were investigated. A small axisymmetric jet facility was used to test the experimental arrangement, data collection and reduction programs, and investigate sources of error (laser speckle, misalignment, laser frequency fluctuations, split image arrangements, etc.). The uncertainty in the measured velocity was found to be less than 4 m/s. PDV was then incorporated in "large scale" wind tunnel facilities studying supersonic (injection into a supersonic free stream) and subsonic (the interaction of a leading-edge vortex and a vertical tail) flows.

ON THE DEVELOPMENT OF PLANAR DOPPLER VELOCIMETRY

Gregory S. Elliott

INTRODUCTION

Recently, techniques employing molecular filters to modify the light scattered from particles or molecules in the flow field have been tested for measuring single or multiple properties in a flow field simultaneously. The molecular filter is simply a cylindrical optical cell which contains a molecule that has absorption lines within the frequency tuning range of the interrogation laser. The molecular filter is placed in front of the receiving optics to modify the recorded intensity based on the frequency of the scattered light. Miles et al.[1992] introduced the use of molecular iodine filters to flow diagnostics, a technique that they called filtered Rayleigh scattering (FRS). By using an injection seeded frequency-doubled Nd:YAG laser (λ = 532 nm) the linewidth is narrow enough (100 MHz) and can be tuned through some of the iodine absorption lines. Using the FRS technique for background suppression, the laser is tuned so that unwanted scattering from walls and windows is absorbed while the Doppler shifted Rayleigh (or Mie) scattering from molecules or particles in the flow field is shifted outside the absorption profile. The magnitude of the Doppler shift is given by

$$\Delta f_D = \frac{1}{\lambda}(\underline{k}_s - \underline{k}_0)\cdot\underline{V} \qquad (1)$$

where k_s and k_0 are the observed and incident unit light wave vectors, respectively, V is the flow velocity vector, and λ is the wavelength of the incident light. The FRS flow visualization technique has been used by Elliott et al.[1992] in the study of compressible free mixing layers.

In addition to using the technique for qualitative flow visualizations, Miles et al. [1992] showed that average properties of the flow at each point in the illuminated plane can be obtained if the scattered signal is collected from molecules. Also Elliott et al. [1997] used FRS to obtain instantaneous temperature measurements in reacting flow fields. By collecting scattering from an individual point instead of a two-dimensional plane, Elliott and Samimy [1996] were able to obtain instantaneous velocity, density, temperature, and pressure information in a technique called Filtered

Angularly Resolved Rayleigh [FARRS] scattering.

Other molecular filter-based techniques have been developed to measure the average velocity (Doppler Global Velocimetry, DGV, [Meyers and Komine, 1991]) or instantaneous velocity (Planar Doppler Velocimetry, PDV, [Elliott et al., 1994]) in a two-dimensional plane from particles in the flow field. With PDV two cameras are utilized; one camera has a pressure broadened iodine filter in front of it which produces gradual slopes in the absorption profile instead of the sharp Doppler broadened filter used by Miles et al. [1992]. A second camera with no filter (similar to DGV) is used to normalize the image from the filtered camera, accounting for variations in laser intensity, and particle size and density in the flowfield. When the laser frequency is tuned to the edge of the sloping region, the Doppler shift will move the scattered intensity into the sloping region of the absorption profile, resulting in an increase in the transmission. Thus when the transmission ratio is calculated from the two cameras, the velocity can be obtained. Instantaneous velocities have been measured in compressible mixing layers [Elliott et al., 1994], while multiple velocity components have been recorded in boundary layers [Arnette et al., 1995].

EXPERIMENTAL ARRANGEMENT

Figure 1 gives a schematic of the laser, and camera geometry for the experiments described below. The frequency-doubled Nd:YAG laser used for these experiments was a Spectra-Physics GCR-150 capability of delivering 400 mJ per pulse. The laser has an injection seeder to provide a narrow line width (~ 100 MHz) and the laser frequency can be tuned through absorption lines of iodine around 532 nm. The Nd:YAG laser used in these experiments has a pulse duration of approximately 10 ns which effectively freezes the turbulent motion in the flow field resulting in instantaneous measurements. A small portion of the laser beam was sampled using fast photodiodes and integrated using a Stanford Research System boxcar integrator to monitor the laser energy, and the reference iodine cell, and a third path where the camera cell could be placed for calibration. The SRS Boxcar integrator provides eight channels of analog input/output and was also used to provide the laser frequency bias voltage and record the Build Up Time (BUT) voltage which indicates if the laser is seeded or not. The laser beam was formed into a sheet by a combination of cylindrical and spherical optics and retro-reflected back on itself for the transverse jet experiments (for reasons which will be shortly discussed).

The scattered signal is collected using two 16-bit PixelVision back-illuminated CCD cameras

fitted with Nikkor 105mm lens except for the transverse jet experiments where Princeton Instruments 14-bit intensified CCD cameras were used. The two camera arrangement was selected to eliminate the cross-talk that occurs when two images are recorded on a single CCD array. In accordance with the arguments of McKenzie [1997], the relatively large camera aperture was selected to minimize the effect of laser speckle. In order to collect both the filtered and reference images the scattered light first passes through a polarizer followed by a nonpolarizing beam splitter cube which separates the scattered light equally to the filtered and reference cameras. A neutral density filter is placed in front of the reference camera to allow the two cameras to have a similar intensity range. The images are stored on a Pentium 133 MHz personal computer providing camera control, laser frequency tuning, and data collection. A MS Windows based data collection program was written in association with William Weaver at ISSI. Two programs are available in the data collection program. The first program (SRS450.EXE) tunes the laser through voltages so that reference and camera filter profiles can be obtained. This program collects the photodiode voltages, BUT voltage, and laser frequency bias voltage outputting single shot or averaged measurements to a file. The second program (PDV.EXE) is used to take the PDV measurements to control the SRS boxcar integrator, cameras, and laser. This PDV program creates a running logfile that contains the image file name, photodiode voltages from the filter reference system, BUT voltage, and laser bias voltage for the image taken. Therefore the laser frequency, and wether the laser was locked is tracked for each image greatly reducing the errors associated with the PDV measurements.

A major component of the PDV system is the iodine filter. The iodine filter is simply a glass cylinder 9 cm in diameter and 24 cm in length with flat optical windows on both ends. Similar iodine filters have been used in other molecular filter-based techniques [Miles et al. 1992 and Elliott et al., 1994]. Iodine vapor is formed in the cell by inserting a small amount of iodine crystals and evacuating the cell. The cell temperature (T_{cell}) is raised above the ambient temperature with electrical heat tape regulated with a digital temperature controller so that no iodine crystallizes on the windows. The coldest point in the cell is set in the side arm (T_{I2}), which is housed in a water jacket and maintained at a constant temperature by a circulation water bath. The temperature of the side arm controls the vapor pressure (number density) of the iodine in the absorption cell. After the cell has stabilized, the cold arm valve is closed so that there is a set amount of iodine vapor in the cell. In order to get a sloping absorption profile, a small amount of nitrogen is added to the cell to pressure broaden the profile. This is done by placing the cold finger in an acetone dry ice bath (70

K) to freeze out the iodine. In this way, one can measure the partial pressure of the nitrogen, since iodine is in the solid state. The cell is then reheated to allow the iodine to return to the gas phase.

METHODOLOGY FOR OBTAINING PDV MEASUREMENTS

Through the development of the PDV system for use in "large-scale" wind tunnels a procedure for taking measurements to provide and evaluate the necessary information has been established. The first step in PDV is to take iodine filter profiles using the SRS450.EXE program of the reference and camera filters simultaneously. This should be done at the beginning of the test each day to insure that the iodine filter has not leaked and is operating at the correct temperature. Second, with the optical arrangement to be used in the tests, a white card with black dots (referred to as a dotcard) is placed in the test section in the plane of the illuminating laser sheet. This step provides the tie-points necessary to map the filtered and unfiltered images to corresponding locations. Third, with the laser tuned to a single frequency outside of the iodine absorption well, images are acquired of a broadly illuminated card. Neutral density filters put in the laser path effectively vary the intensity of the recorded images. This step provides the data necessary to calibrate the filtered and unfiltered cameras to each other. The files acquired in this step will be known as greencard files. Fourth, with the laser tuned to the same out-of-filter frequency as in the previous step, images are acquired of the flow to be investigated. Because the laser is out of the filter, no attenuation of the image seen through the filter. This step provides a check on the validity of the calibration produced in the third step, and the files recorded for this step will be referred to as out-of-filter files. Fifth, with the laser tuned to a frequency appropriate for Doppler sensitivity, images are acquired of the flow to be investigated. These are the files that actually contain the flow velocity data, and they will be referred to as the raw data files. In addition to the five data categories described above, background files are acquired so that the effects of stray light and camera pixel offsets can be subtracted before data processing. Also to check for variations in frequency across the laser sheet, the attenuated laser sheet was focussed on the white card and filter scans were taken. This data has not been incorporated to date.

Once the filter profiles and data images are taken the data reduction software perform the following analysis.

1. Using the dot card and green card images the filtered and reference cameras are aligned and calibrated

2. The alignment and calibration are checked using data taken with the laser frequency tuned outside the iodine absorption line
3. The transmission ratio is calculated from the filter and reference images
4. The velocity is calculated from the transmission ratio using the optical geometry, the iodine absorption profile, and the instantaneous frequency of the laser

The specifics of these data reduction programs are available from the principle investigator.

SYSTEM CALIBRATION MEASUREMENTS

As a first step in making the PDV system usable for velocity measurements in large scale wind tunnel facilities, performance characteristics of the laser, optical system, and image analysis programs were accessed. The optically thick absorption line used in the present experiments was at 18789.28 cm^{-1}. The profiles were taken with the cell operated at T_{cell} = 358 K and T_{I2} = 318 K. Figure 2 gives the absorption profiles for the cell operated with 20 torr of nitrogen and with no nitrogen present. Using more or less nitrogen, the pressure broadened absorption profile can be adjusted to the slope needed for the expected frequency range encountered in the flow.

A significant source of error reported by investigators is laser speckle. Laser speckle results from the interference between coherent wavelets from an irregular target that is illuminated by a coherent light source [McKenzie, 1997]. Since the laser speckle changes in both space and time for a moving scatterer, the reference and filtered images can not can not be matched when speckle is present. One way that the speckle effects can be reduced is by increasing the aperture of the collection system. For f-numbers less than 8 the speckle was found to be significantly reduced. The effect of speckle can further be reduced by operating in the Rayleigh scattering regime, increasing the object distance, or spatial filtering the data. Unfortunately the last two methods decrease the camera resolution. The desire to use a lower f-number to reduce speckle makes the possibility of using a split image arrangement difficult since the region in which images overlap increases at lower apertures.

In order to make accurate velocity measurements additional errors can occur with the image processing algorithms. Several algorithums to align and calibrate the cameras were attempted. The final image alignment algorithum was tested and found to align the cameras with a maximum deviation of 0.15 pixels. The intensity calibration was checked by dividing the images from the two cameras after they had been calibrating. The standard deviation was typically less than 4% for

instantaneous image and less than 1% for averaged images. To improve the image alignment and calibration a spatial filter can be incorporated. This not only helped image alignment and intensity calibration, but also reduces laser speckle effects.

TRANSVERSE JET INJECTION INTO A SUPERSONIC FLOW

The first application of the new PDV software was applied to measure the velocity field of a transverse jet injected into a Mach 2 crossflow. The experiments were performed in the supersonic research facility located at Wright-Patterson Air Force Base. The details of this facility appear in other references [Gruber and Nejad, 1994 and 1995]. The two dimensional nozzle section accelerates the flow to a uniform Mach 1.98 free stream. The tunnel can be run continuously with a stagnation temperature and pressure of 293 K and 317 kPa respectively, resulting in an average free stream velocity of 516 m/s. For the present study, two injector geometries were incorporated with circular and elliptic geometries into removable elements in the bottom wall of the test section. For the elliptical jet the semi-major axis is aligned with the streamwise coordinate. In all cases presented, the jet was oriented 90 degrees to the free stream. Air was injected at the same exit pressure (476 kPa), velocity (317 m/s) and density (6.64 kg/m^3), resulting in a momentum flux ratio of 2.9.

In order to use PDV, the flow field had to be seeded with particles which would rapidly adjust to the turbulent fluctuations encountered in the flow. The particles used in this experiment were formed from the combustion of silane (SiH_4) which forms solid silica dioxide (SiO_2), water, and hydrogen as products of combustion.

When making velocity measurements using molecular filter based techniques, a difficulty is encountered in measuring a velocity component which is "natural" (aligned with the natural orthogonal coordinates) to the flow being studied. To circumvent this problem we chose to retro-reflect the laser sheet back on itself. In this way, the incident laser beam has components in the positive and negative directions which "effectively" cancels out the spanwise component of the velocity. Using Equation 1 and the present spanwise view (laser sheet normal to the flow) results in a sensitivity to the velocity vector essentially in the streamwise direction, and thus instantaneous streamwise velocities can be obtained.

Figure 3 and 4 present average streamwise velocity (U) images for circular and elliptical jets. The velocity images at x/D = -2 (3a, 3b) show the reduced velocity behind the separation shock with

a center velocity of 440 m/s and 410 m/s for the circular and elliptical jets respectively. For the elliptical jet the lateral extent is reduced by approximately 55%. Moving downstream to x/D = -1, the spanwise cross-section is just after the bow shock has started to form and the streamwise velocity for the circular jet is 230 m/s and 210 m/s for the elliptical jet. These lower subsonic values are consistent with flow visualizations that show the flow has passed through a nearly normal shock. Assuming a normal shock results in a velocity of 196 m/s which is confirmed by the measurements. Again the lateral extent of the bow shock for the same downstream location is reduced for the elliptical case. Also, present is the region on each side of the bow shock defining the separation shock region which has a velocity of 430 m/s and 407m/s for the circular and elliptic jets respectively. The streamwise velocity decreases further as the boundary layer is approached.

The next three streamwise locations (Figures 3e-3h, 4a, and 4b at x/D = 0, 1, 2) show the bow shock increasing in size around the transverse jet which is injected into the free stream. At all locations, the lateral extent of the bow shock is reduced for the elliptical jet although the spanwise extent is almost the same. Within the bow shock the velocity increases as the streamwise location increases from 270 m/s at x/D = 0 to 350 m/s at x/D = 2 for the circular jet and 250 m/s to 430 m/s for the elliptic jet. There is a distinct vertical line separating the fluid inside the bow shock and fluid inside the separation shocks observed on both sides of the jet. Although the jet core has no signal, the reduction in the velocity is clearly seen in the shear layer separating the jet from the free stream fluid.

For streamwise locations of x/D = 4, 6, and 8 (Figures 4c-4h) the velocity in the free stream of the bow shock is relatively constant, approximately 440 m/s to 490 m/s for the circular case and 440 m/s to 470 m/s for the elliptical jet. By these streamwise locations, enough of the jet has entrained free stream fluid to allow the streamwise velocity to be resolved in the counter rotating vortices. The streamwise velocity is greatly reduced toward the core of the vortices and has a greater deficit for the elliptical case. The streamwise velocity near the core of the vortices for the circular jet at x/D = 4, 6, and 8 is 150 m/s, 160 m/s, and 220 m/s. Alternatively, the streamwise velocity near the core of the vortices for the elliptical jet at x/D = 4, 6, and 8 is unresolved, 150 m/s, and 220 m/s. Both jet geometries indicate an increase in the streamwise velocity in the vortices as more of the free stream is entrained as they develop downstream.

The streamwise turbulence intensity is defined as the standard deviation of the streamwise velocity fluctuation (σ_u) normalized by the local mean streamwise velocity (U) is given in Figures

5 and 6 for x/D from -2 to 8. When displaying the turbulence intensity, the scale was made logarithmic since there is a wide dynamic range from the free stream to the fluctuations associated with the shear layer between the jet and free stream. The free stream turbulence intensity was found to be approximately 0.04 to 0.06 for both jet cases. At x/D = -2 the separation shock has a turbulence intensity roughly at the free stream value. This is true at all streamwise locations. Figures 5c, and 5d the turbulence intensity is 0.22 and 0.24 for the circular and elliptical jets in the region marking the bow shock fluctuations. Between the turbulence peak in the boundary layer and the peak associated with the bow shock fluctuation the turbulence intensity is 0.17 for the circular jet and 0.1 for the elliptical jet. This fluctuation is caused by the turbulent structures in the jet and the unsteady waves that emanate from them.

For the next three streamwise locations at x/D = 0, 1 and 2 (Figures 5e -5h, 6a, and 6b) there are three regions of interest. First is the streamwise turbulence due to the bow shock fluctuations which decrease slightly traveling downstream from 0.17 (x/D = 0) to 0.09 (x/D = 2) for the circular jet and 0.15 (x/D=0) to 0.11 (x/D = 2) for the elliptical jet. This decrease in turbulence intensity is most likely due to the fact that the bow shock weakens as it develops downstream and the influence of the eddies becomes reduced. The second region of interest is between the jet boundary and the bow shock. At the left and right sides of this region, but still within the bow shock, the turbulence levels are the same as the undisturbed free stream (0.04 to 0.06). In the region above the jet, however, the streamwise turbulence intensity is slightly higher at approximately 0.12 for the circular jet and 0.09 for the elliptic jet. Again this higher turbulence intensity is caused by the turbulent structures in the jet and the unsteady waves that emanate from them. The area this region occupies seems to be constant for the circular jet, but for the elliptic jet the region seems to take up less lateral area and grows in the spanwise direction at greater streamwise distances. This spanwise growth may be indicative of the axis switching which investigators have observed for elliptic jets [Gruber et al., 1997]. The final region in these three images is the shear layer next to the jet core. Obviously the turbulence intensity peaks in the shear layer having a maximum value of 5.0 before there is no signal to analyze (indicated by the totally black region). The mixing layer continues to grow in thickness downstream until the jet collapses forming two streamwise vortices.

The final three images (Figures 6c-6h) show the streamwise turbulence intensity of the two counter rotating streamwise vortices. Again the region of the flow above the streamwise vortices have a slightly higher turbulence intensity for reasons stated previously. The cores of the streamwise

vortices have approximately constant streamwise turbulence levels at x/D = 4 and 6 ($\sigma_u/U \approx 1.5$ for the circular jet and $\sigma_u/U \approx 2.3$ for the elliptical jet). This value decreases as more free stream fluid is entrained to 0.8 for the circular jet and 1.6 for the elliptic jet. Also it is seen that the total area of the streamwise turbulence intensity for the counter rotating vortices are greater for the elliptical jet compared to the circular jet. This would indicate that for the same effective diameter an elliptical jet will better mix the with the free stream when injected transverse to the supersonic free stream. This trend has also been observed in previous investigations [Gruber et al., 1996].

AXISYMMETRIC JET

The supersonic test facility consisted of an axisymmetric, vertically-issuing jet exhausting into ambient air. The stagnation chamber and jet were mounted on a stepper-motor controlled assembly that provided linear positioning in three orthogonal directions. A pressurized vessel containing ethanol was connected to the stagnation chamber to provide seeding. The jet was formed by a converging-diverging nozzle designed by the method of characteristics to operate at Mach 1.36. The exit diameter of the nozzle was 12.7mm and the nominal stagnation temperature was 294 K. The calculated jet exit velocity assuming isentropic flow was 400 m/s. The plane of the laser sheet contained the jet axis, and the laser propagation direction was about 10 degrees off the vertical. The camera view was normal to the laser sheet. The component of the calculated velocity to which the PDV system was sensitive had a magnitude of 283 m/s (Note that this is the portion of the free stream velocity projected onto the vector which the system is sensitive to).

Figure 7 shows the average of 100 instantaneous measurements of the Mach 1.36 jet. The spatially-averaged measurement in a core region of the jet is 270 ± 6 m/s, which underestimates by about 5 percent the value of 283 m/s computed from the isentropic assumption. Velocities found in the jet range from approximately 120 m/s to 270 m/s. Velocity measurements in the shear layer drop out below approximately 120 m/s either because the seed species revaporizes in the shear layer as the static temperature increases or because of a requirement for a minimum signal strength in the processing software. This issue will be addressed further. Figure 7 also shows that the PDV system is sensitive enough to detect a weak shock structure, which was not seen in Schlieren visualizations. Figure 8 gives the component of turbulence intensity in the direction of the PDV system sensitivity. As expected, the turbulence intensity is high in the shear layers and decreases as the core of the jet is approached. Radial diffusion of the turbulence intensity is evident in the thickening of the shear

layer in the downstream direction. This facility was useful in investigating various ideas to reduce and evaluate experimental uncertainties.

LARGE WIND TUNNEL TEST

An opportunity was given to test the PDV system in a large subsonic wind tunnel. The subsonic test facility (Subsonic Aerodynamic Research Laboratory, SARL operated with Dr. Thomas Beutner) consisted of an open circuit, low-speed wind tunnel with a 3.05 x 2.13 m test section. In the inlet, a movable smoke-generating rig utilizing Rosco theatrical fog was available to seed the flow. "Empty tunnel" tests at Mach numbers of 0.2 and 0.3 were conducted to establish the basic effectiveness of the current PDV system in the large scale facility with the selected illuminating and viewing geometry. As seen in Figure 9, the laser sheet was oriented span-wise across the tunnel with a propagation direction normal to the surface of the model. The camera viewed the sheet from upstream and above the tunnel, leading to a PDV system sensitivity to the velocity component in the direction, $-0.209\hat{i}-0.003\hat{j}+0.978\hat{k}$. This velocity component was chosen because it is strongly affected by streamwise vortices. The tests in SARL measured the interaction of the vortices generated by the sharp leading edge of a delta-wing with its twin vertical tails. Leading edge sweep on the model was 70 degrees. The tail shape was chosen to be characteristic of the F-15 platform. Tails were located along a radial line originating at the apex of the delta wing. The tail position was chosen to lie along the vortex core trajectory. Tests were conducted both on the clean wing and on the wing with tails. The model had sharp leading edges on both the wing and the tails in order to fix separation points. The angle-of-attack was 23.2 degrees and the free stream Mach number was 0.2. The Reynolds number based on the root chord was 1.94×10^6. At this condition, no vortex bursting is present over the model surface on the clean delta-wing. However, the presence of the tails causes unsteady vortex bursting near the mid-chord location.

In the subsonic tests, the "empty tunnel" cases at Mach 0.2 and 0.3 were encouraging in that in each case the PDV measured velocity component was found to be within 2 m/s of the component obtained from the velocity measured using pitot probes. Figure 10 shows the averaged DPV-sensitive velocity component above a clean delta-wing based upon 62 images. The imaging plane is normal to the wing and located 1.27 cm behind it. As expected two well defined vorticies are clearly indicated by a horizontal region with positive and negative peaks associated with each side of the vortex. Figure 11 shows the same velocity component in the same plane for a delta-wing with

tails based upon 90 images. In each case, the velocities in the upper region of the field-of-view, away from the wing, approach the known free stream velocity. As anticipated, the vortical structure in the case with tails is weaker due to the effect of vortex bursting. In the case of the clean delta-wing, the positive outboard vortical velocities are smaller in magnitude than the negative inboard velocities. This result agrees with the predictions of a CFD computation [Rizzetta, 1996]. Data was also taken for each wing at three other stations. A description of the flow field based upon the complete data set, including measurements of fluctuating quantities, and a fuller comparison with the numerical solution is planned.

TWO COLOR PLANAR DOPPLER VELOCIMETRY

In the PDV arrangement described above, a monochrome illumination is applied. The filtered and reference images are recorded on separate cameras or on different portions of the same camera in split image arrangements. Accurately normalizing the filtered image by the reference image requires them to be aligned and calibrated accurately. With Dr. Steve Arnette from Sverdrup Technologies a new two-color technique was demonstrated. This has the possibility of simplifying the system and eliminating some sources of uncertainty.

In two color planar Doppler velocimetry (TCPDV) the flow field is illuminated using the red beam from a Dye laser pumped by an injection-seeded Nd:YAG laser and the green beam from a second injection-seeded Nd:YAG laser. The red and green beams are formed into spatially-overlapping sheets which propagate through the center of the Mach 1.5 axisymmetric jet. An iodine cell is placed in front of a Kodak DCS460 color CCD camera which images the flow field. The scattered intensity from the red and green beams can be separated by calibrating the camera for laser sheet intensity variations and pixel color bleed. The red signal provides the reference image (since the red frequency is outside the absorption lines of iodine) and the green signal provides the filtered image as before.

A single instantaneous velocity measurement is presented in Figure 12 with the flow direction from the right to the left. Figure 12a is the raw image with individual red and green images given in Figure 12b, and Figure 12c respectively. The effect of the frequency discrimination is clearly seen observing that the intensity maxima in both images have different shapes, but the outline of the flow is the same. The processed velocity is given in Figure 12d with the measured velocity vector transformed to the streamwise direction. For the perfectly expanded Mach 1.5 jet, the

velocity in the core should be 433 m/s which is in good agreement with the measurements taken so far. The data is currently being analyzed further and measurements were made to better quantify the experimental uncertainty.

ACKNOWLEDGMENTS

The principle investigator would like to give a special thanks to Dr. Campbell Carter for his help in conducting all the above experiments. Also for Dr. William Weaver and Andrew Mosedale for there help in programing the data collection and image processing programs. Finally this work would not be possible without the laboratories, equipment, and direction provided by Dr. Mark Gruber.

REFERENCES

Arnette, S.A., Samimy, M., and Elliott, G.S., "Expansion Effects on Supersonic Turbulent Boundary Layers," *AIAA Journal*, Vol. 33, No. 4, pp. 430-438, 1995.

Elliott, G.S., Samimy, M., and Arnette, S.A., "Study of Compressible Mixing Layers Using Filtered Rayleigh Scattering Based Visualizations," *AIAA Journal*, Vol. 30, No. 10, pp. 2567 - 2569, 1992.

Elliott, G.S., Samimy, M., and Arnette, S.A., "A Molecular Filter Based Velocimetry Technique for High Speed Flows," *Experiments in Fluids*, Vol. 18, pp. 107-118, 1994.

Elliott, G.S., Glumac, N., Carter, C.D., and Nejad, A.S. "The Measurement of Two Dimensional Temperature fields Using Molecular Filter Based Technique," Accepted to the *Combustion Science and Technology*, 1997.

Elliott, G.S. and Samimy, M., "A Rayleigh Scattering Technique for Simultaneous Measurements of Velocity and Thermodynamic Properties," *AIAA Journal*, Vol. 34, No. 11, pp.2346-2352, 1996.

Gruber, M.R., Nejad, A.S., Chen, T.H., and Dutton, J.C., "Mixing and Penetration Studies of Sonic Jets in a Mach 2 Freestream," *Journal of Propulsion and Power*, Vol. 11, No. 2, pp. 315-323, 1995.

Gruber, M.R., Nejad, and Dutton, J.C., "An Experimental Investigation of Transverse Injection from Circular and Elliptical Nozzles into a Supersonic CrossFlow," Wright Laboratory Technical Repot WL-TR-96-2102, 1996.

McKenzie, R.L., "Planar Doppler Velocimetry Preformance in Low-Speed Flows," AIAA 97-0498, 1997.

Meyers, J.F., and Komine, H., "Doppler Global Velocimetry: A New Way to Look at Velocity," *Laser Anemometry*, Vol. 1, 1991.

Miles, R.B., Forkey, J.N., and Lempert, W.R., "Filtered Rayleigh Scattering Measurements in Supersonic/Hypersonic Facilities," AIAA Paper 92-3894, 1992.

Rizzetta, D.P., "Numerical Simulation of the Interaction between a Leading-Edge Vortex and a Vertical Tail," AIAA Paper 96-2012, 27th AIAA Fluid Dynamics Conference, June 17-20, New Orleans, LA.

ARTICLES PUBLISHED FROM SFRP

Arnette, S.A., Elliott, G.S., Mosedale A., Carter, C.D., "A Two Color Approach to Planar Doppler Velocimetry," Accepted to the AIAA 36th Aerospace Sciences Meeting, Reno, Nevada, 1998.

Elliott, G.S., A. Mosedale, Gruber, M., Carter, C.D., and Nejad, A.S. "The Study of a Transverse Jet in a Supersonic Cross-Flow Using Molecular Filter Based Diagnostics," Submitted to AIAA Propulsion and Power, 1997.

Mosedale, A., Elliott, G.S., Carter, C.D., and Beutner, T.J., "On the Use of Planar Doppler Velocimetry," Submittted to the AIAA 29th Joint Fluid Dynamics Conference, June, 1998.

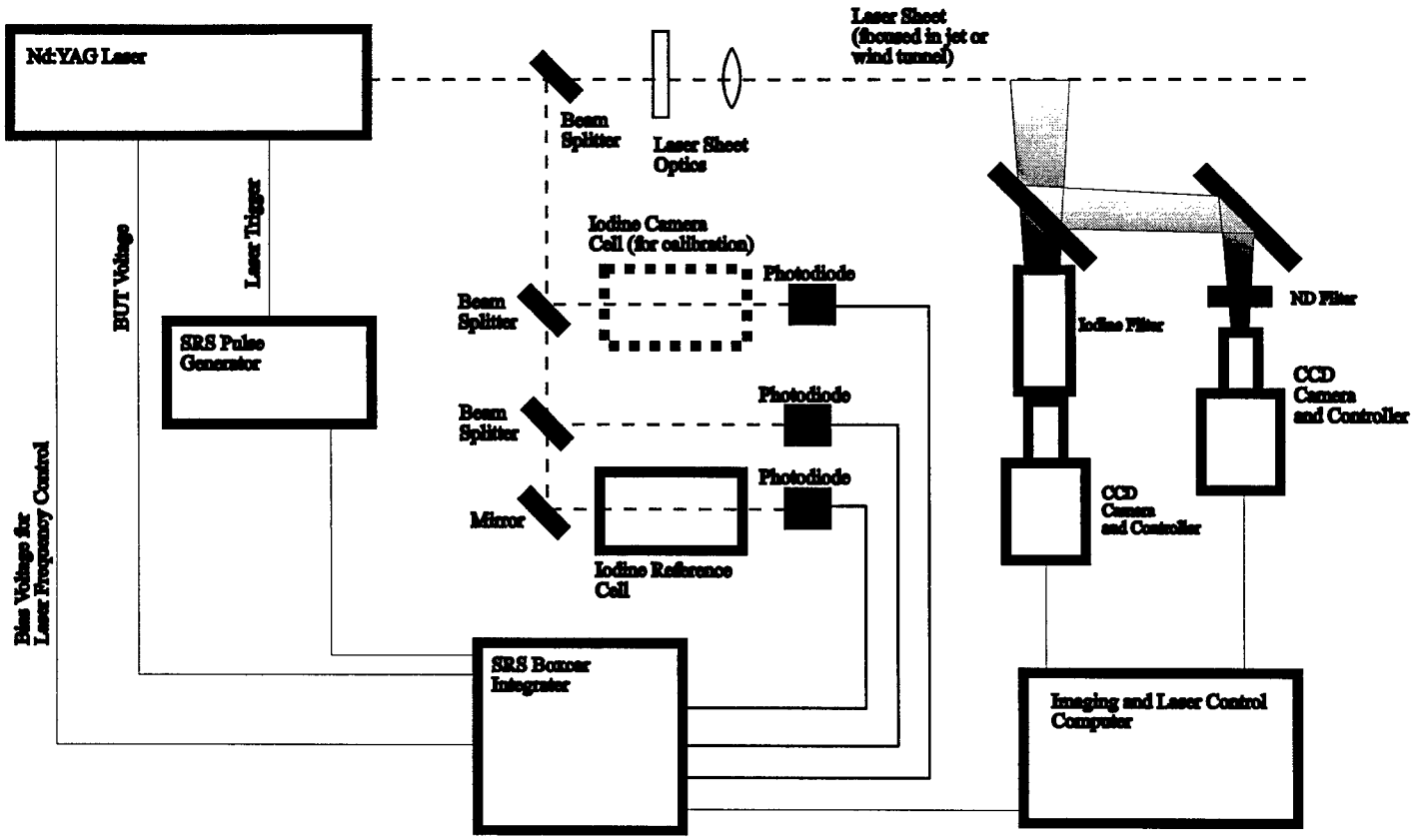


Figure 1. Schematic of laser and optical arrangement for PDV measurements.

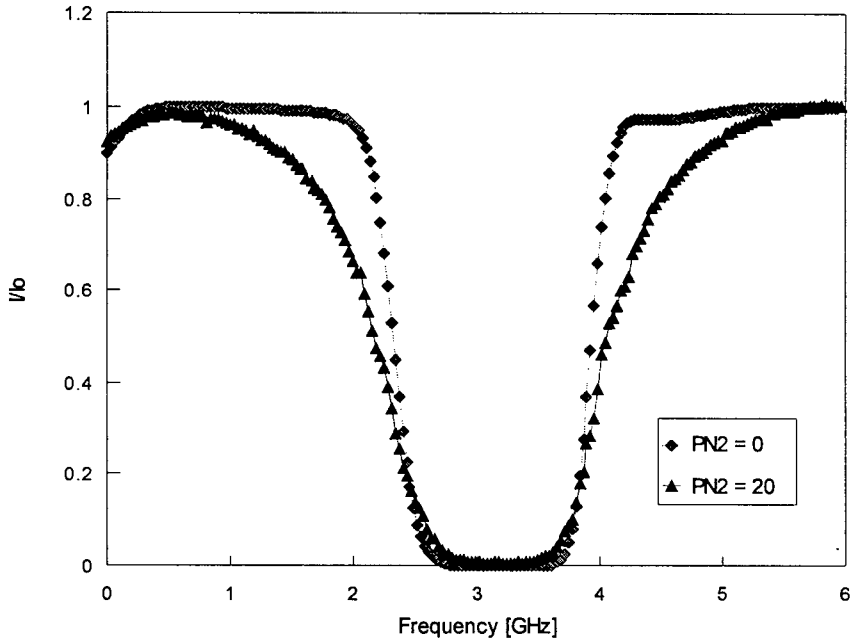


Figure 2. Pressure broadened absorption profile of iodine used in the PDV experiments.

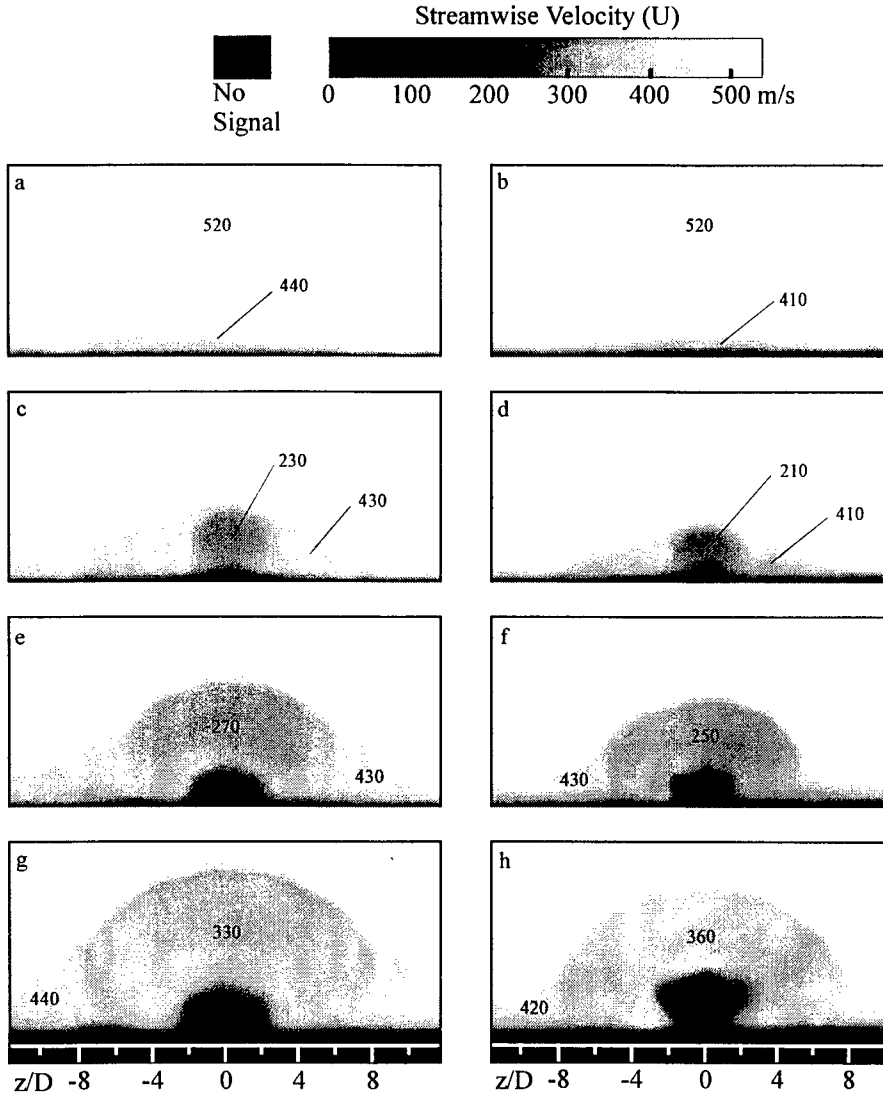


Figure 3. Spanwise views of the mean streamwise velocity for circular (a, c, e, and f) and elliptical (b, d, f, and h) jets. The streamwise locations are x/D = -2 (a and b), -1 (c and d), 0 (e and f), and 1 (g and h).

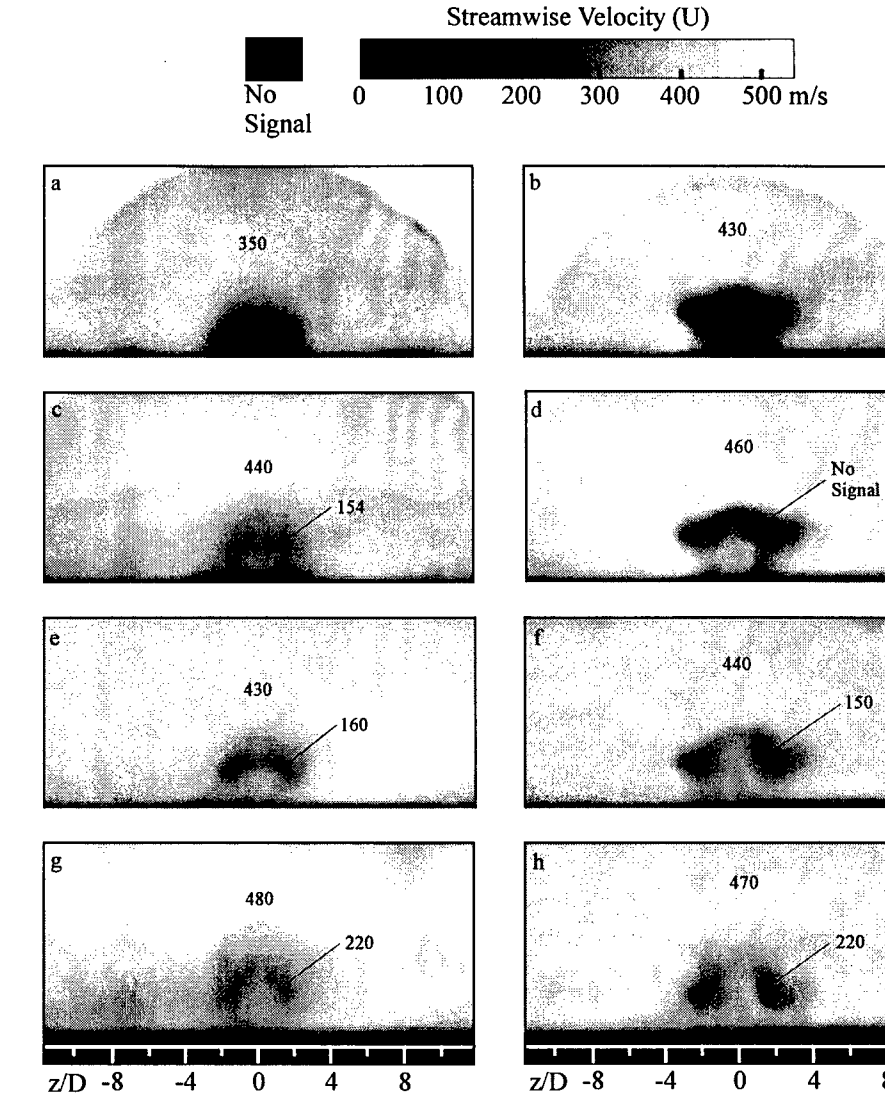


Figure 4. Spanwise views of the mean streamwise velocity for circular (a, c, e, and f) and elliptical (b, d, f, and h) jets. The streamwise locations are x/D = 2 (a and b), 4 (c and d), 6 (e and f), and 8 (g and h).

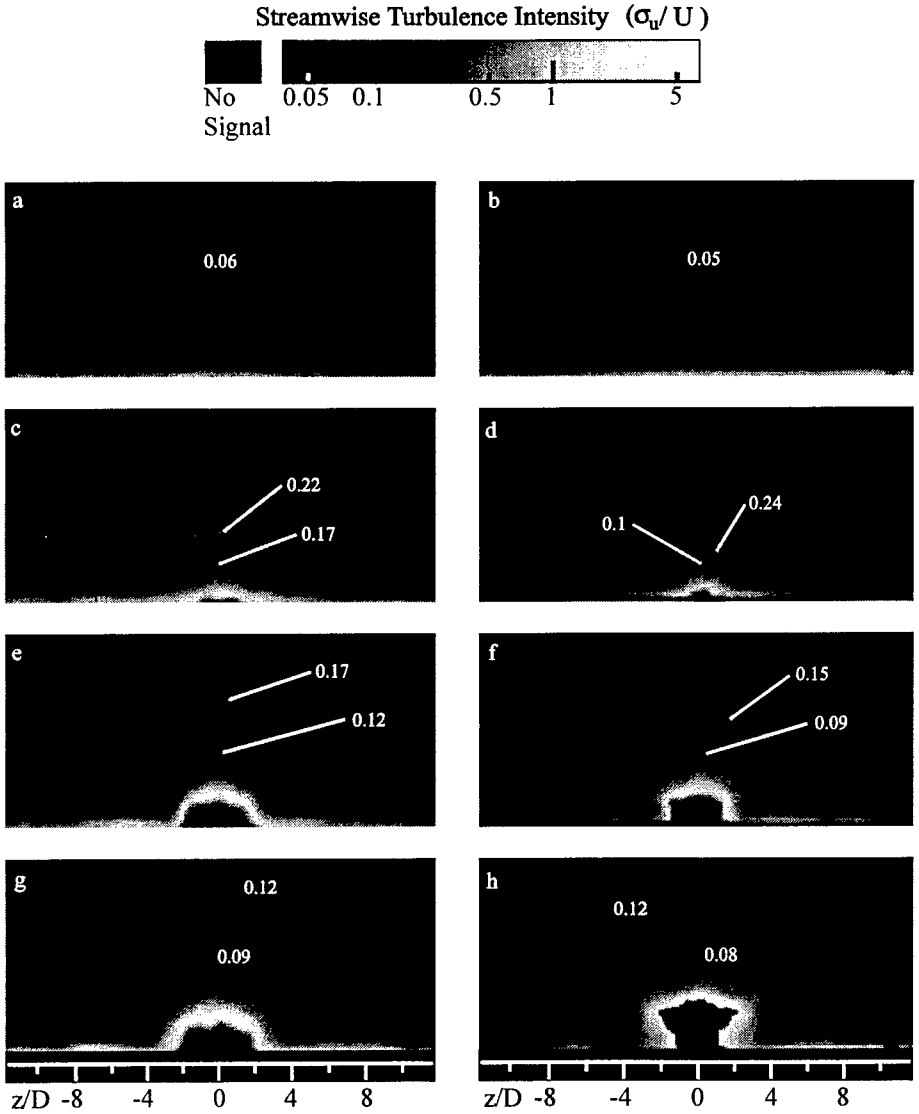


Figure 5. Spanwise views of the streamwise turbulence intensity for circular (a, c, e, and f) and elliptical (b, d, f, and h) jets. The streamwise locations are x/D = -2 (a and b), -1 (c and d), 0 (e and f), and 1 (g and h).

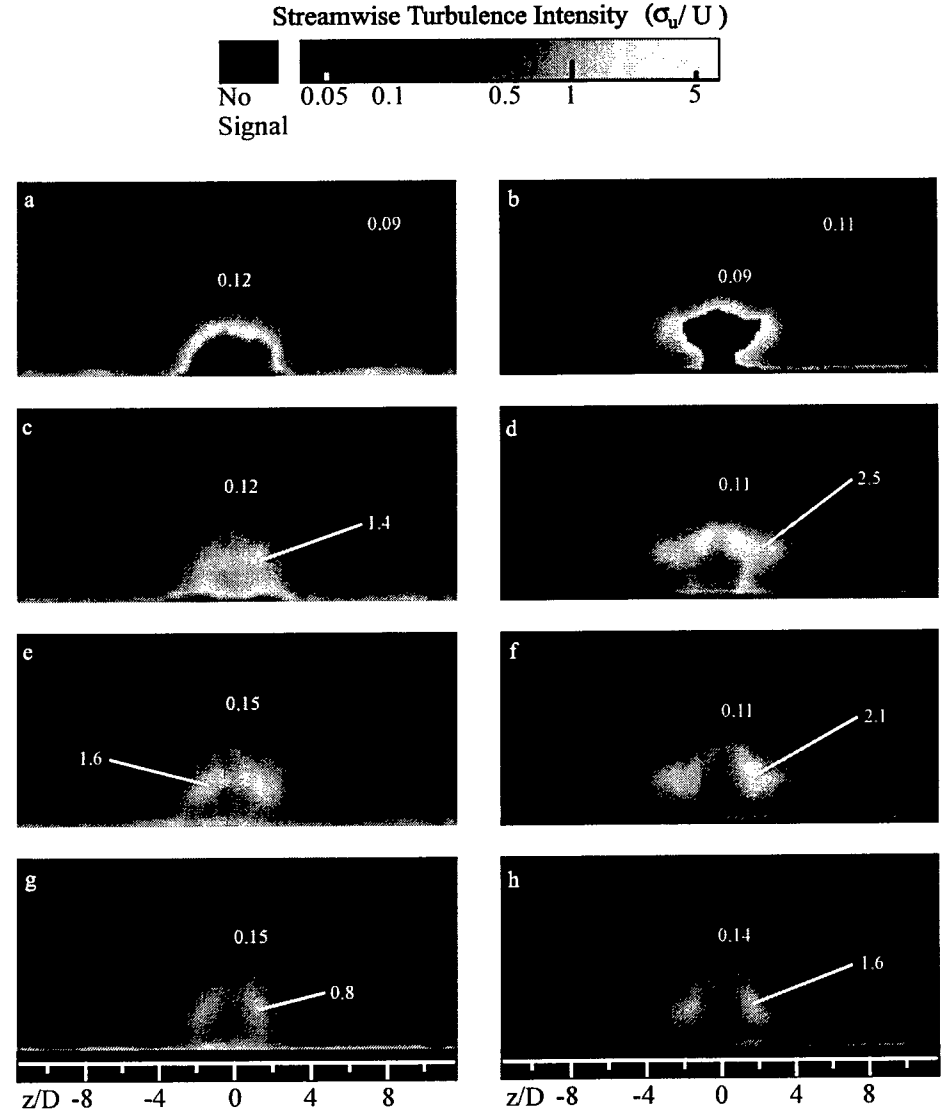


Figure 6 Spanwise views of the streamwise turbulence intensity for circular (a, c, e, and f) and elliptical (b, d, f, and h) jets. The streamwise locations are x/D = 2 (a and b), 4 (c and d), 6 (e and f), and 8 (g and h).

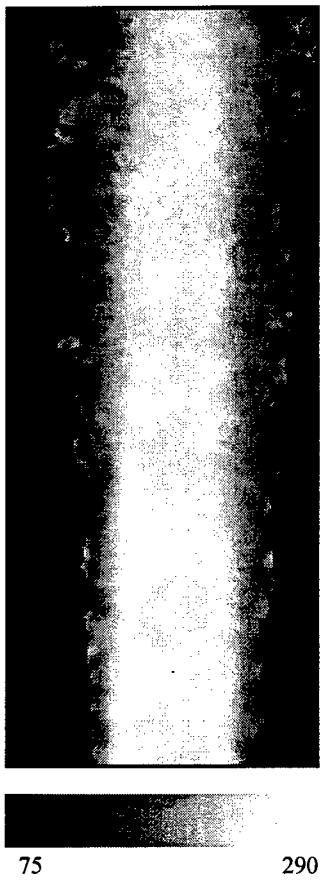


Figure 7. Velocity component in the direction of PDV system sensitivity for a Mach 1.36 free jet. 100 images were averaged.



Figure 8. Turbulence intensity in the direction of PDV system sensitivity for a Mach 1.36 jet.

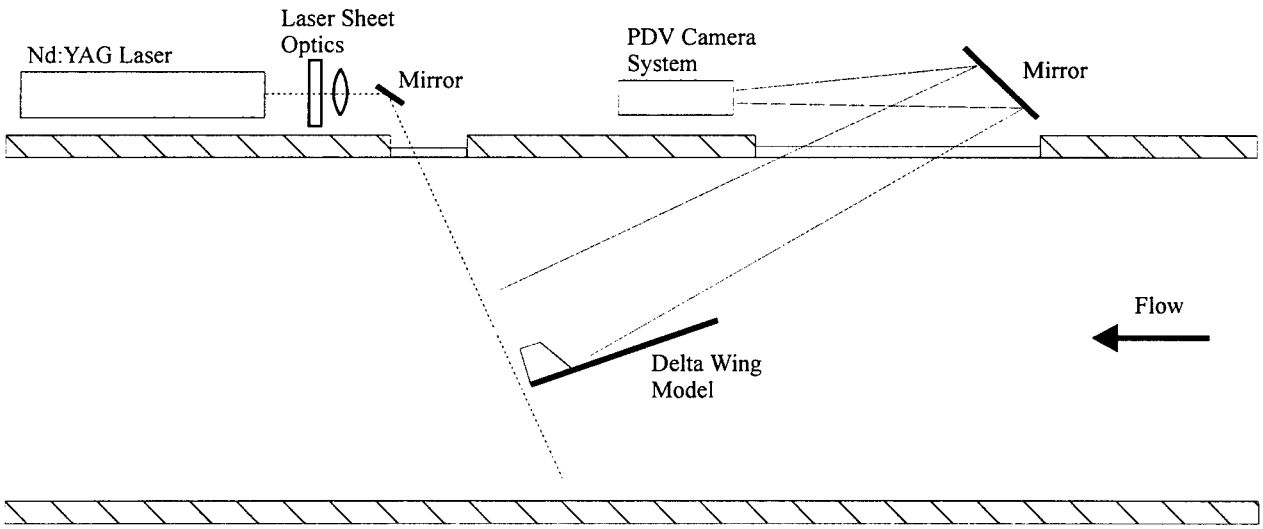
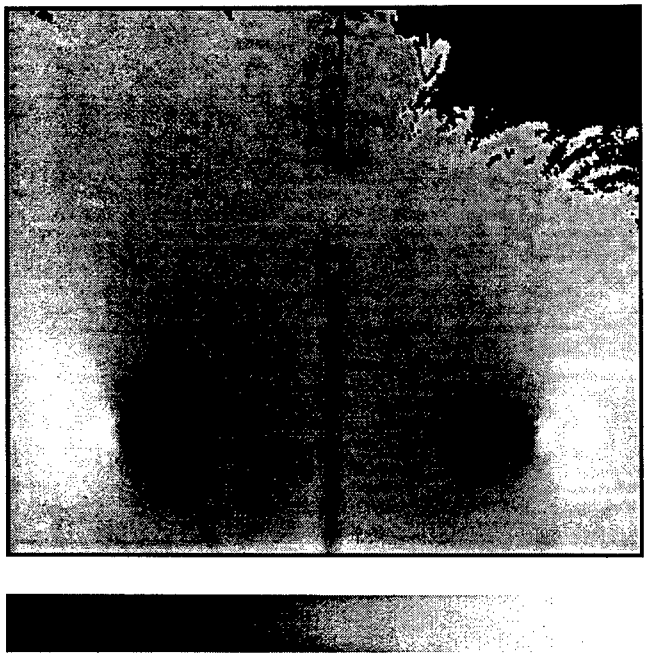
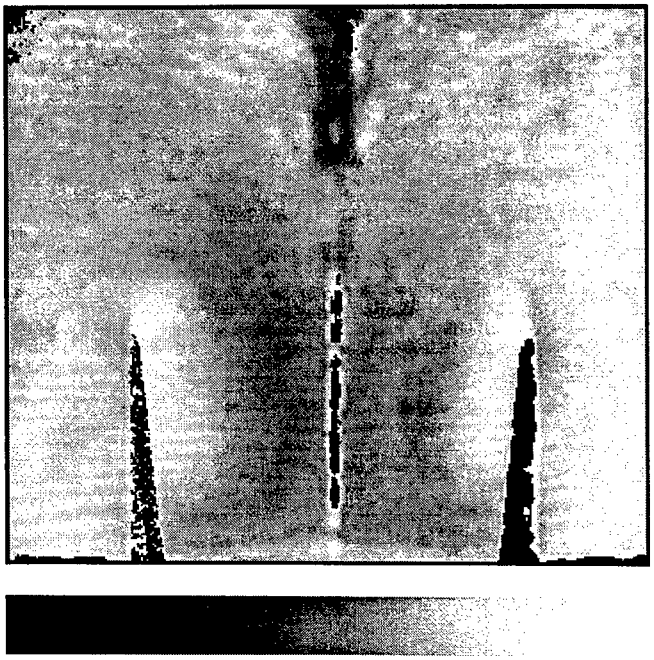


Figure 9. Relative arrangement of the laser, camera, and delta wing model in the wind tunnel.



-120 40 m/s

Figure 10. Average velocity component in the direction of DPV sensitivity above a delta-wing at a 23.2° angle-of-attack in a Mach 0.2 flow. Based upon



-120 40 m/s

Figure 11. Average velocity component in the direction of DPV sensitivity above a delta-wing with tails at a 23.2° angle-of-attack in a Mach 0.2 flow.

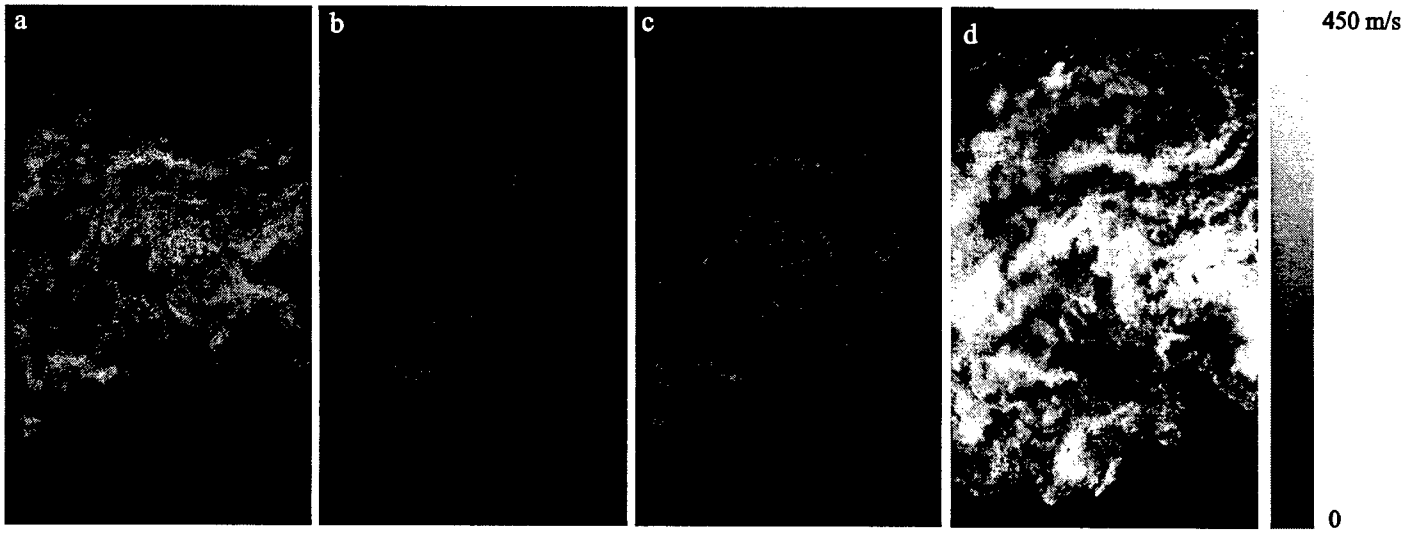


Figure 12. Contents of a color Planar Doppler Velocimetry realizatioin: a) two-color filtered image, b) extracted green image, c) extracted red image, and d) resulting velocity image.

EVALUATION OF THE POINTWISE k-2 TURBULENCE MODEL TO PREDICT TEANSITION AND SEPARATION IN LOW PRESSURE TURBINE

Dr. Elizabeth A. Ervin
Assistant Professor
Department of Mechanical and Aerospace Engineering

University of Dayton
300 College Park
Dayton, Ohio 45469-0210

Final Report for:
Summer Faculty Research Program
Wright Laboratory

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

And

Wright Laboratory

September, 1997

EVALUATION OF THE POINTWISE k-$\mathcal{R}$ TURBULENCE MODEL TO PREDICT TRANSITION AND SEPARATION IN A LOW PRESSURE TURBINE

Elizabeth A. Ervin
Assistant Professor
Department of Mechanical and Aerospace Engineering

Abstract

Low pressure turbines in aircraft experience large changes in Reynolds number as the engine operates from take-off to high altitude cruise. Many low-pressure turbine blades contain regions of strong acceleration and diffusion. As the Reynolds number decreases, these regions develop large unsteady transitional and separation zones. Computational models show limited success in predicting such flow phenomena. The point wise k-$\mathcal{R}$ turbulence model has been recently proposed to represent wall bounded and free shear flows. Thus, it may be conveniently used to represent turbine flows, which are often modeled with a mesh around the airfoil (O-grid) interacting with another mesh, for the outer flow (H-grid). The point wise k-$\mathcal{R}$ model is used here with a meridonial coordinate system to evaluate its ability to predict transition length, as well as boundary layer separation, on a two-dimensional linear low-pressure turbine blade cascade. The new turbulence model is evaluated with experimental results of a low-pressure turbine blade cascade.

Concurrently, a fundamental study of the flow dynamics around a blade with oscillating cooling flow is being investigated. A transient solution of the Reynolds-averaged Navier-Stokes, continuity, and energy equations is being developed to analyze the effects of a pulsing jet on vortex development and interaction with the blade surface. Current studies suggest that an oscillating bleed flow passed through a turbine rotor blade can reduce the friction drag on the blade. Furthermore, the resulting boundary layer structure and possible separation from the blade will decrease the effective available area for the high-speed flow between adjacent blades, improving off-design performance. The status of this effort is discussed.

EVALUATION OF THE POINTWISE k-$\mathcal{R}$ TURBULENCE MODEL TO PREDICT TRANSITION AND SEPARATION IN A LOW PRESSURE TURBINE

Elizabeth A. Ervin

Introduction

Low pressure turbines in aircraft experience large changes in Reynolds number as the engine operates from take-off to high altitude cruise. Many low-pressure turbine blades contain regions of strong acceleration and diffusion. As the Reynolds number decreases, these regions develop large unsteady transitional and separation zones. Computational models show limited success in predicting such flow phenomena. For example, Murawski, et al., (1997) measured separation on a low-pressure turbine blade cascade at an exit Reynolds number (Re), based on suction surface length, from 50,000 to 300,000, typical of the values of a low-pressure turbine. The free-stream turbulence intensity was varied from 1.1 to 8 percent. The size of the separation zone and the velocity deficit in the wake decreased with increasing Re and increasing turbulence intensity. Numerical simulations were performed using a two-dimensional model with the Baldwin and Lomax (1978) algebraic model to account for turbulence. The numerical simulations were not able to capture the separation at Re greater than 80,000, resulting in unrealistically low pressure-loss coefficients.

Walsh et al., (1997) performed a similar study, using a heated coating method to measure heat transfer coefficient profile and thus the separation and transition to turbulence on the blade surface. The authors were able to show that the onset of transition appears to be a linear function of the turbulence intensity over the range that was tested (0.5 to 15 percent). It was shown that the onset of transition moves forward on the blade with increasing turbulence intensity and the location of fully turbulent flow does not vary. Both studies (Murawski, et al., 1997, Walsh et al., 1997) used the Langston blade profile.

The original two-dimensional model used the Baldwin and Lomax (1978) model to account for turbulence. Algebraic turbulence models, such as this, are not able to model the convection and diffusion of turbulent kinetic energy. Goldberg (1994) proposed a promising pointwise k-$\mathcal{R}$ turbulence model that is able to model both wall bounded and free shear flows. This makes it ideal to represent turbine flows, which are often modeled with a mesh around the airfoil (O-grid) interacting with another mesh, for the outer flow (H-grid). It is applied here with a meridonial coordinate system to evaluate its ability to

predict transition length, as well as boundary layer separation, on a two-dimensional linear low-pressure turbine blade cascade. The point wise k-$\mathcal{R}$ model is similar to the low Reynolds number k-ε model of Lam and Bremhorst (1981), where k is the turbulent velocity fluctuation kinetic energy and ε is the viscous dissipation. $\mathcal{R}$ is the undamped eddy viscosity and is equal to k^2/ε. It offers two advantage over previous models:

1. $\mathcal{R}$ is zero at the wall, unlike ε.

2. The model is not based on wall functions, which makes it adaptable for external flows.

In this summer study, the newly developed k-$\mathcal{R}$ model, used in conjunction with the two-dimensional VBI (Vane-Blade Interaction) software, is compared with the experimental work of Murawski, et al., 1997 and Walsh et al., 1997. In addition, a model of the same turbine blade cascade using a commercial CFD software (CFX) is also presented for comparison, using both a low Reynolds number model and the standard form (Launder and Spaulding, 1974) of the k-ε model.

Concurrently, a fundamental study of the flow dynamics around a blade with oscillating cooling flow is being investigated. A transient solution of the Reynolds-averaged Navier-Stokes, continuity, and energy equations is being developed to analyze the effects of a pulsing jet on vortex development and interaction with the blade surface. Recent developments suggest that an oscillating cooling flow through a turbine rotor blade may reduce the friction drag on the blade. The resulting boundary layer structure and possible separation from the blade would decrease the effective available area for the high-speed flow between adjacent blades. This would improve off-design performance, similar to that of a variable area turbine. The oscillating flow would be used to control the vortex development on the blade surface. Vortex development and interaction with a surface are complex processes, and measurements in operating engines are difficult and expensive. A transient solution of the Reynolds-averaged Navier-Stokes, continuity, and energy equations will permit analysis of the effects of the oscillating jet on vortex development and interaction with the blade surface.

A wall jet consists of an outer shear layer and an inner layer that behaves like a viscous boundary layer. Cohen, et al. (1992) calculated two unstable modes: an inviscid mode that depicts the large-scale disturbances in the free shear layer, and a viscous mode that concerns the small-scale disturbances near the wall. They showed that the relative importance of each mode can be controlled by small amounts of blowing and suction.

Studies of an oscillating plane wall jet, in an external flow, have shown 10 to 40 percent reductions in shear drag, with minimal effect on maximum velocity decay and jet spreading rate (Fasel, et al., 1995). Detailed particle image velocimetry (PIV) measurements of an acoustically perturbed laminar plane wall jet showed that the perturbation enhances growth of a vortex in the outer shear layer. This, in turn, interacts with the inner layer, resulting in a counter-rotating vortex pair. (Shih and Gogineni, 1995). The vortex pair remains attached to the wall under the influence of the downstream vortex pair until it is further diffused downstream. When the vortex pair is dislodged from the surface, jet spreading and transition to turbulence follow. The forcing frequency determines the distance between adjacent vortex pairs, which in turn controls the flow field.

Little has been done to numerically simulate the interaction of turbine blade bleed jets with the primary flow. Vogel (1994) developed a steady solution of the Reynolds-averaged Navier-Stokes, continuity, and energy equations to model flow over a turbine blade section with film cooling. Internal coolant geometry was modeled as well as the outer flow region. The model demonstrated vortex development, typical of jets in crossflow, and compared favorably with flow visualization experiments that were conducted. The effect of the steady-state coolant jets on the shear drag, if any, was not reported.

Clearly, more study is needed to see if a periodic jet can reduce the shear drag on a turbine blade and simultaneously control the flow dynamics. No data concerning vortex development and control with oscillating cooling flow on an airfoil appears to have yet been published. Hence, it is believed that the current study will is the first examination of drag reduction and boundary layer structure on a turbine blade with an oscillating coolant flow. The status of this effort is discussed.

Methodology

VBI Code Description

The software used for the simulation of the governing equations was developed by Allison Engine Company, under U. S. Air Force Contract F33615-90-C-2028 for Wright Laboratory, to study vane-blade interaction, as described by Rao, et al. (1994a, 1994b). This software was verified and revised as part of the proposed effort for the addition of cooling as described above.

The conservative forms of the transient Reynolds-averaged Navier-Stokes, continuity, and energy equations are solved on a blade-to-blade stream surface of revolution. A numerical finite difference

technique is used, with central differencing for second order accuracy in space, and a five-stage Runge-Kutta algorithm for second order accurate integration in time. An artificial dissipation model that blends second and fourth order differences is added to damp out non-physical oscillations produced by central differencing. It utilizes pressure as a sensor to capture physical discontinuities such as shock waves and stagnation points.

The code uses a body fitted hyperbolic O-grid embedded in a rectangular H-grid as shown in Figure 1 (O-grid only). The outer H-grid resolves the free stream flow and the O-grid is used in the boundary layer region, with fine grid resolution near the surface.

Non-reflective inflow and outflow boundary conditions are calculated based on the methodology developed by Cline (1977). No-slip conditions are used on the airfoil surface(s) and periodic boundary conditions are used in the polar direction. The interface between the stator exit and the rotor inlet can be modeled with overlapping H-grids and a time-space phase-lag procedure, originally developed by Erdos (1977).

In the numerical simulation, body-fitted curvilinear coordinates are utilized and the flow is mapped to uniformly spaced rectangular coordinate region with the Jacobian matrix of the transformation. Also, the variables are non-dimensionalized. The second viscosity coefficient, λ, is set equal to -2/3 μ, where μ is the dynamic viscosity, and the Prandtl number is constant.

The prior-existing two-dimensional code uses the algebraic Baldwin and Lomax (1978) model to account for turbulence. The new two-equation algorithm is based on the recently developed point wise k-$\mathcal{R}$ model (Goldberg, 1994), discussed in the introduction.

The transport equations for the turbulent kinetic energy, k, and the undamped eddy viscosity, $\mathcal{R}$, are:

$$\frac{\partial Q}{\partial t}+\frac{\partial F_j}{\partial x_j}=\frac{\partial F_{v_j}}{\partial x_j}+\vec{S}_{source}, j=1,2 \qquad (1)$$

where:

$$Q=\begin{bmatrix}\rho k\\ R\end{bmatrix}, F_j=\begin{bmatrix}\rho u_j k\\ u_j R\end{bmatrix}, F_{v_j}=\begin{bmatrix}(\mu+\frac{\mu_t}{\sigma_k})\frac{\partial k}{\partial x_j}\\ (\nu+\frac{\nu_t}{\sigma_\varepsilon})\frac{\partial \mathcal{R}}{\partial x_j}\end{bmatrix},$$

$$\vec{S}_{source}=\begin{bmatrix}P-\frac{(\rho k)^2}{\rho\mathcal{R}}\\ -\frac{1}{\sigma_\varepsilon}(\nabla\nu_t\cdot\nabla\mathcal{R})+(2-C_{\varepsilon 1})\frac{\mathcal{R}P}{\rho k}-(2-C_{\varepsilon 2})k\end{bmatrix} \quad (2)$$

Equation 2 uses the conservative form of the transport equations for k and $\mathcal{R}$ as described by (Goldberg, 1994). The source term, P, is defined as the production of turbulent energy by the work of the main flow against the Reynolds stresses:

$$P=\left[\mu_t\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}-\frac{2}{3}k\delta_{ij}\right)\right]\frac{\partial u_i}{\partial x_j} \quad (3)$$

The latter form of P is based on the Boussinesq approximation (1877) and is the form used by Launder and Spalding (1974). The u_i terms represent the mean velocity components and u_i' is the fluctuating velocity component in the i-direction. The eddy viscosity, μ_t, is defined as:

$$\mu_t=C_\mu f_\mu \rho\mathcal{R} \quad (4)$$

The two-equation k-$\mathcal{R}$ model affects the transport equations for momentum and energy and these effects are summarized here due to lack of completeness in the literature. The eddy viscosity, μ_t, is added to the dynamic viscosity, μ, including the expression for second viscosity coefficient. The μ/Pr terms are replaced with $\left(\frac{\mu}{Pr}+\frac{\mu_t}{Pr_t}\right)$, as in the Baldwin and Lomax (1978) model. Pr_t is typically set to 0.9 for air flows and is the value used in the prior-existing software.

The constants σ_k, σ_ε, $C_{\varepsilon 1}$, $C_{\varepsilon 2}$, and C_μ are chosen to be 1.0, 1.3, 1.44, 1.92 and 0.09, respectively, and are the standard coefficients as recommended by Launder and Spalding (1974) for plane jets, mixing layers and flows near walls. The function f_μ is unity in the Launder and Spalding model, and was modified by Launder and Sharma (1974) and later by Lam and Bremhorst (1981) to extend the model to near-wall regions. Goldberg (1994) used a similar formulation:

$$f_\mu = \frac{1 - e^{-A_\mu R_T^n}}{1 - e^{-A_\mu R_T^n}} \tag{5}$$

where R_T is a form of a turbulence Reynolds number:

$$R_T = \frac{k^2}{\nu\varepsilon} = \frac{\mathcal{R}}{\nu} \tag{6}$$

The constants, A_μ, and A_ε were chosen to be 2.5×10^{-6} and $C_\mu^{3/4}/2\kappa$ ($\kappa = 0.41$), respectively, and $n = 2$, by prior experimentation (Goldberg, 1994).

The boundary conditions are as follows:

1. Set $\mathcal{R} = k = 0$ at solid walls.
2. Set $\mathcal{R} = O(10^{-5})$ at the freestream and initial conditions.
3. Prescribe the freestream k based on a given turbulence intensity, Tu, using: $k = \frac{3}{2}(TuV_\infty)^2$.
4. Extrapolate k and $\mathcal{R}$ from interior points to outflow boundaries.

Incorporating the k-$\mathcal{R}$ model into the original VBI software, required the transformation of Equations (1) and (2) to the 2-D meridonial coordinate system used by Rao, et al. (1994a, 1994b). Furthermore, the meridonial coordinate system equations were mapped to a body fitted coordinate system with the Jacobian matrix of the transformation. The convective terms in the transport equations used first order upwind differencing (for stability), while the diffusive and source terms used the standard central-type discretizations (Ervin, 1996).

Addition of Coolant Mesh

A methodology for the addition of a tangential cooling slot (H-grid) to the existing code has been developed. The required modifications to the software as well as additional pre-processing software are currently in progress. The film cooling is being modeled in a manner similar to that of Vogel (1994). The ejection region requires a separate H-mesh to represent the flow channel (Figure 1, 2). The location and size of the slot are variable. The inlet pressure and velocity will vary periodically to create an oscillating flow at the blade surface. At the coolant exit, the H-mesh will interface with the outer O-grid using the chimera scheme (Benek et al., 1985) that is also used by VBI for the interface of the O-grid

with the outer H-mesh. Mesh edge points that overlap the neighbor mesh are called "fringe" points and are identified for interpretation by the flow code.

CFX Code Description

The commercial CFX software (version 4.1) was developed by AEA Technology, Advance Scientific Computing of Waterloo, Ontario. The Navier-Stokes, continuity, and energy equations are solved on a finite volume mesh, using an iterative process. The mesh is built on a multi-block structure with grid refinement capability within each block. A semi implicit method for linking velocity and pressure, based on SIMPLEC is used and the software features several methods for time dependence, such as adaptive time stepping. Higher order schemes for the advection terms are available and the solver features several iterative methods.

A two-dimensional model was developed for the Langston blade geometry using 8 blocks (Figure 3). An effort was made to simulate the O-grid of the VBI code. Here the grid spacing is some what finer than the grid used for the VBI calculations (the O-grid is shown in Figures 1 and 2). An incompressible model was used due the low Mach number of the flow.

The turbulence models used here are the standard k-ε (Launder and Spalding, 1974) and the low Reynolds number k-ε model (Rhie and Chow, 1983). Standard constants were selected for σ_k, σ_ε, $C_{\varepsilon 1}$, $C_{\varepsilon 2}$, and C_μ. Several other turbulence model models are available.

Results and Discussion

The following plots show results for different turbulence models. The calculations were performed at an axial chord Reynolds numbers of 53,000 and at a pitch to axial chord ratio, $p/c_x = 0.944$. The low Reynolds number was selected to compare the separation calculations. The VBI computations were done with both a laminar model and a Baldwin-Lomax turbulence model ($Tu = 0\%$) and finally, a k-$\mathcal{R}$ model with the initial turbulence level, Tu, equal to 0%. The VBI results are shown in Figures 4 and 5. All three models capture the separation region on the suction surface. In these figures the k-$\mathcal{R}$ model results tend to follow the laminar results, suggesting that the k-$\mathcal{R}$ model is more sensitive to separation. Note that s is the distance along the airfoil and s = 0 corresponds to the minimum axial location of the blade.

The CFX results are shown in Figures 6 through 8. The curves on these plots represent lines of constant pressure while the vectors represent the velocity. The CFX computations were done with a laminar model, the conventional k-ε model and the low Reynolds number k-ε turbulence model. In both of the k-ε models, the initial turbulence level, Tu, is 3.6%. Note that the conventional k-ε model does not capture the separation region on the suction surface, indicating the importance of having a low Reynolds number model for a two-equation turbulence model.

Conclusions

A fundamental study using a computational model of the full transient Navier-Stokes equations was used to examine low Reynolds number flows typical of a low pressure turbine stage. The calculations confirmed the phenomena of separation at low Re, low Tu and $p/c_x = 0.93$. Separation was not seen in the case of $p/c_x = 1.18$, at Re = 50,000, although the experiments did show separation for chord Reynolds number of 67,500 at this wide blade spacing (Tu = 0.5%). The computations showed the interaction of the competing influences of adverse pressure gradient and transition to turbulence on the suction surface, and their relationship to separation.

The point wise k-$\mathcal{R}$ turbulence model shows promise of showing the turbulence transition and separation on a low pressure turbine blade. It is note worthy that this model is similar to a low Reynolds number k-ε turbulence model. The low Reynolds k-ε turbulence model shows the wall separation rather than the conventional k-ε turbulence model. Improvements to the boundary conditions of the implementation of the k-$\mathcal{R}$ turbulence model into the VBI code need to be considered.

Concurrently, a fundamental study of the flow dynamics around a blade with oscillating cooling flow is well in progress. The implementation of the k-$\mathcal{R}$ turbulence model into the VBI code is nearly complete as is the addition of the cooling geometry.

Bibliography

Allen, M. G. and Glezer, A., 1995, "Jet Vectoring Using Zero Mass Flux Control Jets," presented at AFOSR Contractor and Grantee Meeting on Turbulence and Internal Flows, Wright Patterson AFB, Dayton, OH, May, 1995, pp. 95-100.

Baldwin, B. S. and Lomax, H., 1978, "Thin Layer Approximation and Algebraic Model for Separated Turbulent Flows," AIAA Paper 78-0257.

Benek, J. A., Buning, P. G. and Steger, J. L., 1985, "A 3-D Chimera Grid Embedding Technique," AIAA Paper 85-1523.

Boussinesq, Q., 1877, Theorie de l'Ecoulement Tourbillonant, Vol. 23, pp.46-50, Paris: Comptes-Rendus de l'Academie des Sciences.

Cline, M. C., 1977, "NAP: A Computer Program for the Computation of Two-Dimensional, Time-Dependent, Inviscid Nozzle Flow," Los Alamos National Laboratory Report LA-5984.

Cohen, J., Amitay, M. and Bayly, B. J., 1992, "Laminar-Turbulent Transition of Wall Jet Flows Subjected to Blowing and Sucking," *Physics of Fluids A*, Vol. 4, pp. 283-289.

Erdos, J. I., Alzner, E. and McNally, W., 1977, "Numerical Solution of Periodic Transonic Flow through a Fan Stage," *AIAA Journal,* Vol. 15, pp. 1559-1568.

Ervin, E. A., 1996, "Computations of Drag Reduction and Boundary Layer Structure on a Turbine Blade with an Oscillating Bleed Flow," AFOSR Final Report for Summer Faculty Research Extension Program.

Fasel, H., Ortega, A. and Wygnanski, I., 1995, "Convective Flow and Heat Transfer Due to a Forced Wall Jet," presented at AFOSR Contractor and Grantee Meeting on Turbulence and Internal Flows, Wright Patterson AFB, Dayton, OH, May, 1995, pp. 29-33.

Goldberg, U. C., 1994, "Toward a Pointwise Turbulence Model for Wall-Bounded and Free Shear Flows," *Journal of Fluids Engineering*, Vol. 116, pp. 72-76.

Lam, C. K. G., and Bremhorst, K., 1981, "A Modified Form of the k-ε Model for Predicting Wall Turbulence," *Journal of Fluids Engineering*, Vol. 103, pp. 456-460.

Launder, B. E., and Spalding, D. B., 1974, "The Numerical Computation of Turbulent Flows," *Computer Methods in Applied Mechanics and Engineering*, Vol. 3, pp. 269-289. Reprinted in Vol. 81, pp. 269-289 (1990).

Murawski, C. G., Sondergaard, R., Rivir, R. B., Vafai, K., 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," presented at the ASME International Gas Turbine & Aeroengine Congress & Exhibition, Orlando, FL, June 2-5, 1997, paper 97-GT-95.

Rao, K. V., Delaney, R. A., and Topp, D. A., 1994a, "Turbine Vane-Blade Interaction, Vol. 1, 2-D Euler/Navier-Stokes Aerodynamic and Grid Generation Developments," Wright Laboratory Report WL-TR-94-2073.

Rao, K. V., Delaney, R. A., and Dunn, M. G., 1994b, "Vane-Blade Interaction in a Transonic Turbine, Part 1 Aerodynamics," *Journal of Propulsion and Power*, Vol. 10, pp. 305-311.

Rhie, C. M. and Chow, W. L., 1983, "Numerical Study of the Turbulent Flow Past an Airfoil with Trailing Edge Separation," *AIAA Journal*, Vol. 21, pp. 1527-1532.

Shih, C. and Gogineni, S., 1995, "Experimental Study of Perturbed Laminar Wall Jet," *AIAA Journal*, Vol. 33, pp. 559-561.

Vogel, D. T., 1994, "Navier-Stokes Calculation of Turbine Flows with Film Cooling," ICAS-94-2.5.3.

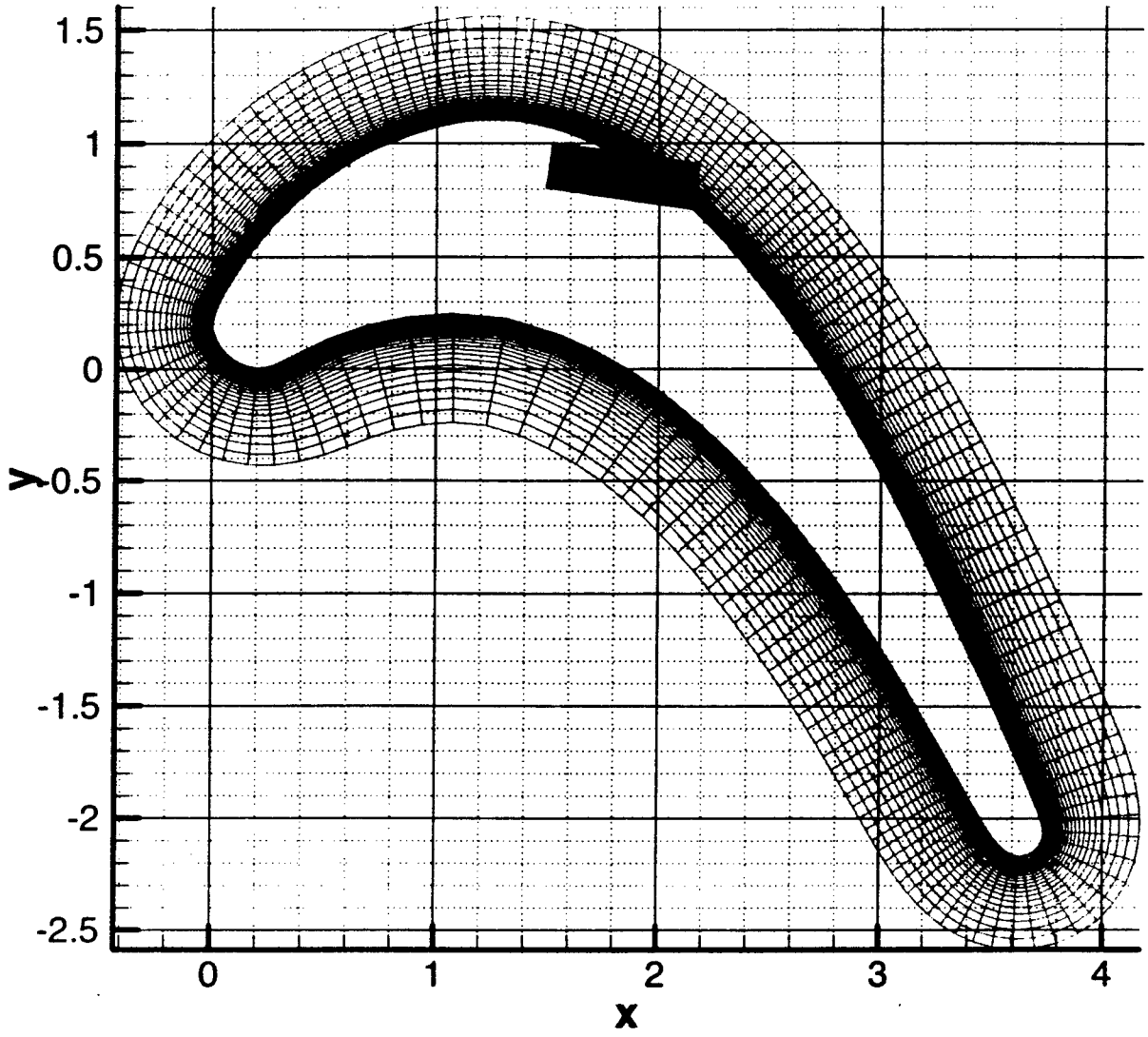


Figure 1. Slot H-grid and blade O-grid for VBI

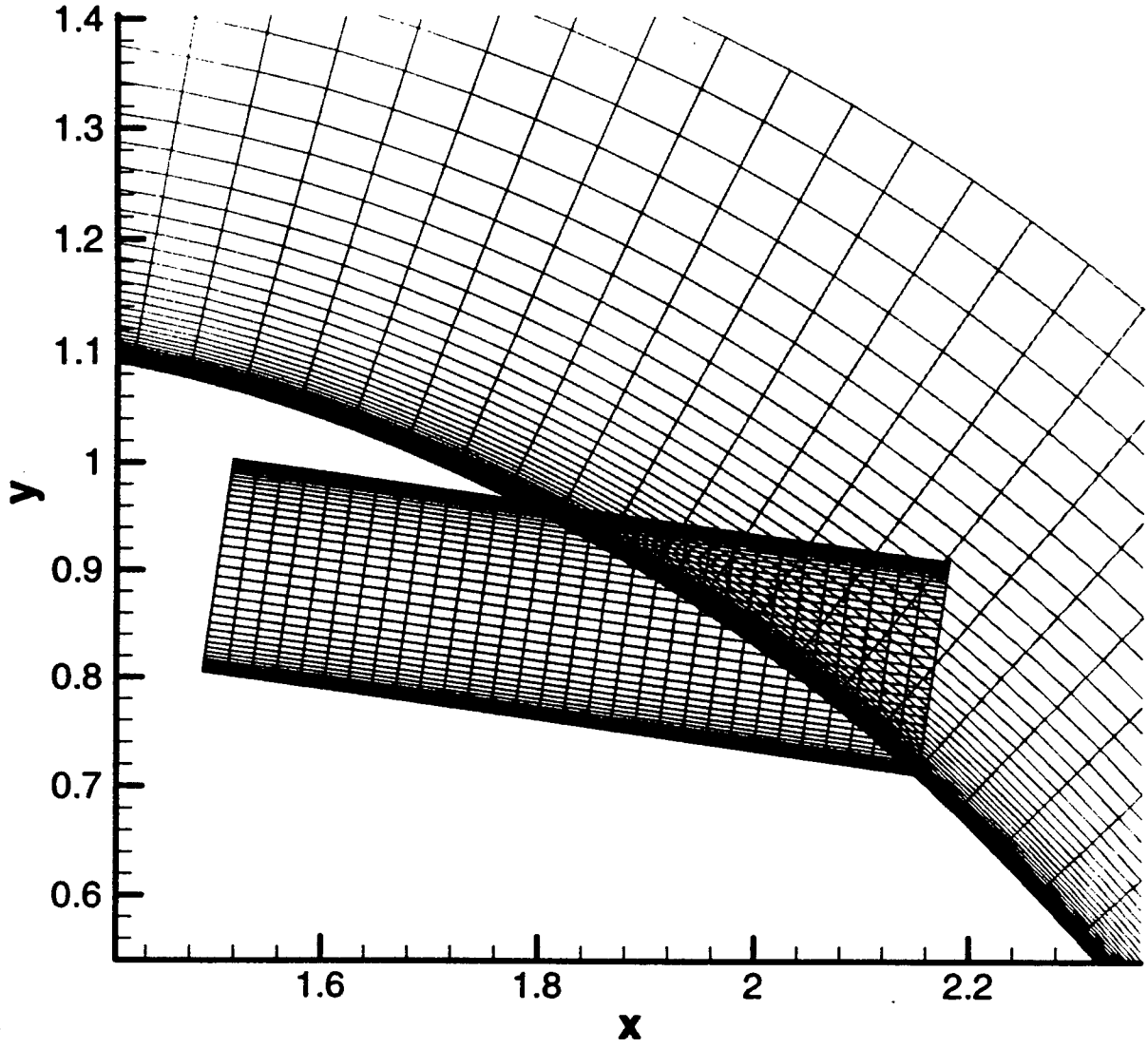


Figure 2. Close-up of slot H-grid and blade O-grid at intersection for VBI

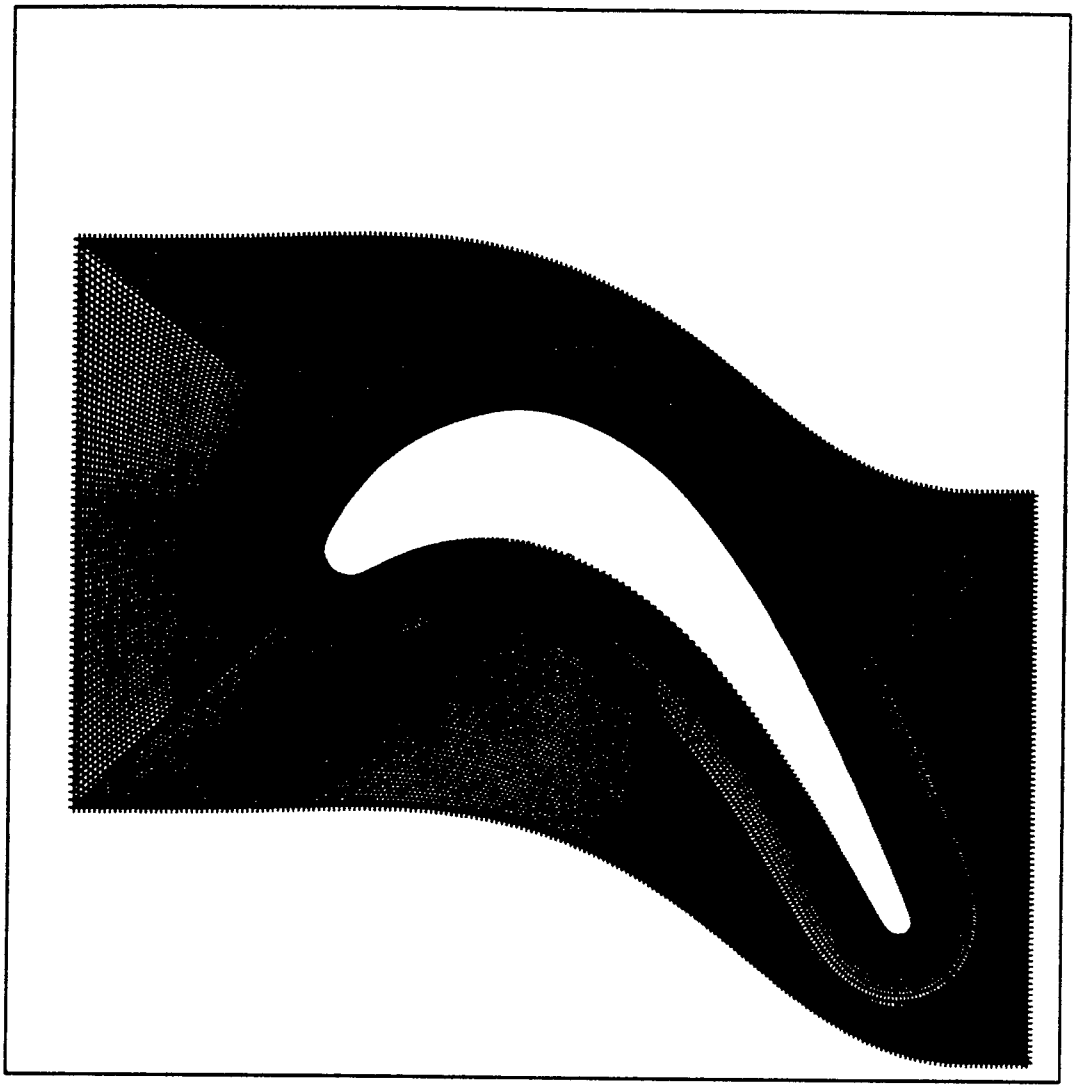


Figure 3. CFX grid for modeling flow over a Langston blade

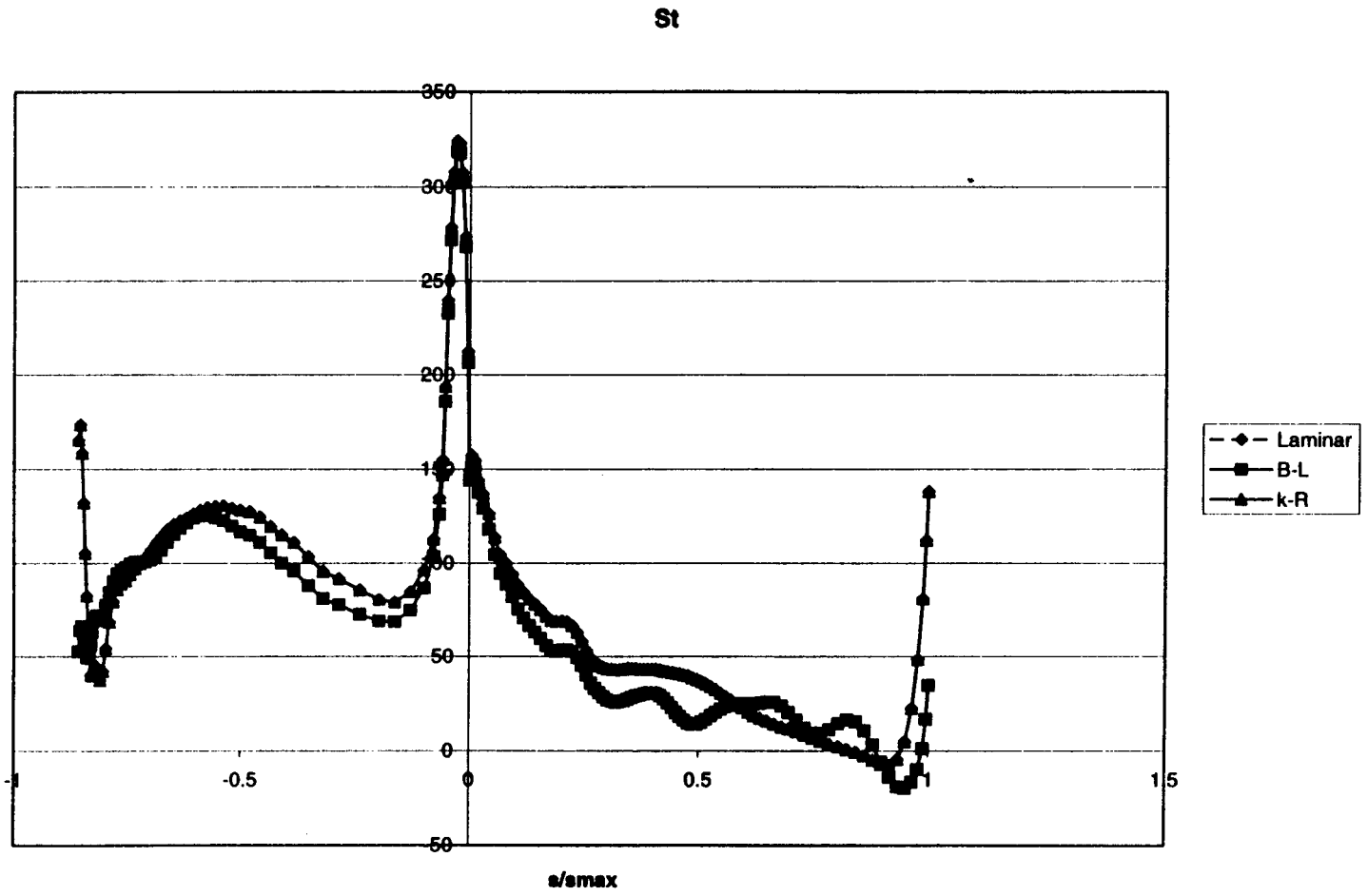


Figure 4. Comparison of VBI results at Re = 53k: Stanton number versus airfoil distance

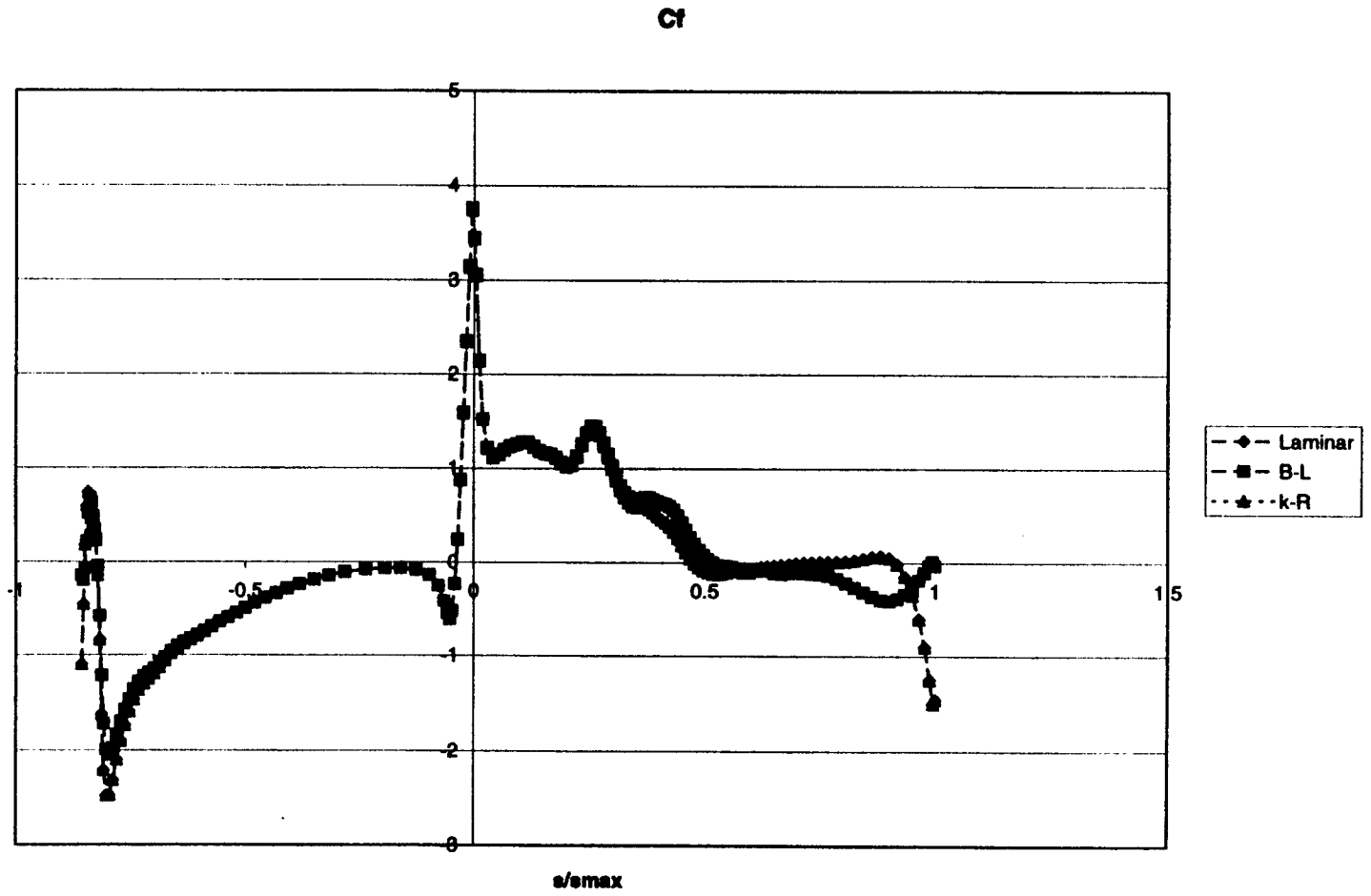


Figure 5. Comparison of VBI results at Re = 53k: Skin friction coefficient versus airfoil distance

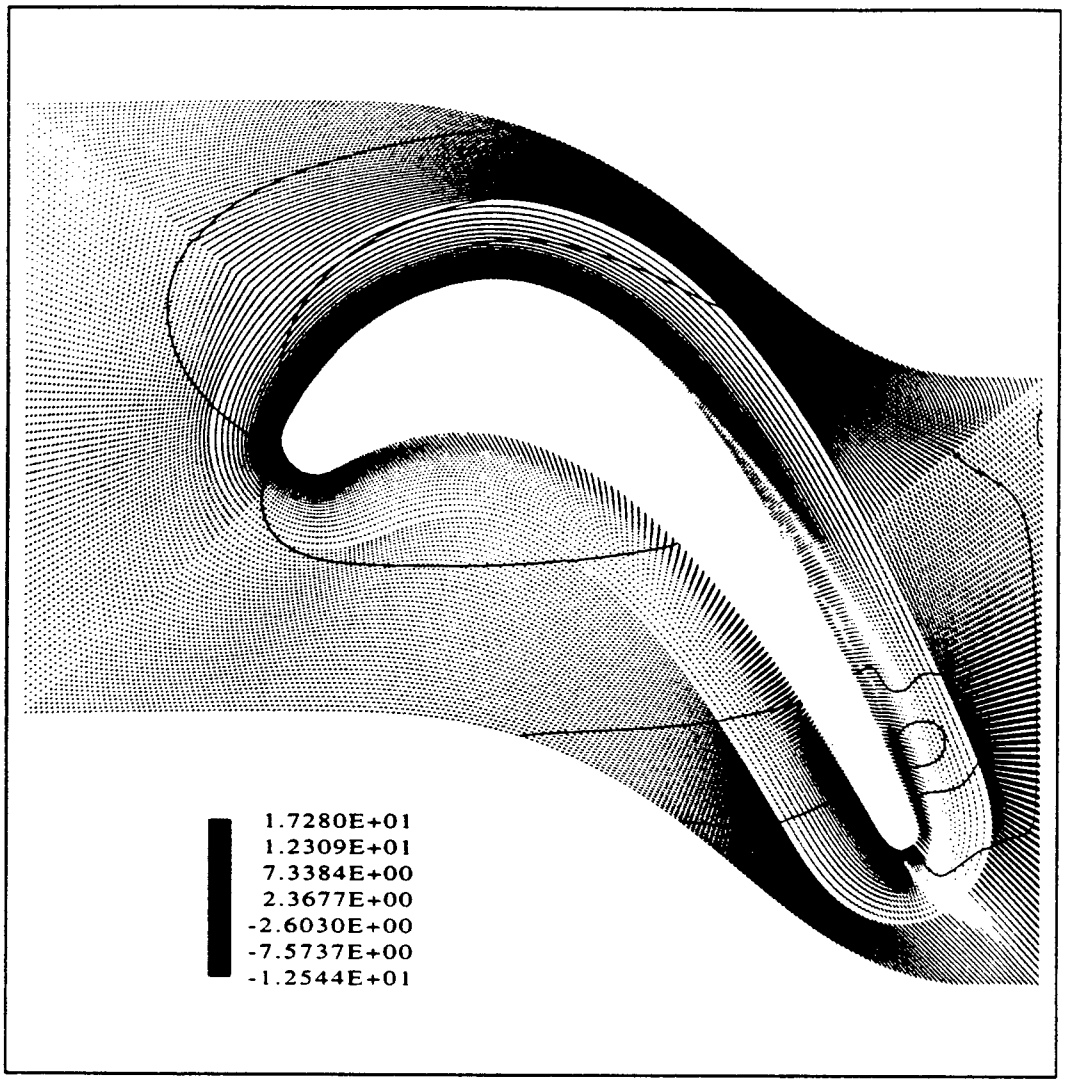


Figure 6. CFX results at Re = 53k: Laminar model

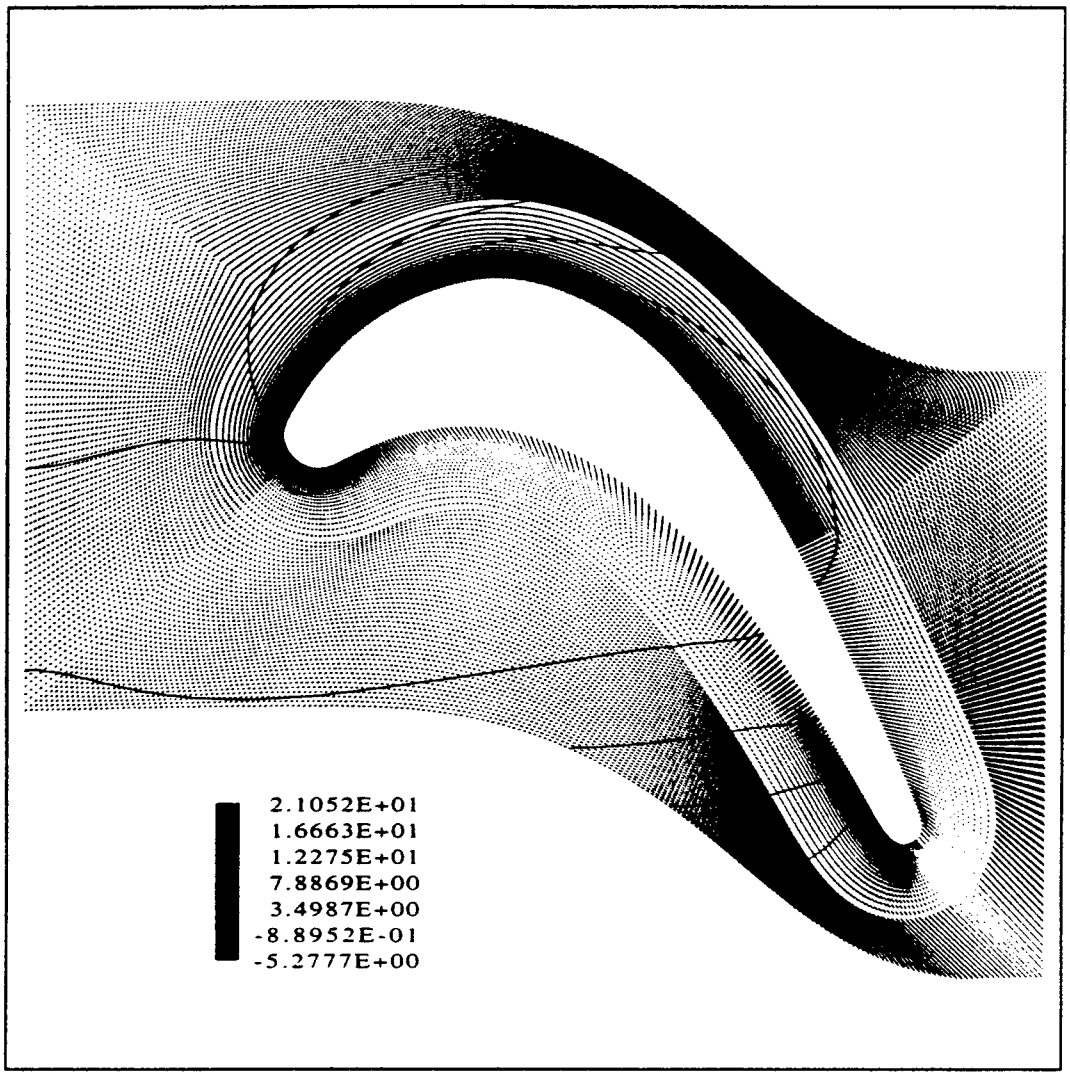


Figure 7. CFX results at Re = 53k: k-ε turbulence model

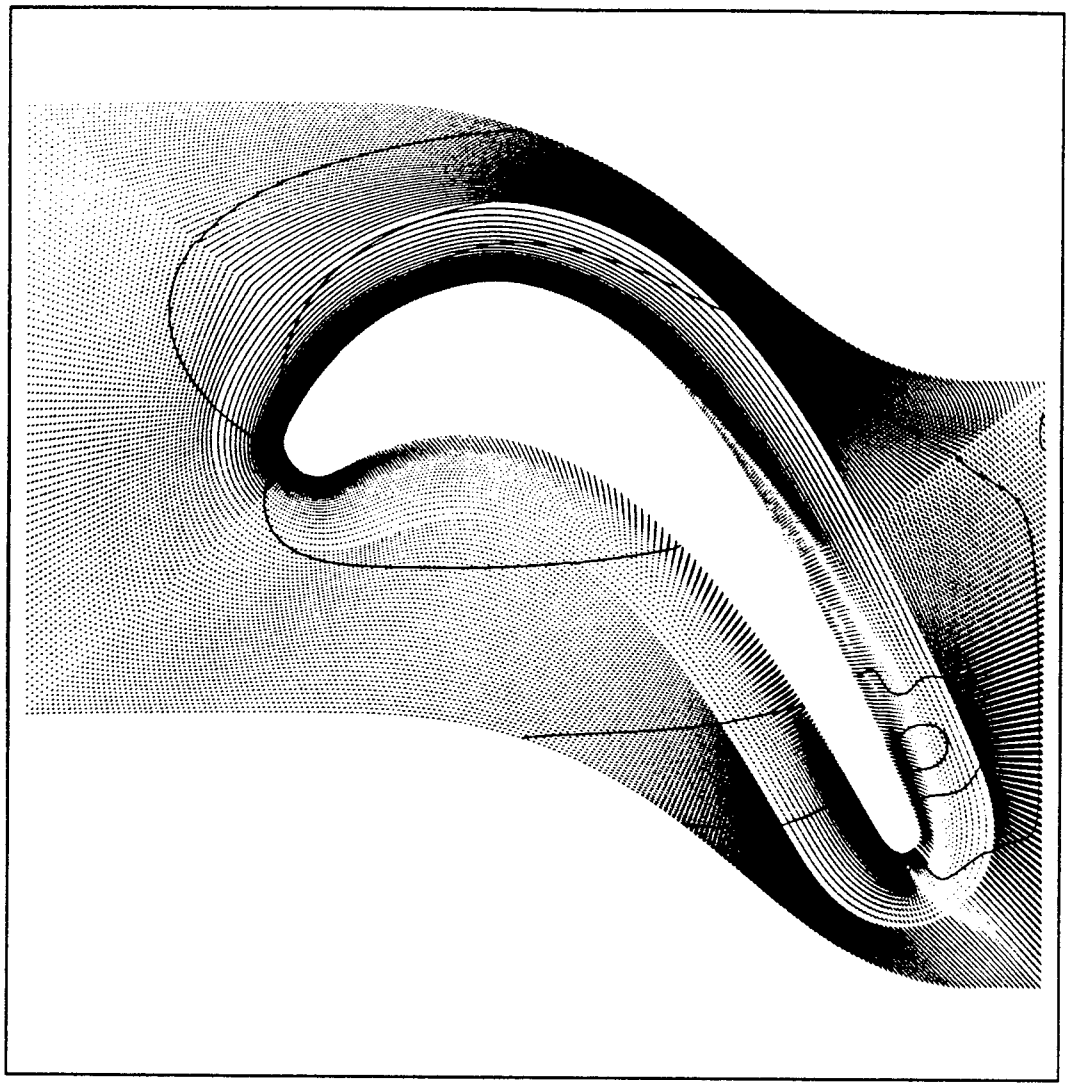


Figure 8. CFX results at e = 53k: Low Reynolds number k-ε turbulence model results

VERTICALLY INTERCONNECTED 3D MMICs WITH ACTIVE INTERLAYER ELEMENTS

Altan M. Ferendeci
Associate Professor
Department of Electrical and Computer Engineering and Computer Science

University of Cincinnati
Cincinnati, Ohio 45221-0030

Final Report for
Summer Faculty Research Program
Wright Patterson Air Force Laboratory

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

and

Wright Laboratory

August 1997

VERTICALLY INTERCONNECTED 3D MMICs WITH ACTIVE INTERLAYER ELEMENTS

Altan M. Ferendeci
Associate Professor
Department of Electrical and Computer Engineering and Computer Science
University of Cincinnati

Abstract

Vertically interconnected 3D monolithic microwave integrated circuit (MMIC) design is extended into a new novel configuration with active layers dispersed between the interlayers. In a conventional 3D MMICs various active devices including transistors, resistors and capacitors are all placed on the base substrate. The upper layers contain the passive elements for necessary matching networks, bias networks and transmission lines connecting various circuit elements. In the modified proposed configuration, high power devices (i.e., HBT developed at WL) are placed on the substrate and additional circuit elements including low power active devices are dispersed between various layers. This is possible with the recent advances made in SOI transistor technology. NMOS transistors can be processed by first deposition of polysilicon on any layer, followed by laser crystallization of the polysilicon islands and followed by transistor processing.

VERTICALLY INTERCONNECTED 3D MMICs WITH ACTIVE INTERLAYER ELEMENTS

Altan M. Ferendeci

Introduction

Electronically steerable phased array antennas have numerous application areas both in military and in the commercial market. At present, electronically steerable phased array antennas are heavy and bulky and they can not conform to all specific surface topologies. This is dictated by the surface to depth aspect ratio of a given unit. The T/R module, power amplifier, filters, phase shifters, antennas and the digital control circuitry are all have to placed in the same package [1]. In addition to increasing the weight of the unit, this makes the depth of these units very long compared to the surface area of the radiating antenna. Since the separation of each unit can not be greater than the free space half wavelength of the operating signal, this large aspect ratio dictates that these units can only be placed on a plane or on a line.

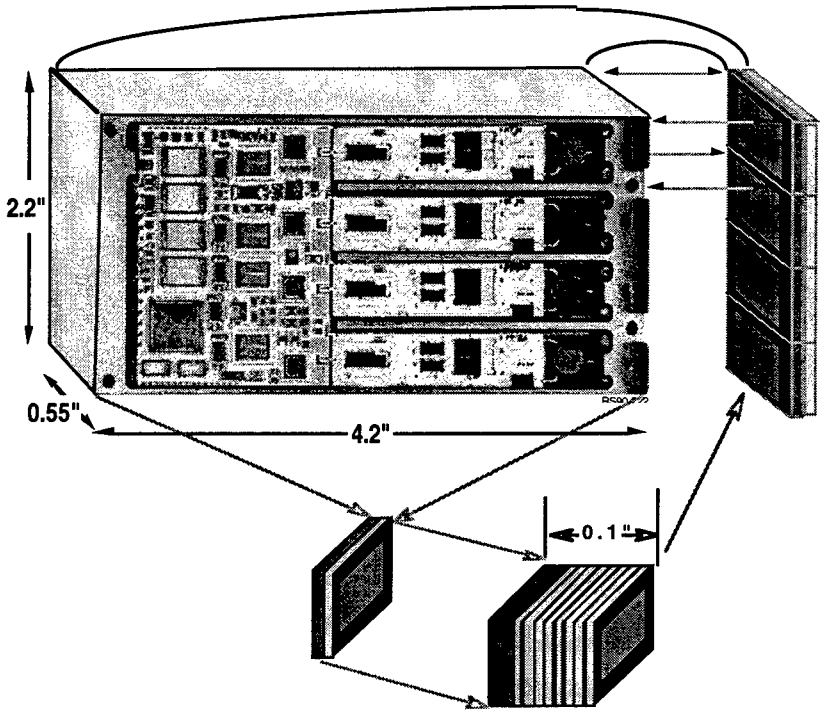


Figure 1. Comparison of present day Quad Phased Array Unit with the proposed 3D unit.

For many applications, conformity of the phased array antenna system units to various other surface topologies such as cylindrical, spherical and arbitrary surfaces will increase enormously the functionality and applicability of these units. In order to realize such a flexible unit,

the depth of these units should be very small compared to the surface area of each unit so that they can be placed in close proximity to each other even on odd surface geometry. With the advances made in patch type planar antennas and 3D vertically interconnected MMIC technology, it is possible to shrink the depth of the individual units to millimeter dimensions. To realize this reduction in geometry, various system circuits and components have to be distributed throughout various layers. In order to remove heat efficiently, high power components such as the power amplifier have to be placed on the first substrate layer. The other system components should be distributed within the unit with caution so that the they are not affected by the heat generated by the power units.

Figure 1 shows a typical present day quad phased array antenna system unit and the comparison of the same functional unit when implemented in a vertically interconnected 3D MMIC. The space and weight savings are enormous.

In this work, a 3D phased array antenna system is presented which greatly reduces the aspect ratio of the individual radiating units. In addition to incorporating all the necessary components of the unit in a small volume, various active circuits in the form of SOI transistors are distributed within the interlayers of the 3D system. In this way, more compact units with all necessary system components are placed within various layers whose overall thickness does exceed a few millimeters. These units can then placed over any conformal surface.

Phased Array Antenna System

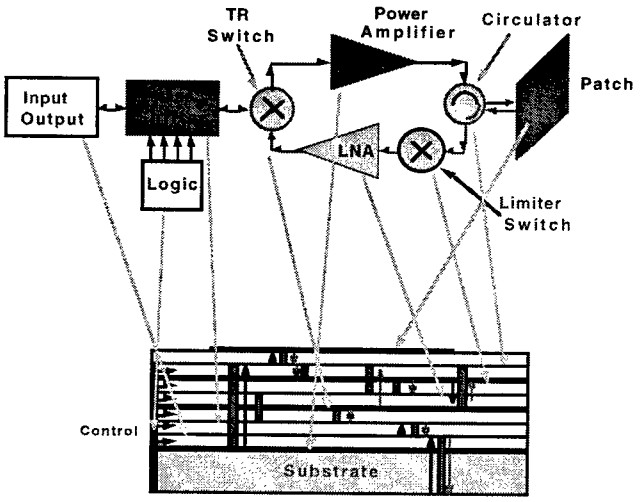


Figure 2, Basic building blocks of a T/R module for phased array antenna unit.

Figure 2 shows the fundamental system components of a phased array antenna unit used in the proposed electronically steerable phased array antenna system. The patch antenna is placed on the upper layer. The power amplifier is placed on the substrate so that the heat generated by the power amplifier can be efficiently removed without drastically affecting the performance of the other active devices. The intermediate levels are made up multiple dielectric and metal layers. Some of the metal layers are processed as circuit planes and are sandwiched between upper and lower metal planes which function as ground planes. Various circuit and ground planes are connected by vertical interconnects in the form of vertical posts processed by via hole metal fills.

Conventional 3D MMIC is comprised of a configuration where all active devices including resistors and capacitors are placed on the substrate and the connection between various system elements and circuit components are made through the vertical interconnects [2,3]. All upper layers contain passive components such as the matching network elements and transmission lines connecting various system components. According to the published literature, only receiver circuitry is integrated in this 3D fashion. At present, no power devices such as power amplifiers and high power oscillators are incorporated into any 3D configuration. One of the major problems associated with the incorporation of power dissipation is the removal of the heat generated by the power devices. All vertically interconnected circuitry that has been developed so far uses GaAs as the substrate material and various low power transistors are processed on this substrate. If power devices are not properly incorporated into such a substrate, it will affect drastically change the operating characteristic of the neighboring active devices due to the increase in the temperature of the substrate.

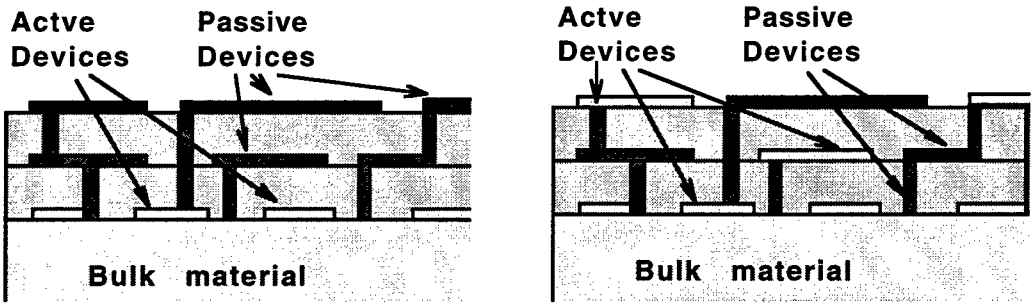


Figure 3. a) Conventional 3D MMIC and b) modified MIMIC with interlayer active devices.

Patch antennas are planar antennas that are replacing the conventional waveguide type antennas. The patch antennas occupy very small space vertically. The radiating element is separated from the ground plane by a dielectric layer whose dielectric constant and thickness determines the radiation efficiency of the antenna. In a phased array antenna system, the spacing between each

antenna element is restricted to be equal to or smaller than one half of free space wavelength at the operating frequency. For X-band operation this is roughly 15 mm between each antenna elements. This distance is the maximum distance allowable in order to prevent high side lobe generation by the phased array antenna system.

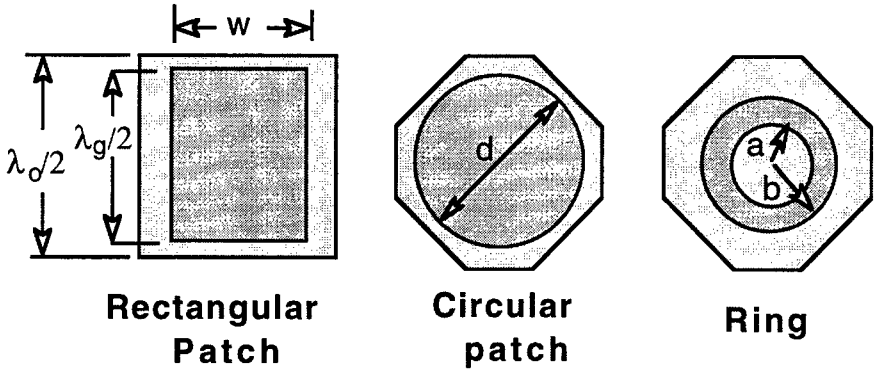


Figure 4. a) Rectangular, b) circular and c) ring patch antennas

The size of the radiating element is determined by the shape of the patch. Figure 4 shows rectangular, circular and ring patch antennas. In each case the size of the radiating element is determined by the restricted resonance conditions and is always smaller than the antenna separation. Therefore, the complete circuit can be placed over an area close to 15 mm square. This seems relatively large area but as can be seen from Figure 2, there are many systems components that have to be placed on such a small area. This is especially true if various filters and couplers are to be placed on such a substrate. It is therefore necessary to build the circuit vertically. The relatively large horizontal surface area that would have been occupied in a 2D integration can be vertically expanded thus shrinking the surface area to the required dimensions.

The ideal approach to the implementation of a 3D MMIC circuit is to distribute the low power active elements especially the active devices such as low power transistors into various interlayers. This will provide additional freedom in the placement of the system components. This ideal configuration is shown in Figure 3. This configuration requires development of a technology of either attaching chip transistors on various interlayers or somewhat directly processing the transistors within these layers.

Both of these technologies have advantages and disadvantages. Chip transistor attachment is ideal. In this system of 3D interconnect, HEMT or MESFET transistors with excellent frequency response characteristics can be processed separately, diced and attached to the proper layer. In order to build up the upper layers, the chip has to be thinned initially to close to a few micron dimensions. In this way, the upper layers that will be subsequently added will not loose their planarity. Otherwise, there will be bumps and valleys as the upper layers are processed. These will make the following upper layer processing very difficult, especially if there has to be transistors

within these layers. The connection between the transistors and the circuit elements on the same layer has to be done using wire bonding or any other hybrid process which will further complicate the overall 3D circuit implementation.

The second alternative approach is to process the transistors directly on a given intermediate level. At present time, growth of compound semiconductors at the intermediate levels that are being considered here is not possible with the present day processing technologies. Growth of good quality thin compound semiconductor layers have only been possible on substrates having the same or close lattice constants. Since the upper layers will contain ether metal or dielectric material, direct growth of compound semiconductors on these materials can be ruled out from such a 3D MMIC application.

AN alternative but a viable possibility is to process NMOS transistors directly on the upper levels [4,5,6]. Similar application of NMOS transistors have been demonstrated in a vertically interconnected 3D circuit implementation. Even though this application is concentrated on the implementation of digital circuitry at relatively low operating frequencies, it can easily be extended to the realization of microwave circuitry operating at higher frequencies [7,8]. There are highly encouraging recent developments in SOI NMOS transistor technology [9,10]. A 4 layer vertically interconnected digital circuitry using NMOS transistors dispersed within the intermediate layers has already been processed [11]. This circuitry demonstrated the feasibility of implementation of such a 3D concept [12].

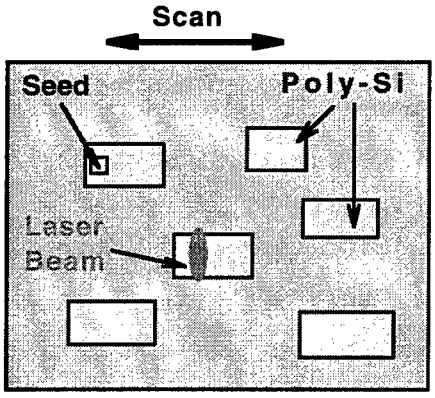


Figure 5. Laser crystallization of polysilicon islands.

The basic building block for this type of 3D circuitry is the laser crystallization of polysilicon islands [13]. Polysilicon is deposited over a given intermediate layer, followed by laser crystallization of polysilicon and finally processing of the NMOS transistor structure. The connection between the transistor and other circuit elements are made through patterned metal deposition.

Figure 5 shows the top view of an intermediate level. Two possibilities exist for the polysilicon deposition. Polysilicon can be either directly deposited over metal surface which has been deposited over a dielectric layer or directly on the dielectric layer itself. The islands where the transistors has to be processed are then located. As shown in Figure 6, using a focused Argon ion laser, the polysilicon islands are crystallized by heating the substrate initially to 350°C followed by scanning the substrate in the required x-y directions.

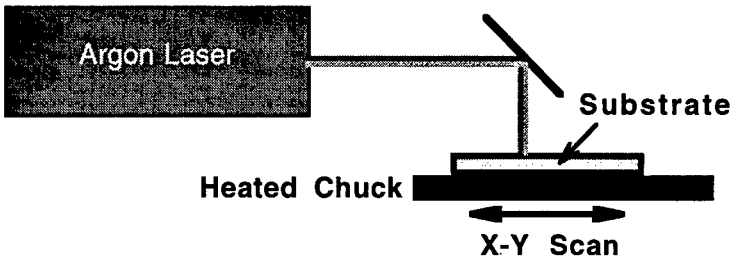


Figure 6. Laser set up for polysilicon crystallization.

The Argon ion laser beam is absorbed by the polysilicon film and as polyilion is crystallized, it's absorption is reduced. thus providing controlled crystallization of the polysilicon.

One of the major problems associated with this technology is the successful crystallization of polysilicon in the preferred crystal direction. Various schemes have been used to achieve this goal.[14]. Thin stripes of Si_3N_4 reflective layers have been deposited and laser beam is scanned in a given direction to crystallize the polysilicon in the <100> plane [15]. In an alternate procedure, extension of the lower substrate layer into the upper layers by Silicon islands or silicon posts grown over the substrate are used as a seed material to initiate crystal growth in the same direction as the substrate material. Although this second technique gives more reliable crystal growth, the first technique is more applicable to multiple layer systems.

If the polysilicon is deposited directly over metal and subsequent crystallization is achieved, O_2 may have to be implanted into Silicon at very high energies to produce a SiO_2 to separate the silicon from the metal. If initially polysilicon is deposited over the dielectric, there may not be need for O_2 implantation and the transistor can be directly processed on the silicon islands.

It is very important that the crystallinity of the islands in each layer is highly reproducible. Once this is achieved, then transistors can be processed on these silicon island. It is also important that subsequent dielectric layer depositions generate planar surfaces for the following upper layers. This is necessary to process submicron gate lengths on the interlayer transistors.

Even without the insertion of active devices in the intermediate layers, there are still many fundamental process technologies that have to be developed in order to successfully realize the vertically interconnected conventional 3D MMIC.

Implementation of 3D MMIC

In order to investigate some of the major processing steps, a simple transmitter circuit will be considered. Successful realization of the various processing steps will provide the basic knowledge base for the implementation of the electronically steerable phased array antenna unit. The basic circuit consists of a power amplifier (or an injection locked oscillator), a filter and a patch antenna as shown in Figure 7.

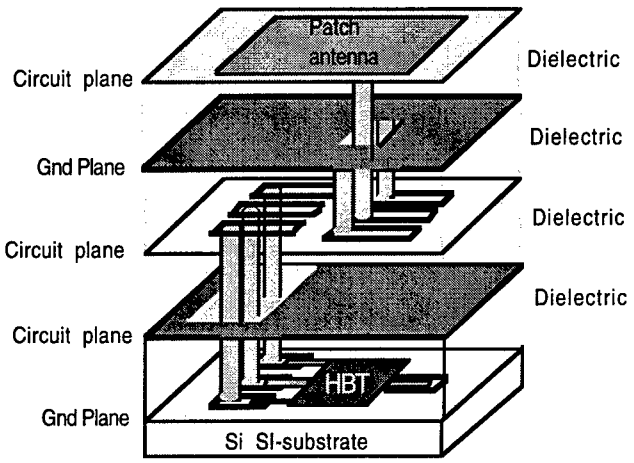


Figure 7. Vertically interconnected simple 3D transmitter unit.

Since high frequency NMOS transistors responding to millimeter wavelengths can be successfully processed on SOI substrates, the question can be raised with regards to using the same SOI substrate for the conventional 3D MMIC. The commercially available best high resistivity Silicon wafer has a resistivity of 10KΩ–cm compared to >MΩ–cm for GaAs substrate. This low resistivity provides lossy circuit elements leading to poor circuit performance. Therefore, use of Silicon as a substrate material for microstrip lines and any other passive components will deteriorate the performance of the circuits containing these elements.

In order to verify the properties of a 4 KΩ-cm substrate, a microstrip line with a side coupled ring resonator is processed on a gold plated substrate Figure 8. The losses were measured as a function of frequency. The result is shown in Figure 7b for a 4 KΩ-cm substrate. These losses become large at higher frequencies thus limiting the usefulness of the Silicon wafers as

substrates. IF SOI substrates are used, they have very low resistivities to start with, therefore they are absolutely unsuitable as substrate materials for MMIC applications

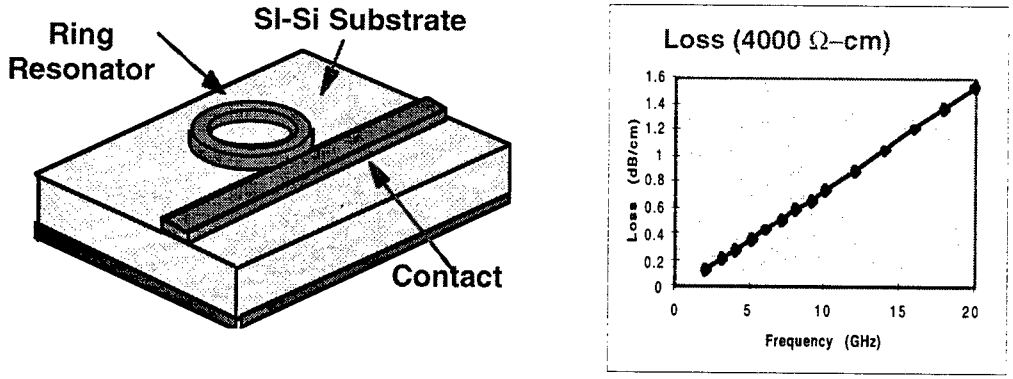


Figure 8. a) Circuit used to measure the losses associated with a Silicon substrate. b) loss (dB/cm) as a function of frequency

In order to verify the effect of the equivalent dielectric losses on the circuit elements, especially on tuned circuit, a simple parallel LC circuit is chosen. A spiral inductor is shunted by a capacitor and the corresponding S parameters are simulated assuming a silicone material as the microstrip substrate. These simulations are repeated under the same conditions by replacing the silicon by a polyimide substrate having the same thickness As can be seen from Figure 9, the Q of the circuit with a Silicon substrate is lot smaller than a polyimide substrate.

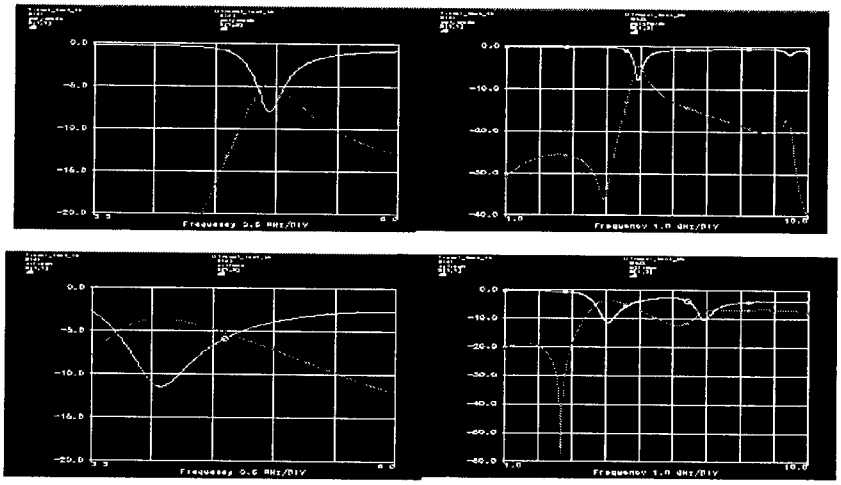


Figure 9. S_{11} and S_{21} for a parallel LC resonant circuit. a) with polyimide and b) 4 KΩ-cm silicon substrate.

On the other hand Silicon has a very high thermal conductivity [16]. Since it is also relatively cheep and can be obtained in large wafer sizes, silicon is still a prime candidate for use as the basic substrate for the 3D MMIC. The power transistor especially the power BJT with thermal

shunt can be attached to the Silicon substrate. By depositing a metal layer over the substrate, silicon is separated from the rest of circuit planes and the upper layers which is made up of low loss polyimide dielectric material can then be processed. This is accomplished by attaching the HBT on the silicon and processing of the rest of the circuit elements including bias and matching networks on the upper layers.

A hybridization technique will be used in attaching the HBT on to the silicon wafer with a novel flip-chip technique. Metal is deposited over the silicon substrate first. Then a portion of metal is etched away from the surface to create an opening equal to the required transistor size.

A well equal to the size of the transistor is then processed by using the well know MEMS technology . Two step process creates the opening shown in Figure 10a. Proper metal strips are then deposited within this area. Thin solder pads (within micron thickness) are then deposited over these stripes within the well. The transistor is flipped over and once it is in place, the substrate is heated so that bonding of the transistor pads to the strips is achieved. The hybridization of this process requires that the HBTs have all consistently the same finished dimensions.

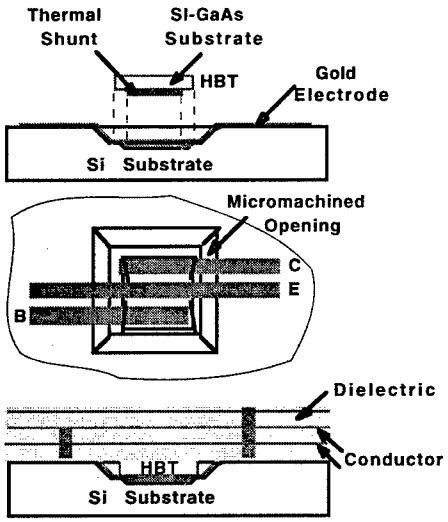


Figure 10. Micromachining of Silicon wafer for HBT attachment.

Micromachining of silicon is a highly developed technology and processing of precise well openings poses no problems [17]. In order to show the feasibility of the proposed novel flip chip type transistor attachment to the silicon substrate, two wells of dimensions 100x80 μm^2 and 100x120 μm^2 square opening on a silicon substrate are processed. Initially SiO_2 is deposited over silicon, Photo resist is then spin coated over SiO_2. It is then exposed using a photographic plate and contact printing. Photoresist is processed and openings with the above sizes are generated. SiO_2 is removed from these islands. Then the rest of photoresist is stripped away from the surface.

The substrate is then etched in a KOH solution. Once proper depth in silicon is reached the etch is stopped. Finally, SiO_2 is then completely removed from the surface. Figure 11 shows the resulting well that is obtained with this process.

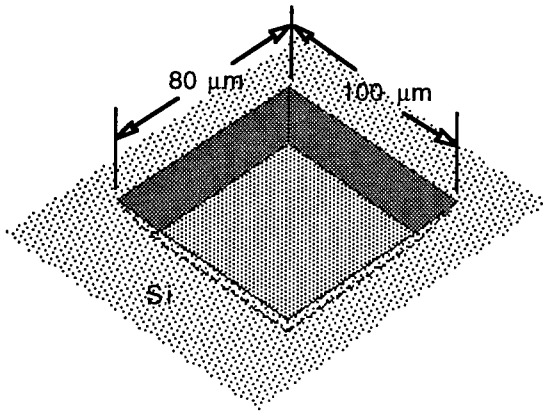


Figure 11. Micromachined well in Silicon. The depth is 23 µm.

Once the HBT is attached to the silicon substrate, the upper layers can then be processed. The next step is to deposit by spin coating the dielectric material (polyimide) over the whole substrate area. This will also fill in the voids generate between the well opening and the HBT transistor.

There are a few fundamental issues that have to be resolved for dielectric deposition process. The minimum dielectric thickness that will provide acceptable line losses and, at the same time, that can be deposited with minimal number of processing steps has to be established. Deciding on the minimum polyimide thickness that can provide sufficiently acceptable low line losses is very important. In order to establish this, loss per unit half wavelength are calculated for both a 50Ω microstrip and a 50Ω stripline transmission lines with different substrate thickness assuming gold metal for the electrodes (0.5 micron thick), loss tangent of 0.001 and ε_r=3.0 are used for the polyimide. As expected, the losses became larger as the polyimide thickness gets thinner, as can be seen from Figure 11.

According to DuPont, manufacturer of the polyimide that will be used as the dielectric material for the 3D, the maximum thickness that can be spin coated at once is 12-13 µm. According to Figure 12, the losses for 10 and 25 µm thick substrates are 1.9 dB/(λ/2) and 0.76 dB/(λ/2) respectively at 10 GHz. Assuming that the same polyimide thickness can be spun on every time, a

single coating of 10 micron thick polyimide can be used for many circuit implementations. If high Q circuits are required, additional polyimide layers can be deposited to increase the layer thickness.

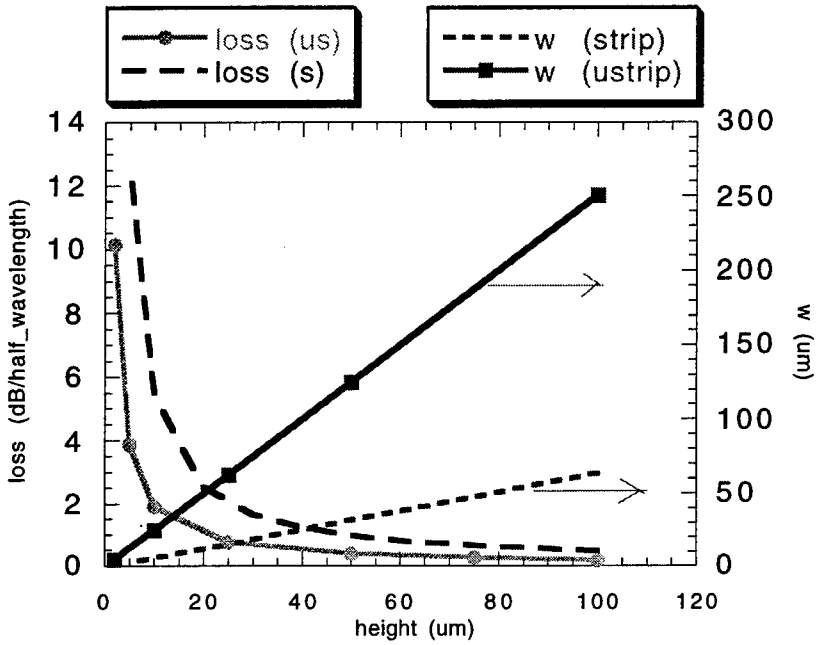


Figure 12. Loss and width of microstrip and stripline as a function of polyimide thickness.

A major problem with polyimide dielectric material is the adhesion of polyimide to metal and metal to the polyimide. But there is ample available literature that addresses the solutions to these problems [18]. Since there are already many 3D digital circuitry that has been realized using polyimide dielectric materials, it is expected that successful deposition of polyimide and subsequent metal depositions will pose no problems.

The first metallization layer over the first polyimide layer will act as the ground plane and the second layer will be the circuit plane [19, 20, 21]. Over the circuit plane another ground plane will be deposited.

The distributed circuits that will be processed may have two possible configurations. If the ground planes cover the whole surface area of the substrate area, the transmission lines will be striplines, otherwise they will be microstrip-like transmission lines.

Figure 13a shows the stripline type configuration. The circuit elements are sandwiched between two metal planes separated with a dielectric material of ε_r. In the second configuration shown if Figure 13b, the ground planes do not completely cover the substrate surface. A finite width ground plane with dimensions a few times greater than the width of the microstrip line is sufficient to confine the fields between the strip and the ground plane [22]. It should be mentioned that in this configuration, the upper surface of the microstrip is not exposed to air but is also

surrounded by the same dielectric material. The upper dielectric thickness should be at least a few times larger than the strip-ground plane height in order to maintain its microstrip_like characteristic. Otherwise, the line parameters will be affected by the presence of the upper finite ground plane. This scheme may minimize the thermal effects due to the differences in the thermal expansion coefficients of the dielectric material and the metal, but has to designed very carefully into the system.

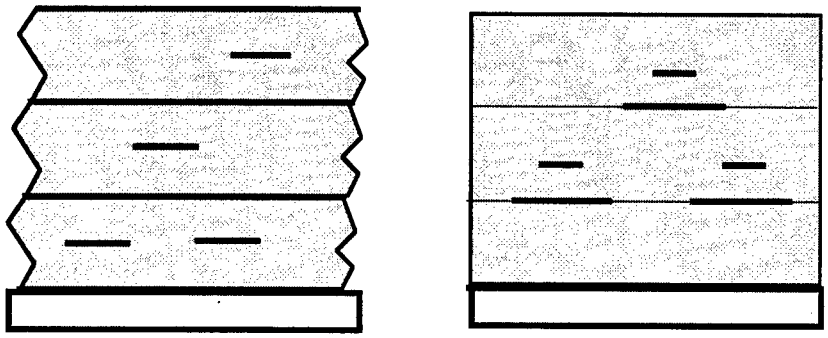


Figure 13. Stripline and microstrip_like (finite ground plane) transmission lines.

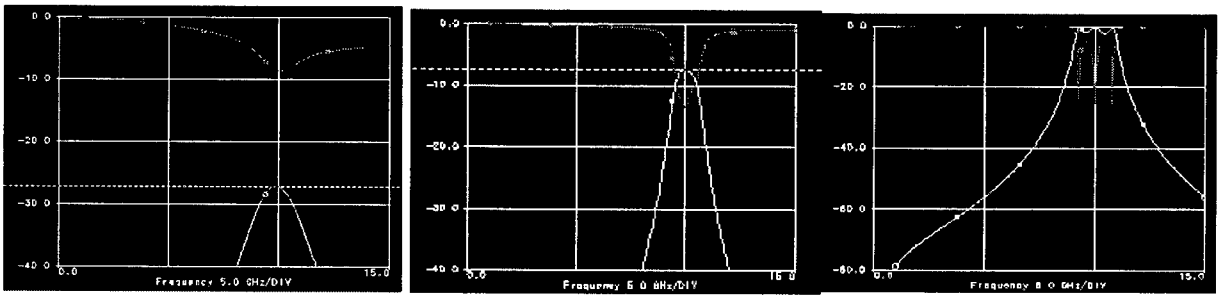


Figure 14. S_{11} and S_{12} for the 3 -section Chebychev filters with metal conductivity a) ground separation a) 20 mm, b)100 mm and c) with infinite conductivity.

In view of the possibility of using stripline type transmission lines, a 3 Section Chebychev filter is designed and simulated assuming an ε_r=3.0 tanδ=0.001 and ground plane separation of 20 and 100 microns. Figure 14 shows the S-parameters of the simulated results. Figure 14a is for ground plane separation of 20 μm and b) for 100 μm. In both cases the insertion loss is very large, -24.2 dB and -7.4 dB respectively. For comparison, same filter response is shown in Figure 14c assuming infinite conductivity for the conductors for the case of 20 μm ground plane separation. The losses for the stripline is too high to be useful as a filter element with Cu conductor.

Circuit processing on a given layer can also be done in two different ways. First technique is to deposit the metal layer over polyimide and process the necessary circuit using the conventional lithographic techniques. In this way, the rest of the metal is etched away and only the metal related to the circuit is left over the polyimide. in the second technique, a lift off process is used. Photoresist is deposited over the complete surface of the polyimide, the photo resist is exposed and processed. The developed resist is removed from the circuit element locations where conductors will reside. Metal is then deposited over the whole substrate surface. Finally, using lift off process, the rest of excess metal is removed.

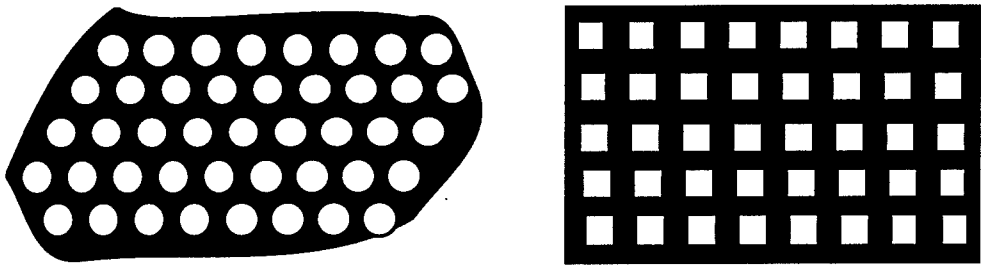


Figure 15, Perforated ground planes, a) circular and b) rectangular

Whole area ground metal deposition may pose thermal expansion problems as a result of subsequent thermal cycling due to the curing process of the upper dielectric layers. Thermal cycling may produce in cracks on the metal coating. This may not be detrimental to the operation of a given circuitry at that level unless adhesion between the metal and the dielectric weakens as a result of the repeated thermal cycling. This problem can be minimized by depositing a perforated ground planes where the perforation sizes are much smaller than the wavelength of operation, as shown in Figure 15. Here either circular or rectangular perforations are used. This technique may also help in adhesion of the upper dielectric layers since the upper and lower dielectric layers will be in contact with each other through the perforations.

Figure 16 shows two possible multilayer processing steps for realizing vertical interconnect metal post depositions. The process shown in Figure 16a is easier to implement and is expected to generate slightly non-uniform vertical posts. This process involves first the deposition of a transmission line or a ground plane. A thin dielectric layer is then deposited over the metal. Using photoresist, the post locations and sizes are exposed. Metal posts are then electrodeposited to the required height. The photo resist is then removed and a thick dielectric layer is spin coated and cured. This process will form mesas over the posts [23]. Mechanical lapping may be required to planarize the upper surface of the dielectric material to the depth of the metal posts. Even though this process seems to be attractive in the overall implementation of the 3D interconnects, required mechanical lapping may not be desirable.

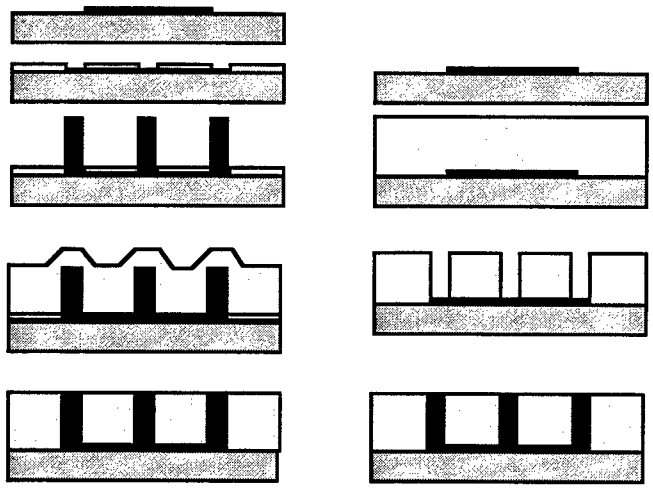


Figure 16, Processing steps for vertical interconnecting posts

The second process shown in Figure 16b begins with metal deposition for the transmission line or a ground plane. The dielectric material is then spin coated and cured to the required final thickness. Photoresist is then used to locate the post locations. Through a wet or dry etching process, the dielectric material is removed from these openings for via holes. Metal is electro-deposited in these holes to the height of the dielectric layer. This process requires anisotropic etching of vertically uniform via holes in a thick dielectric material. If this can be achieved, uniform cross sectional electrodeposition of metal posts will be possible.

Conclusions:

In this report progress toward achieving vertically interconnected 3D MMIC is presented. A novel technique based on SOI NMOS transistor technology is proposed as a means of distributing various transistors in the intermediate levels of the system unit. Preliminary results pertaining to hybridization of the HBT power transistor, vertical post build up, polyimide and circuit metal depositions are presented.

References:

[1] Altan M. Ferendeci, "Vertically Interconnected 3D MMIC," Final Report AFSOR, (Summer 1996).

[2]T. Tokumitsu, et.al., "Three Dimensional MMIC Technology And Applications to Millimeter Wave MMIC's," 1997 Topical Symposium on Millimeter Waves (TSMMW '97) Programs and Abstracts, Kanagawa, Japan, pp.44-45, (July 1997).

[3] A. Hurrich, et.al., "SOI-CMOS Technology with Monolithically Integrated Active and Passive RF Devices on High Resistivity SIMOX Substrates," Proc. 1996 IEEE Int. SOI Conf., pp.130-1(1996).

[4] M. Yoshimi, et.al., "Advantages of Low Voltage Applications and Issues to be solved in SOI Technology," ' Proc. 1996 IEEE Int. SOI Conf., pp.4-5(1996).

[5] Harold J. Hovel, "Silicon-on-insulator substrates: Status and Prognosis," ' Proc. 1996 IEEE Int. SOI Conf., pp.1-3(1996).

[6] P.K.Vasudev, et.al., "Advanced Materials for low power Electronics," Solid State Electronics, Vol.39, 489-497 (1996).

[7] J.P.Colinge, et.al., "A low voltage, Low power Microwave SOI MOSFET," Proc. 1996 IEEE Int. SOI Conf., pp.128-9 (1996).

[8] Mehmet Soyuer, et.al., "a 3-V 4-GHz nMOS Voltage Controlled Oscillator with Integrated Resonator." IEEE Journ. Solid State Circuits, Vol.31, pp.2042-45 (1996).

[9] Y. Yamaguchi, et.al., "Improved Characteristics of MOSFETs on Ultra Thin SIMOX," IEDM Digest, pp.825-28 (1989).

[10] Avid Kamgar, et.al., "Ultra-High Speed CMOS Circuits in Thins SIMOX Films," IEDM Digest, pp 819-32 (1989).

[11] T. Nishimura, et.al., "Three Dimensional IC for high performance Image Signal Processor,' IEEE-IEDM Digest, pp 111-14 (1987).

[12] T. Kunio, et.al., "Three Dimensional ICs having four Stacked Active Device Layers," IEEE-IEDM Digest, pp.837-40(1989).

[13] G.J.Willems, J.J.Poortmans & H.E. Maes, "A semiempirical model for the laser-induced molten zone in the laser recrystallization process," J.Appl.Phys., Vol.62, pp.3408-15 (1987).

[14] D.J.Wouters & H.E.Maes, "Effects of Capping layer material and recrystallization conditions on the characteristics of silicon-on-insulator metal-oxide-semiconductor transistors in laser-crystallized silicon films," J.Appl.Phys. Vol.66, pp.900-9(1989).

[15] K.Sugahara, et.al., "Orientation control of Silicon film on insulator by laser recrystallization," J.Appl.Phys., Vol.62, pp.4178-81(1987).

[16] K.E.Goodson, et.al., "Prediction and Measurement of Temperature Fields in Silicon-on-Insulator Electronic Circuits," Transactions of the ASME, Vol. 117, pp.574-81(1995).

[17] L Ristic (ed.). Sensor Technology and Devices, Artech House, 1994.

[18] R.F.Saraf, et.a., "Tailoring the surface morphology of polyimide for improved adhesion," IBM J. Res.Delep., Vol. 38, 441-56 (1994)

[19] J.P.Raskin, et.al., "An efficient tool for transmission Line on SIMOX Substrates,' Proc. 1996 IEEE Int. SOI Conf., pp.28-9(1996)

[20]Joachim N. Burghartz, et.al., "Microwave Inductors and Capacitors in Standard Multilevel Interconnect Silicon Technology," IEEE Trans. MTT-44, pp.100--3 (1996).

[21] Jor Gondermann, et.al., "Al-SiO2-Al Sandwich Microstrip Lines for High Frequency On-chip interconnects,", IEEE Trans. MTT-41, 2087-91 (1993).

[23] M.J.Kim, et.al., Mo/Cr Metalization for Silicon Device Interconnection, Mat.Res.Symp. Proc., Vol.71, pp.325-331 (1986)

[22] Huang Ho, "3D MMIC," SBIR Interim Report , August 1997.