

REPORT DOCUMENTATION PAGE

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

AFRL-SR-BL-TR-00-

issuing
nation

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE December, 1997		3. REPORT 0754	
4. TITLE AND SUBTITLE 1997 Summer Research Program (SRP), Summer Faculty Research Program (SFRP), Final Reports, Volume 3B, Phillips 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 Phillips Laboratory.					
14. SUBJECT TERMS Air Force Research, Air Force, Engineering, Laboratories, Reports, Summer, Universities, Faculty, Graduate Student, High School Student				15. NUMBER OF PAGES	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL		

UNITED STATES AIR FORCE
SUMMER RESEARCH PROGRAM -- 1997
SUMMER FACULTY RESEARCH PROGRAM FINAL REPORTS

VOLUME 3B
PHILLIPS 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 Licoscas

Program Administrator, RDL
Johnetta Thompson

Program Administrator
Rebecca Kelly-Clemmons

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Bolling Air Force Base
Washington, D.C.
December 1997

20010319 067

AQMO1-06-1192

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.

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

Due to its length, Volume 3 is bound in two parts, 3A and 3B. Volume 3A contains #1-18. Volume 3B contains reports #19-30. The Table of Contents for Volume 3 is included in both parts.

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

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

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 Aircraft	AL/HRT	2- 1
DR Brent D Foy	Wright State University , Dayton , OH Advances in Biologically-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 Mainè 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 Enterococcus 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 Souource 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/LDD	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 Measurement 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 Assessment of an Off-Axis Solar Concentrator	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 separation, Low-Frequency	AFRL/DEH	3- 11

Author	University/Institution Report Title	Phillips Laboratory Directorate	Vol-Page
DR Brian D Jeffs	Brigham Young University , Provo , UT Blind Bayesian 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 elckectronic 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 Electrolytes	PL/VTRP	3- 2
DR Tim C Newell	University of New Mexico , Albuquerque , NM Study of Nonlinear 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

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

SRP Final Report Table of Contents

Author	University/Institution Report Title	Rome Laboratory Directorate	Vol-Pa
DR Micheal A Fiddy	University of Lowell , Lowell , MA Target Identification from Limited Backscattered Field Data	RL/ERCS	4- 1
DR Lili He	Nothorn Illinois University , Dekalb , IL the Study of Caaractreistics of CdS Passivation on InP	RL/ERO	4- 1
DR Edem Ibragimov	Michigan Tech University , Houghton , MI Effects of Surface Scattering in 3-D Optical Mass Storage	RL/IRAP	4- 1
DR Phillip G Kornreich	Syracuse University , Syracuse , NY Analysis of Optically Active Material Layer Fibers	RL/OCPA	4- 1
DR Kuo-Chi Lin	University of Central Florida , Orlando , FL A Study on The Crowded Airspace Self Organized Criticality	AFRL/IFSB	4- 1
Dr. Beth L Losiewicz	Colorado College , Colorado Spring , CO The Miami Corpus Latin American Dialect Database continued Research and Documentation	RL/IRAA	4- 1
DR John D Norgard	Univ of Colorado at Colorado Springs , Colorado Spring , CO Microwave Holography using Infrared Thermograms of Electromagnetic Fields	RL/ERST	4- 1
DR Jeffrey B Norman	Vassar College , Poughkeepsie , NY Gain Spectra of Beam-Coupling In Photorefractive Semiconductors	RL/OCPA	4- 1
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- 2
DR Brajendra N Panda	University of North Dakota , Grand Forks , ND A Model to Attain Data Integrity After System Invasion	AFRL/IFG	4- 2
DR Michael A Pittarelli	SUNY OF Tech Utica , Utica , NY Phase Transitions in probability Estimation and Constraint Satisfaction Problems	AFRL/IFT	4- 2

Author	University/Institution Report Title	Rome Laboratory Directorate	Vol-Page
R Salahuddin Qazi	SUNY OF Tech Utica , Utica , NY Low Data rate Multimedia Communication Using Wireless Links	RL/IWT	4- 24
R Arindam Saha	Mississippi State University , Mississippi State , MS An Implementation of the message passing Interface on Rtems	RL/OCSS	4- 25
R Ravi Sankar	University of South Florida , Tampa , FL A Study of Integrated and Intelligent Network Management	RL/C3BC	4- 26
R Mark S Schmalz	University of Florida , Gainesville , FL Errors inherent in Reconstruction of Targets From multi-Look Imagery	AFRL/IF	4- 27
R John L Stensby	Univ of Alabama at Huntsville , Huntsville , AL Simple Real-time Tracking Indicator for a Frequency Feedback Demodulator	RL/TRAP	4- 28
R Micheal C Stinson	Central Michigan University , Mt. Pleasant , MI Destructive Objects	RL/CAII	4- 29
R 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
R 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- 1
DR Reaz A Chaudhuri	University of Utah , Salt Lake City , UT A Novel Compatibility/Equilibrium Based Iterative Post-Processing Approach For Axisymmetric brittle	WL/MLBM	5- 1

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 Processing 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 Application to Target Tracking	WL/MNAG	5- 17
DR Robert J DeAngelis	Univ of Nebraska - Lincoln , Lincoln , NE Quantitative Description of Wire Textures 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 Neurobiological Principles to Machine Vision For Arms	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 & Separation 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 , Tucson , 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 Complexes 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 Efficienet 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-Pa
DR Junghsen Lieh	Wright State University , Dayton , OH Develop an Explosive Simulated Testing Apparatus for Impact Physics Research at Wright Laboratory	WL/FIV	5- 4
DR James S Marsh	University of West Florida , Pensacola , FL Distortion Compensation and Elimination in Holographic Reocnstruction	WL/MNSI	5- 4
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- 4
DR William S McCormick	Wright State University , Dayton , OH Some Observations of Target Recognition Using High Range Resolution Radar	WL/AACR	5- 4
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- 4
DR Dakshina V Murty	University of Portland , Portland , OR A Useful Benchmarking Method in Computational Mechanics, CFD, adn Heat Tansfer	WL/FIBT	5- 5
DR Krishna Naishadham	Wright State University , Dayton , OH	WL/MLPO	5- 5
DR Serguei Ostapenko	University of South Florida , Tampa , FL	WL/MLPO	5- 5
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- 5
DR Rolfe G Petschek	Case Western Reserve Univ , Cleveland , OH AB INITIO AUANTUM CHEMICAL STUDIES OF NICKEL DITHIOLENE COMPLEX	WL/MLPJ	5- 5
DR Kishore V Pochiraju	Stevens Inst of Technology , Hoboken , NJ Refined Reissner's Variational Solution in the Vicinity of Stress Singularities	AFRL/ML	5- 5

SRP Final Report Table of Contents

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Page
R 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
R Mateen M Rizki	Wright State University , Dayton , OH Classification of High Range Resolution Radar Signatures Using Evolutionary Computation	WL/AACA	5- 57
R Shankar M Sastry	Washington University , St Louis , MO	WL/MLLM	5- 58
R Martin Schwartz	University of North Texas , Denton , TX Computational Studies of Hydrogen Abstraction From Haloalkanes by the Hydroxyl Radical	WL/MLBT	5- 59
R Rathinam P Selvam	Univ of Arkansas , Fayetteville , AR Computation of Nonlinear Viscous Panel Flutter Using a Full-Implicit Aeroelastic Solver	WL/FIMC	5- 60
R Yuri B Shtessel	Univ of Alabama at Huntsville , Huntsville , AL Smoothed Sliding Mode control Approach For Addressing Actuator Deflection and Deflection Rate Saturation	AFRL/VA	5- 61
R Mario Sznajder	Pennsylvania State University , University Park , PA Suboptimal Control of Nonlinear Systems via Receding Horizon State Dependent Riccati Equations	WL/MNAG	5- 62
R Barney E Taylor	Miami Univ. - Hamilton , Hamilton , OH Photoconductivity Studies of the Polymer 6FPBO	WLMLBP	5- 63
R 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
R Krishnaprasad Thirunarayan	Wright State University , Dayton , OH A VHDL MODEL SYNTHESIS APPLET IN TCL/TK	WL/AAST	5- 65

SRP Final Report Table of Contents

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Pa
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 A Study 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 Experimental and Computational Analysis of the Unsteady Blade Row Potential Interaction in a Tr	WL/POTF	5- 6
DR Rama K Yedavalli	Ohio State University , Columbus , OH Improved Aircraft Roll Maneuver Performance Using Smart Deformable Wings	WL/FIBD	5- 6

Author	University/Institution Report Title	Arnold Engineering Development Center Directorate	Vol-Page
R Csaba A Biegl	Vanderbilt University , Nashville , TN Parallel processing for Turbine Engine Modeling and Test Data validation	AEDC/SVT _____	6- 1
R Frank G Collins	Tennessee Univ Space Institute , Tullahoma , TN Design of a Mass Spectrometer Sampling Probe for The AEDC Impulse Facility	AEDC _____	6- 2
R Kenneth M Jones	N Carolina A&T State Univ , Greensboro , NC	AEDC/SVT _____	6- 3
R Kevin M Lyons	North Carolina State U-Raleigh , Raleigh , NC Velocity Field Measurements Using Filtered-Rayleigh Scattering	AEDC/SVT _____	6- 4
R Gerald J Micklow	Univ of Alabama at Tuscaloosa , Tuscaloosa , AL	AEDC/SVT _____	6- 5
R Michael S Moore	Vanderbilt University , Nashville , TN Extension and Installation of the Model-Integrated Real-Time Imaging System (Mirtis)	AEDC/SVT _____	6- 6
R 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
R Nicholas S Winowich	University of Tennessee , Knoxville , TN	AEDC _____	6- 8
R 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
R 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

SRP Final Report Table of Contents

Author	University/Institution Report Title	Arnold Engineering Development Center Directorate	Vol-Pa
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- 1:
DR Devendra Kumar	CUNY-City College , New York , NY A Simple, Multiversion Concurrency Control Protocol For Internet Databases	SAALC/LD _____	6- 10
DR Ernest L McDuffie	Florida State University , Tallahassee , FL A Propossed Exjpert System for ATS Capability Analysis	SAALC/TI _____	6- 1'
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 HBCU/MI applicants. In an effort to achieve AFOSR's goal of 10% of all applicants and selectees being HBCU/MI qualified, the RDL team decided to try other avenues of approach to increase the number of qualified applicants. Through the combined efforts of the AFOSR Program Office at Bolling AFB and RDL, two very active minority groups were found, HACU (Hispanic American Colleges and Universities) and AISES (American Indian Science and Engineering Society). RDL is in communication with representatives of each of these organizations on a monthly basis to keep up with their activities and special events. Both organizations have widely-distributed magazines/quarterlies in which RDL placed ads.

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

In addition to RDL's special recruiting efforts, AFOSR attempts each year to obtain additional funding or use leftover funding from cancellations the past year to fund HBCU/MI associates. 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%
<i>IEEE Spectrum</i>	2%	1%
<i>BIIHE</i>	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)

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

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

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

(Continued on next page)

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

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

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

3. 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%

**LABORATORY STUDIES OF IONOSPHERIC PLASMA EFFECTS PRODUCED
BY LIGHTING-INDUCED WHISTLER WAVES**

Min-Chang Lee
Professor
Department of Plasma Science

Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Final Report for:
Summer Research Program
Phillips Laboratory

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

And

Phillips Laboratory

September 1997

Laboratory Studies of Ionospheric Plasma Effects Produced by Lightning-induced Whistler Waves

Min-Chang Lee

**Plasma Science and Fusion Center
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139**

Abstract

Laboratory experiments have been conducted at MIT, using the student-built Versatile Toroidal Facility (VTF), to investigate the ionospheric plasma effects produced by lightning-induced whistler waves. Lower hybrid waves, generated by the lightning-induced whistler waves, can cause prominent plasma effects, such as the acceleration of electrons and ions and the spectral broadening of plasma waves. Two mechanisms by which whistler waves can generate lower hybrid waves are examined in the VTF experiments. One is the simultaneous excitation of lower hybrid waves and low-frequency mode waves by intense whistler waves (Lee and Kuo, 1984). The other is the nonlinear mode conversion of whistler waves into lower hybrid waves in the presence of short-scale field-aligned density striations (Groves et al., 1988). The effect of lower hybrid waves on the spectra of Langmuir waves are singled out for investigation. The results of laboratory experiments show that lower hybrid waves can beat with Langmuir waves to produce frequency-upshifted and frequency-downshifted Langmuir waves [Lee et al., 1997], broadening the spectra of Langmuir waves. The intensity of these beat waves, however, depends upon the angle of wave propagation with respect to the magnetic field.

Laboratory Studies of Ionospheric Plasma Effects Produced by Lightning-induced Whistler Waves

Min-Chang Lee

1. Introduction

Lightning can generate whistler waves with a broad spectrum of frequencies but with dominant energy in the range of 1.6 - 10 KHz with a peak at 5 KHz (Helliwell, 1965). The lightning-induced whistler waves, having intensities of a few tens of mV/m, have been measured in the ionosphere by rocket-borne instruments (Kelley et al., 1985; M.C. Kelley, personal communication, 1985). It is shown in Lee and Kuo (1984) that whistler waves with intensities of tens of mV/m can excite lower hybrid waves and field-aligned zero-frequency (the so-called purely growing) mode waves simultaneously via a parametric instability. This involves a four-wave interaction process which requires a threshold intensity of whistler waves. Only the VLF (very-low-frequency) band (3 - 30 KHz) of the lightning-induced whistler waves may be intense enough to excite lower hybrid waves directly.

However, weak whistler waves can produce weak lower hybrid waves in the presence of short-scale (i.e., a few meters to tens of meters) field-aligned ionospheric density irregularities (Groves et al., 1988). This process, by contrast, does not involve a plasma instability but a nonlinear scattering process. The "pre-existing" ionospheric density striations can effectively scatter whistler waves into lower hybrid waves via a nonlinear mode conversion process. The short-scale ionospheric density striations can appear in the natural spread F process (e.g., Woodman and LaHoz, 1976) or in the ionospheric heating by HF waves (e.g., see the 1974 November issue of Radio Science). Because the nonlinear scattering process does not require a threshold field of whistler waves, lower hybrid waves with a broad spectrum of frequencies can be produced during lightning in the ionosphere structured with short-scale density striations.

The appearance of lower hybrid waves in lightning can yield prominent ionospheric plasma effects, such as the acceleration of electrons (ions) along (across) the geomagnetic field and the frequency broadening of plasma waves (Kuo and Lee, 1992; Lee et al.,

1997). The energization of electrons and ions, caused by lower hybrid waves, can alter the plasma distribution functions. The density fluctuations, associated with lower hybrid waves can nonlinearly scatter high-frequency plasma waves, e.g., Langmuir waves (produced by photo-electrons or HF waves) into frequency-upshifted (so-called anti-Stokes) Langmuir waves and frequency-downshifted (so-called Stokes) Langmuir waves (Kuo and Lee, 1992; Lee et al., 1997).

Presented in this paper are our laboratory studies of afore-mentioned ionospheric plasma effects produced by lightning-induced whistler waves. The reported work in Section 2 includes the generation of lower hybrid waves by whistler waves and the frequency broadening of Langmuir waves caused by the lower hybrid waves. Discussions and conclusions are finally given in Section 3.

2. Laboratory Experiments

(A) The Versatile Toroidal Facility (VTF)

Our laboratory experiments were carried out at MIT's Plasma Science and Fusion Center, using the student-built Versatile Toroidal Facility (VTF) which has a major radius of about 1 meter, a minor radius of 0.4 meter, and a height of 1.1 meters (R.F. Duraski, M.S. Thesis, MIT, 1991). Presented in Figures 1 and 2 are the characteristic features of the VTF.

As illustrated in Figure 1, the VTF has a helical magnetic field, typically 0.07 tesla (C. Yoo, M.S. Thesis, MIT, 1991), that guides electrons emitted from heated LaB_6 filaments to flow from the bottom side to the topside of the plasma chamber. These upward flowing electrons serve two purposes: (1) creating a background plasma through electron collisions with neutral gases (e.g., H_2 , Ar, O_2), and (2) forming a field-aligned electric current. Displayed in Figure 2 is the measured background plasma density and the toroidal magnetic field (approximately equal to the helical magnetic field) as a function of the major radius of the VTF plasma chamber. The peak density can reach $6 \cdot 10^{17} \text{ m}^{-3}$ in a hydrogen plasma and a field-aligned current of 100 amperes can be consistently achieved.

The VTF plasma machine provides a plasma environment that can reasonably

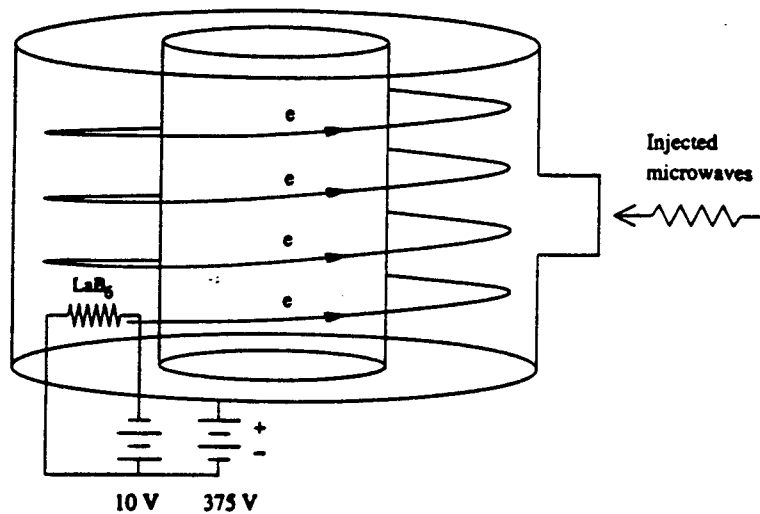


Figure 1. Schematic illustration of the Versatile Toroidal Facility (VTF). The electron beam system is a heated filament coated with LaB_6 to emit electrons.

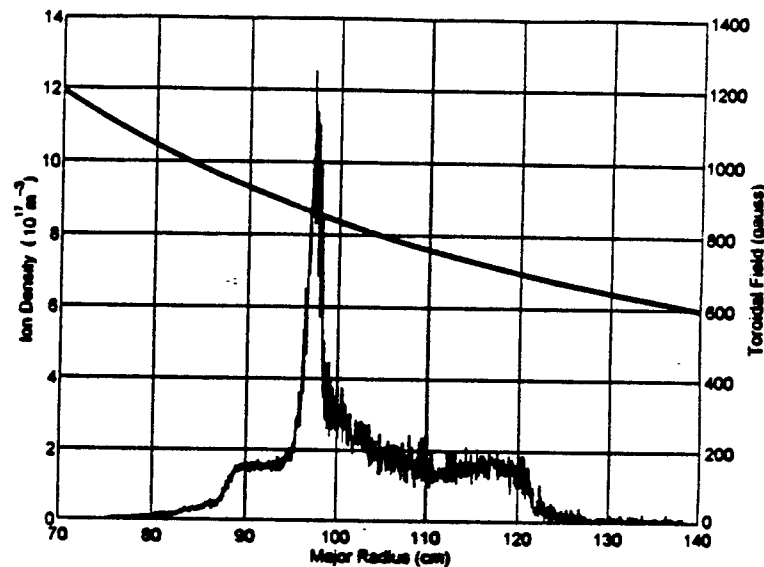


Figure 2. The toroidal magnetic field and the density profile of electron beam-produced plasmas versus the radius of the VTF.

simulate the ionospheric F region: $\omega_{pe}/\omega_{ce} \geq 3$, and $T_e \sim T_i \sim 5$ eV where ω_{pe} , ω_{ce} , T_e , and T_i represent the electron plasma frequency, the electron gyrofrequency, the electron temperature, and the ion temperature in the VTF, respectively. As shown in Figure 1, a microwave at 2.45 GHz with either ordinary or extraordinary polarization is injected into the plasma chamber from a magnetron to study radio wave interactions with ionospheric plasmas.

(B) Generation of Whistler Waves and Lower Hybrid Waves

Whistler waves are produced in the VTF by the injected electron beams moving along the helical magnetic field. As displayed in Figure 3, much more intense whistler waves are generated by weak electric current (1 ampere). It is seen that low-frequency (~ 10 MHz) mode waves are excited under stronger electric current (120 ampere) together with the generation of less intense whistler waves. By contrast, low-frequency mode waves are absent under weak electric current, when intense whistler waves are generated. The whole spectrum of whistler waves in the VTF, partially shown in Figure 3, is quite flat with a sharp cut-off slightly exceeding half of the electron gyrofrequency, in consistence with the spectrum for bremsstrahlung [see, e.g., Panofsky and Phillips, 1962].

During the operation of the electron beam system for the period of 3 seconds, the electric current reaches the steady-state condition in the VTF. A stronger electric current indicates that a higher plasma density is produced in the VTF by the energetic electron beam. As shown in Figure 2, the electron-beam-produced plasma has a sharp density gradient in the radial direction of the VTF plasma machine. An intense electric current, flowing along the helical magnetic field, together with a sharp radial density gradient forms a source of free energy, driving the so-called current convective instability (see, Ossakow and Chaturvedi, 1979 and references therein). This instability is responsible for the excitation of low-frequency mode waves, as seen in Figure 3, with a growth time less than the confinement time ($\approx$ a fraction of a milli-second) of the VTF by one order of magnitude.

What cannot be seen in Figure 3 is the excitation of lower hybrid waves. The lower hybrid resonance frequency, $f_{lhr} = f_{pi}/(1 + f_{pe}^2/f_{ce}^2)^{1/2}$ calculated to be 30 MHz, is marked

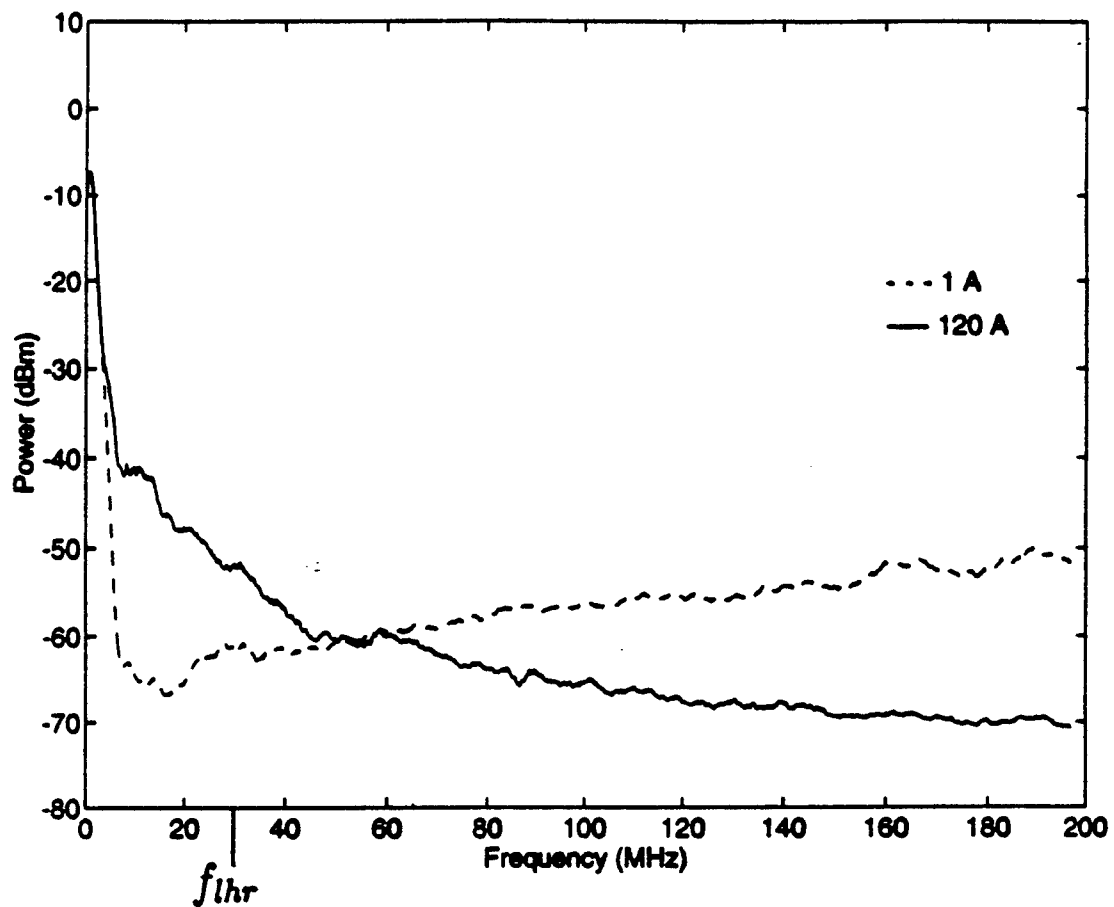


Figure 3. The low-frequency spectrum of whistler waves generated by the injected electron beams under the conditions of a high electric current (120 ampere) and a low electric current (1 ampere), respectively.

on the horizontal axis in Figure 3, where f_{pe} , f_{pi} , and f_{ce} are the electron plasma frequency, the ion plasma frequency, and the electron gyrofrequency, respectively. The two mechanisms whereby whistler waves generate lower hybrid waves, as discussed in Lee and Kuo (1984) and Groves et al. (1988), respectively, are examined in the VTF experiments. The lower hybrid waves and whistler waves are measured by a Langmuir probe and a loop probe, respectively. The electric current-excited low-frequency mode waves, measured by a dipole probe, are found to be highly field-aligned. If E_0 , ϕ_1 , and ϕ_2 denote the wave field intensity of whistler waves, the electric potential of low-frequency mode waves, and the electric potential of lower hybrid waves, respectively, the results of the VTF experiments are given in Figures 4(a), 4(b), and 4(c). Presented for discussions in Figure 4(a) is ϕ_2 versus the product of E_0 and ϕ_1 for lower hybrid waves with frequencies near the lower hybrid resonance frequency. This process involves whistler waves with frequencies higher than those of lower hybrid waves by a few KHz, which are approximately equal to the frequencies of low-frequencies mode waves.

Under the condition of high electric currents, low-frequency mode waves (ϕ_1) can be favorably excited by the current-convective instability (Ossakow and Chaturvedi, 1979). Meanwhile, weak whistler waves (E_0) and lower hybrid waves (ϕ_2) appear simultaneously. It is found that ϕ_2 increases with the product ($E_0\phi_1$) of E_0 and ϕ_1 , and nearly $\phi_2 \propto E_0\phi_1$ (see Regime B of Figure 4(a)) in good agreement with Equation (9) of Groves et al. (1988). The VTF results support Groves et al.'s mechanism, which generates weak lower hybrid waves (ϕ_2) via the nonlinear scattering of whistler waves (E_0) off the pre-existing low-frequency field-aligned mode waves (ϕ_1). Two things should be noted. One is that, while the nonlinear scattering process discussed in Groves et al. (1988) is based on waves with single wave numbers (k_0 , k_1 , k_2), the measured waves in the VTF experiments have a finite bandwidth of wave numbers. The other is that the measured low-frequency mode waves are not zero-frequency mode waves. Rather, they have small Doppler frequencies due to moving electrons in the electric current.

Under low electric currents (typically, a few ampere), the current convective instability cannot be excited and thus, low-frequency mode waves as intense as those shown in Figure 4(a) are absent. However, intense whistler waves (E_0) and lower hybrid waves (ϕ_2) as well as relatively weak low-frequency mode waves (ϕ_1) appear concurrently. The

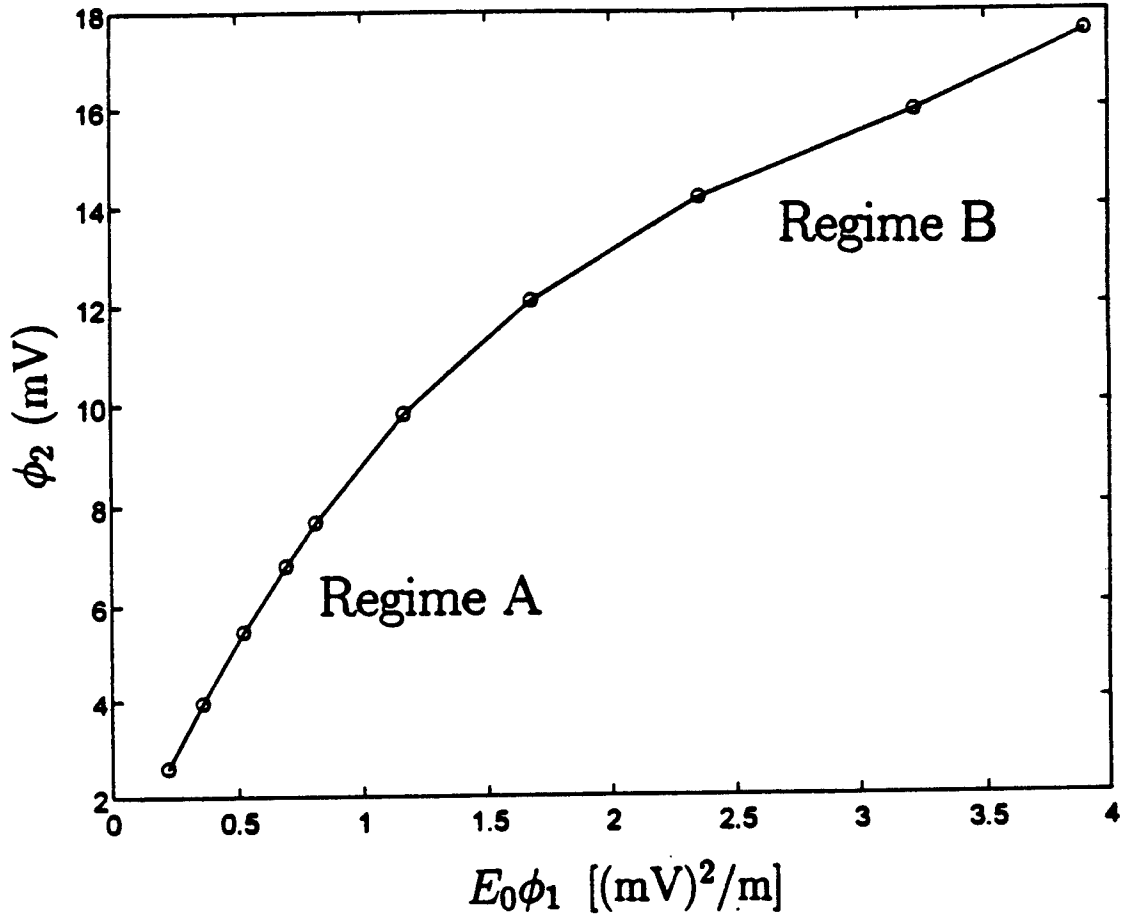


Figure 4(a). The electric potential (ϕ_2) of lower hybrid waves as a function of the product ($E_0\phi_1$) of whistler wave field intensity (E_0) and the electric potential (ϕ_1) of low-frequency mode waves. Regimes (A) and (B) correspond to the conditions of high electric currents and low electric currents, respectively.

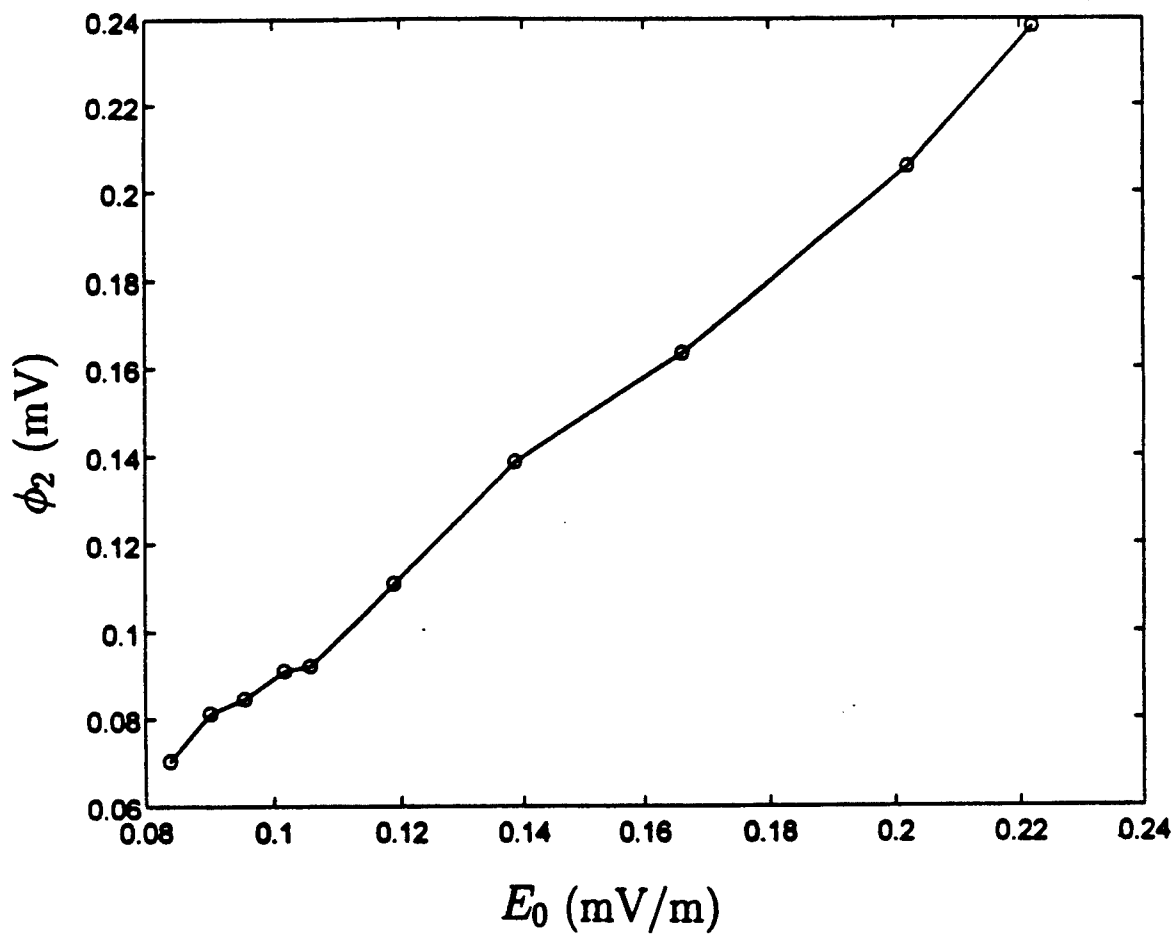


Figure 4(b). The electric potential (ϕ_2) of lower hybrid waves as a function of the whistler wave field intensity (E_0).

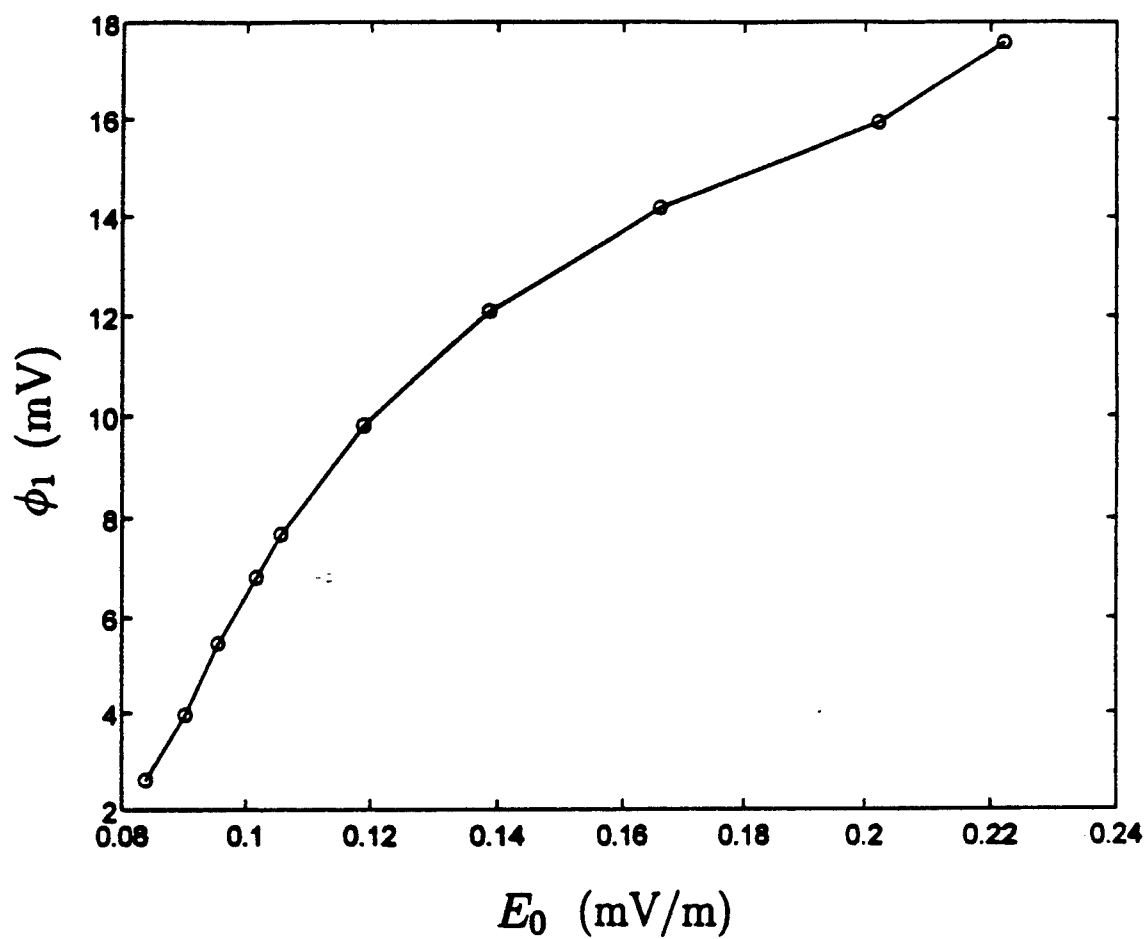


Figure 4(c). The electric potential (ϕ_1) of low frequency mode waves as a function of the whistler wave field intensity (E_0).

measured wave field intensity (E_0) of whistler waves versus the electrical potential (ϕ_2) of lower hybrid waves, and the wave field intensity (E_0) of whistler waves versus the electrical potential (ϕ_1) of low-frequency mode waves are plotted in Figures 4(b) and Figure 4(c), respectively. It is seen that the intensities of both the lower hybrid waves (ϕ_2) and low-frequency mode waves (ϕ_1) increase with that of whistler waves (E_0) at decreasing rates. In other words, when the intensity of whistler waves is high, those of low-frequency mode waves and the lower hybrid waves become saturated. This trend can also be seen clearly in Regime A of Figure 4(a). It can be understood in terms of Lee and Kuo's mechanism (1984): lower hybrid waves and low-frequency mode waves are excited simultaneously by whistler waves via an instability, this instability is seen to saturate in the VTF experiments when the intensity of whistler waves is high. The two regimes in Figure 4(a) show the distinctive difference between Lee and Kuo's mechanism (1984) and Groves et al.'s mechanism (1988) in generating lower hybrid waves by whistler waves. The transition region seems to occur at the intensity of whistler waves exceeding 1 V/m, as seen in Figures 4(b) and 4(c): while the low-frequency mode waves are saturated, the lower hybrid waves are not. This is because the instability-excited low-frequency mode waves (Lee and Kuo, 1984) can scatter whistler waves into lower hybrid waves (Groves et al., 1988) superposed to the instability-excited lower waves.

(C) Spectral Broadening of Plasma Waves

The presence of lower hybrid waves can broaden the spectra of high-frequency plasma waves, for instance, the photoelectron-induced or HFF heater-induced Langmuir waves (see, e.g., Kuo and Lee, 1992; Lee et al., 1997). The density fluctuations associated with the lower hybrid waves can nonlinearly scatter Langmuir waves to produce frequency-upshifted (i.e., anti-Stokes) and frequency-downshifted (i.e., Stokes) langmuir waves as observed in the ionospheric heating experiments at Arecibo, Puerto Rico (Lee et al., 1997) and proved theoretically by Kuo and Lee (1992).

It should be pointed out that frequency-upshifted but not frequency-downshifted Langmuir waves were detected in Arecibo experiments by a 430 MHz radar transmitting diagnostic waves vertically. Because the geomagnetic dip angle at Arecibo is about 40° , the radar only picks Langmuir waves upgoing or downgoing at an angle of 50° with respect

to the geomagnetic field. In the laboratory (VTF) experiments reported here, both the frequency-upshifted and frequency-downshifted Langmuir waves were detected by a dipole probe along the magnetic field. How we can simulate the modifications of ionospheric Langmuir waves caused by the lightning-induced lower hybrid waves is briefly described below.

The electron beam-produced plasmas in the VTF have a density profile, as shown in Figure 2, similar to the ionospheric F region near the F peak. Except near the edge of the density profile, the ratio of electron plasma frequency to electron gyrofrequency can exceed unity under the condition of high electric currents, simulating the ionospheric plasma environment reasonably well. As afore-mentioned in Section 2(B), under high electric currents, low-frequency mode waves and weak whistler waves are generated concurrently. The subsequent scattering of whistler waves off the low-frequency mode waves leads to the nonlinear mode conversion of whistler waves into lower hybrid waves (see Figure 3). However, no high-frequency plasma waves, such as Langmuir waves and upper hybrid waves, appear inherently in the VTF plasmas produced by intense electron beams.

In order to investigate the effect of lower hybrid waves on high-frequency plasma waves, an intense O-mode microwave is injected into the VTF vacuum chamber from a magnetron, operated at the frequency of 2.45 GHz, as illustrated in Figure 5. The O-mode microwave interacts resonantly with the VTF plasmas at its reflection point where the wave frequency is slightly less than the local electron plasma frequency. The microwave-plasma interaction leads to the decay of the O-mode microwave into a Langmuir wave and an ion acoustic wave via the so-called parametric decay instability. The excited Langmuir wave can grow at the expense of the microwave and, then, decay into a daughter Langmuir wave and an ion acoustic wave. The successive decay of Langmuir waves yields a cascading spectrum of Langmuir waves as shown in Figures 6(a), 6(b), and 6(c). These Langmuir waves were measured by a dipole probe (see Figure 5). Shown in Figure 6(a) is the spectrum of Langmuir waves in the absence of lower hybrid waves (D.T. Moriarty, Ph.D. Dissertation, MIT, 1996), which is distinctively different from those in the presence of lower hybrid waves.

Figure 6(b) is the spectrum of Langmuir waves measured by the dipole probe along the magnetic field, while Figure 6(c) is that at an angle of 20° with respect to the magnetic

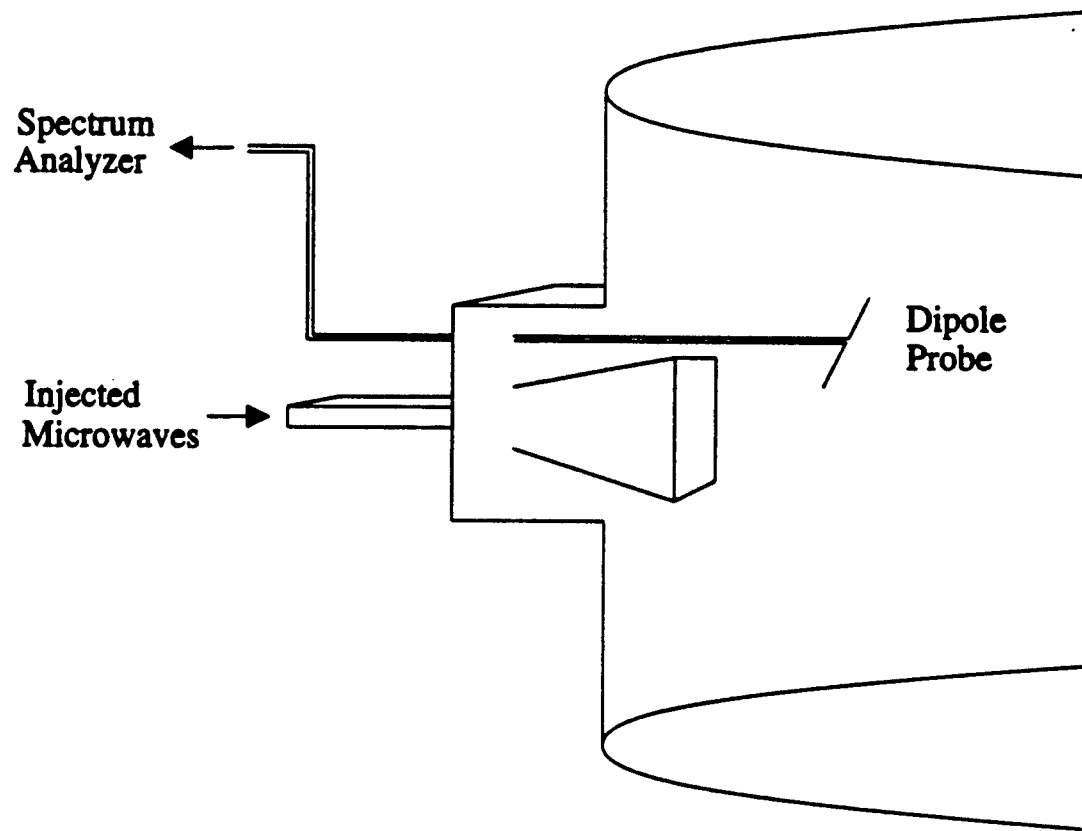


Figure 5. Illustration of VTF experiments on microwave-excited Langmuir waves and their spectral modifications by lower hybrid waves.

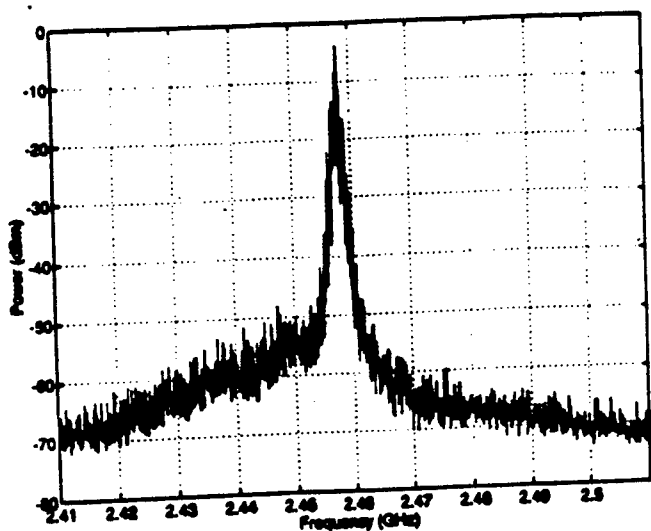


Figure 6(a). Spectrum of microwave-excited Langmuir waves in the absence of lower hybrid waves.

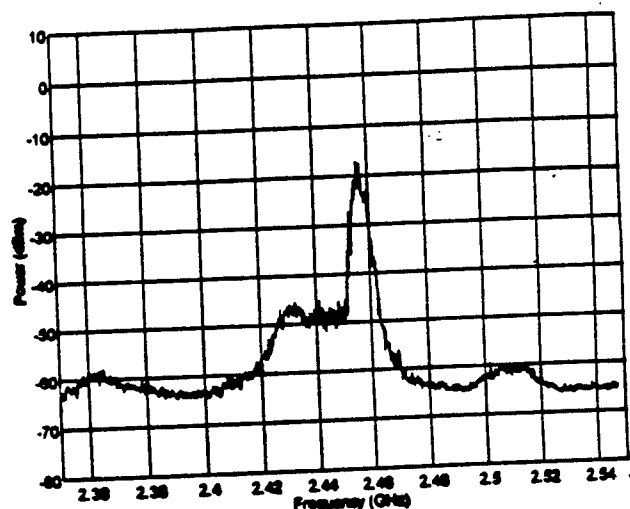


Figure 6(b). Spectrum of microwave-excited Langmuir waves measured by a dipole probe along the magnetic field.

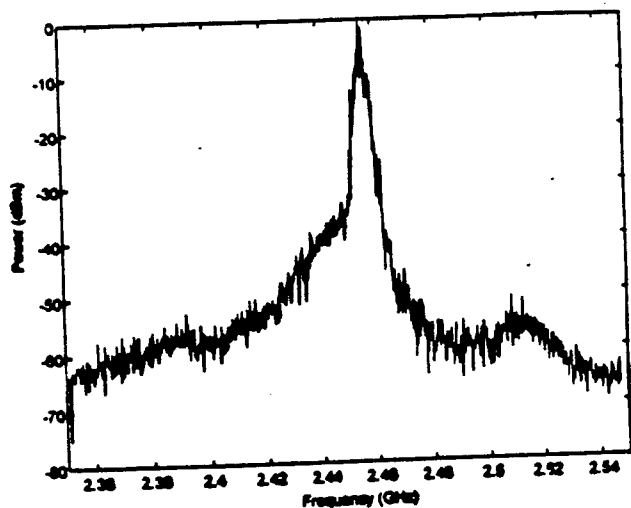


Figure 6(c). Spectrum of microwave-excited Langmuir waves measured by a dipole probe at an angle of 20° with respect to the magnetic field.

field. It is interesting to note that both the frequency-upshifted (i.e., anti-Stokes) and frequency-downshifted (i.e., Stokes) Langmuir waves were measured by the dipole probe along the magnetic field. By contrast, only the frequency-upshifted Langmuir waves were detected by the dipole probe at 20° across the magnetic field. As mentioned before, the Arecibo radar, transmitting signals at 50° with respect to the geomagnetic field during the ionospheric heating, could only detect the cascading and frequency-upshifted Langmuir waves, whose spectrum is very analogous to that shown in Figure 6(c) (Lee et al., 1997).

When the radar and the dipole probe are tilted at an angle with respect to the magnetic field, they preferentially pick obliquely propagating Langmuir waves. Langmuir waves propagating along the magnetic field satisfy the dispersion relation: $\omega^2 = \omega_{pe}^2 + 3k^2v_{te}^2$, where ω_{pe} , k , v_{te} , are the angular electron plasma frequency, wave number, and electron thermal velocity, respectively. Whereas, Langmuir waves propagating obliquely across the magnetic field in ionospheric plasmas or VTF plasmas follow the dispersion relation: $\omega^2 = \omega_{pe}^2 + 3k^2v_{te}^2 + \Omega_e^2 \sin^2\theta$, where Ω_e and θ are the angular electron gyrofrequency and the wave propagating angle with respect to the magnetic field. Thus, at the same location, Langmuir waves propagating obliquely have larger frequencies than those propagating parallel to the magnetic field.

When parallel-propagating Langmuir waves encounter lower hybrid waves, they beat together to produce Langmuir waves with sum frequencies (viz., frequency-upshifted or anti-Stokes mode waves) and difference frequencies (viz., frequency-downshifted or Stokes mode waves). If these beat waves propagate along the magnetic field, both the frequency-upshifted and frequency-downshifted Langmuir waves are "off-resonance" modes. They are equally likely detected, as we have demonstrated in the VTF experiments with a dipole probe oriented along the magnetic field. By contrast, if the beat waves propagate obliquely across the magnetic field, the frequency-upshifted (viz., anti-Stokes) Langmuir waves are "normal" or "eigen" modes, while the frequency-downshifted (viz., Stokes) Langmuir waves are "off-resonance" modes. Consequently, the intensities of anti-Stokes Langmuir waves are greater than those of Stokes Langmuir waves by several orders of magnitude as shown in Kuo and Lee (1992). This explains why only frequency-upshifted Langmuir waves were detected by the Arecibo radar and by the dipole probe tilted off the magnetic field.

3. Summary and Conclusions

Laboratory experiments on the Versatile Toroidal Facility (VTF) have been conducted to investigate the ionospheric plasma effects produced by lightning-induced whistler waves. Prominent plasma effects result from the generation of lower hybrid waves by the lightning-induced whistler waves. Intense lower hybrid waves and zero-frequency mode waves can be excited by whistler waves simultaneously in a four-wave interaction process via an instability (Lee and Kuo, 1984). Relatively weak lower hybrid waves can be generated by whistler waves in the presence of short-scale field-aligned density irregularities. The scattering of whistler waves off density irregularities renders the nonlinear mode conversion of whistler waves into lower hybrid waves (Groves et al., 1988). Lower hybrid waves may subsequently interact with the background plasmas to cause acceleration of electrons and ions along and across the magnetic field, respectively, and broaden the spectra of background plasma waves (Kuo and Lee, 1992; Lee et al., 1997).

The ionospheric plasma environment can be reasonably simulated in the Versatile Toroidal Facility (VTF) by electron-beam produced plasmas. The electron beam moving along the helical magnetic field produces field-aligned electric currents. Under the condition of high electric currents, intense low-frequency mode waves are excited by the current convective instability (Ossakow and Chaturvedi, 1979) in the VTF plasmas which have a sharp density gradient. Concurrently generated are weak whistler waves, which are scattered nonlinearly by the intense low-frequency mode waves into weak lower hybrid waves, in agreement with Groves et al.'s mechanism (1988). The scattering of whistler waves by ionospheric density irregularities, rendering the nonlinear mode conversion of whistler waves into lower hybrid waves, have been observed in recent rocket experiments on lightning-induced ionospheric effects (M.C. Kelley, personal communication, 1997).

Under the condition of low electric currents in the VTF, intense whistler waves are generated, but intense low-frequency mode waves comparable to those excited by the current convective instability are absent. However, intense whistler waves can excite lower hybrid waves and zero-frequency mode waves simultaneously as discussed in Lee and Kuo (1984) and as observed in the ionosphere (Berthelier et al., 1982; Liao et al., 1989; Sotnikov et al., 1991; Dalkir et al., 1992). As the intensity of whistler waves increases, the excited

lower hybrid waves and zero-frequency mode waves (having a Doppler frequency in the VTF) are seen in Figures 4(a) and 4(b) to saturate as expected. After the low-frequency mode waves are excited, they can scatter whistler waves and render the nonlinear mode conversion of whistler waves into lower hybrid waves (Groves et al., 1988). These lower hybrid waves, originating from the nonlinear mode conversion process, are superposed to those excited by the instability. Consequently, the saturation level of measured lower hybrid waves (see Figure 4(b)) is not as prominent as that of the low-frequency mode waves (see Figure 4(c) in the VTF).

The density fluctuations associated with the lower hybrid waves can effectively scatter high-frequency ionospheric plasma waves, e.g., the photoelectron-induced or EM wave-excited Langmuir waves, and thus, modify the spectra of Langmuir waves. As shown in the VTF experiments (see Figure 6(a)), Langmuir waves are excited by an O-mode microwave via the parametric decay instability (PDI). The PDI-excited Langmuir waves, propagating primarily parallelly to the magnetic field, can beat with lower hybrid waves to produce frequency-upshifted (anti-Stokes) and frequency-downshifted (Stokes) Langmuir waves along the magnetic field (see Figure 6(b)), but only frequency-upshifted Langmuir waves across the magnetic field (see Figure 6(c)). For parallel-propagating Langmuir waves, both the frequency-upshifted (anti-Stokes) and frequency-downshifted (Stokes) Langmuir waves are off-resonance modes, which have equal intensities. By contrast, for obliquely propagating Langmuir waves, the frequency-upshifted Langmuir waves are eigen modes, while the frequency-downshifted are off-resonance modes. The intensity of frequency-upshifted Langmuir waves exceeds that of frequency-downshifted Langmuir waves by several orders of magnitude (Kuo and Lee, 1992). This explains why only frequency-upshifted, obliquely propagating Langmuir waves are observed in the VTF experiments and in the ionosphere (Lee et al., 1997).

Acknowledgements

The author thanks his graduate student, Ryan Riddolls for assistance in the VTF experiments. Fruitful discussions with Dr. William J. Burke of the Air Force Phillips Laboratory, especially about Bremsstrahlung radiation, is appreciated.

References

- Berthelier, J.J., et al., Measurements of the VLF electric and magnetic components of waves and DC electric field on board the AUREOL-3 spacecraft: The TBF-ONCH experiment, *Ann. Geophys.*, 38, 643, 1982..
- Dalkir, Y.R., M.C. Lee, K.M. Groves, and S.P. Kuo, A mechanism responsible for the observation of symmetric lower hybrid sidebands and a low frequency mode in the upper ionosphere, *J. Geophys. Res.*, 97, 17,195, 1992.
- Duraski, R.F., Design and construction of the Versatile Toroidal Facility for ionospheric chamber research, Department of Nuclear Engineering, M.S. Thesis, Adviser: M.C. Lee, M.I.T., 1991.
- Groves, K.M., M.C. Lee, and S.P. Kuo, Spectral broadening of VLF radio signals traversing the ionosphere, *J. Geophys. Res.*, 93, 14,683, 1988.
- Helliwell, R.A., *Whistlers and Related Ionospheric Phenomena*, Stanford University Press, Palo Alto, Calif., 1965.
- Kelley, M.C. et al., Electrical measurements in the atmosphere and the ionosphere over an active thunderstorm, 1. Campaign overview and initial ionospheric results, *J. Geophys. Res.*, 90, 9,815, 1985.
- Kuo, S.P., and M.C. Lee, A source mechanism producing HF-induced plasma lines (HFPLs) with up-shifted frequencies, *Geophys. Res. Lett.*, 19, 249, 1992.
- Lee, M.C., and S.P. Kuo, Production of lower hybrid waves and field-aligned density striations by whistlers, *J. Geophys. Res.*, 89, 10,873, 1984.
- Lee, M.C. et al., Laboratory reproduction of Arecibo experimental results: HF wave-enhanced Langmuir waves, *Geophys. Res. Lett.*, 24, 115, 1997.
- Liao, C.P., J.P. Freidberg, and M.C. Lee, Explosive spread F caused by lightning-induced electromagnetic effects, *J. Atmos. Terr. Phys.*, 51, 751, 1989.
- Moriarty, D.T., Laboratory studies of ionospheric plasma processes with the Versatile Toroidal Facility (VTF), Department of Nuclear Engineering, Ph.D. Dissertation,

Adviser: M.C. Lee, M.I.T., 1996.

Ossakow, S.L., and P.K. Chaturvedi, Current convective instability in the diffuse aurora, *Geophys. Res. Lett.*, 6, 4, 1979.

Panofsky, W.K.H., and M. Phillips, Classical Electricity and Magnetism, 2nd Ed., Addison-Wesley, Reading, Massachusetts, 1962.

Sotnikov, V.I. et al., Excitation of sidebands due to nonlinear coupling between a VLF transmitter signal and a natural ELF emission, *J. Geophys. Res.*, 96, 11,363, 1991.

Woodman, R.E., and C. LaHoz, Radar observations of F region equatorial irregularities, *J. Geophys. Res.*, 81, 9,447, 1976.

Yoo, C., Plasma confinement optimization of the Versatile Toroidal Facility for ionospheric plasma simulation experiments, Department of Nuclear Engineering, M.S. Thesis, Adviser: M.C. Lee, M.I.T., 1989.

MICROCONTROLLER-BASED IMPLEMENTATION OF ADAPTIVE STRUCTURAL CONTROL

Donald J. Leo
Professor
Department of Mechanical, Industrial and Manufacture Engineering

University of Toledo
Nitschke Hall
Toledo, OH 43606

Final Report for:
Summer Research Program
Phillips Laboratory

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

And

Phillips Laboratory

October 1997

Microcontroller-Based Implementation of Adaptive Structural Control

Donald J. Leo
Mechanical, Industrial, and Manufacturing Engineering Department
University of Toledo
Toledo, OH 43606-3390

September 30, 1997

ABSTRACT

Microcontroller-based feedback algorithms were examined for applications in adaptive structural control. The feedback control algorithms consisted of tunable second-order filters that adaptively suppressed vibration by tracking the frequency of resonant modes. Two circuit implementations were considered: a purely digital design implemented on a microcontroller, and a digitally-programmable analog design implemented using a microcontroller and an analog active filter. The purely digital design, although requiring less components, was limited by the speed of the microcontroller and the lack of floating point arithmetic. An initial design consisting of one second-order filter was implemented at a sampling rate of 2 kHz, but the use of fixed-point processing limited the filter frequency to above 40 Hz. The limitations of the purely digital approach motivated the use of a digitally-programmable analog filter as the feedback compensator. Using a digitally-programmable analog filter as the feedback compensator eliminated problems due to processor speed and fixed-point arithmetic while maintaining the reconfigurability of the algorithm. A frequency-tracking filter was implemented using a frequency identification algorithm implemented with a phase-locked loop and a microcontroller. A circuit was built and several tests were performed on a beam with a variable tip mass. The control experiments demonstrated that the active damping achieved with a fixed filter would be reduced by a factor of three as the natural frequency of the beam varied 24 Hz to approximately 11 Hz, whereas the adaptive filter maintained closed-loop damping ratios between 14% and 20% over the same frequency range. Furthermore, the frequency-tracking algorithm was able to adapt to vibration in approximately 500 μ s, making it a practical algorithm for transient vibration suppression.

INTRODUCTION

Active damping is a proven solution for the suppression of resonant structural vibration. In contrast to passive damping methods which utilize energy-dissipating materials to suppress vibration, active damping methods utilize feedback control systems to dissipate structural energy. The feedback control systems consist of vibration sensors, control actuators, and electronic circuits that are designed to couple to the structural vibration and dissipate energy in lightly-damped resonant modes.

Active damping methods are practical approaches to vibration suppression due to their simplicity and inherent robustness with respect to uncertainty in the structural dynamics. Early work in the control of flexible structures proved that, in theory, collocated velocity feedback is unconditionally stable in the presence of unmodelled resonant modes [1]. Although theoretically valid, velocity feedback suffers from several practical problems due to its sensitivity to sensor-actuator noncollocation and unmodelled phase lags [2].

More practical methods of active damping account for unmodelled dynamics at the expense of reduced performance. Feedback control systems consisting of tuned resonant filters are more robust with respect to unmodelled dynamics, but typically only add damping to, at most, a few structural modes [3]. Broadband active damping is achieved by designing multiple control loops and tuning the individual filters to different frequency ranges [4].

To date most of the basic theory for active damping has been developed and more attention is focused on the implementation of active structural control. Representative space structure experiments have demonstrated that low-order control systems have the ability to control structural vibration, and hardware has been developed to implement a wide variety of active damping control laws. Even more recently an embedded control device consisting of integrated sensors, actuators, and control electronics was developed for active structural control applications [5].

As these references demonstrate, much of the recent work in active structural control has focused on implementing microprocessor-based structural control systems that are easily reconfigurable. In the former work, a standalone microprocessor-based control system with a graphical user interface was developed for implementing active control. Control designs were performed on commercially available software and downloaded directly to the microprocessor via a standard PC interface. In a similar fashion, the embedded control device developed by CSA Engineering, Inc., although not microprocessor based, could be reconfigured easily by downloading feedback control parameters directly from a personal computer. In both systems a major emphasis was placed on developing hardware that was easily changed through standard microprocessor interfaces.

The majority of work in active structural control has focused on the use of digital signal processors for implementing control algorithms. Digital signal processors (DSPs) are the microprocessor of choice due to the fact they are optimized for real-time computation. Not only do DSPs have additional hardware that increases the computational speed of the processor, their programming architecture is optimized to minimize the amount of time required to perform floating point arithmetic. These two features give DSPs the ability to perform accurate real-time computation at high sampling rates, making them an excellent choice for applications in active structural control.

In recent years the increasing power of DSPs has increased the practicality of active control algorithms. Currently, state of the art DSPs have instruction times on the order of 40 ns and can perform up to 120 million floating point operations per second. With the use of fast analog-to-digital and digital-to-analog converters, it is possible to implement high-order, multi-input-multi-output control laws with sampling rates on the order of 10 to 30 kHz.

The increase in power of DSPs has been paralleled by an increase in power of another class of microprocessors known as microcontrollers. In contrast to DSPs, which are optimized for real-time computation, micro-

controllers are low-end microprocessors that typically include on-board memory and a variety of peripherals. Although not nearly as powerful as DSPs – instruction times are on the order of 100 to 200 ns and an on-chip hardware multiplier is typically not included in the device – the primary advantages of microcontrollers are their cost and wide availability. Currently, one-time programmable microcontrollers are approximately an order of magnitude less expensive than low-end, fixed-point DSPs. Furthermore, microcontrollers are widely used in a number of price constrained industries, such as automotive and telecommunications, making them easily available and very cost effective.

The increasing power of cost effective microcontrollers motivates their use as processors for active damping applications. Although microcontrollers are not suitable for high performance applications that require fast sampling rates, it is possible that they are useful for implementing active damping control laws for structural vibration suppression. The fact that active damping control laws are typically low order and do not require excessively high sampling rates means that digital signal processors might not be required for digital implementation. A more cost effective approach might be the use of microcontrollers specifically designed for implementing active structural damping.

The objective of this work was to investigate the use of a certain class of microcontrollers for implementing active structural damping, with a particular emphasis on the use of adaptive structural control. The first section of this report discusses the theory of active damping and develops an analytical basis for the adaptive control algorithm. The second section examines a purely digital implementation of the feedback compensator and discusses the limitations that arise due to the use of a microcontroller. The third section examines a hybrid approach to the problem in which the feedback compensator is implemented on a digitally-programmable analog filter and the frequency-tracking algorithm is resident on a microcontroller. The final section compares the two approaches and discusses the major conclusions.

Active Damping for Structural Vibration Suppression

An examination of the equations of motion for a vibrating structure motivates the use of tuned second-order filters for active damping applications. Consider a structure modelled by a linear set of matrix equations,

$$M\ddot{x} + D\dot{x} + Kx = B_w w + B_u u, \quad (1)$$

where M , D , and K are the mass, damping, and stiffness matrices of the dynamic system; x is a time-dependent vector of physical coordinates; w and u are the exogenous and control forces, respectively; and B_w and B_u are the influence matrices of the two sets of forces.

If the mass-normalized mode shapes are denoted ψ and we assume that the damping matrix is uncoupled by the undamped mode shapes, we can diagonalize the system through the coordinate transformation

$$x = \psi\eta, \quad (2)$$

where η is the time-dependent vector of modal coordinates [7]. Substituting the previous expression into equation 1 and premultiplying by ψ^T yields

$$I\ddot{\eta} + \Delta\dot{\eta} + \Lambda\eta = \psi^T B_w w + \psi^T B_u u. \quad (3)$$

The outputs of the dynamic system are separated into the regulated variables, z , and feedback sensors, y . Assuming that the outputs are a linear combination of the physical states and the inputs, we have the set of equations

$$\begin{aligned} z &= C_{zd}x + C_{zv}\dot{x} + D_{zw}w + D_{zu}u \\ y &= C_{yd}x + C_{yv}\dot{x} + D_{yw}w + D_{yu}u, \end{aligned} \quad (4)$$

which can be transformed into modal coordinates through equation 2:

$$\begin{aligned} z &= C_{zd}\psi\eta + C_{zv}\psi\dot{\eta} + D_{zw}w + D_{zu}u \\ y &= C_{yd}\psi\eta + C_{yv}\psi\dot{\eta} + D_{yw}w + D_{yu}u. \end{aligned} \quad (5)$$

The open-loop transfer matrices for the dynamic system can be obtained by transforming equations 3 and 5 into the Laplace domain and solving for z and y in terms of the inputs w and u . This results in a set of four equations which are denoted $P_{zw}(s)$, $P_{zu}(s)$, $P_{yw}(s)$, and $P_{yu}(s)$:

$$\begin{aligned} P_{zw}(s) &= C_{zd}A(s)B_w + sC_{zv}A(s)B_w + D_{zw} \\ P_{zu}(s) &= C_{zd}A(s)B_u + sC_{zv}A(s)B_u + D_{zu} \\ P_{yw}(s) &= C_{yd}A(s)B_w + sC_{yv}A(s)B_w + D_{yw} \\ P_{yu}(s) &= C_{yd}A(s)B_u + sC_{yv}A(s)B_u + D_{yu} \end{aligned} \quad (6)$$

The matrix $A(s)$ is defined as

$$A(s) = \psi (s^2 + \Delta s + \Lambda)^{-1} \psi^T. \quad (7)$$

The closed-loop system is obtained by implementing the linear feedback control law, $K(s)$:

$$y = K(s)u, \quad (8)$$

which results in the following set of closed-loop equations

$$H(s) = P_{zw}(s) + P_{zu}(s)K(s)(I - P_{yu}(s)K(s))^{-1}P_{yw}(s). \quad (9)$$

Specific expressions in terms of the modal matrices and input-output matrices are obtained by substituting the expressions in equation 6 into equation 9.

The use of tuned second-order filters is motivated by the decoupled equations of motion. Equation 3 states that, under certain assumptions regarding the structural damping, the equations of motion can be decoupled into a set of second-order oscillators. This is a well known result in vibration theory, and has interesting implications for the active structural control. It implies that we can examine the structural control problem on a mode-by-mode basis and, to a certain degree, gain insight into the problem by examining the control of a simple second-order oscillator. This is discussed in the following section.

Single-Mode Control

Consider the example of a single structural mode where the regulated variables and feedback sensors are simply the physical displacement of the mode. In a practical application, this would mostly likely be a simple model of a complex structure in which the structural mode of interest is well separated from the remaining flexible dynamics. In this case, the matrices C_{zv} and C_{yv} as well as all of the direct transmission matrices, are identically equal to zero. Furthermore, the remaining matrices reduce to

$$A(s) = \frac{\omega_s^2}{s^2 + 2\zeta_s\omega_s s + \omega_s^2} \quad (10)$$

Assuming that

$$B_w = a \quad B_u = b \quad C_{zd} = c \quad C_{yd} = d, \quad (11)$$

where a , b , c , and d are constants, the open-loop matrix partitions can be written as

$$\begin{aligned} P_{zw}(s) &= \frac{ac}{\Delta_s(s)} \\ P_{zu}(s) &= \frac{bc}{\Delta_s(s)} \\ P_{yw}(s) &= \frac{ab}{\Delta_s(s)} \\ P_{yu}(s) &= \frac{b^2}{\Delta_s(s)}, \end{aligned} \quad (12)$$

where,

$$\Delta_s(s) = \frac{s^2 + 2\zeta_s\omega_s s + \omega_s^2}{\omega_s^2}, \quad (13)$$

and ω_s and ζ_s are the damping and frequency, respectively, of the structural mode. If the control law, $K(s)$, is a positive position feedback filter expressed in terms of the feedback gain, g , and the filter damping and frequency, ζ_f and ω_f :

$$K(s) = \frac{g^2\omega_f^2}{s^2 + 2\zeta_f\omega_f s + \omega_f^2} = \frac{g^2}{\Delta_f(s)}, \quad (14)$$

then the expression for the closed-loop system is reduced to

$$H(s) = ac \frac{\Delta_f(s)}{\Delta_s(s)\Delta_f(s) - b^2g^2}. \quad (15)$$

Equation 15 illustrates that the closed-loop poles of the system are a function of the structural parameters, the filter parameters, and the open-loop dc gain. The gain ac is irrelevant to the closed-loop damping since it simply multiplies the closed-loop transfer function. The previous expression can be nondimensionalized by substituting $s = \omega_s\sigma$, dividing through by ω_s^4 , replacing b^2g^2 by the nondimensional variable γ^2 , and replacing ω_f/ω_s by the nondimensional ratio r_f . This results in the following nondimensional expression

$$H(\sigma) = \frac{\sigma^2 + 2\zeta_f r_f \sigma + r_f^2}{(\sigma^2 + 2\zeta_f r_f \sigma + r_f^2)(\sigma^2 + 2\zeta_s \sigma + 1) - \gamma^2 r_f^2}, \quad (16)$$

which can be expanded into

$$H(\sigma) = \frac{\sigma^2 + 2\zeta_f r_f \sigma + r_f^2}{\sigma^4 + a_3\sigma^3 + a_2\sigma^2 + a_1\sigma + a_0}, \quad (17)$$

where

$$\begin{aligned} a_3 &= 2\zeta_s + 2\zeta_f r_f \\ a_2 &= 1 + r_f^2 + 4\zeta_f \zeta_s r_f \\ a_1 &= 2\zeta_s r_f^2 + 2\zeta_f r_f \\ a_0 &= (1 - \gamma^2) r_f^2 \end{aligned} \quad (18)$$

Pole Allocation Approach to PPF Design

The objective of the PPF design is to choose the feedback filter parameters that achieve a desired level of closed-loop damping. As equation 17 demonstrates, the closed-loop system has a fourth-order characteristic equation, which is equivalent to two oscillating poles when the roots of the polynomial are complex. One complex pair corresponds to the poles of the feedback filter, while the other pair of complex poles corresponds to the structural mode of vibration. For a given set of filter parameters, ζ_f and ω_f , and structural parameters, ζ_s and ω_s , the closed-loop poles are simply a function of the nondimensional gain γ .

One method of designing the feedback filter is to set the closed-loop characteristic equation equal to a set of desired closed-loop poles:

$$\begin{aligned} (\sigma^2 + 2\zeta_{cl}r_{cl}\sigma + r_{cl}^2)(\sigma^2 + 2\zeta_{cl}\alpha r_{cl}\sigma + \alpha^2 r_{cl}^2) = \\ (\sigma^2 + 2\zeta_f r_f \sigma + r_f^2)(\sigma^2 + 2\zeta_s \sigma + 1) - \gamma^2 r_f^2, \end{aligned} \quad (19)$$

where ζ_{cl} is the damping ratio of the closed-loop poles,

$$r_{cl} = \omega_{cl}/\omega_s, \quad (20)$$

α is a scalar constant, and ω_{cl} is the desired closed-loop natural frequency.

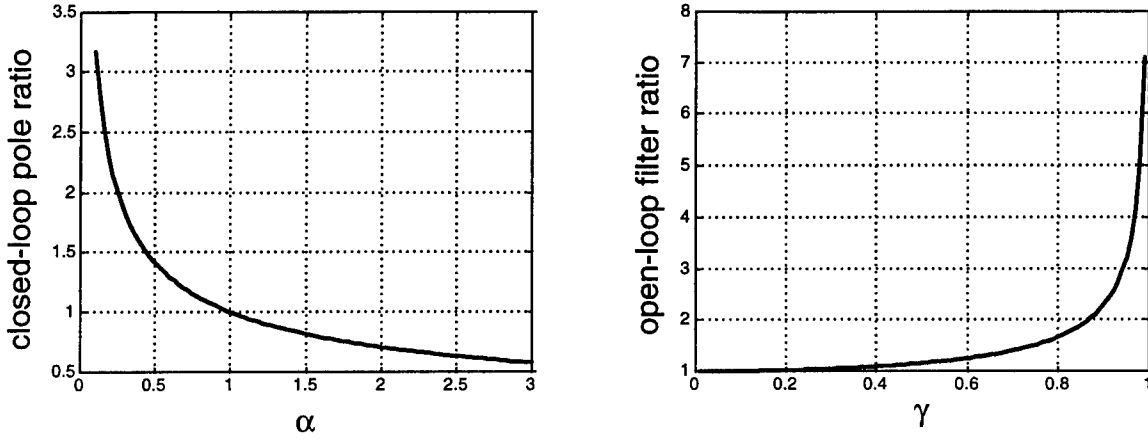


Figure 1: Open-loop filter ratio and closed-loop pole ratio as a function of the design parameters of a positive position feedback filter.

If equation 19 is expanded, we can write a set of four equations that must be satisfied for the closed-loop poles to be at the desired location:

$$\begin{aligned}
 \sigma^3 \quad & 2\zeta_{cl}r_{cl}(1+\alpha) = 2\zeta_fr_f + 2\zeta_s \\
 \sigma^2 \quad & r_{cl}^2(\alpha^2 + 1) + 4\zeta_{cl}^2\alpha r_{cl}^2 = r_f^2 + 1 + 4\zeta_f\zeta_sr_f \\
 \sigma^1 \quad & 2\zeta_{cl}\alpha^2 r_{cl}^3 + 2\zeta_{cl}\alpha r_{cl}^3 = 2\zeta_fr_f + 2\zeta_sr_f^2 \\
 \sigma^0 \quad & \alpha^2 r_{cl}^4 = r_f^2 - \gamma^2 r_f^2
 \end{aligned} \tag{21}$$

Assuming that the structural damping is negligible ($\zeta_s \approx 0$), we can solve these four equations and write a set of design rules for the PPF filter:

$$r_{cl} = \sqrt{1/\alpha}, \tag{22}$$

$$r_f = \sqrt{\frac{1}{1-\gamma^2}} \tag{23}$$

$$\zeta_{cl} = \frac{1}{2} \sqrt{\frac{2-\gamma^2}{1-\gamma^2} - \frac{\alpha^2+1}{\alpha}} \tag{24}$$

$$\zeta_f = \zeta_{cl} \sqrt{\frac{(1+\alpha)^2(1-\gamma^2)}{\alpha}} \tag{25}$$

Equations 22 through 25 can be used to choose the filter parameters as a function of the nondimensional gain, γ , and the ratio of the closed-loop natural frequencies, α . Equations 22 and 23 demonstrate that the natural frequency ratios are simply a function of either α or γ , while the closed-loop damping ratios are a function of both the nondimensional gain and the ratio of closed-loop natural frequencies. Referring to Figure 1, we see that the ratio of closed-loop poles, r_{cl} , is simply equal to 1 if $\alpha = 1$. Also, the figure indicates that the open-loop filter ratio, ω_f/ω_s , that achieves the desired closed-loop pole locations has an asymptote of 1 as the control gain decreases towards zero.

The achievable closed-loop damping ratio is a more complicated function of the design parameters α and γ . Examining equation 24, we see that the closed-loop damping ratio is a function of both the nondimensional gain and the ratio of closed-loop natural frequencies. For a particular value of α we can plot the variation of ζ_{cl} as a function of the nondimensional gain and determine the achievable performance.

Consider the case plotted in Figure 2 for $\alpha = 1$. For this case, nondimensional gains less than approximately 0.5 yield a linear variation in the closed-loop damping ratio. As the gain becomes larger than 0.5, the closed-loop damping ratio increases more rapidly with respect to γ than in the region less than 0.5.

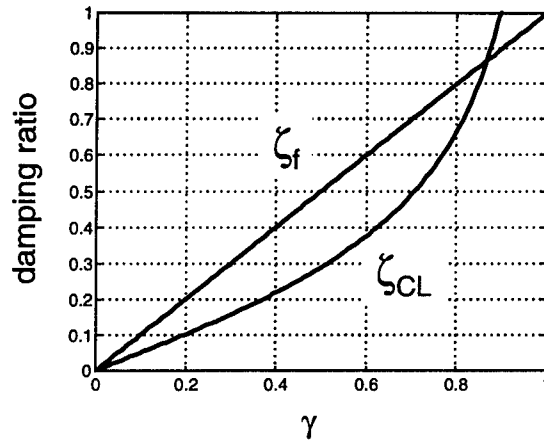


Figure 2: Closed-loop damping ratio, ζ_{cl} , and required filter damping ratio, ζ_f , as a function of the nondimensional gain for $\alpha = 1$.

Also plotted in Figure 2 is the filter damping ratio required to achieve the desired closed-loop damping. As plotted in the Figure, and highlighted in equation 25, the required damping ratio is approximately equal to twice the desired closed-loop damping ratio in the low-gain region ($\gamma < 0.5$).

Combining the information from Figures 1 and 2 yields a set of practical rules for the design of PPF compensators:

1. For a gain margin of greater than 2.0 (6 dB), the desired closed-loop damping ratio is achieved by setting the natural frequency of the filter approximately equal to the frequency of the target structural mode.
2. When the nondimensional gain is less than 0.5 and $\alpha = 1$, the closed-loop damping in the structural mode is approximately equal to 0.6γ and the required damping ratio of the feedback compensator is approximately equal to twice the desired closed-loop damping.

These design rules trade off performance in the interest of increasing robustness. Figure 2 illustrates that a larger amount of closed-loop damping is achievable if the gain of the feedback compensator is increased. Of course, this increase in observable performance occurs at the expense of gain margin in the control system. Reducing the gain margin at low frequency increases the likelihood that changes in the DC gain of the structural response will lead to closed-loop instability. Keeping the gain margin greater than 2.0 (or 6 dB) means that the DC gain of the structural response can change by up to a factor of two before closed-loop instability will occur.

Digital Filter Implementation

The first half of the work concentrated on a digital implementation of the positive position feedback filter. The primary objective of the work was to determine the following limitations on the use of the microcontroller:

- Maximum sampling rate of the digital filter,
- Maximum number of second-order filters that could be implemented on the microcontroller,
- The limits on filter accuracy using fixed-point arithmetic.

The digital filter implementation of the vibration suppression algorithm consisted of three components: an LTC 1298 12-bit analog-to-digital converter (ADC), a PIC16C65A microcontroller, and a 12-bit LTC 1451

Table 1: Measured operation times for implementing a second-order IIR filter on the PIC16C65A with a 20 MHz clock speed

Operation	Number of Instructions	Time @ 20 MHz (μ s)
Read ADC	637	127.4
Compute output	1307	261.3
Output to DAC	555	111.0
Total		499.7

digital-to-analog converter (DAC). Code for interfacing with the LTC 1298 was already available, and custom code was written to interface with the LTC 1451. Code was developed on the PC and downloaded to the microcontroller via the PicStart Plus hardware.

Code was developed to implement a second-order Infinite Impulse Response (IIR) filter using the finite difference equations:

$$\begin{aligned} d(k) &= u(k) - a_1d(k-1) - a_2d(k-2) \\ y(k) &= b_0d(k) + b_1d(k-1) + b_2d(k-2) \end{aligned} \quad (26)$$

where $u(k)$ and $y(k)$ are the filter input and output, respectively, at the k^{th} time step; $d(k)$ is an intermediate variable; and a_1 , a_2 , b_0 , b_1 , and b_2 are the coefficients that define the frequency response of the filter. Three microcontroller operations were required to implement the IIR filter:

1. Read a value from the ADC
2. Solve finite difference equation (equation 26) to compute filter output
3. Output value to DAC

The speed and accuracy at which the microcontroller performed each of these steps determined the sampling rate and frequency limitations of the filter.

Maximum Sampling Rate

The sampling rate of the digital filter was set by the speed at which the microcontroller could read the ADC, solve the finite difference equation, and output a value to the DAC. Table 1 lists the measured values at which the PIC microcontroller could perform each of these operations. As Table 1 indicates, the total time to complete all of the operations was $\approx 500 \mu$ s, setting the filter sampling rate at 2 kHz. Approximately 52% of the computation time was dedicated to computing the filter output, while the remaining time was spent reading and writing to the analog I/O.

There are two ways to increase the sampling rate of the digital filter: replace the current ADC and DAC hardware and optimize the code that performs the finite difference equation. The largest gains could be obtained by changing the hardware for the A/D and D/A conversions. Both the LTC 1298 and LTC 1451 have 12-bit serial interfaces, which reduces the number of I/O lines required to interface with them – 3 in the case of the ADC and 4 in the case of the DAC – but significantly increases the time required to transfer data to and from the processor.

A much faster option would be to replace the serial I/O with a parallel interface. Parallel communication requires a larger number of I/O ports, but requires much less time due to the fact the data transfer between the microprocessor and ADC/DAC takes place almost simultaneously. Although no timing studies were performed to directly compare a serial interface with a parallel interface, it is probable that the time required for reading and writing to the analog I/O would be reduced by a factor of 5 to 10: from roughly 250 μ s to between 50 and 25 μ s. Thus, simply replacing the serial interface with a parallel interface could increase

Table 2: Estimated computation time and sampling rate for an optimized filter implementation.

Operation	Current code (Number of cycles)	Optimized code (Number of cycles)	Optimized Computation Time at 20 MHz (μ s)	Optimized Computation Time at 33 MHz (μ s)
Read ADC	637	60	12	7.3
Compute output	1307	800	160	97
Write to DAC	555	60	12	7.3
		Total time (μ s):	184	111
		Sampling rate (kHz):	5.44	9.01

the sampling rate from 2 kHz to between 3.33 and 3.63 kHz, although it would also increase the number of required digital I/O lines from 7 to between 25 and 30.

Another way of increasing the sampling rate would be to optimize the code that solves the finite difference equation of the filter and computes the output signal. The current generation of code was compiled from a C program and downloaded into the microcontroller. It is probable that the program could be optimized by working directly with the assembly code; possibly by handcoding loops and reducing the amount of overhead associated with reading and writing intermediate variables.

The gains associated with code optimization are estimated to be between 25% and 50%, much less than those obtained by replacing the current suite of digital-to-analog I/O. That estimate is based on results published in the *Embedded Control Handbook*, in which a second-order IIR filter was implemented in approximately 800 instruction cycles by thoroughly optimizing the assembly code [8]. Assuming a 20 MHz clock frequency and a 200 ns instruction time, the time required to solve the finite difference equation would be 160 μ s and the sampling rate of the filter would be set at 6.25 kHz. These estimates are summarized in Table 2.

The final method of increasing the sampling is, of course, to increase the clock speed of the processor. Although the fastest processors presently available from Microchip were used for these studies (20 MHz), Microchip is currently developing a 33 MHz version of their PIC16CXX line of microcontrollers. Upgrading to a 33 MHz clock speed would result in a 65% increase in speed due to the fact that the instruction cycle time would be reduced from 200 ns to 121 ns. An estimate of the computation time and sampling rate of a second-order filter running on a 33 MHz processor is also included in Table 2.

Maximum Filter Order

The maximum number of filter sections that could be implemented on the PIC microcontroller was interrelated with the maximum sampling rate of the digital filter. If we assume that the digital filter is simply a parallel implementation of N second-order sections and the fixed overhead required for the multiple sections is identical to that required for one section, then it is relatively straightforward to determine an expression for the time required to solve N finite difference equations:

$$\begin{array}{llll}
 \text{Current code:} & 238.4 + 261 * (\text{Number of filter sections}) & \mu\text{s} \\
 \text{Optimized using a 20 MHz clock (estimate):} & 24 + 160 * (\text{Number of filter sections}) & \mu\text{s} \\
 \text{Optimized using a 33 MHz clock (estimate):} & 14.6 + 97 * (\text{Number of filter sections}) & \mu\text{s}
 \end{array}$$

The sampling rate of the digital filter can be plotted as a function of the section order to compare the performance of different implementations. Figure 3 demonstrates that, as expected, increasing the filter order decreases the achievable sampling rate. A rule of thumb for digital filter design is that the filter dynamics should be no higher than roughly 1/10 of the filter sampling rate. With this rule in mind,

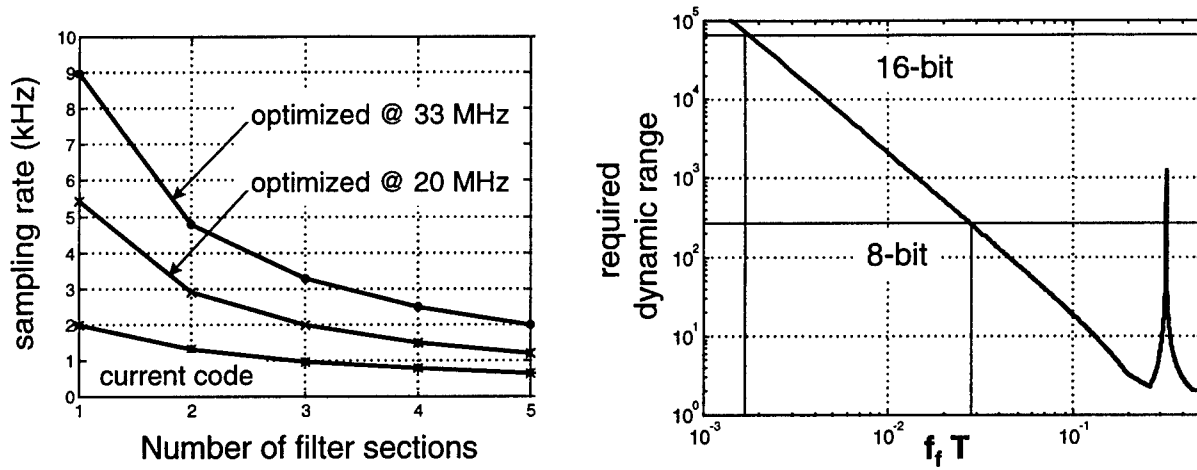


Figure 3: Achievable sampling rate of the digital filter as a function of section order (left) and required dynamic range of the filter coefficients (right).

Figure 3 can be used to approximate the highest filter frequency that can be achieved with the different implementations: simply divide the sampling rate by 10 and that is a rough estimate of the maximum achievable filter frequency.

Limitations due to Fixed-Point Arithmetic

The previous two sections discussed the limitations on the filter sampling rate as a function of the code implementation and the section order. Another limitation was imposed on the digital filter implementation due to the fact that the PIC microcontroller is not optimized to perform floating point arithmetic operations. The lack of a fast floating point processor – like those found in DSPs – requires the use of fixed-point mathematics in the real-time solution of the finite difference equation. The interesting problem with fixed-point arithmetic is that, unlike the limitations due to code optimization and section order, the constraints limit the *lowest frequency* achievable with the digital filter.

The performance limitations due to fixed-point processing were studied by discretizing the expression for the continuous time filter (equation 14). Using the Tustin transformation [6],

$$s = \frac{2}{T} \frac{z - 1}{z + 1}, \quad (27)$$

on where T is the sample time of the filter, the continuous-time representation of the compensator yields the z -domain polynomial of the discretized filter:

$$K(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}, \quad (28)$$

where the coefficients are expressed in terms of ζ_f and $\phi = 2\pi f_f T$:

$$\begin{aligned} a_1 &= \frac{2\phi^2 - 8}{4 + 4\zeta_f \phi + \phi^2} \\ a_2 &= \frac{4 - 4\zeta_f \phi + \phi^2}{4 + 4\zeta_f \phi + \phi^2} \\ b_0 &= \frac{\phi^2}{4 + 4\zeta_f \phi + \phi^2} \\ b_1 &= 2b_0 = 2 \frac{\phi^2}{4 + 4\zeta_f \phi + \phi^2} \\ b_2 &= b_0 = \frac{\phi^2}{4 + 4\zeta_f \phi + \phi^2} \end{aligned} \quad (29)$$

For a filter with a specified damping ratio, the only parameter that affects the value of the coefficients is the ratio of the filter frequency to the sampling frequency. Figure 3 is a plot of the ratio of the maximum value

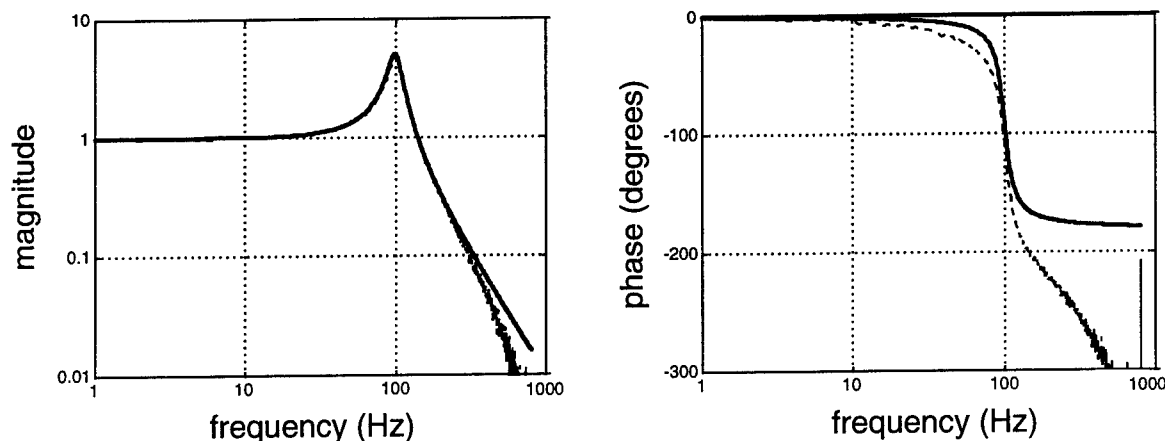


Figure 4: Digital filter implementation (solid) and the frequency response of the continuous-time representation (dashed).

of the coefficients to the minimum value over a range of $f_f T$ for $\zeta_f = 1$. The spike in the dynamic range is caused by the fact that one of the coefficients passes through zero and changes sign, thus causing a division by zero error. The spike is irrelevant, though, due to the fact it occurs at a ratio greater than $1/10$.

The important part of the plot is in the region below a filter frequency to sampling frequency less than $1/10$. As Figure 3 indicates, the required dynamic range in the filter coefficients increases with a slope of 40 dB/decade on a log-log plot. Thus, reducing $f_f T$ by a factor of 10 increases the required dynamic range by a factor of 100.

Also plotted in Figure 3 is the maximum dynamic range achievable with 8-bit and 16-bit arithmetic. Due to the fact that the dynamic range requirements increase as $f_f T$ decreases, a hard constraint on dynamic range limits the lowest frequency that is achievable with the discretized filter. In the case of 8-bit arithmetic, which has a dynamic range of 256:1, the limit is ≈ 0.03 and in the case of 16-bit arithmetic, which has a dynamic range of 65536:1, the limit is ≈ 0.002 . Combining this lower limit with the higher limit imposed by sampling frequency yields bounds on the achievable filter frequency for 8-bit and 16-bit fixed-point arithmetic.

Digital Filter Implementation

Several digital filters were implemented on the PIC16C65A to determine the performance of the microcontroller. Code was downloaded onto the microcontroller and the input-output frequency response was obtained with a spectrum analyzer. Figure 4 is a magnitude and phase plot of a digital filter implemented with a DC gain of 1, a 100 Hz filter frequency, and a damping ratio of 0.10; also overlayed on the plot is the frequency response of the continuous-time filter with the same parameters. The digital filter was implemented with a sampling rate of 2 kHz.

The measurements illustrate that at $f_f T = 0.05$, the digital implementation is able to accurately match the magnitude of the continuous time filter but significant lags exist in the phase response. The magnitude is matched well across the whole frequency range; the only significant differences occur at approximately 1 kHz – or half the sampling rate. Significant differences occur in the phase response due to delays in the zero-order hold of the DACs and processing delays in the microcontroller.

The errors that occurred due to fixed-point processing were studied by implementing several filters with varying center frequencies. Figure 5 is a magnitude plot of several different digital filter implementations, each with different center frequency. Although all of the DC gains were set equal to 1 in the discretized transfer function, the inability to accurately represent the continuous-time filter caused magnitude errors at

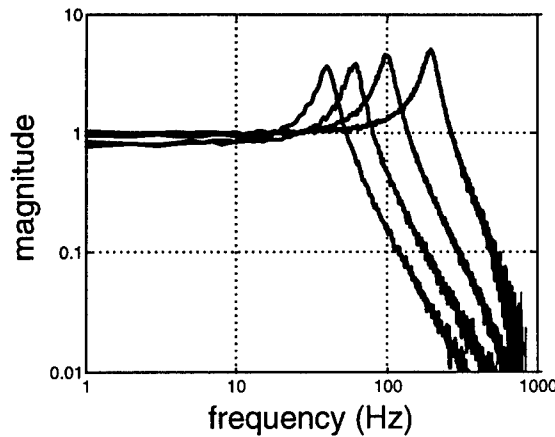


Figure 5: Several digital filter implementations demonstrating the errors that occur at low frequency due to fixed-point arithmetic.

low frequencies. The problem became more pronounced as $f_f T$ was decreased to approximately 0.02 – or a filter frequency of 40 Hz. For this filter, the DC gain error was approximately 20% due to the fact that the fixed-point arithmetic did not have the dynamic range to accurately represent the discrete filter coefficients.

Digitally-Programmable Analog Implementation of the Adaptive Filter

The limitations imposed by digital sampling rate and fixed-point processing motivated the use of a ‘hybrid’ approach to the adaptive filter implementation. In the hybrid approach, the PPF compensator was implemented on a digitally-programmable analog filter and the microcontroller acted as a ‘supervisor’ circuit that determined the filter frequency and set the filter parameters. Unlike the previous implementation, which had no method of determining the settings for the feedback filter, the hybrid implementation also incorporated a phase-locked loop to measure the natural frequency of the vibrating structure.

A block diagram of the digitally-programmable analog implementation is shown in Figure 6. The circuit consists of four main elements:

1. **Phase-Locked Loop** The vibration of the beam was measured using a structural sensor and input to the phase-locked loop to produce a square wave that had the same frequency as the beam vibration.
2. **Microcontroller with a Mechanical Trigger** The microcontroller measured the output of the phase-locked loop and determined the correct clock signal frequency for the digitally-programmable analog filter. A mechanical trigger was implemented so that the microcontroller could be reset by simply holding the beam back for approximately two seconds and releasing it.
3. **Digitally-Programmable Analog Filter** The positive position feedback compensator was implemented on a switched-capacitor filter that could be reprogrammed with the microcontroller. An analog smoothing filter was placed in series behind the digitally-programmable filter to reduce audible noise due to discrete steps in the filter output.
4. **Bridge Amplifier** The actuators were driven with a simple bridge amp implemented on a standard operational amplifier.

Each of these components will be discussed in more detail in the following subsections.

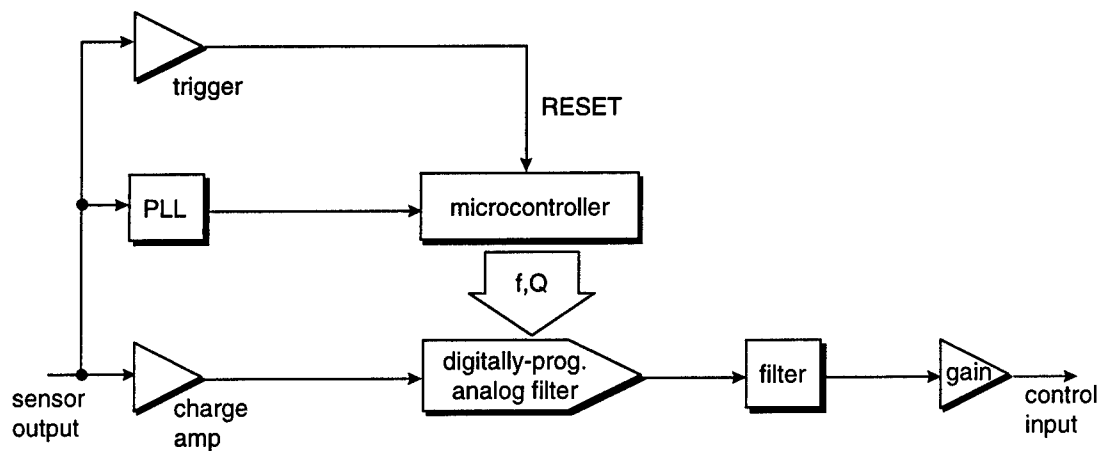


Figure 6: Block diagram of the digitally-programmable analog implementation of the adaptive filter.

Phase-Locked Loop

The main function of the phase-locked loop was to track the motion of the beam and output a logic signal of the same frequency as the natural frequency of the beam. The two primary reasons to use a phase-locked loop to perform this function were:

1. The phase-locked loop can track a frequency that is heavily buried in noise, even when the signal to noise ratio is less than or equal to one. This 'noise' can be actual random noise in the circuit, or the 'noise' could actually be other structural modes of vibration.
2. The phase-locked loop outputs a square wave signal that can be easily measured by the microcontroller for the purpose of setting the correct parameters of the feedback filter.

The phase-locked loop is itself a feedback circuit that has three main components: a signal multiplier that outputs a square wave that has a duty cycle proportional to the difference in the frequency of the two input signals, an analog filter, and a voltage controlled oscillator (VCO) that outputs a square wave whose frequency is proportional to the dc input. Many (many, many) types of phase-locked loops are commercially available, but all the different types essentially can be categorized into analog or digital phase-locked loops, where type of phase-locked loop refers to the type of signal multiplier that is used in the feedback circuit.

Although a great number of phase-locked loops exist, the principles on which they operate are essentially the same. The reference signal is compared to the output of the voltage-controlled oscillator through what amounts to an analog multiplication. The output of the signal multiplier has two components: a dc offset that is proportional to the phase difference between the two signals, and an oscillating component whose frequency is twice the frequency of the input signals. The multiplier output is passed through a lowpass analog filter that reduces the oscillating component and passes the dc component of the signal. The dc signal then sets the oscillator frequency of the VCO. The feedback loop minimizes the error between the phase of the reference signal and the VCO output.

The dynamics of phase-locked loops are rather complicated, but they are relatively easy to implement in practice. For small perturbations about the center frequency of the VCO, the control system is approximately linear and can be analyzed using standard frequency-domain design techniques [10]. Nonlinear analysis methods are required to analyze large deviations from the center frequency of the VCO, but, in general, complex analyses were not required for implementing phase-locked loops [11]. In practice, all of the information required for implementing the PLL was on the datasheet that came with the device [12].

A digital phase-locked loop was used in the adaptive filter circuit to simplify the microcontroller interface. Although the operating principles were essentially the same as describe previously, the VCO output was a

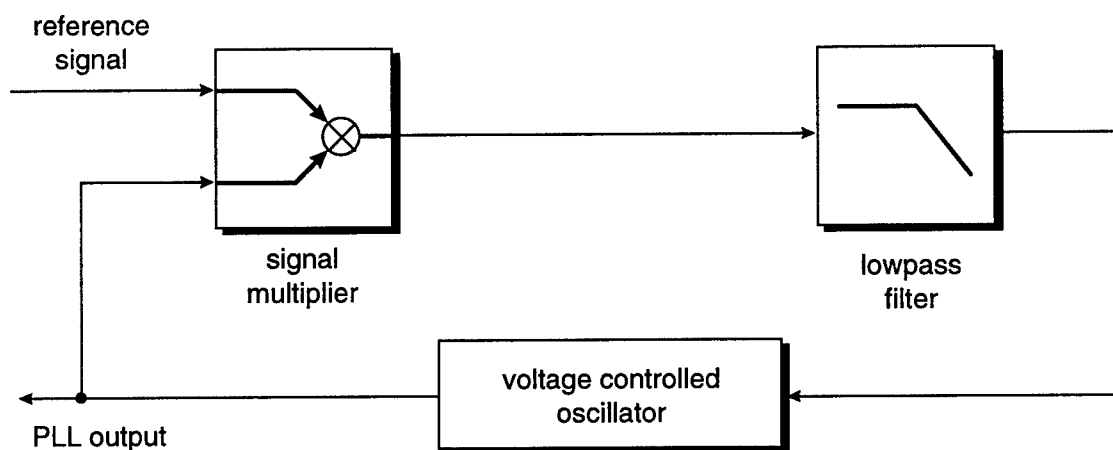


Figure 7: Block diagram of a phase-locked loop.

TTL logic signal that could be easily interpreted with the microprocessor for adapting the filter parameters. The center frequency and operating range of the phase-locked loop were set at 8.35 Hz and 5 Hz to 24 Hz, respectively, for all of the control experiments.

Digitally-Programmable Analog Filter

The PPF compensator was implemented on a digitally-programmable analog (DPA) filter that could be reprogrammed via the digital I/O ports of the microcontroller. The DPA was a switched-capacitor filter whose switching frequency, f_{clk} , was set by a square signal applied to its oscillator input [9]. The ratio of the filter frequency to the oscillator frequency was set via a 6-bit digital word that was downloaded serially from the microcontroller to the DPA filter. For all of the present analyses, the ratio of f_f/f_{clk} was set at 200 to minimize the phase lags that occur due to the holds at the filter output. The damping ratio of the filter was set via a 7-bit digital word also downloaded serially in conjunction with the frequency ratio settings.

The use of a DPA filter to implement the PPF filter eliminated two of the problems that occurred when using a purely digital implementation of the compensator:

1. Problems due to fixed-point processing were nonexistent due to the fact that the filter, although discretized due to the switched-capacitor implementation, was an analog device that did not use a finite difference equation to compute the filter output.
2. Phase lags due to sample and hold were negligible due to the fact that the switching frequency was 200 times greater than the cutoff frequency of the filter.

The discrete nature of the device (due to the switched-capacitor implementation) necessitated the use of a smoothing filter in series with the filter output. Without the smoothing filter, the actuators would emit audible noise in the 2 kHz to 6 kHz frequency range when the beam was moved. For this reason, a second-order Chebyshev filter with a break frequency of approximately 1 kHz was placed in series with the filter output. The smoothing filter eliminated a significant portion of the audible noise without adversely affecting the performance of the feedback compensator.

Microcontroller with Mechanical Trigger

Implementing the PPF compensator on a digitally-programmable analog filter offloaded a significant portion of the real-time computation from the microcontroller. In this implementation, the microcontroller acted primarily as a timer/counter that used the output of the phase-locked loop to determine the parameters

of the switched-capacitor filter. As described in the previous subsection, the center frequency of the DPA filter, f_f , is set by inputting a clock signal with a frequency of $200 f_f$. Therefore, the main function of the microcontroller was to measure the vibration frequency and output a clock signal that had a period of $1/200^{th}$ the period of natural vibration of the beam.

The period of the beam vibration was determined with a simple algorithm that used external interrupts to count the time between successive rising edges of the phase-locked loop output. The output of the phase-locked loop was input into the external interrupt pin of the microcontroller. On the first low-to-high transition the microcontroller started incrementing a dummy variable that increased by one on each successive rising edge. On the fifth rising edge the microcontroller began incrementing a counter every $1.6 \mu s$ using a 16-bit timer with a prescaler of 8. On the next rising edge the timer variable was read and the period of the beam vibration (as measured by the output of the phase-locked loop) was computed. The computed period was then divided by 200 and the microcontroller began outputting a square wave with a period of $1/200^{th}$ of the period of the beam vibration.

The five-cycle delay was incorporated into the algorithm to improve the accuracy of the measurement. Initial tests indicated that without a short delay the algorithm would often measure inaccurate values of the vibration period. The errors were attributed to the fact that the first rising edge of the phase-locked loop output did not always correspond to a full cycle of the vibration period, therefore the microcontroller would not be able to accurately calculate the period. Placing a five-cycle delay in the algorithm significantly improved the accuracy of the period measurement, although it is probable that a one or two cycle delay would have also had the same beneficial effects.

Once the period had been computed, the microcontroller would output the DPA clock frequency until it was reset by the mechanical trigger. The mechanical trigger consisted of a lowpass filter with a relatively low break frequency – approximately 2 Hz – and a Schmitt Trigger that went from high (+5 V) to low (≈ 0 V) when the lowpass filter output reached its maximum value. The lowpass filter served two purposes: it filtered out the oscillating component of the charge amp output, which eliminated false resets when the beam was vibrating, and it provided a controlled step response that reached its maximum value approximately 1 or 2 seconds after the beam was held back at the tip. Holding the beam back approximately 1 or 2 seconds reset the microcontroller and enabled the filter adapt to the vibration frequency of the beam. This provided a simple method of adapting the filter parameters without a significant change in the beam demonstration.

Bridge Amplifier

The piezoceramic actuators were driven with a bridge amplifier to increase their control authority. Due to the fact that the beam vibration was relatively low frequency (< 30 Hz), standard operational amplifiers with approximately 20 mA of current output could be used to drive the actuators. The bridge amp was implemented on a LM324 quad op-amp that was operated from a ± 18 VDC supply, therefore the bridge amplifier was able to deliver a maximum of ± 36 to the actuators.

Digitally-Programmable Analog Implementation of the Adaptive Filter

The effectiveness of the digitally-programmable adaptive filter was demonstrated in a series of experiments on a flexible composite beam. The experimental setup, as shown in Figure 8, consisted of a 0.256 m long flexible clamped-free beam constructed from 6 plies of graphite epoxy. The width of the beam was 4.7 cm and the thickness was 1.52 mm. Piezoceramic actuators were bonded on both sides of the beam near the fixed end and a single piezoceramic sensor was bonded at approximately $1/3$ the length of the beam. The

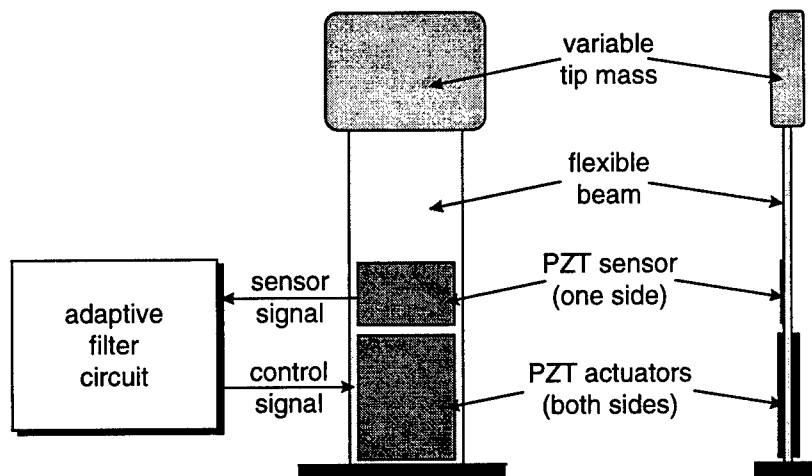


Figure 8: Schematic of the experiment demonstrating the digitally-programmable adaptive filter.

piezoceramic sensor was a single 0.254 cm layer while each piezoceramic actuator consisted of two 0.254 cm thick layers bonded on top of one another. Two layers were bonded on top of one another to increase the control authority of the structural actuators. A clip was placed at the free end of the beam so that mass could be easily added to the beam.

The control circuit was connected by hooking up the output of the piezoceramic sensor to the input of the charge amplifier and connecting the output of the bridge amplifier to the piezoceramic actuators (see Figure 8). The circuit itself was powered by a set of six 9 V batteries; four of the batteries provided the ± 18 VDC for the bridge amplifier and the remaining two batteries powered the ± 5 VDC for the digital components of the circuit.

The performance of the beam was measured by exciting the structure with an impact hammer and measuring the output of the surface bonded piezoceramic sensor. The input-output transfer function was measured with a spectrum analyzer over the frequency range 0 to 200 Hz. Closed-loop damping ratios were measured by curvefitting the measured frequency response and extracting the structural parameters from the best fit of the data.

The experiments were performed in two stages. First, an off-line identification procedure was used to determine the gain of the feedback filter. An appropriate choice for the filter gain was obtained using the measured frequency response between the piezoceramic actuators and the piezoceramic sensor. The DC asymptote of the frequency response was measured from the experimental data, and a filter gain/DC gain ratio of 0.4 was chosen so that approximately 20% critical damping would be obtained. This gain was then fixed throughout the remainder of the control experiments.

The performance of the adaptive filter was examined with three sets of experiments. In the first set of experiments, the filter was adapted to the natural frequency of the beam without any added tip mass and then additional mass was placed at the free end. For each value of the added mass, a frequency response between the impact hammer and the piezoceramic sensor output was obtained to determine the performance of the nonadaptive filter. Five values of mass ranging from 8.21 g to 38.1 g were added to the beam, which changed the natural frequency of the beam from 24.1 Hz (without any added mass) to approximately 11.3 Hz – over a factor of two variation in the frequency.

The second set of experiments was performed by adapting the filter to the 11.3 Hz natural frequency and removing mass from the free end. As with the previous experiment, the frequency response between the impact hammer and the piezoceramic sensor was measured for each value of removed mass. This set of experiments was performed to analyze the performance of the nonadaptive filter that has been designed to add damping to the 11.3 Hz mode of the beam.

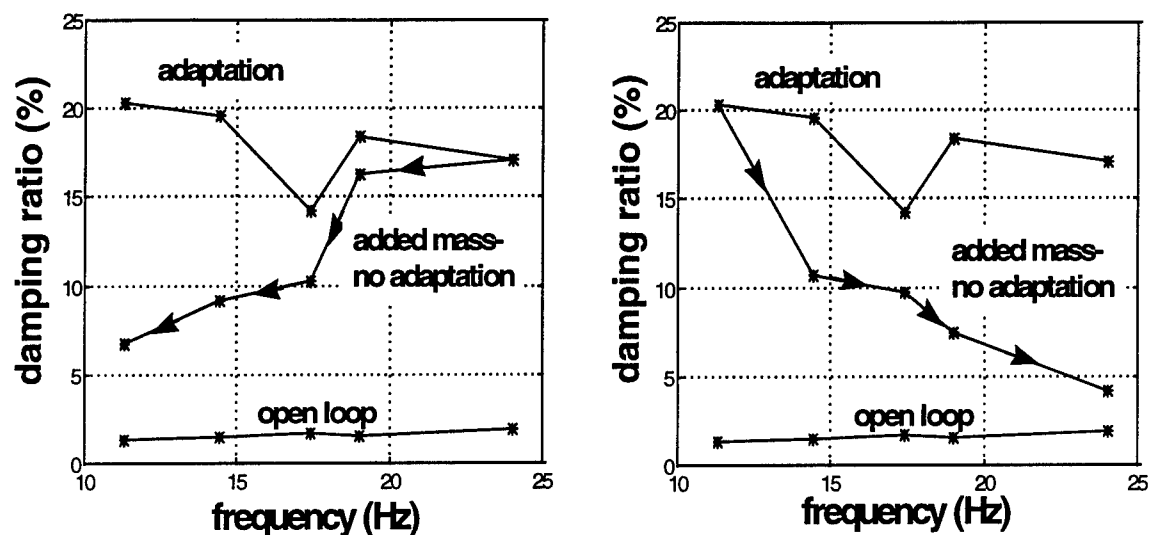


Figure 9: Measured damping ratios from the nonadaptive and adaptive filter tests: filter adapted to 24 Hz mode (left), and filter adapted to 11.3 Hz mode (right).

The final set of experiments analyzed the effectiveness of the adaptive filter that tracked the frequency of the beam. Starting with no additional mass on the free end of the beam, the filter was adapted to the first natural frequency of the beam and the performance transfer function was measured. Next, the smallest weight was added to the free end and the filter was adapted by mechanically triggering a reset in the microprocessor. A second frequency response was obtained of the beam with the adapted filter. This procedure was repeated for the remaining three values of added mass, and the resulting set of data was analyzed to determine the achievable damping ratio for a filter that tracks the first natural frequency of the vibrating beam.

The complete set of eighteen measurements, along with a set of six open-loop measurements, were analyzed to determine the closed-loop structural damping for each value of the added mass. The results are plotted in Figure 9. The plot on the left illustrates the closed-loop damping ratios achieved for the nonadaptive filter tuned to the 24 Hz mode and the closed-loop damping ratios of the adaptive filter. As expected, the damping ratios achieved with the nonadaptive filter drop sharply as the natural frequency of the beam decreases. Approximately a factor of three reduction (from 17% critical to roughly 6% critical) in the closed-loop damping ratio occurs as the beam frequency changes from 24 Hz to 11.3 Hz. In contrast, the adaptive filter maintains between 14% critical and 20% critical damping as the beam frequency drops. For the lowest frequency mode, 11.3 Hz, the adaptive filter has approximately three times more closed-loop damping as the nonadaptive filter.

Similar trends occur when the filter is adapted to the lowest frequency structural mode. The plot on the right in Figure 9 demonstrates that the drop in performance is even more pronounced when the structural mode increases in frequency and the filter parameters are fixed. In this case, the closed-loop structural damping reduces from 20% critical to less than 5% critical as the beam frequency increases from 11.3 Hz to 24 Hz. As with the previous case, the adaptive filter maintains a more constant level of performance over the range of added mass.

Figure 10 is a more detailed illustration of the effects of filter adaptation. The open-loop frequency response between the impact hammer input and the piezoceramic sensor output is compared to the frequency response after the filter has adapted to the 24 Hz mode (curves 1 and 2). Comparing these curves, we see that adapting the filter to the 24 Hz natural frequency reduces the peak response by approximately a factor of 10

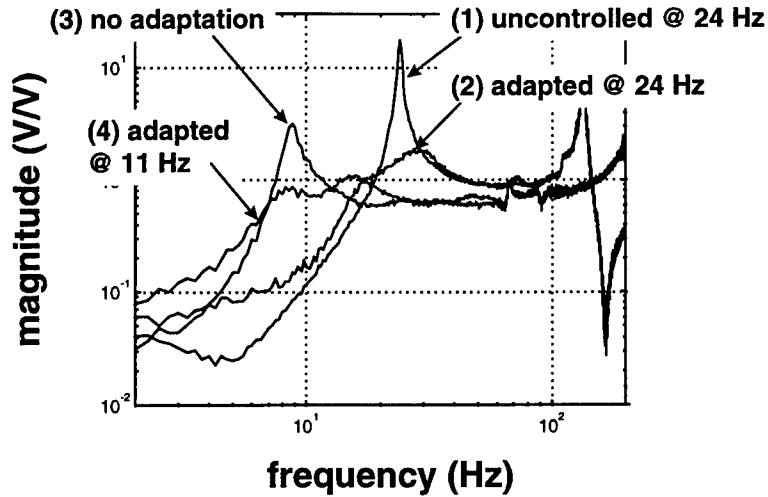


Figure 10: Frequency responses for no control, no added mass (1); 24 Hz filter, no added mass (2); 24 Hz filter, 38.1 g added mass and an 11.3 Hz structural mode (3); and 11 Hz filter and 38.1 g added mass (4).

and increases the damping ratio from $\approx 1\%$ to 17%. Adding the maximum tip mass reduces the resonance of the mode to 11.3 Hz and reduces the performance of the fixed-parameter filter (curve 3). Adapting the filter to the 11.3 Hz mode recovers the optimal performance and increases the structural damping from $\approx 5\%$ to slightly greater than 20% (curve 4).

Filter Adaptation

As previously described, the frequency adaptation algorithm was restarted with a mechanical trigger connected to the sensor output of the beam. This adaptation occurred over approximately a 500 μs interval after the microcontroller had been reset.

Although the frequency adaptation algorithm was a reliable method of determining the vibration frequency of the beam, there were instances in which the algorithm identified an incorrect natural frequency. These anomalies were attributed to the timing of the beam vibration with respect to the start of the identification algorithm. The algorithm was sensitive to incorrect timing and would identify the wrong vibration frequency approximately 5 to 10% of the time.

Making the algorithm more reliable would involve a more detailed analysis of the timing of the mechanical trigger and the frequency identification algorithm. It is probable that a more reliable triggering system, one less prone to the user vibrating the beam in the right manner, could be coded into the microcontroller. This would involve examining the output of the phase-locked loop and determining whether the microcontroller is counting the correct cycle of the PLL output.

Conclusions

PIC microcontrollers are viable processors for implementing active damping control laws, but certain limitations exist due to the current processor architecture. The speed of the processor limits digital filter sampling rates to roughly $5 \text{ kHz}/N$, where N is the number of second-order filter sections. Optimizing the current code and using a 33 MHz processor (as opposed to the 20 MHz processor used for these studies) would increase the sampling rate to approximately $9 \text{ kHz}/N$. Assuming the maximum achievable filter frequency to be 10 times less than the sampling rate, these limitations restrict the highest filter frequency to between

500 Hz and 900 Hz per second-order section. Fixed-point processing limits the lowest frequency achievable with the PIC microcontroller. For 8-bit arithmetic this limitation is approximately 0.03 times the sampling rate, whereas for 16-bit arithmetic this limitation is approximately 0.002 times the sampling rate. Thus, sampling rate limitations and the use of fixed-point arithmetic places an upper and lower bound on the frequency of the feedback compensators.

More design freedom is achieved by offloading the filter implementation to a digitally-programmable analog filter and using the PIC microcontroller in a 'supervisory' role. This 'hybrid' approach eliminates the problems due to sampling rate and fixed-point arithmetic by reducing the number of real-time computations that must be performed by the microcontroller. A circuit consisting of a digitally-programmable analog filter, a PIC16C65A microcontroller, and a digital phase-locked loop was designed and tested in a series of control experiments. The feedback compensator was implemented on the digitally-programmable analog filter, and a frequency-tracking algorithm was implemented using the phase-locked loop and the PIC microcontroller. The structure was a flexible clamped-free beam with a variable tip mass that changed the natural frequency from 24 Hz to approximately 11 Hz. A series of tests demonstrated that the closed-loop damping achieved with a fixed-parameter filter would be reduced by approximately a factor of three as the frequency varied from 24 Hz to 11 Hz. A second set of control experiments demonstrated that the adaptive filter could reliably track the natural frequency of the beam in approximately 500 μ s and maintain between 14% and 20% damping in the first structural mode.

One possible extension of this work would focus on the design of an adaptive filter that suppresses forced vibrations in the beam. The current algorithm is essentially a frequency-tracking approach that assumes that the gain and damping of the filter has been predetermined through an off-line identification. A more general algorithm would also include on-line identification of the optimal control gain and filter damping ratio. It is likely that this algorithm could be implemented using the current circuit design due to the fact that there is sufficient memory available in the microcontroller.

References

- [1] Balas, M.J., "Active control of flexible systems," *Journal of Optimization Theory and Applications*, vol. 25, no. 3, pp. 425-436, 1978.
- [2] Goh, C.J., Caughey, T.K., "On the stability problem caused by finite actuator dynamics in the collocated control of large space structures," *International Journal of Control*, vol. 41, no. 3, pp. 787-802, 1985.
- [3] Fanson, J.L., Caughey, T.K., "Positive position feedback control for large space structures," *AIAA Journal*, vol. 28, no. 4, pp. 717-724, 1990.
- [4] Dosch, J.J., Leo, D.J., Inman, D.J., "Modeling and control for vibration suppression of a flexible active structure," *Journal of Guidance, Control, and Dynamics*, vol. 18, no. 2, pp. 340-346, 1995.
- [5] Anderson, E. H., Holcomb, M. D., Leo, D. J., Bogue, A. X., and Russo, F.L., "Integrated Electromechanical Devices for Active Control of Vibration and Sound," 1997 International Mechanical Engineering Conference, Dallas, TX, Volume AD-54.
- [6] Franklin, G.F., Powell, J.D., Workman, M.L., *Digital Control of Dynamic Systems*, Addison-Wesley Publishing Company, Reading, MA, 1990.
- [7] Inman, D.J., *Engineering Vibration*, Prentice-Hall, Englewood Cliffs, NJ, 1994.
- [8] Palacherla, A., "AN540: Implementing IIR Digital Filters," *Embedded Control Handbook*, Microchip Corporation.
- [9] Maxim Integrated Products, "MAX260/261/262 Microprocessor Programmable Universal Active Filters," Product Data Sheet.
- [10] Best, R.E., *Phase-Locked Loops*, McGraw-Hill, New York, 1984.
- [11] Lindsey, W.C., Simon, M.K., Editors, *Phase-Locked Loops and their Applications*, IEEE Press, New York, 1978.

Associate did not participate in the program.

**EXPERIMENTAL VALIDATION OF THREE-DIMENSIONAL
RECONSTRUCTION OF INHOMOGENEITY IMAGES IN TURBID MEDIA**

**Hanli Liu
Assistant Professor
Joint Program in Biomedical Engineering**

**The University of Texas at Arlington
P.O. Box 19138
Arlington, TX 76019**

**Final Report for:
Graduate Student Research Program
Kirtland AFB
Phillips Laboratory**

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

And

Phillips Laboratory

August 1997

EXPERIMENTAL VALIDATION OF THREE-DIMENSIONAL RECONSTRUCTION OF INHOMOGENEITY IMAGES IN TURBID MEDIA

Hanli Liu
Assistant Professor
Joint Program in Biomedical Engineering
The University of Texas at Arlington and
The University of Texas Southwestern Medical Center at Dallas

Abstract

Near infrared radiation for imaging tumors in tissue has been recently explored and investigated. In particular, to validate a novel FFT imaging reconstruction algorithm, we have performed experiments on tissue phantoms using a frequency-domain photon migration system. The system includes an amplitude-modulated laser diode, two radio frequency (RF) signal generators, an avalanche photo detector, and a lock-in amplifier. The tissue phantoms were made with inhomogeneities imbedded and were then scanned with the system at various RF frequencies. The data from the amplitude and phase measurements were analyzed using the FFT reconstruction algorithm. The reconstructed data show clear validation of the FFT algorithm and afford to localize inhomogeneities hidden in turbid media in three dimensions. In addition, based on the results, we present preliminary analysis on optimization of experimental parameters to obtain good-quality, reconstructed images with best experimental efficiency.

EXPERIMENTAL VALIDATION OF THREE-DIMENSIONAL RECONSTRUCTION OF INHOMOGENEITY IMAGES IN TURBID MEDIA

Hanli Liu

1. Introduction

In the last few years, numerous research efforts and technology developments have been made towards using near infrared (NIR) radiation for human tissue quantification and imaging. Three forms of light used in the field include short-pulsed light in time-domain (TD), intensity-modulated light in frequency-domain (FD), and a continuous wave form (CW). Among them, the frequency-domain system has its advantages over the other two: a FD system can acquire data almost in real time in an order of ms, much faster than a TD system, which utilizes a single photon counting method and requires much longer data acquisition time (in multi seconds). In addition, with the help of diffusion theory, the FD system allows us to extract the optical properties of tissues/organs under study and to image inhomogeneities hidden in tissues/organs. In comparison, CW method is in lack of quantification unless close-distanced (2-5 mm) measurements are performed.

The light source used for a FD system is intensity modulated at a radio frequency (RF) ranging from several MHz up to GHz at the wavelength range from 650 nm to 900 nm. In principle, NIR light illuminating tissues becomes very much scattered about 1 mm after the light travels in tissue. In the FD case, the intensity-modulated light travelling in tissue will form diffuse photon density waves (DPDW's), and the amplitude and phase of DPDW's carry information on the optical properties of the tissue. By extracting amplitude, phase, or optical properties of thick tissue, we can reconstruct tissue images that show inhomogeneities imbedded inside the thick tissue. This provides a new methodology for tumor imaging in medical applications.

Over last few years, common geometries used for imaging studies are 1) source(s)-detector pairs scanning horizontally or 2) circularly.¹ In both cases, mechanical movement is required, and it provides only two-dimensional images. Recently, Matson et al proposed a new approach to three-dimensional tumor localization in turbid media with the use of FD measurements in a single plane.² The theoretical basis of this novel reconstruction algorithm stems from Fourier Optics: applying Fast Fourier Transform (FFT) to the diffusion equation that DPDW's obey, and then extending the solution of Fourier Optics to DPDW's. This FFT reconstruction algorithm has several advantages, such as, 1) the light source does not need to move mechanically, 2) it has a potential to avoid mechanical movement, which has been shown to produce significant errors in imaging reconstruction. The theory has yet to be verified by experiments, although it was supported by computer simulations at one RF frequency of 1 GHz with selected scanning parameters.

The goal of this summer research project is to experimentally validate and optimize this three-dimensional reconstruction algorithm to image inhomogeneities in turbid media. In this report, we will first introduce the theoretical foundation for the FFT reconstruction algorithm, followed by a brief description on experimental setups, materials and methods. Then we will show experimental results and reconstructed images, and then investigate the correlation among experimental parameters that affect the quality of the reconstructed images. Furthermore, computer simulations are performed to confirm the findings that a larger scanning area is more crucial than smaller scanning pixels to obtain a higher resolution image. Finally, we will draw conclusions from this study and indicate future work.

2. Theoretical Background

Based on Fourier Optics³, we define the following relations:

$$A_z(f_x, f_y) = \iint U(x, y, z) e^{-i2\pi(f_x x + f_y y)} dx dy, \quad (1)$$

$$U(x, y, z) = \iint A_z(f_x, f_y) e^{i2\pi(f_x x + f_y y)} df_x df_y, \quad (2)$$

where $U(x, y, z)$ is the complex field of a monochromatic wave across the plane at z , and A_z is a two-dimensional Fourier transform of the function $U(x, y, z)$. It can be shown that a free-space solution to the source-free Helmholtz equation exists in the form of spatial frequency spectrum across a plane parallel to the xy plane but at a distance z from it.³

In addition, it has been proved⁴ that the diffusion equation adequately represents photon propagation in tissue, and the source-free diffusion equation in FD can be written as

$$\nabla^2 U(x, y, z) + k^2 U(x, y, z) = 0, \quad (3)$$

where $U(x, y, z)$ is the spatial distribution of a DPDW in a scattering medium, and $k^2 = (-v\mu_a + i\omega)/D = 3\mu_s'(-\mu_a + i\omega/v)$, v is the speed of light travelling in tissue, ω is the modulation frequency of the light multiplied by 2π , and μ_a and μ_s' are absorption and reduced scattering coefficients of the scattering medium, respectively.

Now inserting the inverse Fourier transform definition, eq. (2), into the diffusion equation, eq. (3), results in:

$$\frac{d^2}{dz^2} A_z(f_x, f_y) + \left[k^2 - (2\pi f_x)^2 - (2\pi f_y)^2 \right] A_z(f_x, f_y) = 0. \quad (4)$$

In analogy to the solution obtained in Fourier Optics for regular plane waves, we can find a solution to eq. (4) using the Fourier transform definition:

$$A_{z_1}(f_x, f_y) = A_{z_0}(f_x, f_y) e^{i\Delta z \sqrt{k^2 - (2\pi f_x)^2 - (2\pi f_y)^2}}, \quad (5)$$

where $\Delta z = z_0 - z_1$, is the distance between the detection plane at z_0 and the reconstruction plane z_1 . Furthermore, we can define a transfer function, $H_{\Delta z}(f_x, f_y)$, describing how a DPDW propagates in a homogeneous medium,

$$H_{\Delta z}(f_x, f_y) = \frac{A_{z_1}(f_x, f_y)}{A_{z_0}(f_x, f_y)} = \exp \left\{ i \Delta z \sqrt{\left(\frac{-v\mu_a + i\omega}{D} \right)^2 - (2\pi f_x)^2 - (2\pi f_y)^2} \right\}. \quad (6)$$

It is worthwhile to notice several important points:

- 1) in using diffusion theory, we neglect boundary effects and assume the medium to be homogeneous with spherical inhomogeneities;
- 2) in applying Fourier Optics to obtain equation (5), we extend the wave vector, k , from a real number to a complex number to include scattering effects, as the diffusion equation does;
- 3) eq. (6) describes the change in the Fourier transform of the photon density wave as it propagates, and thus eq. (6) can be used to invert the propagation and to reconstruct the photon density wave behind the detection plane.

3. Necessity of Experiments

The primary goal of this summer research project is to experimentally validate the FFT reconstruction algorithm, as introduced above. Taking experiments is necessary because

- 1) experiments permit studying boundary effects on the reconstructed images by using finite sizes of tissue phantoms;
- 2) we can study instrument detectability and sensitivity in locating hidden objects with different sizes and optical properties;
- 3) we can optimize scan parameters by performing experiments with various scan dimensions, pixel sizes, and modulation frequencies;
- 4) measurement on homogeneous samples provide feasibility of obtaining correct background information for the reconstruction and demonstrate the importance of such information.

4. Experimental Setup

As shown in Figure 1, the electronic components of our FD system include two RF signal generators, (Rhode & Schwarz, SMY01) ranging from 9 kHz to 1.04 GHz, two 50% power splitters (Mini-Circuits, ZFSC-2-1), two RF amplifiers (Mini-Circuits, ZFL-1000H), a NIR laser diode (LaserMax Inc, LSX-3500) modulatable up to 100 MHz at 780 nm and 680 nm with laser powers of 3 mW (at 116 mA) at 9 mW, an Avalanche Photo Diode (APD) (Hamamatsu, C5331), two frequency mixers (Mini-Circuits, ZLW-2), two low pass filters with the corner frequency at 1 MHz (Mini-Circuits, SLP-1.9), a lock-in amplifier (Stanford Research Systems, SR850), and a computer-controlled, two-dimensional positioning stage (Aerotech, Inc. ATS100).

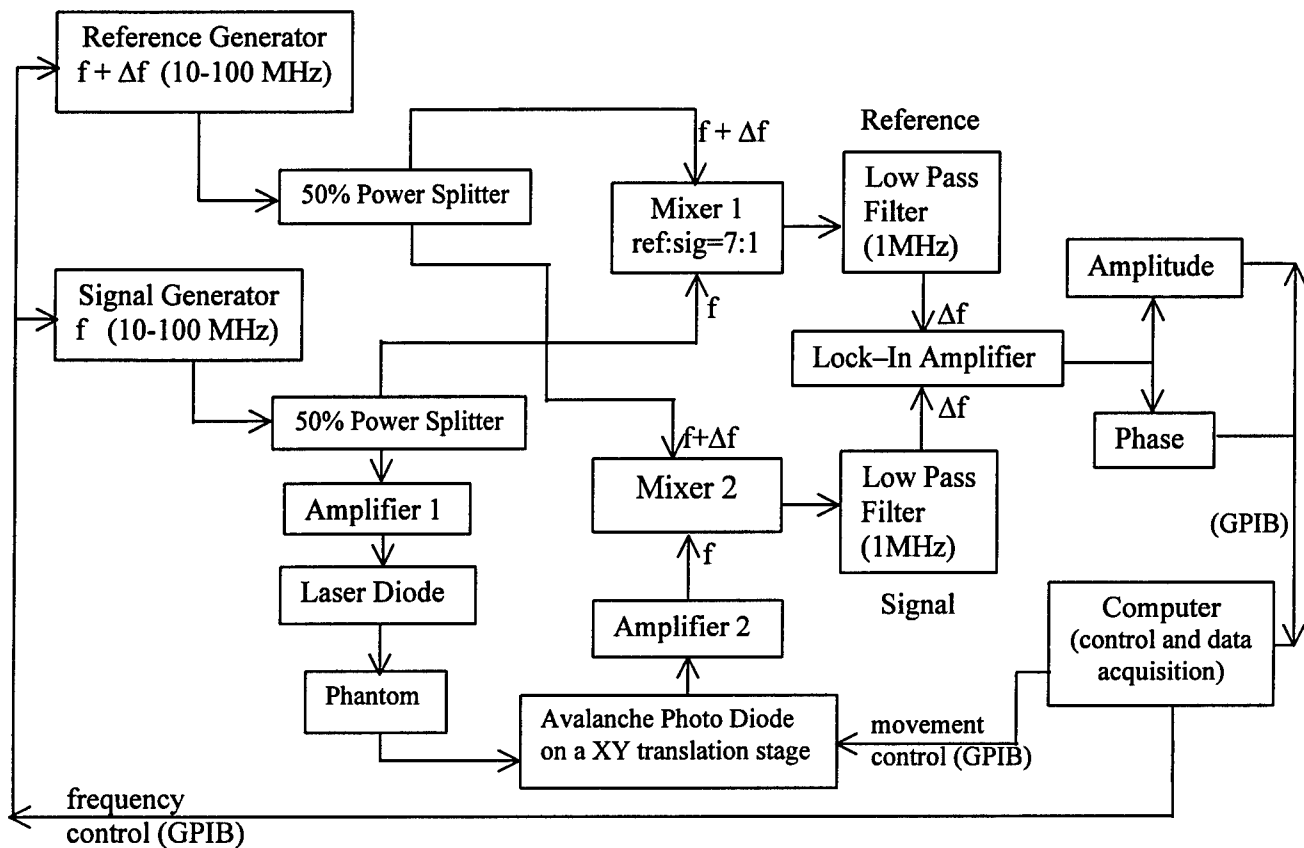


Figure 1. Experimental Setup of a FD system

The frequencies of the two signal generators were chosen from 10 to 100 MHz with an offset $\Delta f = 25$ kHz for the reference signal with respect to the measurement signal in order to achieve a heterodyne system. Both RF signals for the reference at $f + \Delta f$ and for the measurement at f were 50% power split, and a branch from each the two RF signal generator was mixed by Mixer 1 to generate a reference signal for a lock-in amplifier. The output levels of the two signal generators were set up with a ratio of 7 to 1 between the reference and measurement signals to optimize the mixing efficiency. The second branch of the measurement signal at f was amplified to gain enough power to modulate the intensity of a NIR laser diode at 780 nm. Then the intensity-modulated, manufacture-collimated laser light illuminated a tissue phantom and then was detected by a movable APD to convert the optical signal to electrical signal. A second amplifier amplified the output of APD before the electrical signal was mixed with the second branch of the reference signal at $f + \Delta f$ using Mixer 2. The output frequencies of Mixer 1 and Mixer 2 consisted of $2f + \Delta f$, $f + \Delta f$, f , Δf , and higher harmonic components. By utilizing two low pass filters with a corner frequency of 1 MHz, only the signal with frequency of Δf at 25 kHz passes to the lock-in amplifier. The corresponding readings for amplitude and phase were recorded

automatically by a computer; in turn, feedback from the computer controlled the sensitivity and time constant settings for the lock-in amplifier based on the readings. In addition, through a GPIB interface, the computer set the RF frequencies for the two signal generators and controlled a XY positioning stage, which held the APD for 2-dimensional imaging scans.

5. Materials and Methods

Tissue phantoms were made out of plastic resin that hardens in molds by adding catalyst. The resin is combined with TiO_2 powder for scattering property and sometimes with NIR dye for absorption property.⁵ The optical properties of homogeneous phantoms were determined using the slope algorithm⁶ in transmission geometry. Since the algorithm requires the information on the homogeneous background medium that contains an object/objects, multiple samples were made as a set at the same time to ensure that the samples of the set have the same background optical properties. In general, a set of phantoms included a homogeneous sample, and two others with different objects implanted inside. The phantom sizes were typically 14cm x 14cm x 5cm, and absorption and scattering coefficients of background media were within a range of 0.01 to 0.1 cm^{-1} and 3 to 17 cm^{-1} , respectively. The objects used as inhomogeneities were plastic black, white, and clear beads with various diameters and shapes, as shown and listed in Figure 2.

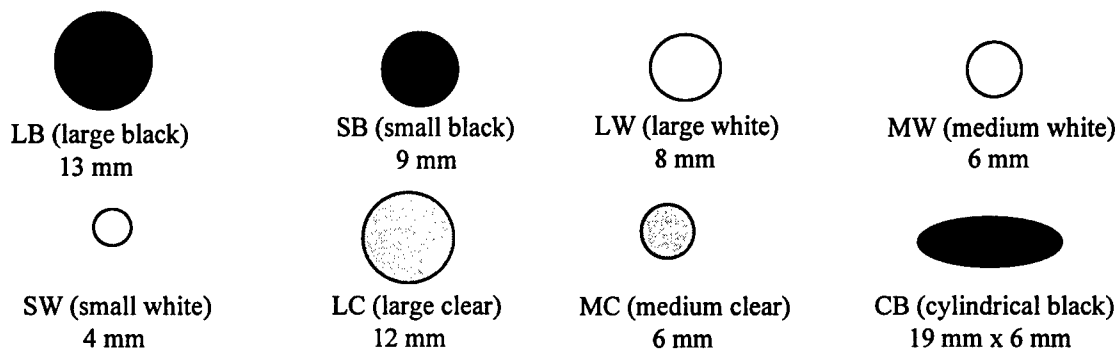
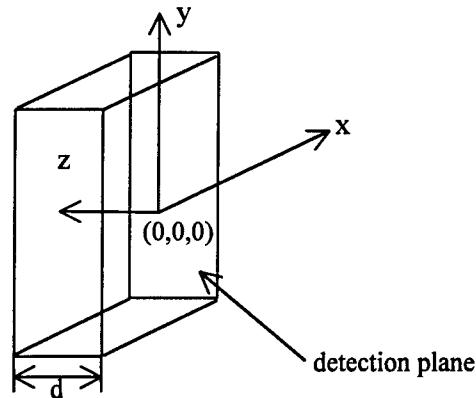


Figure 2. Beads used as tumor objects imbedded in tissue phantoms

We will use a XYZ coordinate system, shown in Figure 3, in this report to refer to the positions of hidden objects in the phantoms. The XY plane is in parallel with the detection plane, the origin is at the center of the detection plane, and the Z axis is pointing from the detection plane into the phantom. Normally, a bead was placed at the center of a phantom with a coordinate of $(0, 0, d/2 \text{ cm})$, where d is the thickness of the phantom. For multiple hidden tumors, two beads (the same or different kinds) were placed at $(0, -2 \text{ cm}, -d/2 \text{ cm})$ and $(0, 2 \text{ cm}, -d/2 \text{ cm})$ inside of a phantom. In making inhomogeneous phantoms, we positioned the object(s) on top of a semi-dried, background material which has been cast earlier, and then added equal amount of ready-to-cast resin on top of the first part to complete the

phantom. In this way, we implanted the objects inside the phantoms with good 3-dimensional coordinate references, and X-ray measurements have confirmed the positions of imbedded objects.

Figure 3. the xyz coordinate system used to label locations of hidden objects inside a phantom.



All of the scans used for this report were performed on areas of 8 cm x 8 cm having 64x64, 32x32, or 16x16 pixels with pixel spacing of 0.125 cm, 0.25 cm, and 0.5 cm, respectively. The modulation frequencies were scanned from 10 MHz to 100 MHz in various steps in actual experiments, but all the data shown in this report were taken mainly at 20 or 40 MHz. An average measurement time was about 2.5 hours for a 32x32 scan with two modulation frequencies (20 and 40 MHz). The experiments were performed in transmission geometry with the light source fixed at one side of the sample and the movable detector located on the other side of the sample. The software package Interactive Data Language (IDL) was used to process the measured data and to perform the reconstruction algorithm. In addition, the PMI⁷ simulation code was used to predict and confirm the experimental data.

Because the transfer function contains an exponential decay term, the reconstructed DPDW's are not stable, containing ringing patterns. A better understanding for this artifact is under investigation. However, at present stage, to stabilize the reconstructed data or eliminate ringing patterns, we utilized a low-pass filter (pillbox) to cut off high spatial frequency components. Specifically, the cutoff frequency was ranged from only a few pixels, such as 1-4, in the spatial frequency domain. We multiplied the Fourier spectra by this cutoff filter before inverse Fourier transforming the spatial frequency spectra to reconstruct the DPDW's. For convenience, in the rest of the report, we will use "filter" as a short notation to represent the pixel number in Fourier domain of cutoff spatial frequencies.

After the amplitude, A , and phase, θ , were measured as a function of positions in a XY plane at $z=z_0$, we followed the procedures given below to reconstruct three-dimensional images:

- 1) based on A and θ , calculate complex photon density waves $U_h(x,y,z_0)$ and $U_{inh}(x,y,z_0)$, respectively, for both homogeneous sample and the sample containing objects;
- 2) subtract the homogeneous wave from the inhomogeneous wave, $U(x,y,z_0)=U_{inh}(x,y,z_0)-U_h(x,y,z_0)$;

- 3) FFT $U(x,y,z_0)$ to obtain its Fourier transform $A(f_x, f_y, z_0)$;
- 4) multiply $A(f_x, f_y, z_0)$ by the transfer function to execute the reconstruction process at z_1 ;
- 5) multiply $A(f_x, f_y, z_1)$ by a low-pass filter with a selected filter number (cutoff pixels);
- 6) inverse FFT $A(f_x, f_y, z_1)$ to the spatial domain for the reconstructed image at z_1 ;
- 7) repeat steps 4) to 6) for different values of z_1 to back propagate the reconstruction;
- 8) determine the x, y coordinates of the hidden object at the detection plane, and then plot amplitude versus z to localize the tumor in Z direction.

6. Experimental Results and Image Reconstructions

Figure 4 shows an example of amplitude images of DPDW's measured at the detection plane using a 680 nm laser scanned with 32x32 pixels on a 8 cm x 8 cm area of a phantom containing one black object. The phantom was 4.7 cm thick, and the black bead had a diameter of 9 mm imbedded at (0, 0, 2.3 cm). The illumination spot on the phantom was located at (-1.5 cm, 0, 2.3 cm), 1.5 cm off-centered in x axis with respect to the center of the phantom. The modulation frequency used for this set of measurements was 20 MHz, and the filter used for the reconstruction was 2. The absorption and reduced scattering coefficients of the background medium were 0.14 ± 0.06 and $3.1 \pm 1.4 \text{ cm}^{-1}$, respectively, which were determined using a homogeneous sample. Three amplitude plots of the complex DPDW's in Figure 4 are obtained at the detection plane for a homogeneous field [4(a)] measured from the homogeneous sample, perturbed field [4(b)] from the inhomogeneous sample, and the deviation between the two fields [4(c)], i.e., perturbed field (pfield) - homogeneous field (hfield). Notice that in obtaining the image of 4(c), we first subtracted the two complex fields (pfield-hfield), which were calculated from amplitude and phase measurements of the two samples, and then plotted the amplitude of the subtracted field.

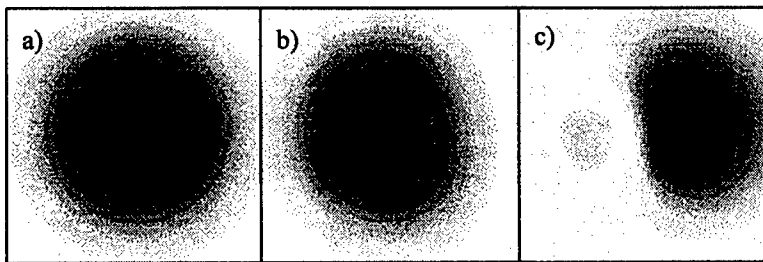


Figure 4. Amplitude images of the DPDW's measured at the detection plane from a phantom with one hidden high-absorbing object, scanned at 20 MHz with an array size of 32 x 32. a) homogeneous field=hfield, b) perturbed field=pfield, c) pfield hfield.

Figure 5 uses the same set of data as those for Figure 4. In the case of Figure 5, the FFT reconstruction algorithm was utilized to obtain the images at different depths behind the detection plane. The filter used here was 2. Let's first look at Figures 5(c) to 5(e), which are amplitude images at the

detection plane, 1.25 cm, and 2.5 cm behind the detection plane, respectively. The center of the actual black object is 2.3 cm behind the detection surface and 1.5 cm off from the center of the scanned area in X direction. These three images clearly and consistently show the location of the center of a hidden inhomogeneity in the XY plane at different depths, indicating that the object is a few centimeters deep behind the detection plane. For more quantitative localization of the object in Z direction, we need to have a XZ/YZ image or a plot in Z direction at the selected x and y coordinates of the detected object determined from the previous XY images. Figures 5(a) and 5(b) are cross-section images at XZ and YZ planes passing through the center of the scanned area. The top of Figure 5(a) corresponds to the detection plane along x axis, and the bottom represents the illumination plane. In this case, this XZ plane goes through the center of the object horizontally at $y=0$, and one can see that the imaged object is 1.4 cm off from the center in x direction. A plot of amplitude versus Z at the corresponding x coordinate provides the location of the object in Z direction (not shown). Same principle applies to Figure 5(b), in which case the YZ plane is at $x=0$, 1.5 cm apart from the plane passing through the center of the object. Because of this, this image does not give accurate location of the object in Z direction.

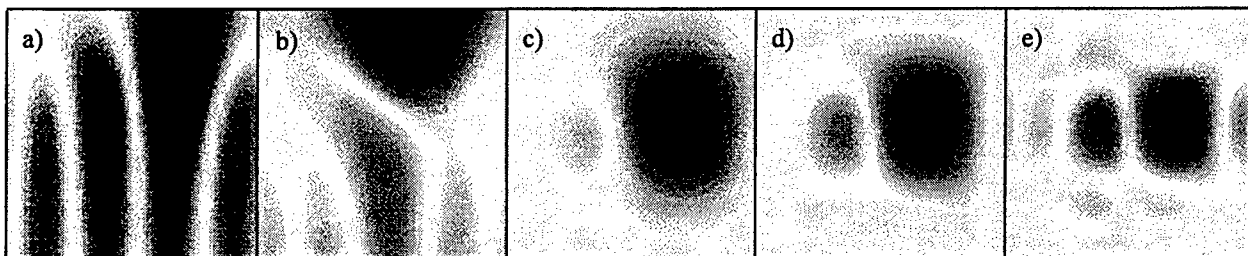


Figure 5. Reconstructed images for a phantom with one object, scanned at 20 MHz with an array size of 32 x 32. a) x-z plane: the plane through the center of the measured volume horizontally, b) y-z plane: the plane through the center of the measured volume vertically, c) x-y plane: the detection plane, d) x-y plane: 1.25 cm behind the detection plane, e) x-y plane: plane near the object's center, 2.5 cm behind the detection plane.

Figure 6 is obtained from another set of two phantoms, one of which is homogeneous and the other having two high-absorbing objects with diameters of 9 mm. With the use of a 780 nm laser modulated at 20 MHz, the absorption and reduced scattering coefficients of the background medium are found to be 0.015 ± 0.001 cm and 15.1 ± 0.3 cm, respectively. The thickness of the inhomogeneous phantom is 4.2 cm. The two objects were imbedded inside the phantom with a separation of 4 cm apart at $(0, -2$ cm, 2.1 cm) and $(0, 2$ cm, 2.1 cm). The phantom was set 0.5 cm off centered with respect to the laser and the scanning plane along x axis. Similarly to Figure 5, Figures 6(a) and 6(b) are amplitude images measured from the homogeneous and inhomogeneous phantoms, and subtraction of the two complex fields of the two samples gives Figure 6(c). These three images are obtained at the detection

plane and show clearly two hidden objects after the simple subtraction. Furthermore, if the FFT reconstruction algorithm is applied, we can obtain the reconstructed amplitude images of DPDW's at different depths behind the detection plane. Figures 7(c) to 7(f) show such amplitude images reconstructed at the detection plane, 1 cm, and 2 cm behind the detection plane, respectively. All three images indicate consistently and unambiguously two hidden objects at about (0.5 cm, 2cm) and (0.5 cm, -2 cm) in the XY plane at different depths. As the reconstructed plane goes deeper and closer to the center of the hidden objects, the resolution for the imaged objects becomes better. In this case, a YZ image, shown in Figure 7(b), supports the findings from the XY images, and it also provides locations of the two objects in depth, which occur at the peaks of the two dark, cylindrical spots. Since the image plane for Figure 7(a) is in between the two objects, not passing through either center of the hidden objects, this XZ image does not give us very useful information in locating objects.

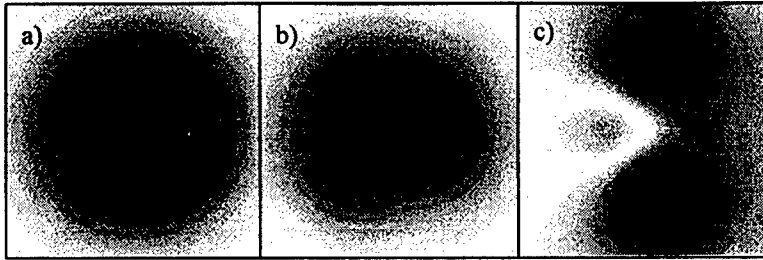


Figure 6. Amplitude images of the DPDW's measured from a phantom with two high-absorbing objects, scanned at 20 MHz with an array size of 32 x 32. a) hfield, b) pfield, and c) pfield-hfield.



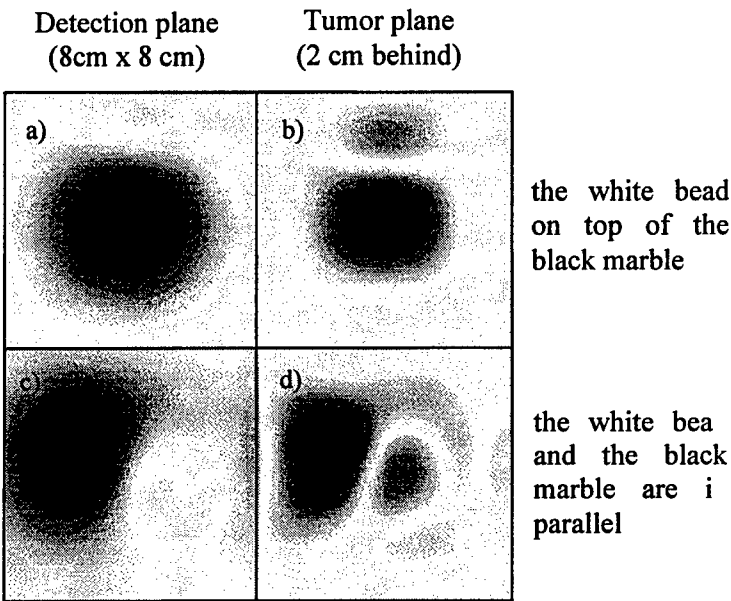
Figure 7. Reconstructed images for a phantom with two objects, scanned at 20 MHz with an array size of 32 x 32 for a scan area of 8 cm x 8cm with Filter 2 used for the reconstruction. a) x-z plane: the plane through the center of the measured volume horizontally, b) y-z plane: the plane through the center of the measured volume vertically, c) x-y plane: at the detection plane, d) x-y plane: 1 cm behind the detection plane, and e) x-y plane: plane near the object's center, 2 cm behind the detection plane.

Notice that there are multiple bright, ringing patterns/loops in Figure 7(b); they are artifacts due to Fourier transform. Higher filters used in the Fourier domain for the reconstruction algorithm give more ringing patterns. This limits the bandwidth of the spatial frequency components in the reconstruction and

thus limits the spatial resolution of the reconstructed images. If higher spatial frequency components are included, we have to be careful to differentiate the real hidden objects from FFT reconstruction artifacts. In the next section, we will discuss how to select filter numbers, or cutoff spatial frequencies, to optimize the reconstructed images. Further studies and understanding of the cause of ringing patterns are underway.

The phantom used to obtain Figure 8 contains two objects, one black marble having 13 mm diameter and one white bead of 8 mm diameter. They were implanted at the middle of a 4.5 cm thick phantom in Z direction and about 2.5 cm apart vertically (or horizontally) symmetric to the center of the phantom. Similarly, 32x32 pixels were chosen to scan a 8cm x 8cm area with a modulation frequency of 20 MHz. The absorption and reduced scattering coefficients of the background are $0.017\pm0.001\text{ cm}^{-1}$ and $15.9\pm0.6\text{ cm}^{-1}$, respectively. Figures 8(a) and 8(b) are amplitude plots, respectively, at the detection plane and 2 cm behind the detection plane near the centers of the objects. It is seen that at the detection plane, the image for the white object does not appear very well, whereas the image for the black object is very obvious and clear. The reconstructed image on the XY plane near the centers of the objects (Figure 8(b)), however, displays two clear imaging spots with different magnitudes. The big, dark spot is corresponding to the black marble, and the small, light spot stems from the white bead. We can easily obtain the separation distance of 2.7 cm between the spots from Figure 8(b). To confirm the result, we turned the phantom 90° around the Z axis (the axis perpendicular to the detection plane) and rescanned the phantom. The reconstructed images for this case are shown in Figures 8(c) and 8(d), and the results are very consistent with those observed in Figures 8(a) and 8(b).

Figure 8. Amplitude images of a phantom with two objects, a 13 mm black marble and a 8 mm white bead, scanned at 20 MHz with an array size of 32 x 32. a) and c): at the detection (x-y) plane, b) and d): near the center of the samples, 2 cm behind the detection plane. Filter of 2 was used for this reconstruction.



In order to know how different wavelengths and powers of the laser affect the quality of the reconstructed images, we performed two measurements on the same phantom using a 780 nm, 3 mW and a 680 nm, 9 mW laser diode. In this case, the phantom contained two objects, one 12 mm black marble and 12 mm clear bead. These two objects were imbedded in the middle of a 5 cm thick phantom, and other parameters were the same as those used for Figure 8. The background medium has absorption and reduced scattering coefficients of $0.016 \pm 0.00 \text{ cm}^{-1}$ and $10.0 \pm 0.7 \text{ cm}^{-1}$ at 780 nm, and $0.015 \pm 0.001 \text{ cm}^{-1}$ and $13.7 \pm 0.3 \text{ cm}^{-1}$ at 680 nm. With a 3 mW laser at 780 nm, we reconstructed amplitude images of the phantom, shown in Figures 9(a) and 9(b), at the detection plane and at the plane passing the centers of the objects, 2.5 cm behind the detection plane. It is seen that the clear bead does not give a very distinguished, resolvable feature in either case. Furthermore, after we switched to a higher (9 mW) power laser at 680 nm, the reconstructed images (Figures 9(c) and 9(d)) look similar to those with the 780 nm laser, except that the 680 nm, higher power laser gives a little better spatial resolution than the other case. This may be because a higher power laser gives a better signal-to-noise ratio. But overall, there is no significant difference between these two lasers in imaging large-mismatched objects in our phantoms.

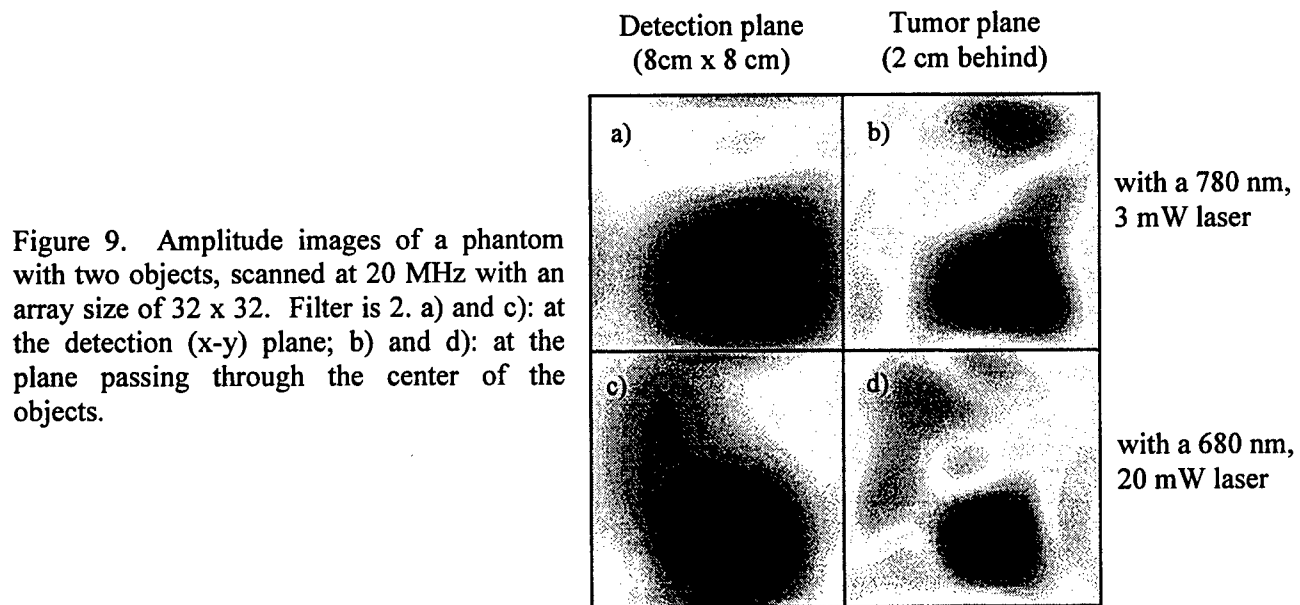
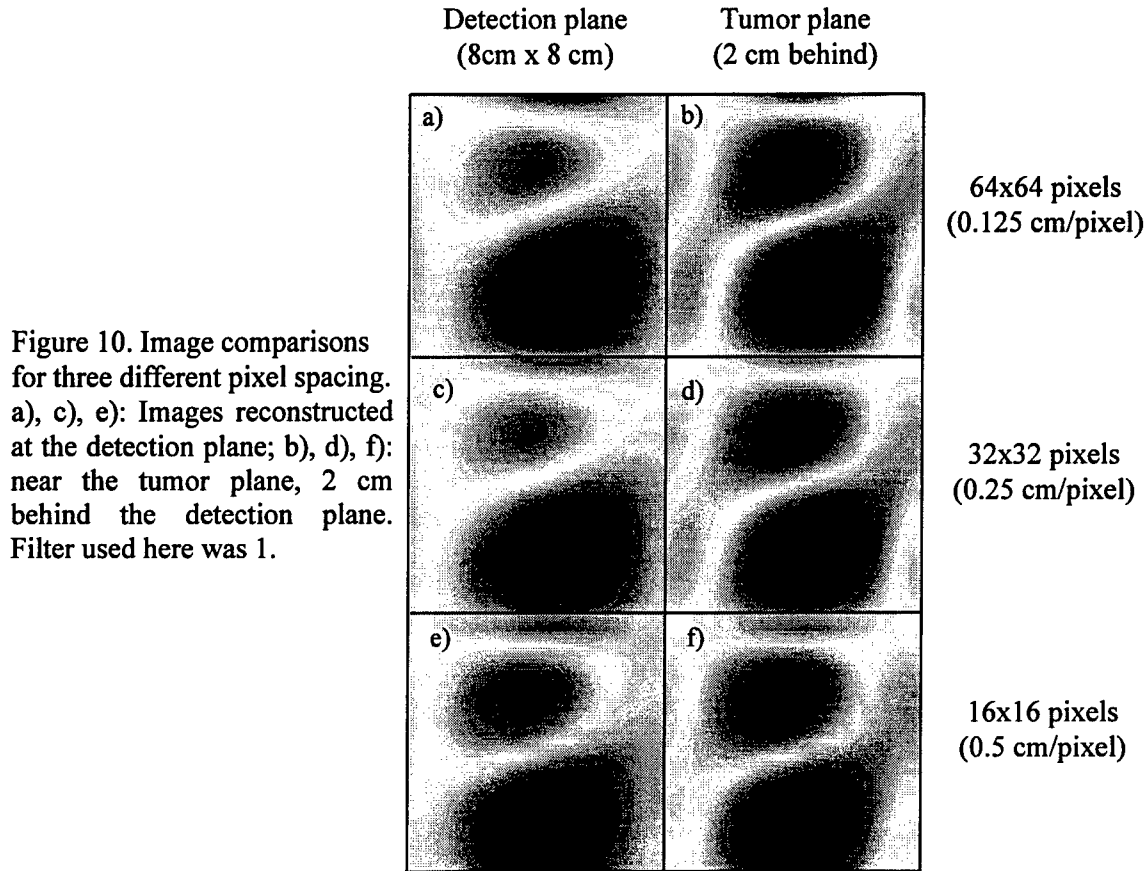


Figure 9. Amplitude images of a phantom with two objects, scanned at 20 MHz with an array size of 32 x 32. Filter is 2. a) and c): at the detection (x-y) plane; b) and d): at the plane passing through the center of the objects.

To study the relationship between the quality of the reconstructed images and pixel spacing, we also performed a series of scans on the same phantom with array sizes of 64 x 64, 32 x 32, and 16 x 16 having pixel spacing of 0.125 cm/pixel, 0.25 cm/pixel, and 0.5 cm/pixel, respectively. In this case, the phantom was 4.4 cm thick and contained a 8 mm black and a 9 mm white bead. They were located at the middle depth of the sample and about 4 cm apart along Y axis with the white bead on top of the black bead. Figure 10 presents a set of reconstructed amplitude images at both the detection plane and at the

plane near the centers of the objects, 2 cm below the detection plane. The filter used here was 1, and the laser was at 780 nm. Figure 10 shows clearly that there is not much difference in image quality among the reconstructed images using either smaller (0.125 cm/pixel) or larger (0.5 cm/pixel) pixel spacing.



To obtain the object location in Z direction, we first determine the x and y coordinates of the detected object(s) based on XY images, such as those shown in Figure 10, and then plot the amplitude versus Z to obtain the peak position. The z coordinate at the peak of the amplitude will be the location of the hidden object in Z direction. To show this strategy, we plot a z plot in Figure 11(a) based on Figure 10(d) at $X_{\text{pixel}}=20$, $Y_{\text{pixel}}=6$, which are the pixel coordinates of the observed, dark spot for the hidden black bead. In this case, the pixel step in Z direction is the same as that used in X and Y directions, i.e., 0.25 cm/pixel. Figure 11(a) displays a peak at pixel number=8 behind the detection plane, implying that this peak is at $8 \times 0.25 \text{ cm}=2 \text{ cm}$ behind the detection plane. Similarly, Figure 11(b) plots the amplitude versus Z pixels at $X_{\text{pixel}}=20$, $Y_{\text{pixel}}=22$, for the other observed spot resulting from the hidden white bead. This figure shows a dip at $Z_{\text{pixel}}=8$ behind the detection plane, giving the consistent location of the hidden

object in Z direction. A dip, rather than a peak, in amplitude might stem from a possible fact that the absorption of the white bead is lower than the background material.

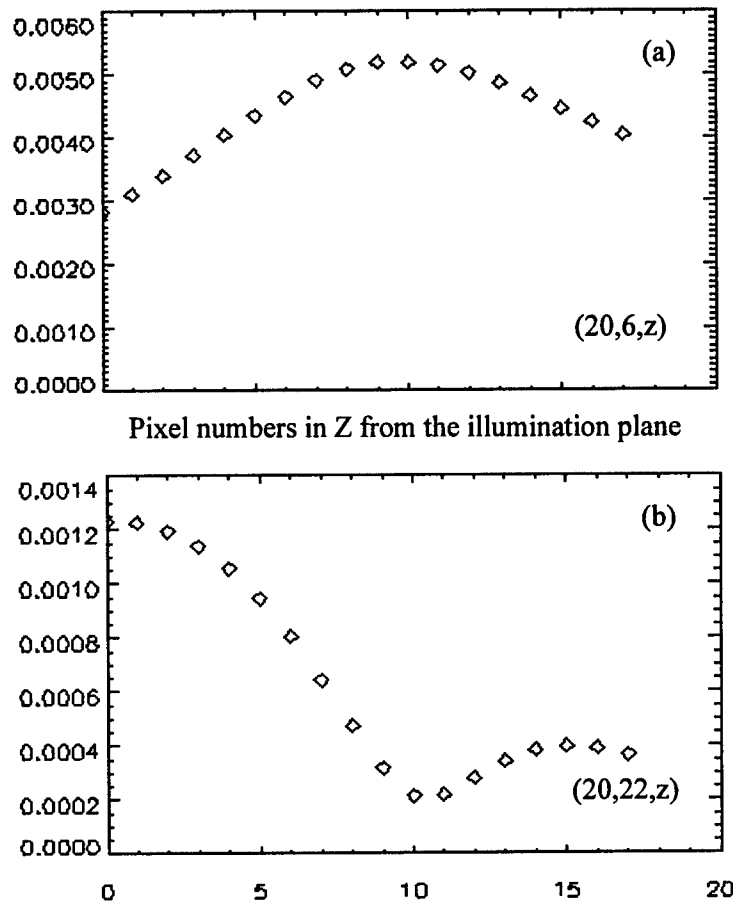


Figure 11. Z plots along the lines perpendicular to the detection plane and containing the black bead (a) and white bead (b). The pixel spacing in Z direction is 0.25 cm/pixel, corresponding to that shown in Figures 10(c) and 10(d). The filter used here was 2.

To confirm the object location in depth obtained from Figure 11, we plot two Z plots for the data with 16 x 16 array sizes, as shown in Figures 10(e) and 10(f), at the corresponding x and y coordinates of the two implanted objects. These two Z plots are given in Figures 12(a) and 12(b), respectively, for the black and white bead. In this case, the pixel spacing in Z is 0.5 cm/pixel. One can see that a peak and a dip appear at Z pixel number=4 behind the detection plane, i.e., $4 \times 0.5 \text{ cm} = 2 \text{ cm}$ behind the detection plane. Thus, the object location in Z direction given in figure 11 is in good agreement with that in figure 11.

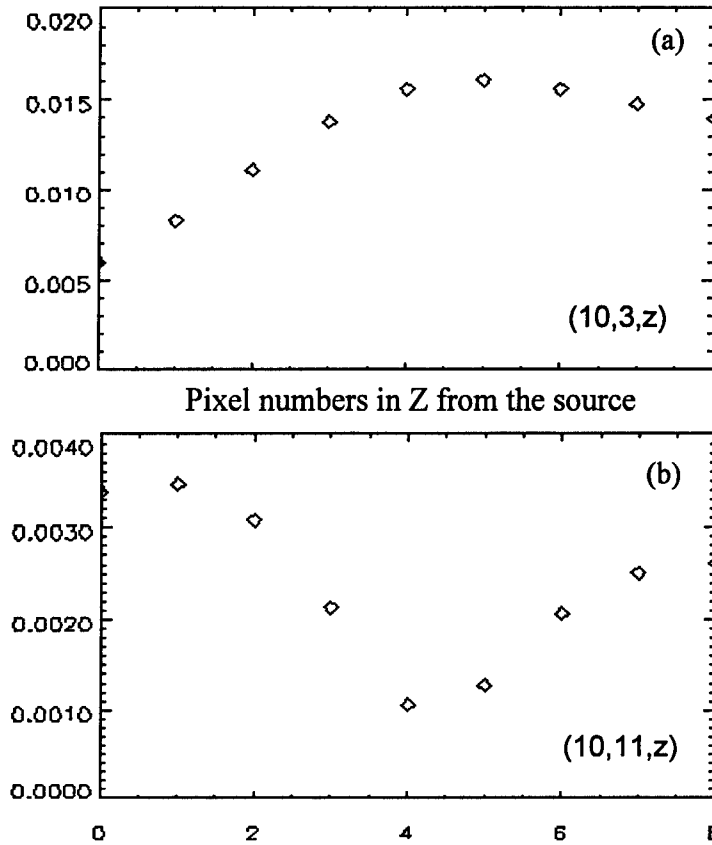


Figure 12. Z plots along the lines perpendicular to the detection plane and containing the black bead (a) and white bead (b). The pixel size in Z direction is 0.5 cm/pixel, corresponding to that shown in Figures 9(e) and 9(f). The filter used here was 2.

7. Data Analysis

Parameters used to scan and reconstruct images are as follows: 1) scanning dimensions/areas, such as 8x8 or 4x4 cm², 2) pixel spacing, such as 0.125, 0.25, or 0.5 cm/pixel, 3) modulation frequencies at 20, 40, or 100 MHz, and 4) filters numbers or cutoff spatial frequencies. What is the relationship among those parameters, and how do they affect the resolution of the reconstructed images? The answer to these questions may be helpful to optimize the scanning and reconstruction parameters for good-quality reconstructed images with best efficiency.

When a FFT of a 2-dimensional spatial domain function is performed, the number of pixels (or elements) in the spatial frequency domain is equal to the same number of pixels in the spatial domain. The 2-d spatial frequencies, f_x and f_y , will be multiple increments of Δf_x and Δf_y . Thus, we can rewrite the transfer function (TF) as follows:

$$H_z(f_x, f_y) = \frac{A_z(f_x, f_y)}{A_0(f_x, f_y)} = \exp \left\{ i \Delta z \sqrt{\left(\frac{-v\mu_a + i\omega}{D} \right)^2 - (2\pi f_x)^2 - (2\pi f_y)^2} \right\}$$

$$= \exp \left\{ -\Delta z \sqrt{\left(\frac{\nu\mu_a}{D}\right) + (2\pi n\Delta f_x)^2 + (2\pi n\Delta f_y)^2} - i\frac{\omega}{D} \right\}, \quad (7)$$

where Δz is the distance from the detection plane to the reconstructed plane behind the detection plane, and Δz is a positive number. n is an integer, ranging from 0 to $n_{\max}$, where $n_{\max}$ is equal to one half of the number of pixels in a row in the spatial domain. In the reconstruction algorithm, n corresponds to the filter number for the cutoff spatial frequency. It is known that the increments in the 2-d spatial frequency domain, Δf_x and Δf_y , are related to the dimension of the spatial domain as:

$$\Delta f_x = \frac{1}{L_x} = \frac{1}{\text{dimension}(x)}, \Delta f_y = \frac{1}{L_y} = \frac{1}{\text{dimension}(y)}. \quad (8)$$

Therefore, the TF becomes

$$H_z(f_x, f_y) = \exp \{ -\Delta z B [\cos(\frac{\theta}{2}) + i \sin(\frac{\theta}{2})] \} = e^{-\Delta z B \cos(\frac{\theta}{2})} e^{i \Delta z B \sin(\frac{\theta}{2})}, \quad (9)$$

where B is a real number and

$$B = \left\{ [3\mu_a\mu_s' + (\frac{2\pi n}{L_x})^2 + (\frac{2\pi n}{L_y})^2]^2 + (\frac{\omega}{D})^2 \right\}^{\frac{1}{4}}, \quad (10)$$

$$\theta = \tan^{-1} \left[\frac{\omega/D}{3\mu_a\mu_s' + (\frac{2\pi n}{L_x})^2 + (\frac{2\pi n}{L_y})^2} \right]. \quad (11)$$

Eqs. (9) to (11) show quantitative dependence of the TF on the spatial dimension, L_x and L_y , modulation frequency, ω , and cutoff pixels, n , for spatial frequencies. It is seen from eq. (9) that the TF, H_z , has an decaying amplitude of $e^{-\Delta z B \cos(\frac{\theta}{2})}$ with a phasor of $e^{i \Delta z B \sin(\frac{\theta}{2})}$. For a given phantom or tissue with fixed values of absorption and reduced scattering coefficients, μ_a and μ_s' having typical values of 0.1 and 10 cm^{-1} , B and θ , and thus H_z , are functions of ω and n , if the scan dimensions, L_x and L_y , are selected. Figure 13 demonstrates that $\theta/2$ is relatively small with a range of modulation frequencies less than 100 MHz. This means that the transfer function is approximately a pure attenuation function with the phasor equal to 1 when the modulation frequencies are within 100 MHz. Using the same parameters as those for Figure 13, we plot the amplitude of $e^{-\Delta z B \cos(\frac{\theta}{2})}$ with $\theta=0$, as shown in Figure 14. It is seen that at lower modulation frequencies, such as 20 MHz, the amplitude of the transfer function decays much faster than that at higher modulation frequencies, such as 1 GHz. This plot also shows that to obtain the same

magnitude of the normalized amplitude of TF, the number of pixels for Δf would be much smaller at 20 MHz than that at 1GHz. This makes the cutoff filter in the spatial frequency domain to be more limited at lower modulation frequencies than at higher modulation frequencies, i.e., only to be 1 or 2 pixels at 20 MHz but to be up to 5 more pixels at 1GHz. In addition, based on eq. (10), the cutoff filters can be equivalently improved by increasing the scanning dimensions, L_x and L_y .

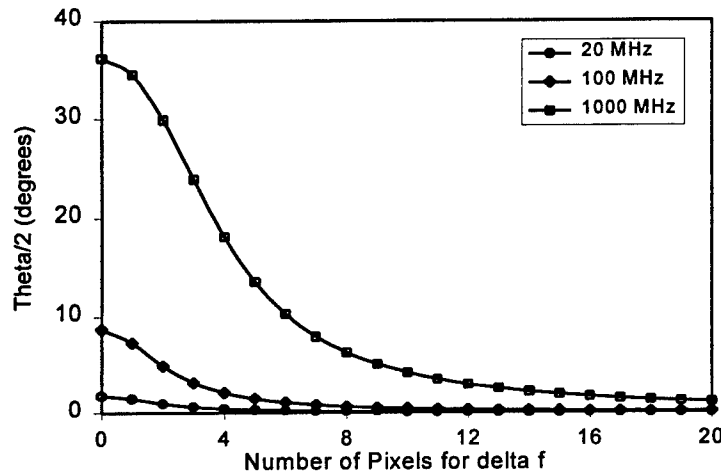


Figure 13. Dependence of $\theta/2$ in degrees on the number of pixels for Δf in the spatial frequency domain. The parameters used for this plot are: $\mu_a=0.1 \text{ cm}^{-1}$, $\mu_s'=10 \text{ cm}^{-1}$, $L_x=L_y=8 \text{ cm}$.

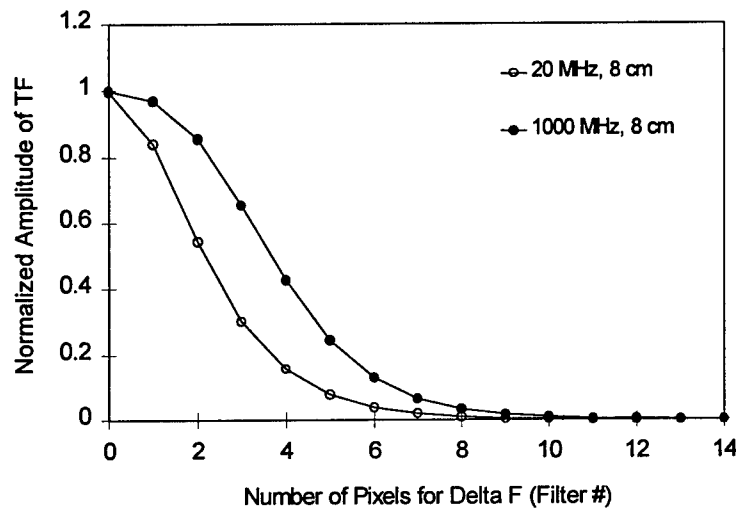


Figure 14. Dependence of normalized amplitude of transfer function on the number of pixels for Δf in the spatial frequency domain. The parameters used for this plot are: $\mu_a=0.1 \text{ cm}^{-1}$, $\mu_s'=10 \text{ cm}^{-1}$, $L_x=L_y=8 \text{ cm}$.

Thus, to improve the resolution of the reconstructed images, we can use higher modulation frequencies or increase the sizes (dimensions) of the detection plane. Decrease in pixel spacing would not be a crucial factor at this point, as long as the sampling frequency (number of pixels per unit length) is larger than the spatial frequency of the imbedded object, i.e., $1/(2 \times \text{diameter of the object})$, according to Nyquist's law.

To confirm the findings mentioned above, we have also conducted theoretical simulations using similar parameters. The results are very consistent with the experimental data.

In the case of biomedical imaging, let us discuss two extreme cases as follows: the first case is where the medium is highly absorbing in comparison to the modulation frequency. For example, at 20 MHz, we have μ_a (typically $0.1 \sim 0.3 \text{ cm}^{-1}$) $\gg \omega/v$ ($=0.006 \text{ cm}^{-1}$ for 20 MHz signal), and the TF becomes a purely exponential decay function. We can approximately use the following relationship to optimize the parameters for the cutoff filter and the scanning dimension, n , L_x and L_y .

$$(2\pi n/L_x)^2 + (2\pi n/L_y)^2 = 3\mu_a \mu_s' \sim 3 \text{ cm}^{-2}, \text{ or } n/L_x \sim 0.2 \text{ cm}^{-1}.$$

In the second case, where the modulation frequency is comparable to or higher than the absorption factor of the medium, we will have μ_a (typically $0.04 \sim 0.1 \text{ cm}^{-1}$) $\ll \omega/v$ ($=0.3 \text{ cm}^{-1}$ for 1 GHz signal). The TF has an exponential decay term and phase shift. In this case, we will use the following relationship to estimate the parameters for the cutoff filter and the scanning dimension, n , L_x and L_y .

$$(2\pi n \Delta f_x)^2 + (2\pi n \Delta f_y)^2 = 3\mu_s' \omega \sim 3\pi \text{ cm}^{-2}, \text{ or } n \Delta f_x \sim 0.6 \text{ cm}^{-1}.$$

8. Conclusions and Future Work

What we have learned from this summer research project is as follows:

- 1) the FFT reconstruction algorithm works in imaging and positioning inhomogeneities 3-dimensionally with good accuracy;
- 2) background information is necessary and important for accurate reconstruction;
- 3) At present stage, for high sensitivity, the imbedded object should be off-centered with respect to the light source and the detection plane. Further studies are underway to improve the reconstruction algorithm in order to eliminate this requirement.
- 4) large scanning dimension/area is more important than small scanning pixel size to achieve better resolution;
- 5) high modulation frequencies ($>500 \text{ MHz}$) may be helpful to achieve better resolution; low modulation frequencies can provide adequate information for tumor screening;
- 6) this summer project has demonstrated that the experimental set-up works and the results generated from it are in good agreement with the theoretical simulations. Inhomogeneities can be detected and located in three-dimensions. However, we are not able to characterize the optical properties of the inhomogeneities at this point. This remains for further study.

Our future work includes:

- 1) make phantoms containing objects which are small, purely absorptive, perturbations from the background medium;
- 2) take measurements on the new phantoms;

- 3) complete a diffraction tomographic model for quantitative reconstruction of optical property imaging;
- 4) apply the diffraction tomographic model to the data taken from the new phantoms for μ_a , μ_s image reconstruction;
- 5) extend the modulation frequencies to KHz range, and explore the possibility at using CCD cameras for image capture;
- 6) extend the system to clinical tests.

9. Acknowledgement

The authors are grateful to Dr. Chuck Matson for providing adequate laboratory equipment and the bulk of the reconstruction programs, as well as for his theoretical insight and useful discussions. In addition, the authors also wish to thank Grant Denton and Scott Newey for all of their helpful technical support. Without all of them, this research project would not have been possible.

10. References

- ¹ M. A. O'Leary, D. A. Boas, B. Chance, and A. G. Yodh, "Simultaneous Scattering and Absorption Images of Heterogeneous Media Using Diffusive Waves within the Rytov Approximation," SPIE, **2389**, p.328-339 (1995).
Brain W. Pogue, Michael S. Patterson, and Tom J. Farrell, "Forward and Inverse Calculations for 3-D Frequency-Domain Diffuse Optical Tomography," SPIE, **2389**, p.328-339 (1995).
- ² C. L. Matson, N. Clark, L. McMackin, and J. S. Fender, "Three-dimensional tumor localization in thick tissue with the use of diffuse photon-density waves," Appl. Opt. **36**, 214-220 (1997).
- ³ Joseph W. Goodman, Introduction to Fourier Optics, (San Francisco, McGraw-Hill Book Company, 1968).
- ⁴ M. S. Patterson, B. Chance, and B. C. Wilson, "Time resolved reflectance and transmittance for the non-invasive measurement of tissue optical properties," Appl. Opt. **28**, 2331-2336 (1989).
- ⁵ M. Firbank, M. Hiraoka, and D. T. Delpy, "Development of a stable and reproducible tissue equivalent phantom for use in infrared spectroscopy and imaging," SPIE, **1888**, 264-274 (1993).
- ⁶ J. B. Fishkin and E. Gratton, "Propagation of photon-density waves in strongly scattering media containing an absorbing semi-infinite plane bounded by a straight edge," J. Opt. Soc. Am. A **10**, 127-140 (1993).
- ⁷ See the PMI (Photon Migration Imaging) Code Home Page: <http://www.osa.org/homes/BIOOPTIC/Resource/softwar.htm>

**THERMOELECTRIC ENERGY CONVERSION WITH SOLID
ELECTROLYTES**

M.A.K. Lodhi
Professor
Department of Physics Engineering

Texas Tech University
Lubbock, Texas 79409

Final Report for:
Summer Research Program
Phillips Laboratory

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

And

Phillips Laboratory

July 1997

THERMOELECTRIC ENERGY CONVERSION
WITH SOLID ELECTROLYTES

M.A.K. Lodhi

Professor

Department of Physics and Engineering Physics
Texas Tech University

Abstract

Alkali-Metal Thermal-to-Electrical Converter (AMTEC) is a high temperature regenerative concentration cell which converts thermal energy directly into electrical energy. The efficient operation of AMTEC cell involves several challenging heat and mass transfer which have been investigated. The AMTEC cell consists of several sodium vapor tubes. The pressure drop is calculated for each node in order to calculate the total pressure drop through the tube. The flow equations are derived in continuum, free molecular motion and transition regimes. As expected the molecular flow regime is more challenging. We studied this regime with Dusty Gas Model (DGM). This approach is expected to remove the discrepancy between the observed data and the results predicted by this model. Some alternate to sodium as working fluids have been surveyed. Potassium seems to a viable candidate.

THERMOELECTRIC ENERGY CONVERSION WITH SOLID ELECTROLYTES

M.A.K. Lodhi

1. Introduction

Direct conversion of heat energy to electrical energy—that is conversion without the use of moving mechanical parts—has both aesthetic and practical appeal. Several devices for thermoelectric direct conversion have been developed. The Seebeck effect, thermionic and magnetohydrodynamic generators are familiar examples. None of these direct converters has been widely adopted because of practical problems such as parasitic heat loss or lack of a critical material with the physical properties necessary for high efficiency and good durability.

Among the less well know direct thermoelectric converters are the thermally regenerative electrochemical systems (TRES). These devices are closed electrochemical systems which include electrochemical cells that produce electrical power. The reactants for these cells are regenerated within the device by thermal energy from a heat source which flows through the device to a heat sink. TRES have also been called electrochemical heat engines. By analogy with the well-known heat engines in which a working fluid is carried around a thermodynamic cycle. They are Carnot-limited in efficiency. Many types of TRES were investigated in the 1950s and 1960s in the search for improved methods of converting the thermal output of nuclear reactors to electrical power.

All of the early TRES were plagued by practical problems such as inefficient heat exchange, electrode polarization, slow chemical regeneration kinetics. Materials separation problems and corrosion. Power densities of these early systems were usually limited to a few tens of milliwatts per square centimeter of electrode area and thermoelectric efficiencies were below 5%. Some comprehensive review articles on TRES are now available in literature. The development of beta-alumina solid electrolytes has made possible a new TRES. The alkali metal thermoelectric converter which is capable of high power and high efficiency and avoids the practical difficulties of the earlier systems.

With the advent of Faraday's law of electromagnetic induction and the steam engine, dynamic conversion systems evolved rapidly and were perfected with engineering details mostly for terrestrial and aircraft devices. As a potential alternative to dynamic conversion systems, static thermal to electrical conversion systems have been investigated, particularly for space power systems. With the discovery of the Seebeck effect, the thermoelectromotive force, which occurs in materials under the effect of a temperature gradient,

has been the driving force for static thermoelectric generators. A large number of static devices have been evolved and studied. The goal for static conversion systems has been to identify static converters with efficiencies that are competitive with dynamic systems. To this end currently, NASA Lewis Research Center (LeRC) has an interest in developing thermal energy storage (TES) for the solar dynamic ground test demonstration (SDGTD) program. Phillips Laboratory has an interest in developing an electrical power system for the payload in Low Earth Orbit (LEO) for a duration of 10 years. For conversion technology, Jet Propulsion Laboratory (JPL) is currently interested in alkali metal thermal to electrical conversion (AMTEC) with respect to the Fast Flyby Pluto Express Mission. In all these endeavors the effort is aimed at developing advanced power conversion and thermal management and control subsystems for space applications. As part of this effort, Alkali-Metal Thermal-to-Electric Converter (AMTEC) device is being developed and tested at USAF Phillips Lab with the goal of providing feedback or improved efficiency, durability, and long life time. To reach this goal, thorough and accurate mathematical modeling and analyses as well as experimental testing of AMTEC device is required. Our efforts therefore, on theoretical side have been to develop an accurate mathematical model of AMTEC at Phillips Lab this summer.

2. Background of AMTEC

AMTEC is a relatively new type device, based on the principle of sodium concentration cell, conceived in late sixties [1,2]. AMTEC devices are a leading power conversion candidate for a large number of applications, particularly low mass, low cost, long life, and high-efficiency power systems for satellites and exploratory space probes. AMTEC systems are very attractive for these applications, because they are static conversion devices which can provide efficiencies close to the theoretical Carnot efficiency. Efficiencies as high as 19% have been measured in laboratory experiments. Optimized AMTEC devices of a mature design can provide theoretical conversion efficiencies between 20 and 40% when operating at a moderate hot-side temperature between 900 K and 1300 K and a cold-side temperature between 400 K - 800 K. Other advantages of AMTEC devices include low maintenance, high durability, high specific power, modular construction, and the ability to operate in conjunction with a high-temperature combustion, a nuclear or a solar heat source. Therefore, AMTEC devices are attractive not only for space, but also for many terrestrial and aerospace applications.

2.1 Operation Principles of AMTEC Cell

To understand the working of AMTEC, its operating cycle is illustrated in Fig 1 [3] A closed vessel is divided into a high temperature (T_2) and high pressure (P_2) region in contact with a heat source and a low temperature (T_1), low pressure (P_1) region in contact with a heat sink. AMTEC is a high-temperature regenerative concentration cell for elemental sodium, which directly converts heat to electrical energy.

The critical material in the operation of the AMTEC is the beta"- alumina solid electrolyte (BASE), a sodium-ion conductor whose ionic conductivity is much larger than its electronic conductivity. In an AMTEC cell, the BASE separates the hot (high-pressure) region filled with a small quantity of liquid sodium (typically <10%) in contact with the heat source, from the cold (low-pressure) region occupied with sodium vapor. A porous electrode covers the low-pressure (outer) side of the BASE. Electrical leads in contact with the porous electrode and the high-temperature liquid sodium exit through the wall of the device and are connected to an external load. The pressure differential across the BASE forces ionization of sodium atoms on the hot-side. The ions diffuse through the BASE toward the low-pressure side in response to the pressure differential gradient of free Gibbs energy while the electrons circulate through the external load, producing electrical work [4]. Electrons and sodium ions recombine at the interface between the BASE and the porous electrode. The resulting sodium atoms absorb their heat of vaporization, move through the electrode and the vapor space, then release their heat of vaporization at the low-temperature condenser surface. A wick structure or an electromagnetic pump return the liquid condensate to the high temperature side of the BASE to complete its circulation. More details on AMTEC are in scarce literature [1-6]. AMTEC possesses conversions efficiencies much higher than other direct thermoelectric devices. It was reported at 19% in 1981 [7], with a power density of 1 W/cm² [6]. Its hot region temperature is in the range of 900-1300 K [8]. This makes it compatible with the general purpose heat source (GPHS) [9] or the projected SP-100 nuclear space power reactor [10]. AMTEC is limited to 1350K because of sodium interactions with the BASE. The lower temperature region is limited to 500 K by the need to maintain the sodium in the liquid state. As the condenser temperature increased above 700 K the efficiency decreases [11]. Besides, AMTEC has high efficiency near Carnot [12], low maintenance and high durability, modular construction and the ability to use high temperature combustion, nuclear, or solar heat source [13].

3. Heat and Mass Flow

The operation of an AMTEC cell involves several challenging heat and mass transfer processes requiring thorough investigations, namely: (a) the diffusion of sodium ions through the BASE tube; (b) the evaporation of sodium at the BASE-porous electrode interface (c) the sodium vapor flow through the porous electrode and the vapor space in the free-molecular regime; and (d) the recirculation of liquid sodium through the capillary tube. Any of these processes can limit the operation of the AMTEC cell. Because sodium is frozen at room temperature, the startup of the AMTEC cell from a frozen state and the effect of the initial distribution of the sodium working fluid in the cell, is of particular concern.

The performance and the conversion efficiency of an AMTEC cell depend on the internal and external parasitic heat and electrical losses. The internal heat losses are made up of 1) conduction through the output current leads and 2) supporting structure of the BASE tube, and of internal radiation heat losses from the 3) porous electrode surface to the condenser surface (through the vapor space). The latter

depends on the a) geometry, b) axial temperature distribution in the condenser shell and the c) emissivity of both surfaces. 4) Of particular importance is the fraction of condenser surface covered with the sodium condensate. If the porous electrode and condenser surfaces exchange radiative energy at rates comparable to those of ideal bodies, the maximum efficiency of the cell would be less than 10%. Fortunately, the reflectivity of the sodium liquid film is greater than 98% in the infrared. Thus, if the surface of the film is smooth, the radiative losses could be as low as 2% of black body emission, even if the porous electrode is an ideal emitter. To take advantage of the high reflectivity of the sodium condensate, the shape of the condenser surface must be designed to reflect most of the radiation emanating from the porous electrode back to it.

The external heat losses are made up of conduction through the 5) support structure of the condenser (or the outer shell of the cell), and 6) radiation from the cell outer shell to the shield surrounding the cell. 7) Heat leakage through the radiation shield may also be important. (In the case of a system comprised of several AMTEC cells, radiation exchange between the shells of the different cell must be considered.) Due to the high operating temperature of the AMTEC radiation is a major heat transfer mechanism. 8) A major challenge involved in calculating radiation heat transfer is determining the **geometric view factor**. 9) Except for the sodium vapor there is no convection heat loss. The sodium vapor heat transfer through convection only affects the internal degree of freedom of AMTEC.

4. Multi-tube Model Code

Two different types of cells are being considered for testing, single-cell units and multi-cell units for the Pluto Express Mission. For the purpose of this research period we were primarily concerned with the multi-tube unit.

We formulated a mathematical model of AMTEC which takes into consideration all those aspects, discussed in section 3, thus leading to a prototype designing, fabricating, testing and further improving of the cell. The three flow regimes, just mentioned above, are arbitrarily determined by Knudsen number, Kn , as follows [14]:

(a)	Continuum (viscous)	$Kn \leq 0.01$
(b)	Slip (transition)	$0.01 < Kn < 1$
(c)	Free Molecular (molecular)	$Kn \geq 1.$

This necessarily requires to divide the AMTEC cell, under investigation into appropriate parts as accurately as possible where those regimes dominate and derive equations necessary to calculate the vapor flow pressure drop including those regimes.

This model has four options for the effect of the molecular collisions with the wall particularly in the rarefied flow regime. In order to incorporate the effect of molecule- wall collisions more realistically, the so called Dusty Gas Model (DGM) is used as the fourth approach [15-18]. The DGM treats this effect as a diffusion problem involving the original gas molecules, in motion, and the giant wall molecules at rest and uniformly distributed. The problem is thus formally treated as the one component flow of a multicomponent gas mixture.

A computer code using TK solver has been developed for heat and mass flow which calculates the internal electrical and thermal losses, and predicts the multi-tube AMTEC performance. The pressure drop is calculated using DGM. Chevron type of shielding is an option available. The vapor is assumed to vary linearly. An input and output set of parameters on the variable sheets are given in the Appendix.

5. Alternate Working Fluid

To look for viable alternate to sodium as working fluid we have collected all those elements which have negative electrode potential with reference to hydrogen electrode. They are given in the accompanying table. Though lithium has the highest negative potential of -3.05 V and is the lightest metal, it has high melting and boiling points.

As a working fluid, Potassium (K) is a viable alternative to sodium. K ion conducting B" alumina (K-BASE) has been made[19] though its conductivity is not known at AMTEC operating temperatures. As with optimized electrodes, analyzing the module performance with K working fluid (a K-module) requires several assumptions. First, the R_{BASE} of K-BASE probably differs from Na-BASE, and K-BASE is probably not as good of a conductor. The K-BASE conductivity is assumed one half the Na-BASE conductivity. Similarly the exchange current, J_0 is probably smaller if; use a value that is one half the J_0 for Na. Finally, the calculations must be corrected for such fluid specific values as the vapor pressure, the molecular weight, the heat capacity, and the latent heat of vaporization.

A K-module is expected to outperform a similar Na-module at lower hot zone temperatures because K has a higher vapor pressure (P_K than Na (P_{Na})). However, at higher temperatures, the change in P_{Na} with temperature is greater than the change in P_K , and the Na-module performance overtakes the K-module performance. For the tube bundle operating between 1000 K and 500 K, the change from Na to K working fluid increases the peak power from 200 to 250 W/l improves the efficiency from 29 to 34%. For hot zone temperatures greater than about 1100 K, the Na-module outperforms the K-module. For the flat plate device, the K-module performance is greater than the Na-module even at 1200 K because of the thinner ionic conductor. The conversion efficiency of the flat plate K-module is 4 to 9% better than the Na-

module over the temperature range with a efficiency of 35% at 1100 K. The peak power approximately doubles in the low temperature range (800-1000 K), and at 1100 K the K-module still has a 27% greater power than the Na-module.

6. Conclusion

A multi-tube AMTEC system has been designed and a code in TK solver has been prepared with DGM. The input-output parameters are given on the variable sheets, ~~in the Appendix~~. An alternate working fluid to sodium has been investigated. It seems that developing a K-electrolyte and using potassium as the working fluid could improve both efficiency and power density for AMTEC.

References

1. R.C. Dahlbert et.al, "Review of Thermionic Technology 1983 to 1992," "A Critical Rev. of Space Nucl Power and Propulsion 1984 - 1992", Ed. M. S. El-Genk (AIP Press, NY 1994) pp 121-161
2. J.T. Kummer and Weber, U .S. Patent 3,458,356 (1968), Assigned to Ford Motor Co
3. G. A. Johnson (1994) "Study of the Rarefied Sodium Vapor Flow in the Pluto Fast Flyby AMTEC Cell," 11th Symposium on Space Nuclear Power and Propulsion , held in Albuquerque, NM, January 1994, AIP Conference Proceeding N°. 301, M. S. El-Genk and M. D. Hoover, Eds., 2: 581-585
4. T. Cole, "Thermoelectric Energy Conversion with Solid Electrolyte", Science, 221,915 (1983)
5. N. Weber "A Thermoelectric Device Based on Beta-Alumina Solid Electrolyte", Energy Conversion 14, 1 (1974)
6. T.K. Hunt, N. Weber, and T. Cole, "Research on the Sodium Heat Engine", Proc 13th Intersoc Energy Conversion Engineering Conf., SAE, Warrandale, PA (1978) p201
7. T.K. Hunt, N. Weber, and T. Cole, "High Efficiency Thermoelectric Conversion with Beta" - Alumina Solid Electrolyte: The Sodium Heat Engine", Fast Ionic Transpost in Solids, ed J.B. Bates, G.C. Ferrington (N-Holland Co., 1981), P263
8. C.P. Bankston, T. Cole, S. K. Khanna and A. P. Thakoor, "Alkali Metal Thermodynamic Conversion (AMTEC) for Space Nuclear Power Systems", Space Nucl Power Sys. Ed. M. S. El-Genk and M. D. Hoover, (Orbit Co., Malabar, Fl, 1985) pp 398-402

9. A. Shock, "Design Evolution and Verification of the General Purpose Heat Source," Proc 15th Intersocieties Energy Conversion Engineering Conf Vol. 2 (AIAA, NY 1980) p 1032
10. D. Birden and F.A. Angela, "Space Reactor-Past, Present and Future", Proc 18th Intersocieties Energy Conversion Engineering Conf, Vol. (AIChE, N. Y., 1983) p 61
11. R. Ewell and J. Mondt "Static Conversion Systems", Space Nuclear Power Sys., Ed. M.S. El-Genk and M. D. Hoover (Orbit Book Co. Malabar, Fl., 1985) pp 385-391
12. C. B. Vinning, R.M. Williams, M. L. Underwood, M. A. Ryan, and J.W. Suiter, "Reversible Thermodynamic Cycle for AMTEC Power Conversion", Proc 27th Intersocieties Energy Conversion Engineering Conf., Society of Automobile Engineers, 1992, pp 3: 123-128
13. M. L. Underwood, et al, "Recent Advances in AMTEC Recalculating Test Cell Performance, "Conf 930103 (AIP, 1993) pp. 885-890
14. S. Dushman and J. M. LAFFERTY, *Scientific Foundations of Vacuum Technique* (John Wiley & Sons, Inc., New York, 2nd Edition) 1962, p
15. R. B. Evans III, G. M. Watson and. A. Mason, " Gaseous Diffusion in Porous Media at Uniform Pressure," J. Chem. Phys., 35, 2076-2083, (1962)
16. R. B. Evans III, G. M. Watson and E. A. Mason, " Gaseous Diffusion in Porous Media - II. Effect of Pressure Gradients, " J. Chem. Phys., 36,1894-1902 (1962).
17. E. A. Mason, A. P. Malinauskas and R. B. Evans, J. Chem. Phys., 46, 3199-3216 (1967).
18. R. E. Cunningham and R. J. J. Williams, "Diffusion in Gases and Porous Media, " (Plenum Press, New York, 1980) p 131
19. G. M. Crosbie and G. T. Tennonhouse, " Potassium Beta"-Alumina Membranes, " J. Amer. Ceramic Soc. 26, 187 (1982)

Anatomic Number	Element Name	Chemical Symbol	Electrode Potential (V)	Atomic Weight	Melting Point (°C)	Boiling Point (°C)	Specific Heat (J/gK)	Specific Gravity (g/cm ³)	Thermal Conductivity (W/cmK)	Ionization Potential (eV)
3	Lithium	Li	-3.05	6.939	180.5	1342	3.582	0.534	0.847	5.392
19	Potassium	K	-2.93	39.102	63.38	759	0.757	0.89	1.024	4.341
56	Barium	Ba	-2.90	137.34	727	1897	0.204	3.62	0.184	5.212
20	Calcium	Ca	-2.87	40.08	842	1484	0.647	1.54	2.00	6.113
11	Sodium	Na	-2.76	22.9898	97.72	883	1.228	0.97	1.41	5.139
12	Magnesium	Mg	-2.37	24.312	650	1090	1.023	1.74	1.56	7.646
13	Aluminum	Al	-1.66	26.9815	660.32	2519	0.897	2.70	2.37	5.986
30	Zinc	Zn	-0.76	65.37	419.53	907	0.388	7.14	1.16	9.394
26	Iron	Fe	-0.44	55.847	1538	2861	0.449	7.87	0.802	7.870
48	Cadmium	Cd	-0.40	112.41	321.07	76	0.232	8.69	0.968	8.993
28	Nickel	Ni	-0.25	58.7	1455	2913	0.444	8.90	0.907	7.635
50	Tin	Sn	-0.14	118.69	231.93	2602	0.228	7.28	0.666	7.344
82	Lead	Pb	-0.13	207.19	327.46	1749	0.129	11.3	0.353	7.416

Table. Properties of the elements which have negative potential

APPENDIX

These are the parameters that you would need to change if you were modeling a different working fluid in the AMPS tk model:

(1) On the input sheet:

- Qs - sensible heat of working fluid
- Ql - latent heat of working fluid
- rNa - electrical resistivity of Na
- kNav - thermal conductivity of Na vapor
- RhoNa - Sodium density
- eNa - emissivity of sodium
- M - sodium molecular weight

(2) On the rule sheet:

under "; Saturation Pressure" the correlation for the saturation pressure needs to be changed for the calculation of Pc, Pa, and Pb to the correct working fluid correlation

(3) On the function sheet:

the function "ut" needs to be changed to a viscosity correlation for the new working fluid (the current one is for sodium)

STUDY OF NONLINEAR DYNAMICS IN A DIODE PUMPED Nd:YAG LASER

Tim Newell
Professor
Department of Physics Engineering

University of New Mexico
1313 Goddard SE
Albuquerque, NM 87106

Final Report for:
Summer Research Program
Phillips Laboratory

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

And

Phillips Laboratory

August 1997

Study of Nonlinear Dynamics in a Diode Pumped Nd:YAG laser

Tim Newell
University of New Mexico

Abstract

Nonlinear dynamics are investigated in a specially designed and constructed pump modulated Nd:YAG laser. The laser operates in single longitudinal and transverse mode as well as in one polarization eigenmode. Investigations reveal that a chaotic state can be obtained. The route into chaos is not characterized by a sequence of period doubling cascades but rather as an apparent crisis. Coexisting attractors are also observed.

Study of Nonlinear Dynamics in a Diode Pumped Nd:YAG laser

Tim Newell

Introduction

Traditional studies of chaotic dynamics in experimental systems have usually focused on electronic circuits {1...4} or mechanical nonlinear oscillators {5}. The advantage of such systems is typically a low dimensional phase space orbit (the attractor), the minimal amount of system noise, and the ready access to the state variables. In contrast, fundamental investigations of nonlinear dynamics in optical systems have not been pursued intensely due to their usually high dimensional attractors as well as relatively large amounts of noise {6}. However, optical systems have certain advantages. Semiconductor lasers, in particular, intrinsically possess very large bandwidths that is particularly appealing in regards to communication purposes. Unfortunately, their high oscillation frequencies means that real time acquisition of intensity time series can only be observed for very short intervals. Hence detailed analysis of the signal cannot be made using time domain measurements.

The purpose of this investigation is study nonlinear dynamics in an optical system as a prelude to active control of chaotic behavior. In this project, nonlinear dynamics is studied in a neodymium doped yttrium-aluminum-garnet laser (Nd:YAG). Due principally to the long fluorescence lifetime of the excited atomic levels ($\tau_f = 240\mu s$), the fundamental relaxation oscillation frequency of this device is typically less than 100 kHz. Consequently, long intensity time traces of this laser may be captured then subsequently examined using techniques in chaotic data characterization {7}.

Furthermore, this laser when operating in a single longitudinal and single transverse mode as well as in a single polarization eigenstate can be modeled using two ordinary differential equations [8]. This allows for concurrent theoretical and experimental studies and quantitative comparison between the two.

The Nd:YAG laser

Historically Nd:YAG lasers have been pumped using rare earth containing arc lamps or tungsten-filament lamps [9]. However, due to the large thermal load and small efficiency level, these sources place great demands on the engineering of the laser design. In contrast, laser diode pumping of the cavity is preferable due to excellent spectral matching with the neodymium absorption bands, good mode matching with the Nd:YAG laser field, and reduced thermal load on the crystal. Additionally, modulation of the laser diode's intensity is a simple matter of modulating its driving current.

The experimental arrangement is shown in Fig. 1. The laser diode pump (SDL-5410-G), LD, produces a single longitudinal and single transverse mode beam at $\lambda=810\text{nm}$. In spite of the highly elliptical beam, good mode matching between the pump and YAG cavity can be achieved if the arrangement is carefully aligned. Additionally, an anamorphic prism may be incorporated into the design increasing the mode matching. Unfortunately, even minute reflections from different optical components were sufficient to stimulate nonlinear behavior in the diode lasers output. The consequences of such feedback is mode hopping and high frequency oscillations. A discussion concerning the suppression of this effect is below. A Faraday isolator (Optics for Research NIR-5), FI, measured at 39.6 dB isolation follows the collimating lens (Newport FL-10B), CL. After the isolator, an acousto-optic modulator, AOM, was sometimes used to modulate the pump by

deflecting the light sinusoidally. In the course of the experiments, it was found that the current modulated diode pump was more noisy than the current stabilized laser diode. Also, the acousto-optic modulator was not limited to a useful frequency range of 50 kHz as was the laser diode's current source. An iris, I, and a $f=75$ cm focusing lens, FL, was placed between the AOM and the YAG crystal, Y. The crystal measures 5mm x 2mm. The absorption efficiency at 810nm is over 80% [10]. The laser cavity is in a hemispherical configuration. The input side is formed by the YAG crystal. It is HR at the Nd:YAG lasing wavelength of 1064 nm and HT at the pump wavelength. A Brewster window, B, was sometimes placed inside the cavity. Its purpose is to insure oscillation of only polarization mode. However, its usage required a larger cavity size meaning that more than one longitudinal mode could lase for large pump intensities. However, without this window, the cavity length could be reduced to less than 7mm (hence the FSR of the cavity is almost 12 GHz). In this case, it was found that with proper mirror tilting, not only would the laser be single longitudinal mode, but it was also single polarization as well for low pumping powers. A curved output coupling mirror, OC, with $R=98\%$ completed the lasing cavity. The YAG light was collimated and stray pump light was blocked using a high bandpass filter. Another Faraday isolator, FI, was employed to prevent reflections back into the cavity from the Fabry-Perot interferometer, FP. A beamsplitter directed a fraction of the YAG light into a New Focus 1811 125 MHz photodetector. The electrical signal was then conditioned using a Stanford Research Systems SR-560 low noise preamplifier then input into a Bittware BAT-301 digitizer (see the Appendix) and a Tektronix DSA-601 oscilloscope.

The laser is operated close to threshold in order to prevent multiple longitudinal modes from lasing as well as two polarizations. Depending on the alignment and mode matching of the laser diode pump, high order transverse modes can oscillate as well as other curious shaped far field

intensities. However, under optimal alignment, the first Gaussian mode dominates. The relationship between the pump and relaxation oscillations follows the expected curve of a 4-level laser [11], namely

$$\omega_r^2 = \frac{1}{\tau \tau_f} \left(\frac{P}{P_{th}} - 1 \right) \quad (1)$$

where ω_r is the relaxation oscillation, P is the pump, P_{th} is the pump at threshold, τ_f is the fluorescence lifetime, 240 μ s, and τ is the photon lifetime. Figure 2 shows this relation between pumping and the square of the relaxation oscillations. From this figure, τ is found to be 3.28 ns.

Stabilization of the laser diode. At pump currents up to 2.5 times the threshold ($I_{th}=30$ mA), this particular laser is prone to mode hopping. At larger currents, mode hopping still occurs but the laser operates in a relatively stable manner. The linewidth of the laser diode can be reduced and mode hopping prevented by effective use of feedback into the semiconductor laser cavity. This is accomplished in the following manner, see Fig. 1. Immediately after the collimating lens, a diffraction grating, DG, is placed so that the pump beam intersects the surface at an angle of approximately 45°. The first order diffracted beam is then directed towards the Nd:YAG crystal. The second order diffracted beam is reflected back into the laser cavity by a nearby mirror, M. This alignment creates a compound cavity in which the external cavity modes couple with the internal cavity modes producing a more stable laser diode.

Heat dissipation and thermal lensing. The crystal rests lengthwise in a small groove drilled in the aluminum mount and is covered with a grooved cap. It is clamped into place from the top using a small plastic tipped set screw. Thermal contact is enhanced around the entire circumference using thermally conductive jelly. Though the temperature is controlled from below using a thermo-electric (TE) cooler, the high thermal conductivity of the aluminum surrounding the crystal insures that for temperatures close to

room temperature, there will be a large degree of symmetry in the thermal cooling around its outer periphery. This produces a radial symmetric index of refraction profile which is advantageous for compensation of thermal lensing. At cooler temperatures deviations from the radial symmetry is expected.

Mechanical stability. The Nd:YAG crystal is clamped into an aluminum block that is cemented to a thermo-electric cooler which is cemented to a large aluminum block. This piece is bolted onto a rail. The output coupling mirror is mounted to a New Focus single mode fiber coupler. This coupler is designed for the high degree of stability necessary to launch light into a single mode fiber, and should not be prone to mechanical vibrations or drift. The fiber coupler is bolted to a large aluminum block that is bolted to the rail.

For completeness, Table 1 summarizes properties of the Nd:YAG crystal.

Table 1
Physical Properties of Nd:YAG laser crystal

Chemical Formula	Nd:Y ₃ Al ₅ O ₁₂	
Density of Nd ions	1.38 x 10 ²⁰	cm ⁻³
Emission cross section	4.1 x 10 ⁻¹⁹	cm ²
Fluorescence lifetime	240 x 10 ⁻⁶	s
Saturation intensity	2.0	kW cm ⁻²
Index of refraction, n	1.82	
Linear thermal dispersion	9.86 x 10 ⁻⁶	
Heat conductivity	13	Wm ⁻¹ K ⁻¹
Density	4.55	g cm ⁻³
Specific heat, c _p	607	J kg ⁻¹ K ⁻¹
Linear Heat transfer	27.5	Wm ⁻² K ⁻¹

Dynamical equations of motion

Two nonlinearly coupled second order differential equations have been used to successfully describe the laser {8}. These are

$$\dot{E}(t) = \frac{1}{\tau} [G(t) - \alpha(t)]E(t) + i\omega E(t) + \sqrt{\epsilon}\eta \quad (3)$$

and

$$\dot{G}(t) = \frac{1}{\tau_f} \left[P - G(t)(1 + |E(t)|^2) \right]. \quad (4)$$

In Eq. 3, E represents the electric field, τ is the photon lifetime, α is the internal losses, ω is the frequency detuning from the nearest cavity mode, ϵ is the spontaneous emission noise strength, and η is a Gaussian random number. In Eq. 4, G is the gain, τ_f is the fluorescence lifetime, and P is the pumping. In the literature, lasers that can be described by this type of equations are designated "Class-B". Without the noise term, these equations will predict only the steady state operation the laser. Chaotic behavior does not occur in such systems due to the fact there is simply not enough room in two dimensional system for the exponential divergence of adjacent trajectories along with the folding of orbits on themselves. In order to enlarge the phase space, the pump parameter or the internal losses are be modulated and the pump term becomes:

$$P = P_o + P_i \sin(\omega_d t) \quad (5)$$

The result of Eq. 5 is to create a 2 1/2 dimensional system. In this driven system nonlinear dynamics can take place. In this investigation, we will pump with a frequency nearby the relaxation oscillations: $0.2\omega_r < \omega_d < 2\omega_r$.

Experimental Results

The laser is pumped very close to threshold in order to obtain the single mode behavior described above. The lasing intensity without intended pump modulation is characterized by very low amplitude relaxation

oscillations that are typically between 20 kHz and 100 kHz depending on the pump level. This undesirable trait can be traced to fluctuations in the laser diode pump beam. Complete stabilization of the intensity, unfortunately, was never obtained. Consequently, the results are tainted and the nonlinear dynamics are obscured by the omnipresent noise. When the laser pump is slightly modulated at frequencies in the vicinity of the relaxation oscillations, the CW output of the Nd:YAG is replaced by large amplitude spiky pulses. This is, in a sense, a form of Q-switching by gain pumping. Here the nonlinear interplay between the electric field and the carrier inversion level cause the amplitude of the pulses and the time spacing between them to be chaotic. If the modulation frequency becomes too great, at least 1.5 times the relaxation oscillations, then the Nd:YAG laser is unable to respond and this pulse train will damp out. Hence the focus is on driving the laser between 0.2 and twice the relaxation oscillation frequency. Time series of the intensity are shown in Fig 3. In these figures, the modulation frequency, ω_d , is kept constant just below ω_r , while the amplitude, P_1 , is different. In Fig. 3(a), the laser fluctuates chaotically. The amplitude of the pulses varies extensively and in addition, the intervals between pulses is irregular. Though the time series is aperiodic with no evident correlation between successive pulses, this is not the case. In Fig. 4, a first return map is constructed. It plots the $N+1st$ peak versus the Nth peak. There is an arch like structure to the plot indicative of a deterministic process. It is smeared due to the insidious presence of noise. (The smearing may also be due in part to the fact that the system may be high dimensional. In this case, the two dimensional plot will have compressed the points so that neighboring points may *be false nearest neighbors* [7]). In contrast to the time series shown in Fig. 3, for time series of a stochastic processes there would be no structure and the return map would appear as a blob of points. Figure 3(b) is measured with a higher level of pumping. We find an interesting phenomenon, coexisting attractors. Here

the laser is seen to be first lasing with large amplitude pulses, but then falls to a train of low amplitude pulses. In terms of the dynamical equations, such a scenario occurs when multiple solution sets exist for the given parameter space. The noiseless system follows one set or the other depending on the initial conditions. In the experimental arrangement, the inherent noise causes a switching from one state to the other. This will have deleterious consequences if the desired result is a strongly gain switched laser.

Conclusions

In this investigation, a specially constructed diode pumped Nd:YAG laser was constructed and pumped modulated in order to unleash its rich nonlinear dynamical behavior. High quality time series were acquired using a state of the art 3 MHz 16-bit vertical resolution digitizer and digital signal processor. The laser exhibits chaotic behavior for a wide range of parameters of the modulation depth and frequency of modulation. While a particular route into chaos was not verified, a period doubling bifurcation from P1 to P2 was recorded. The large amount of noise in the system may have prevented high bifurcations from being observed directly. In spite of this bifurcation, it appears that the laser enters a chaotic state via a crisis. An important result is the observation of coexisting attractors, i.e. two existing solutions of the dynamical equations exist. The system follows one solution or the other depending on the initial conditions. Noise destabilizes the stability of the solution and the phase space trajectory may visit either solution set. In this regime, the laser jumps between large and small amplitude spiky oscillations. The importance of the result is that very small stabilizing signals may be implemented in order to steer the trajectory from one attractor to the other. In doing so, a binary communication sequence can be created.

Future investigations would likely involve employing an Argon laser to pump the Nd:YAG crystal. This laser may be preferable to the diode laser

due to the low amount of noise, low responsivity to feedback, and circular fundamental Gaussian beam profile. The focus remains on mapping the route into chaos then to stabilize existing unstable periodic orbits in order to use as a communications channel.

In conclusion, though solid state lasers represent the oldest type of these devices, their nonlinear dynamics has yet to be fully researched and characterized. Thus, exciting new possibilities for their usage come into being. Chaotic systems are characterized by (a) a wide range of behavior and (b) an extreme dependence to initial conditions and to perturbations of a system parameter. Hence a judicious choice of a steering signal [12] could be implemented in order to choose what type of behavior the laser will follow. This new form of control can lead to many diverse applications such as secure communications.

Appendix Data Acquisition

At the time of this writing, digitizing oscilloscopes are inherently limited by either their sampling rate or the vertical resolution of the acquired signal. Generally, the faster the sampling rate, the lower the resolution. While the sampling rates may be as high as 2 gigasamples per second, the signal is vertical resolved in usually 2^8 (8-bit resolution) or 2^{10} (10-bit resolution) intervals. This experimental restraint often comes as a disconcerting surprise to theorists who commonly work with double wide machine precision floating values. To bridge this precision gap, we have employed the combination of a 3 MHz 16-bit resolution digitizing card (Bittware BAT-301) installed in a personal computer slot with a digital signal processor (Bittware Omega DSP) as a data receiver. The sampling rate of the digitizer is sufficient for successfully capturing time series from the Nd:YAG laser. FIFO memory in the digitizer allows for the high speed capture of 1024 data samples. The arrangement is shown in Fig. 5.

The digitizer DSP combination must be programmed so that data is properly transferred from one board to the other while the digitizer continuously operates. To achieve this, after a predefined number of samples are acquired, the digitizer issues a hardware interrupt. This interrupt, designated IRQ3 by the DSP, signals the DSP that it should read the samples into its data memory. Consequently the digitizer is programmed for the sampling rate and the number of samples to acquire before an interrupt signal is issued. In turn, the DSP must be programmed to process the incoming request, write the samples to memory, and keep a tally of the total number of samples. When the total number of samples are obtained, they are written to a file on the computer's hard disk.

As the DSP card is constructed to be Harvard architecture, the address location of the program code and the data variables must be specified in the physical memory banks. This is accomplished by use of an *architecture* file whose purpose is to map out the program memory and the data memory. If this architecture file does not allocate sufficient resources for either the program code or the data variables, then the C source code will not compile. The process of specifying the architecture file is not too difficult, but problems did arise (and were not successfully resolved) when data arrays larger than 64K words were desired.

A second issue concerns synchronization between the two boards. While the digitizer may go along acquiring data at a 3 MHz rate, the DSP must process this data on a 30 MHz clock cycle. This means that the DSP must acknowledge the interrupt, handle the interrupt overhead time, retrieve a data sample, store it in memory, then reset the interrupt before the subsequent interrupt is issued by the digitizer. If this process cannot be completed in time, the digitizers FIFOs wrap around and data is overwritten destroying the continuity of the time series. In practice the process works well up to 1 MHz sampling speeds, but can break down above these rates. This is due to the fact that naive implementation of the generic interrupt procedures

that are part of the DSP software library will result in unacceptably large overhead times since they can demand between 35 and 120 CPU clock cycles. However, in our implementation, we rewrite the interrupt handler software so that an interrupt is handled in only a few clock cycles. This savings in time allows data to be acquired at the full 3 MHz rate.

The Omega DSP also possesses a pair of digital-to-analog converters (DAC). This allows us the opportunity to acquire data, analyze it, then issue a controlling signal. This feature can be used to control a chaotic trajectory or apply steering and communications signals to a chaotic carrier wave.

References

- [1] J. Testa, J. Perez, and C. Jeffries, Phys. Rev. Lett. **48**, 714 (1982).
- [2] T.C. Newell, P.M. Alsing, A. Gavrielides, and V. Kovanis, Phys. Rev. Lett. **72**, 1647 (1994).
- [3] T.C. Newell, V. Kovanis, and A. Gavrielides, Phys. Rev. Lett. **77**, 1747 (1996).
- [4] T.C. Newell, V. Kovanis, A. Gavrielides, and P. Bennett, Phys. Rev. E **54**, 3581 (1996).
- [5] W. L. Ditto, S. N. Rauser, and M. L. Spano, Phys. Rev. Lett. **65**, 3211 (1990).
- [6] T.C. Newell, V. Kovanis, A. Gavrielides, Phys. Rev. E, to be published August 1997.
- [7] H.D.I. Abarbanel, Analysis of Chaotic Data {Springer-Verlag New York}, 1996.
- [8] L. Fabiny, P. Colet, R. Roy, and D. Lenstra, Phys. Rev. A **47**, 4287 (1993).
- [9], W. Koechner, "Solid-State Laser Engineering," {Springer-Verlag Berlin}, 1996.
- [10] N. P. Barnes, M. E. Storm, P. L. Cross, M. W. Skolaut, IEEE J. QE-26, 558 (1990).
- [11] A. Yariv, "Introduction to Optical Electronics," {Holt, Rinehart, and Winston, New York}, 1976.
- [12] E. Ott, C. Grebogi, J. A. Yorke, Phys. Rev. Lett **64**, 1196 (1990).



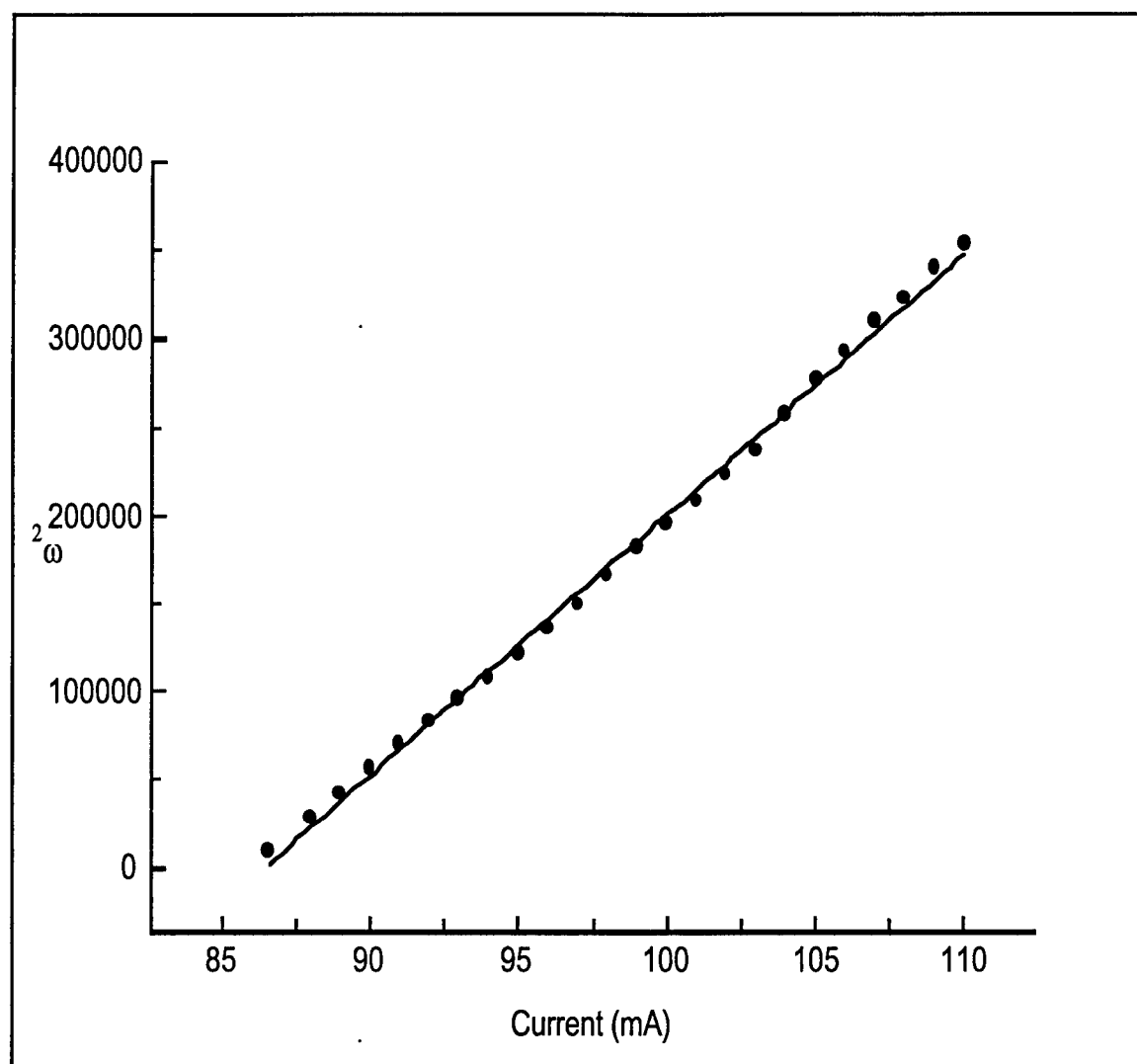


Figure 2

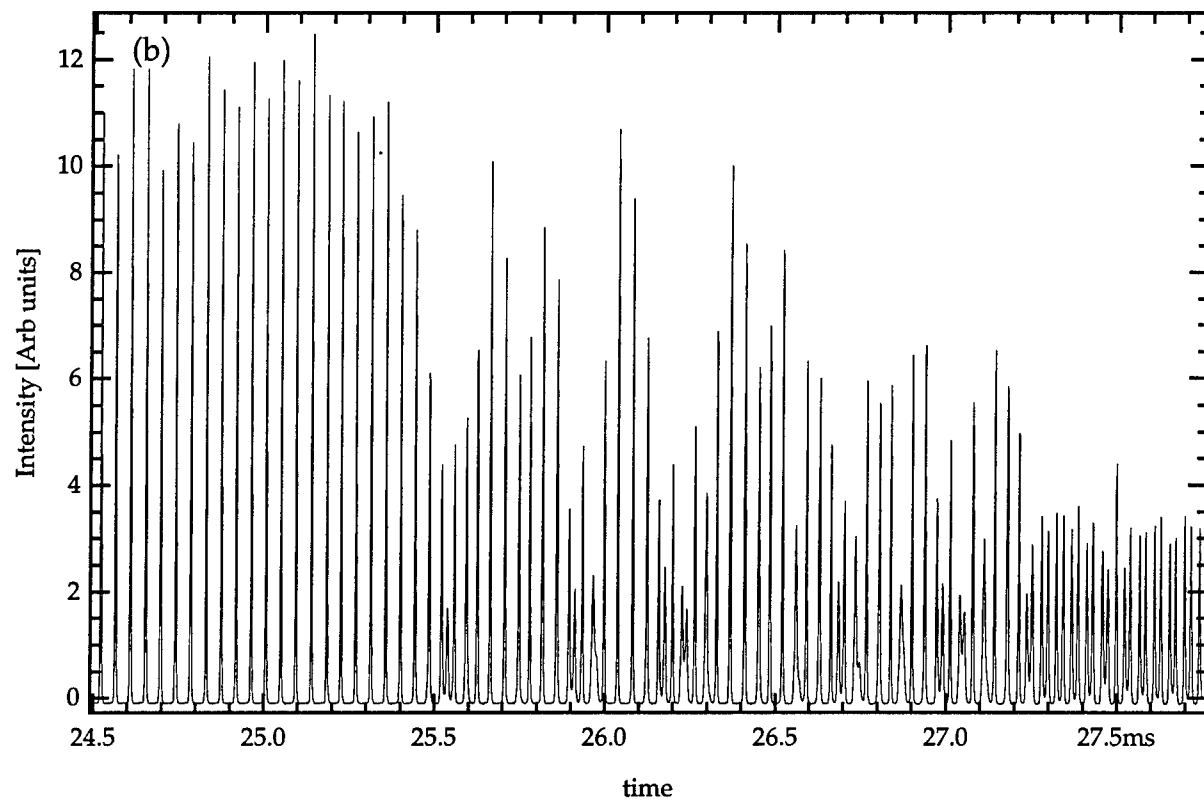
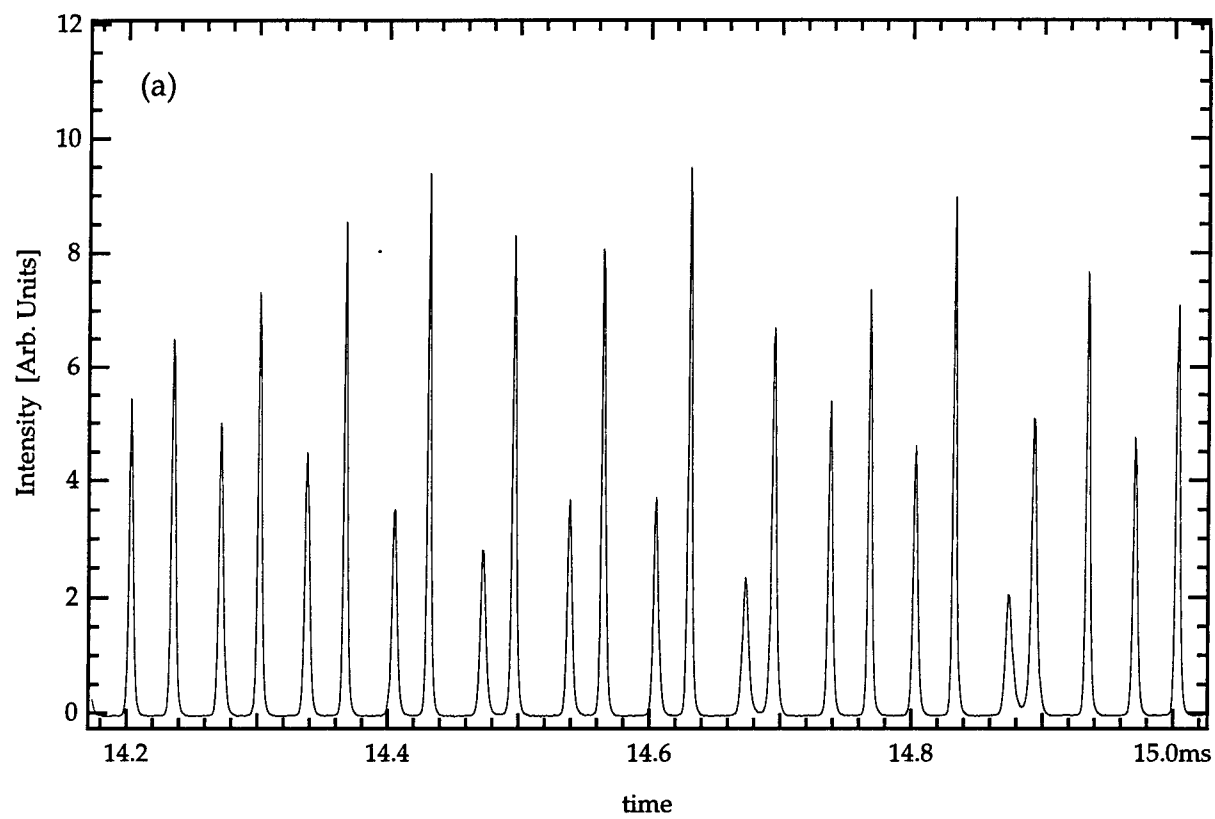


Figure 3

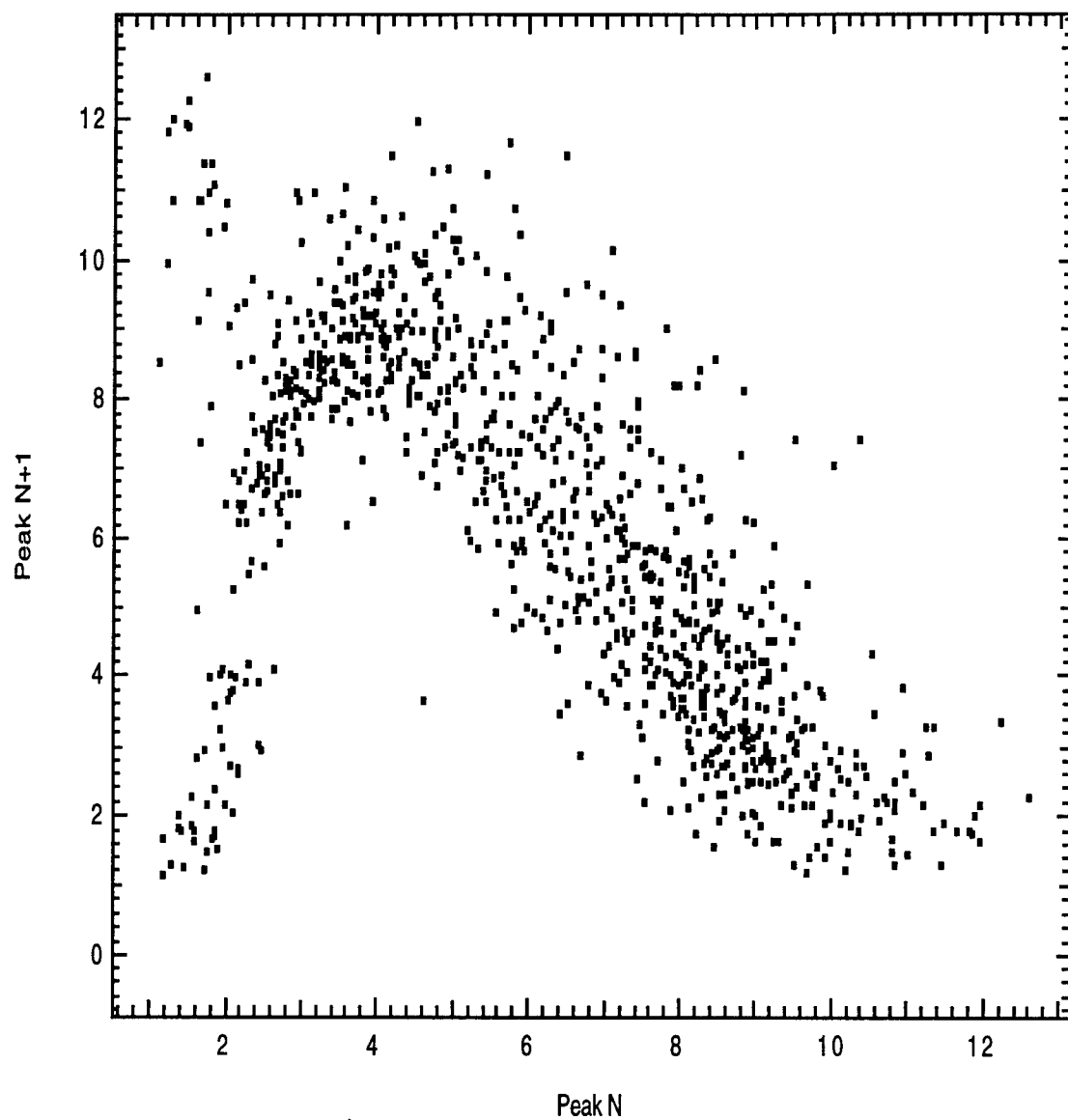


Figure 4

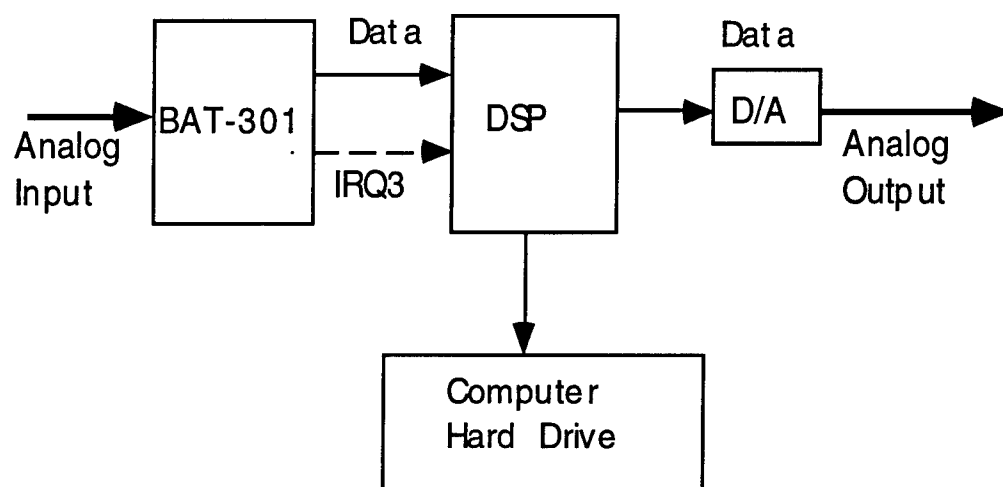


Figure 5

**PREPARATORY WORK TOWARDS A COMPUTER SIMULATION OF
ELECTRON BEAM OPERATIONS ON TSS 1**

Michael J. Pangia
Associate Professor
Department of Geology & Physics Engineering

Georgia Southwestern State University
800 Wheatley Street
Americus, GA 31709-4693

Final Report for:
Summer Research Program
Phillips Laboratory

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

And

Phillips Laboratory

September 1997

PREPARATORY WORK TOWARDS A COMPUTER SIMULATION OF ELECTRON BEAM OPERATIONS ON TSS 1

Michael J. Pangia
Associate Professor
Department of Geology & Physics
Georgia Southwestern State University

Abstract

This work is the first stage towards a Particle-In-Cell simulation that will be used to explain the low kHz frequency electron modulations observed by the Space Particle Correlator Experiment (SPACE) during beam operations on the Tethered Satellite System (TSS 1). The approach is to work in the frame of reference of the orbiter with the ambient plasma flowing past it. A 2-D (two position and two velocity components) treatment is chosen based on the geometry of the problem and computer limitations. The main emphasis of this work is to test that the free-flowing plasma is stable before including the effects of a beam.

FURTHER ANALYSIS OF KILOHERTZ ORDER WAVES ASSOCIATED WITH ELECTRON BEAM OPERATIONS ON STS 46

Michael J. Pangia

I. Introduction

Strong electron modulations in the 0.5 to 2.0 kHz frequency range observed by the Space Particle Correlator Experiment (SPACE) during beam operations on the Tethered Satellite System (TSS 1) have been identified by *Gough et al.* [1997]. The purpose of this research is to begin work on a Particle-In-Cell (PIC) simulation of beam operations to determine the mechanisms responsible for generating the observed plasma waves. Progress was made in testing the stability of algorithms when just the free-flowing plasma is included, i.e., without any interaction with the beam taking place.

II. Basics of Model

Figure 1 shows the basic simulation geometry that was chosen. Its features are:

- (1) The frame reference is that of the orbiter.
- (2) It is two dimensional (two position and two velocity components).

The magnetic field is perpendicular to Figure 1, so that the beam region corresponds to a cylindrical region with the electrons gyrating

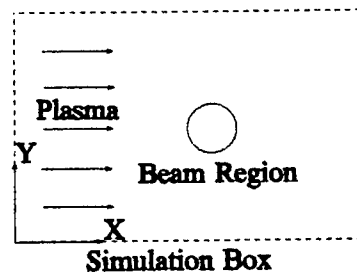


Fig. 1

around the field lines. It is believed that the major effect of the electron beam is to heat the plasma. However, the problem would have to be treated in 3-D to correctly account for the heating process. To simplify the problem so that a two dimensional treatment may be used, the heating process is to be modeled using some, as of yet undecided, algorithm.

Although the electrons act like a fully magnetized fluid and treating the electrons as a fluid would be computationally efficient, the equations of state and transport coefficients for both the ambient and beam electrons are unknown. Therefore, it was decided to treat both the electrons and the ions as particles using explicit time stepping.

III. Stability of Free-Streaming Plasma

The first step is to insure that the free-streaming plasma is stable before adding a beam. A question arose as to how to handle particles that leave the simulation box. The geometry of Figure 1 permits the use of period boundary conditions at the top and bottom boundaries (the Y-limits of the simulation box). However, since the plasma would be slowing down in the presence of a beam, periodic conditions do not apply at the left and right boundaries (the X-limits of the simulation box).

The boundary conditions chosen for the electric field is to set the electric potential to zero at the left and right boundaries. As for the particles, simply deleting particles that exit either the left or right boundaries of the simulation box would result in an instability, because particles that would eventually gyrate back into the simulation box once having exited the left and right boundaries would be getting deleted. The remedy was to do two things. (1) Initially add particles in the regions to the left and right sides of the simulation region having the same initial thermal and drift speeds as the simulation region. These regions act as buffers that allow particles to exit the simulation box and return at some later time step. (2) Use the criterion that particles would be removed from the simulation only if they were several grid steps outside of the left or right boundaries. Particles that are in the buffers are allowed to continue gyrating and drifting, but not to experience any electric field.

A test case was run using 10,000 particles of both species uniformly filling the simulation box. Other parameters are listed in the following table with f_e and f_{∞} being the electron plasma and cyclotron frequencies, m_a and v_a being the mass and thermal speed for species a ($a=i,e$), and V_d being the drift speed of the plasma. The mass ratio is set artificially high as a simplification.

Parameter	Value
f_{∞}/f_e	1
m_i/m_e	10000
v_i/v_e	0.003
V_d/v_e	0.01

The initial phase space plot V_y vs. X proved to be useful in predicting stability. Figure 2 shows the initial electron configuration. The X values are in electron Debye lengths, λ_e , and V_y values are in terms of v_e . The simulation box extends from $X=0$ to $X=20 \lambda_e$. Notice that electrons are permitted to extend outside of the box. (The length of the buffers were $5\lambda_e$.) The skewness in the phase plot is due to placing the electrons at their actual position once their guiding center and velocity components were randomly chosen. By looking at Figure 2 one can say that the size of the buffers must be at least as large as was chosen, otherwise the simulation box would have a shortage of electrons of large positive V_y at $X=0$ (of large negative V_y at $X=20\lambda_e$) which would lead to an artificial instability.

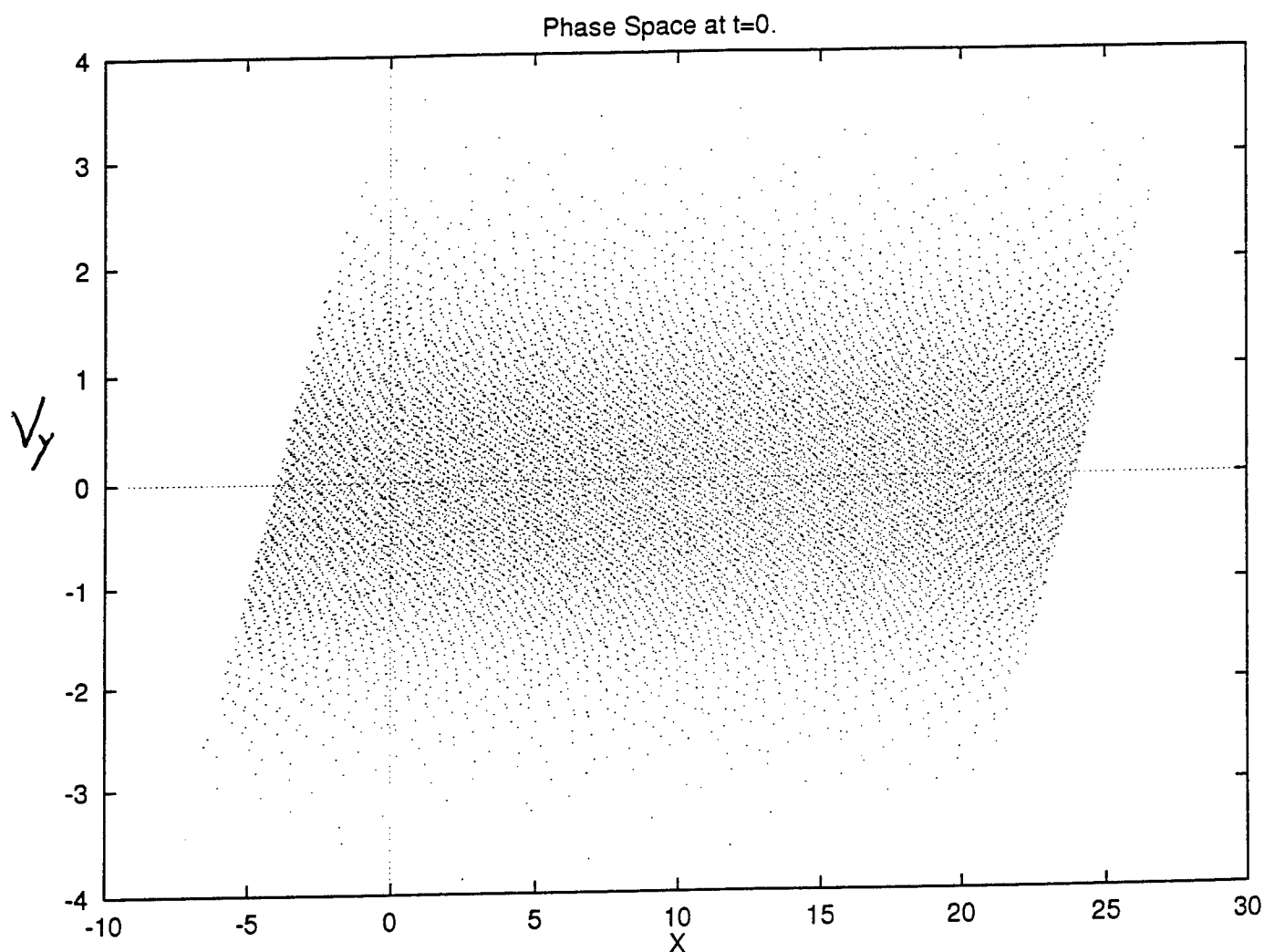


Fig. 2

Figures 3 and 4 show contours of electric potential (in dimensionless form $e\phi/kT_e$) at $t=0$ and $\omega_e t=200$, where ω_e is the electron angular frequency. Although the background level did grow over the time period, it was basically stable. It is believed that the number of particles can be increased and the grid spacing can be decreased to such an extent so as to reduce the noise to an acceptable level when the simulation is run on a faster computer. This will occur after having fully tested the simulation.

Phi Contours at $t=0$.

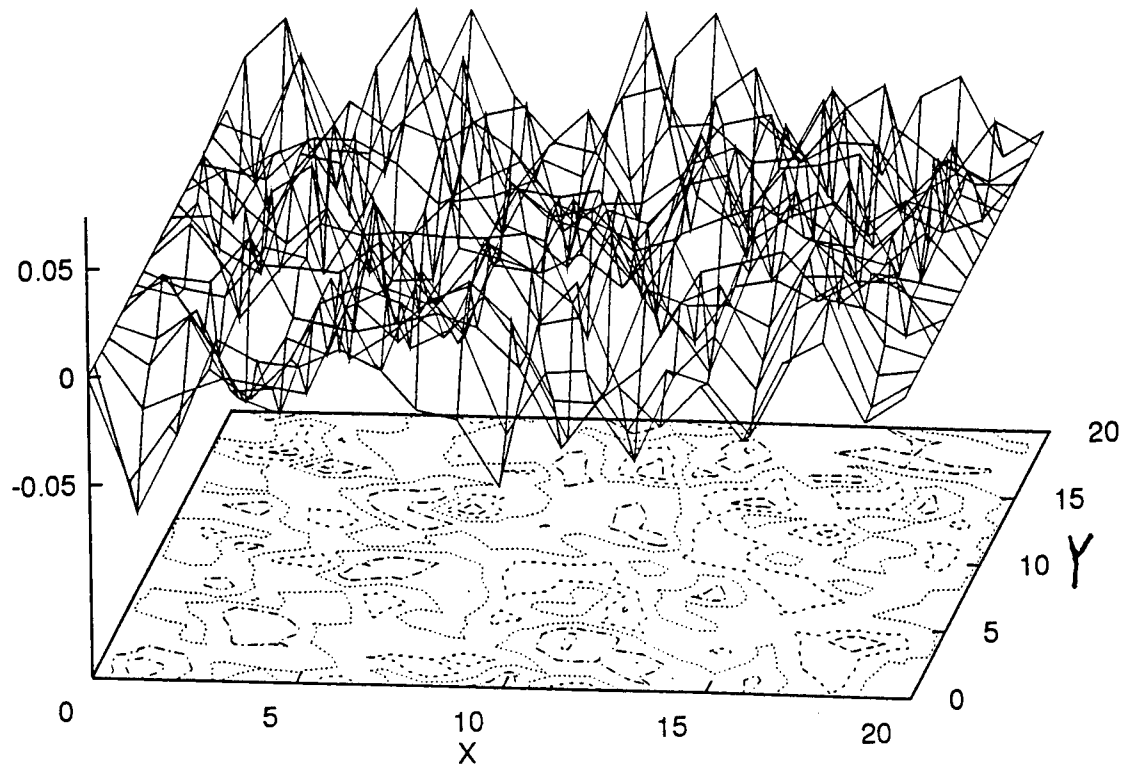


Fig. 3

Conclusions

It is believed that the basic model of free-streaming plasma is stable. This represents completion of the first stage towards modelling the beam operations on TSS 1. The next stage would be inclusion of a heating region to model the effect of the beam.

Phi Contours at $t=200$.

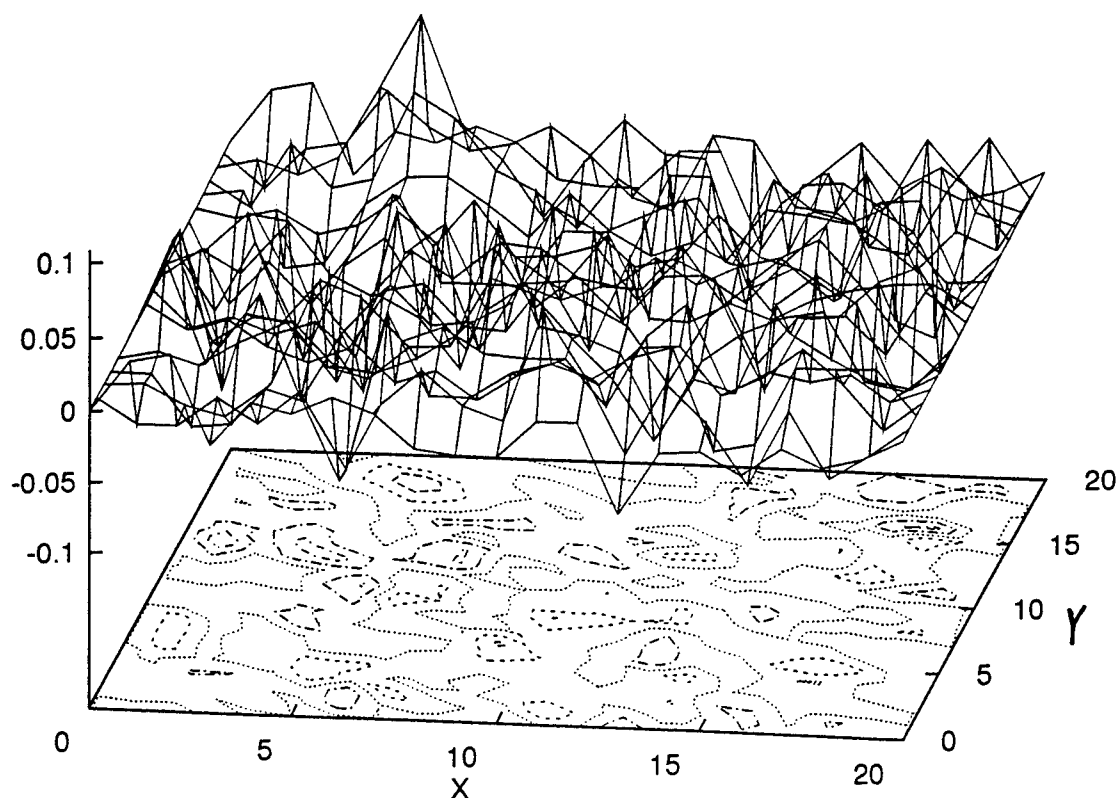


Fig. 4

Reference

Gough, M. P., D. A. Hardy, W. J. Burke, M. R. Oberhardt, L. C. Gentile, C. Y. Huang, D. L. Cooke, W. J. Raitt, D. C. Thompson, W. McNeil, and K. Bounar, Heating and low-frequency modulation of electrons observed during electron beam operations on TSS 1, *accepted to J. Geophys. Res.*, 1997.

MODELING OF IONOSPHERIC CONVECTION
FROM THE IMF AND SOLAR WIND DATA

Vladimir O. Papitashvili
Associate Research Scientist
Space Physics Research Laboratory

The University of Michigan
2455 Hayward Street
Ann Arbor, MI 48109-2143

Final Report for:
Summer Faculty Research program
Phillips Laboratory

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

and

Phillips Laboratory

September 1997

MODELING OF IONOSPHERIC CONVECTION FROM THE IMF AND SOLAR WIND DATA

Vladimir O. Papitashvili
Associate Research Scientist
Space Physics Research Laboratory
The University of Michigan

Abstract

It has been demonstrated in many studies that the large-scale, quasi-steady ionospheric convection at high latitudes is primarily controlled by the interplanetary magnetic field (IMF) and solar wind (SW) parameters. In this study we applied a linear regression analysis technique to the Defense Meteorological Satellite Program (DMSP) electrostatic potential database available at Phillips Laboratory/Geophysics (PL/GPS). We binned satellite data by every 1° of the corrected geomagnetic (invariant) latitude and 0.5 hour of magnetic local time (MLT). Then the regression analyses has been applied to the data in each bin in a similar way as it was done for construction of the IZMIRAN Electrodynamic Model (IZMEM) from ground magnetometer data. This allowed us to calibrate the IZMEM model against satellite data for a future attempt of constructing the DMSP electric potential model for prediction of the ionospheric convection over both the Northern and Southern polar regions solely from the IMF/SW parameters measured near the Earth's orbit.

MODELING OF IONOSPHERIC CONVECTION FROM THE IMF AND SOLAR WIND DATA

Vladimir O. Papitashvili

Introduction

The understanding of complex interactions that couple the solar wind (SW) plasma, the interplanetary magnetic field (IMF), and the Earth's magnetosphere and ionosphere is of fundamental interest of solar-terrestrial physics. In order to integrate what has been learned through a number of national and international disciplinary-oriented research programs for the enhancement of prediction technology, the U.S. National Science Foundation (in cooperation with NOAA, NASA, and DoD) announced a challenge to the space physics community through the establishment of the National Space Weather Program [Maynard, 1995]. The program will be responsible for the development of a science-based prediction technology which could help society to reduce economic losses due to sudden disturbances on the Sun, related geomagnetic storms, and the resulting "blackouts" in electric power, communication, navigation, technological systems, etc.

It has been found in numerous studies that changes in the strength and direction of IMF near the Earth's orbit are important factors in causing the geomagnetic activity that can greatly perturb the magnetosphere and ionosphere. The coupled solar wind-magnetosphere-ionosphere system is most highly energized when the IMF turns south and becomes anti-parallel to the direction of the Earth's magnetic field. A variety of mechanisms are involved in the energy transfer down to the ionosphere and thermosphere. The geomagnetic disturbances, particle precipitation and aurora in the polar caps are generally controlled by the directly driven processes because the interplanetary electric field $E_{IMF} = [-V_{SW} \times B_{IMF}]$ is continuously imparted to the magnetosphere. The disturbances in the auroral zone, middle and low latitudes (i.e., on a global scale) are mainly controlled by the loading-unloading processes when energy is initially stored in the magnetospheric tail and then released in the form of magnetic storms and substorms. The neutral atmosphere circulation is affected significantly by the global magnetic storms; at the same time, the neutral wind "fly-wheel" effect causes a significant post-storm ionospheric ion convection.

There has been a steady evolution in the production of maps of the ionospheric electric po-

tential or plasma convection patterns that are associated with various orientations of the IMF. These maps have consisted of both empirical and theoretical models [e.g., *Heppner, 1972; Heelis et al., 1982; Foster, 1983; Heppner and Maynard, 1987; Rich and Maynard, 1989; Hairston and Heelis, 1990; Papitashvili et al., 1994; Rich and Hairston, 1994; Weimer, 1995; Ruohoniemi and Greenwald, 1996*]. More often than not, the maps consist of pictorial sketches; while useful for the knowledge that they provide, the lack of a flexible or portable, numerical representation may limit the maps' utility.

For the operation purposes, it is desirable to be able to measure continu-

ously the IMF parameters as well as the solar wind velocity and density near the Earth's orbit at the first Lagrangian point (L1, approximately 200 Earth's radii upstream). Having such kind of observations (for example, from the NASA's WIND and/or ACE spacecraft), we can imagine a predictive scheme of the space weather system based on coupled theoretical models that describe the resulting flow of mass, energy, and momentum through the solar-terrestrial system. Figure 1 shows results from the on-going development of such a forecasting prototype.

Brief description of the IZMEM model

Specifically, from the ground magnetometer data we developed a semi-empirical linear regression model of the solar wind-magnetosphere-ionosphere interaction parameterized by the IMF strength and direction: the IZMIRAN Electrodynamics Model, IZMEM [*Feldstein and Levi-*

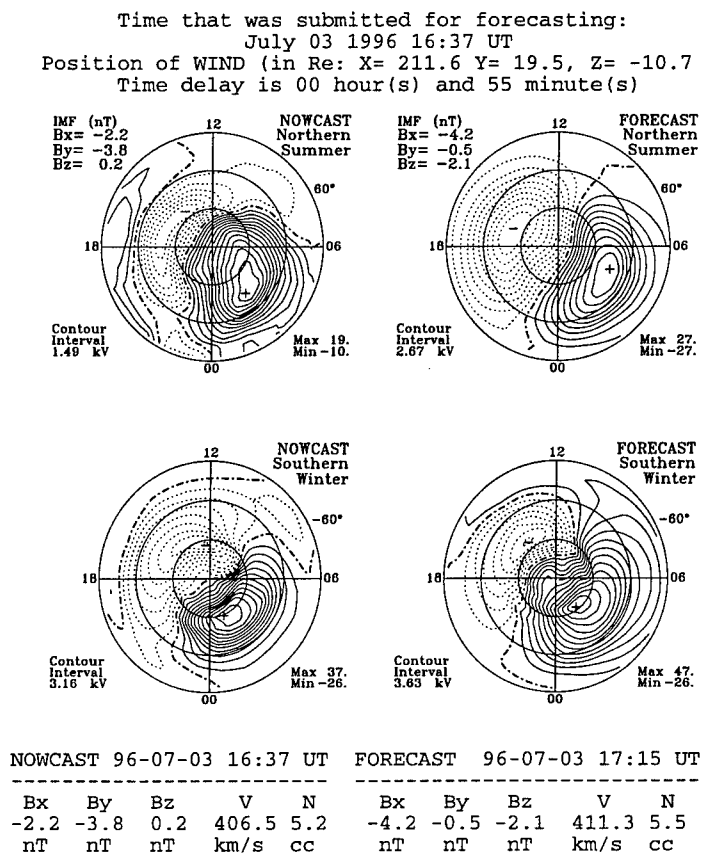


Figure 1. An example of the ionospheric convection specification and forecast obtained from the IZMEM model (http://www.sprl.umich.edu/space_weather).

tin, 1986; Papitashvili et al., 1994, 1995]. We postulate that the solar wind-magnetosphere-ionosphere coupling can be considered, in a first approximation, as a linear "black box". This "box" accepts changes in the IMF and solar wind plasma parameters (B_x , B_y , B_z , velocity V , and density n) as an input signal, and then induces ground-based geomagnetic perturbations as the output signal. This approach has already been used in other studies, in particular, those employing the linear prediction analysis [e.g., Clauer, 1986, and references therein].

In this linear model the regression coefficients relate any ground geomagnetic field component, for example, H , to changes in the corresponding IMF parameter [Levitin et al., 1982; Papitashvili et al., 1990, 1994]:

$$H^i(\theta, \phi) = K_{HBx}^i + K_{HBy}^i + K_{HBz}^i + H_0^i \quad (1)$$

The equation (1) shows only the relationships with the IMF, but its free term can be expanded for the solar wind parameters:

$$H_0^i = K_{HBx}^i + K_{HV2}^i + K_{HnV2}^i + H_{00}^i \quad (2)$$

Here K_{Hxx}^i are regression coefficients for $i = 1, \dots, 24$, where i is the universal time (UT) hour and H_0^i is a residual part of (1) for average solar wind parameters (e.g., $n = 4 \text{ cm}^{-3}$, $V = 450 \text{ km/s}$). H_{00}^i represents the geomagnetic variations that are free from the IMF and solar wind impact.

Developing this model, we utilized the hourly mean, total values of the IMF and ground geomagnetic data for each season of the year (summer, winter, and equinox). Data from both the Northern and Southern polar regions above $\Phi_{\text{CGM}} = \pm 57^\circ$ latitude have been subjected to the multi-regression analyses. Then, utilizing the numerical inversion technique [Kamide et al., 1981; see also Papitashvili et al., 1994], we can obtain the distribution of ionospheric electrostatic potentials over the entire polar regions as a superposition of the IMF related terms accounting for different interplanetary conditions (e.g., positive and negative B_z and B_y , various V and n). A number of ionospheric electrodynamic parameters (such a current function, electric potentials, Joule heating, magnetic disturbances, electric fields, ionospheric, Hall, Pedersen or field-aligned currents) can rapidly be computed from the IZMEM model and mapped over both the Northern and Southern polar regions. Even the ground-induced currents (GIC - important for the electric power companies) can successfully be modeled for given IMF conditions [e.g., Walker and Papitashvili, 1994].

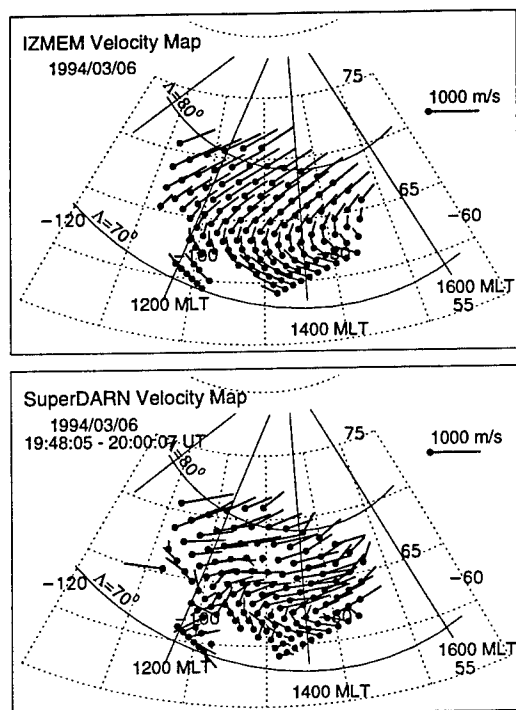


Figure 2. The IZMEM output in geographic coordinates: A comparison of the modeled ion convection velocities at 300-km altitude with the SuperDARN observations.

Figure 2 shows an example where the ionospheric electric fields modeled by the IZMEM are converted into the ion drift velocities at 300 km altitude (top panel) and mapped in geographic coordinates for the points collocated with the SuperDARN radars observations (lower panel) [Kustov *et al.*, 1997]. One can see that both sets of velocity vectors exhibit the same trends, namely the westward flow at lower latitudes and the eastward flow at higher latitudes. Certainly, the model shows smoother flows, but in general the prediction and measurements agree well. The IZMEM model can be accessed via the Web (<http://www.sprl.umich.edu/MIST>).

Thus, we conclude that the IZMEM model is able to provide structural elements of the high-latitude ionospheric convection, that is, different

terms in Eqs. (1) and (2) represent corresponding contributions (elements) from different sources to the total convection pattern. For the first time, the similar approach was utilized for the DMSP electrostatic potential data.

Methodology

The research undertaken at the Phillips Laboratory was two-fold. First, we comprehensively studied the DMSP ion driftmeter database (total 6.8 Gigabytes) sorting four years of data (1993-1996) from 5 satellites (F08 - 1.0 GB, F10 - 1.1 GB, F11 - 1.6 GB, F12 - 1.8 GB, and F13 - 1.3 GB) according to the IMF, corrected magnetic (invariant) latitude and magnetic local time (MLT). According to *Hairston and Heelis* [1993], the DMSP ion drift data have already been averaged every 4 seconds during the database construction. The every record in the data sets consists of the time mark, horizontal and vertical ion drift velocities, geographic and geomagnetic

coordinates, and electrostatic potential calculated along the satellite track over the high latitudes (the latter is corrected for the corotation potential).

The DMSP ion drift database has been sorted and reduced to total 460 Mbytes of data measured above 40° invariant latitudes in both the Northern and Southern Hemispheres and matched with the available IMF records (hourly values, OMNI database obtained from the NASA's National Space Science Data Center). The IMF data for a current hour were used for the matching if the DMSP record had the time mark between 30:00 and 59:59 minutes of each UT hour; the previous IMF UT hour was matched with the DMSP records taken between 00:00 and 29:59 minutes of each hour. This allows us to suppose that the DMSP measurements were generally affected by the averaged IMF conditions pre-existed during previous ~ 30 or more minutes.

Then we sorted out all of these records according to the season of the year (northern summer and southern winter: May-August; northern winter and southern summer: November-February; equinox: March, April, September, and October) and certain IMF conditions: (1) $B_z \leq 0$ for all B_y , (2) $B_z > 0$ for all B_y , (3) $B_y \leq 0$ for $B_z \leq 0$, (4) $B_y > 0$ for $B_z \leq 0$, (5) $B_y \leq 0$ for $B_z > 0$, and (6) $B_y > 0$ for $B_z > 0$. For example, total 1.7 million records were processed for all seasons of the year only to satisfy the 1st condition.

For each of the above-mentioned conditions, the DMSP electrostatic potential data can be binned in every 1° (from 49.5° to 90° , centering on 50° , 51° , etc.) of invariant latitude and 0.5 hr of magnetic local time (00:00-29:59, 00:30-59:59, etc.). The last latitude bin is extended from 88.5° to 90° . If a few 4-sec consecutive DMSP data records fell in the same bin, the electrostatic potential was averaged for these records. This improved the reliability of statistics, though the number of points in each bin reduced. After reduction, the number of points in each bin varies from a few ones to a few hundreds (maximum ~ 1500 in certain bins). As expected, the maximum coverage was achieved for the dawn-dusk local times; the lesser coverage (even zero coverage) is achieved for noon-midnight regions at lower latitudes.

Because such a data processing consumes a lot of the CPU time, we were able to process fully only the data satisfied the 1st condition. According to Eq. (1), we applied the linear regression technique to the data in each bin (if the number of points in the bin was less than 10, this bin was skipped). Due to significant scattering in the DMSP potential data fallen in the same bin for

a certain IMF condition, we applied two selection criteria to the results of the regression analyses. All analyses where the correlation coefficients between the DMSP potentials and the IMF $B_z \leq 0$ data were less than 0.3 were excluded from consideration. Also, calculating background (quasi-viscous) potential from the free term in Eq. (1), we threw out the bins where the background potential exceeded ± 35 kV. In most cases, the number of "good" bins were about 50% out of the total number of bins: $40 \times 48 = 1920$. To ensure that the residual terms and regression coefficients for $B_z(-1\text{-nT})$ in Eq. (1) obtained from DMSP data really represent the background potentials and responses on changing IMF, we compared the IZMEM and DMSP potential values in each bin where the DMSP data analyses were available. This allows us to calibrate the IZMEM model against the DMSP potentials obtained for the 1st IMF condition.

Results

Figure 3 shows the correlation between the DMSP and IZMEM background potentials and the regression coefficients for $B_z(-1\text{-nT})$ step) for the northern summer (only points satisfied the selection criteria are shown). With the equation

$$y = ax + b \quad (3)$$

we can compare and evaluate these two models. Here y stands for the DMSP potentials obtained as result of the regression analyses with the IMF data, x stands for the original IZMEM potentials. b is an offset between two model data sets, and a is a calibration factor which can be applied to the IZMEM potentials [see below]. It is obvious that even applying such strict selection criteria, the certain number of outliers still survive in the data set.

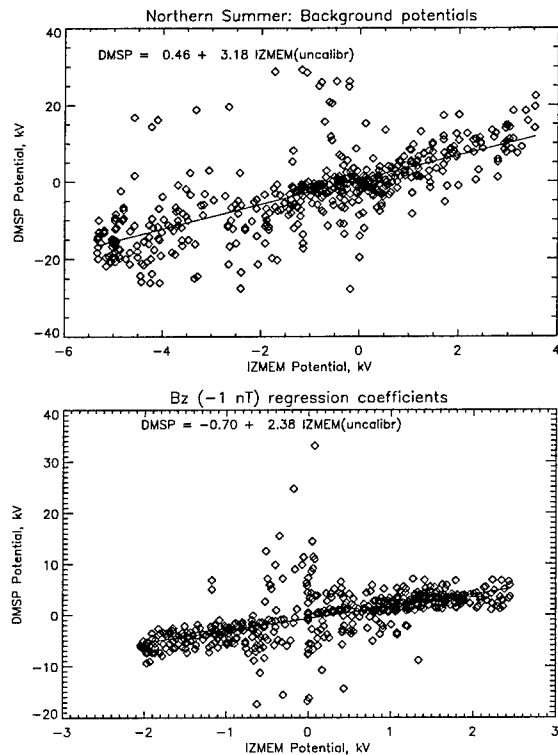


Figure 3. Correlation between the DMSP and IZMEM potentials for northern summer (all data).

The DMSP potential outliers generally group near zero IZMEM potentials. Because the latter mainly represent the boundary condition on $\pm 57^\circ$ invariant latitude where the magnetospheric potential is zero, we suppose that the DMSP potentials in the original data sets have not fully been corrected for the corotation potential.

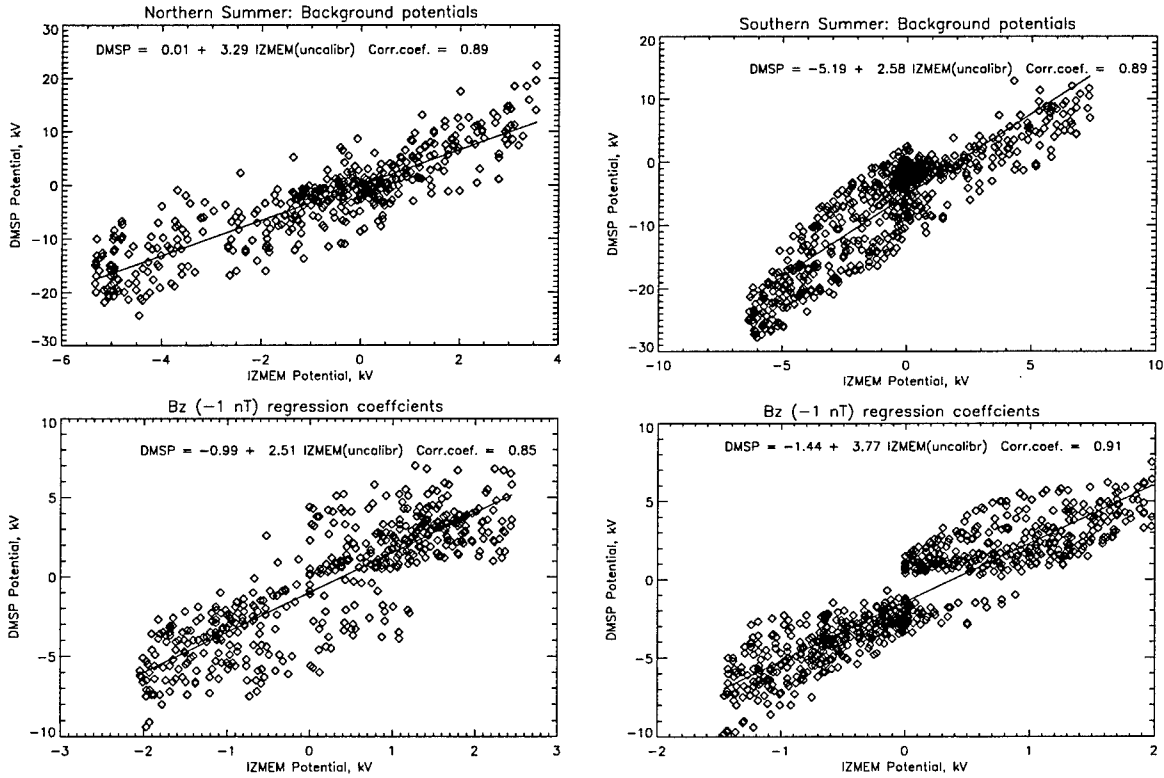


Figure 4. Correlation between the DMSP and IZMEM potentials for the northern (left panels) and southern (right panels) summers (here the DMSP data are cleaned from outliers).

Figure 4 (left panels) show the same data as on Figure 3 but now the DMSP potentials are cleaned from outliers by applying the $\pm 1.5 \sigma$ criterion (σ is a standard deviation obtained for all data in each panel of Figure 3). One can see that this improved the correlation between two data sets. Figures 4–5 show plots for cleaned (from outliers) data for all seasons of the year and for both the Northern and Southern polar regions.

Table 1 summarizes correlation parameters between the DMSP and IZMEM potentials and shows comparison with the calibration factors obtained earlier for IZMEM by *Papitashvili et al.* [1994]. Note that both the correlation coefficients for all data (k_o) and for cleaned ones (k_c) are presented in Table 1.

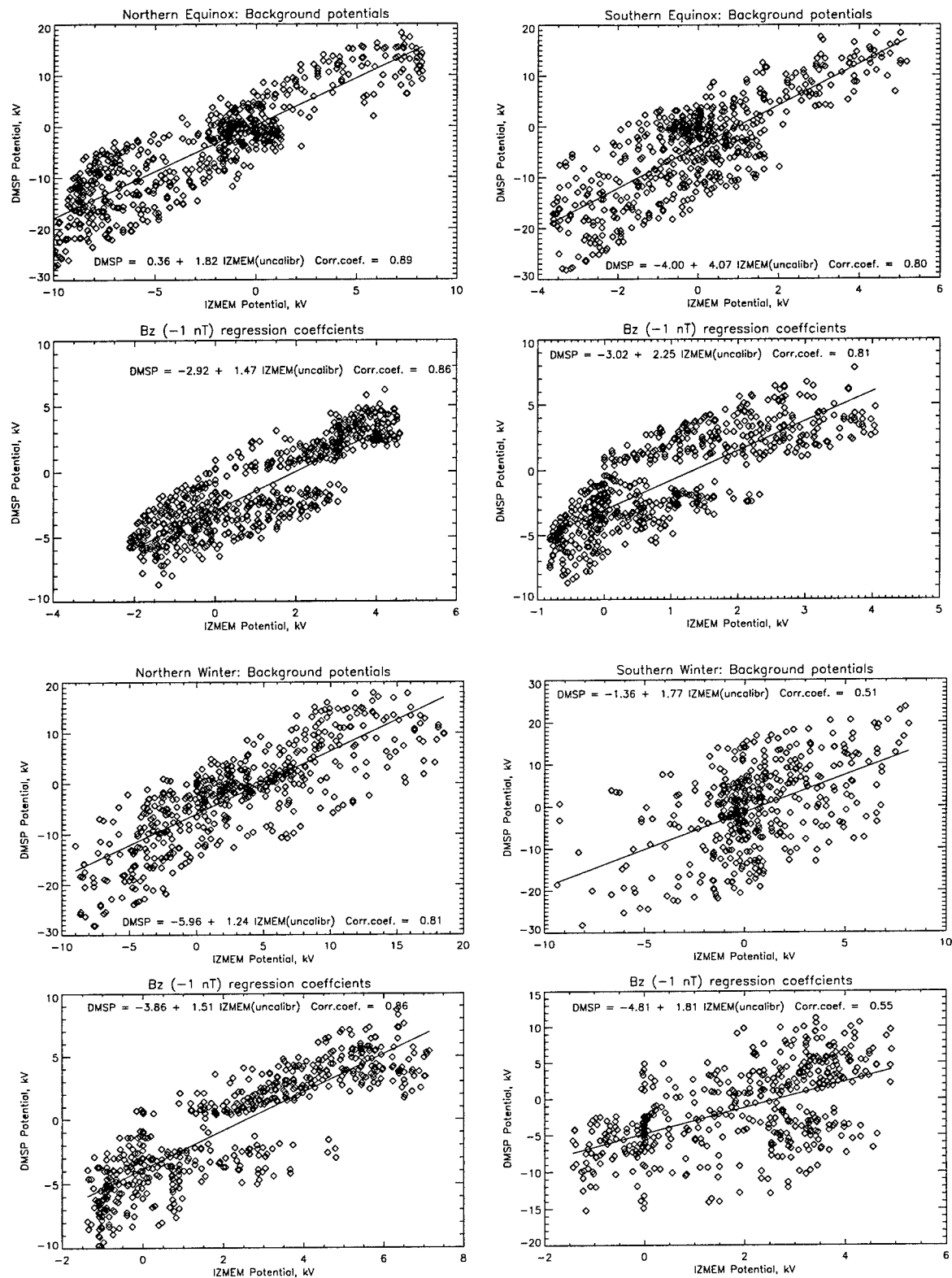


Figure 5. Correlation between the DMSP and IZMEM potentials for the northern (left panels) and southern (right panels) equinoxes (two top panels) and winters (two lower panels).

The number of outliers ignored for the second round of correlation analyses by applying the $\pm 1.5 \sigma$ criterion was only about 10-15% of the total number of points in the analysis. Figures 4-5 show clearly that the DMSP potentials generally have the same trends as the corresponding IZMEM potentials. Better correlation between two data sets is achieved for the summer and equinox seasons in both hemispheres. The winter potentials correlate better in the Northern Hemisphere; the southern potentials correlate fairly.

The coefficients b in Table 1 can be compared with the amplification factors f obtained by *Papitashvili et al.* [1994] solely from the assumption that the background (viscous) cross-polar potential drop should not exceed ~ 35 kV [*Reiff et al.*, 1981]. The latter value was revised in a number of studies [e. g., *Sergeev and Kuznetsov*, 1981; *Mozer*, 1984; *Reiff and Luhmann*, 1986; *de la Beaujardiere et al.*, 1987; *Papitashvili and Papitashvili*, 1996]. It was suggested that this potential might be as low as 6-10 kV, however, the accepted maximum values are between 15 and 25 kV.

Table 1.

Season	Hemi-Sphere	Background Potentials				Bz (-1 nT) Coefficients			IZMEM	
		k_o	k_c	a	b	k_o	k_c	a	B	F
Summer	North	0.71	0.89	0.01	3.29	0.60	0.85	-0.99	2.51	3.2
	South	0.79	0.89	-5.19	2.58	0.80	0.91	-1.44	3.77	3.0
Equinox	North	0.79	0.89	0.36	1.82	0.71	0.86	-2.92	1.47	2.3
	South	0.67	0.80	-4.00	4.07	0.69	0.81	-3.02	2.25	2.7
Winter	North	0.69	0.81	-5.96	1.24	0.65	0.86	-3.86	1.51	1.5
	South	0.32	0.51	-1.36	1.77	0.41	0.55	-4.81	1.81	2.5

As seen from Table 1, most of the coefficients b for background potentials are in order with the IZMEM amplification factors f . Smaller b are obtained for the Bz (-1 nT) regression coefficients than for the background potentials. This suggest that both interaction mechanisms (quasi-viscous and reconnection) behave differently producing the ionospheric responses on changing

IMF and solar wind conditions. We conclude that now these new coefficients a and b can be utilized for re-calibration of the original IZMEM model [see discussion for the need of such a calibration in *Papitashvili et al.*, 1994].

Table 2 shows maximum cross-polar potentials estimated from the DMSP data (Figure 4-5) in comparison with the earlier calibrated IZMEM model. One can see that DMSP potentials are comparable with earlier estimates; the average potentials in two models differ only by 3.6 kV, that is, well within the standard deviation for both data sets. Similar difference for the B_z (-1 nT) regression coefficients is 1.8 kV, that is, within the standard deviation.

Table 2.

Season	Hemisphere	Background Potentials, kV		B_z (-1 nT) Coefficients (kV)	
		IZMEM	DMSP	IZMEM	DMSP
Summer	Northern	30.4	29.5	14.1	11.4
	Southern	37.8	35.4	10.5	12.9
Equinox	Northern	41.4	33.5	15.4	10.8
	Southern	30.5	35.8	13.0	11.0
Winter	Northern	40.5	34.1	13.0	12.8
	Southern	40.5	31.0	15.6	11.6
Average		36.8 ± 5.1	33.2 ± 5.3	13.6 ± 1.9	11.8 ± 2.0

Conclusions

Two major conclusions can be drawn from the presented analyses. First, the cross-polar potential drop estimated from DMSP data is in order of ~ 33 kV. This results provides a strong experimental confirmation that the background (viscous) potential is considerably larger than the currently accepted values (e.g., maximum ~ 15 - 20 kV), though it requires further confirmation for the northward IMF conditions. Second, the DMSP data show that the ionospheric potential response on the 1 -nT step in the southward IMF is ~ 12 kV. This number is close to the earlier results obtained for the IZMEM model.

We have to conclude this study with the presented results though the DMSP data have been processed far enough to undertake similar regression analyses for the northward IMF conditions,

as well as for estimating the responses on changing azimuthal component of the IMF. We recognized that duration of our research term (8 weeks) under the auspices of Air Force Summer Faculty Research Program was simply not enough for this huge computer time-consuming effort. This prevented us to accomplish the entire study "at-once" and to attempt the construction of a fully operational model of the DMSP potentials parameterized by the IMF and solar wind conditions. We hope that we can find a possibility to continue this interesting study in near future.

Acknowledgements

I gratefully thank Frederick Rich for his invaluable assistance with the DMSP data processing. Many useful discussions with Fred and Michael Heinemann (both at PL/GPS) helped me better understand the DMSP database and find venues for the progress.

References

- de la Beaujardiere, O., D. S. Ewans, Y. Kamide, R. P. Lepping, Response of auroral oval precipitation and magnetospheric convection to changes in the interplanetary magnetic field, *Ann. Geophys.*, 5A, 519, 1987.
- Clauer, C. R., The technique of linear prediction filters applied to studies of solar wind-magnetosphere coupling, in *Solar Wind - Magnetosphere Coupling*, edited by Y. Kamide and J. A. Slavin, p. 39, Terra Scientific, Tokyo, 1986.
- Feldstein, Ya. I., and A. E. Levitin, Solar wind control of electric fields and currents in the ionosphere, *J. Geomagn. Geoelectr.*, 38, 1143, 1986.
- Foster, J. C., An empirical electric field model derived from Chatanika radar data, *J. Geophys. Res.*, 88, 981, 1983.
- Hairston, M. R., and R. A. Heelis, Model of high-latitude ionospheric convection pattern during southward interplanetary magnetic field using DE 2 data, *J. Geophys. Res.*, 95, 2333, 1990.
- Hairston, M. R., and R. A. Heelis, High-latitude electric field studies using DMSP data, Preprint PL-TR-93-2036, 45 pp., Phillips Laboratory/Directorate of Geophysics, Hanscom AFB, Massachusetts, 1993.

- Heelis, R. A., J. K. Lowell, and R. W. Spiro, A model of the high-latitude ionospheric convection pattern, *J. Geophys. Res.*, 87, 6339, 1982.
- Heppner, J. P., Polar cap electric field distribution related to the interplanetary magnetic field direction, *J. Geophys. Res.*, 77, 4877, 1972.
- Heppner, J. P., and N. C. Maynard, Empirical high-latitude electric field models, *J. Geophys. Res.*, 92, 4467, 1987.
- Kamide, Y., A. D. Richmond, and S. Matsushita, Estimation of ionospheric electric fields, ionospheric currents and field-aligned currents from ground magnetic records, *J. Geophys. Res.*, 86, 801, 1981.
- Kustov, A. V., V. O. Papitashvili, G. J. Sofko, et al., Dayside ionospheric plasma convection, electric fields, and field-aligned currents derived from the SuperDARN radar observations and predicted by the IZMEM model, *J. Geophys. Res.*, 102, in press, 1997.
- Levitin, A. E., R. G. Afonina, B. A. Belov, and Ya. I. Feldstein, Geomagnetic variations and field-aligned currents at northern high-latitudes and their relations to solar wind parameters, *Phil. Trans. R. Soc. London Ser. A*, 304, 253, 1982.
- Maynard, N. Space weather prediction, *Rev. of Geophys.*, 33, *U.S. Nat. Rep. 1991-1994 IUGG*, Suppl., 547, 1995.
- Mozer, F. S., Electric field evidence on the viscous interaction at the magnetopause, *Geophys. Res. Lett.*, 11, 135, 1984.
- Papitashvili, V. O., and N. E. Papitashvili, Cross-polar electric potentials and AE index, in *The Evaluation of Space Weather Forecasts: Proceedings of a Workshop at Boulder, Colorado, June 19-21, 1996*, edited by K. Doggett, pp. 39-50, NOAA/SEC, Boulder, Colo., 1996.
- Papitashvili, V. O., Ya. I. Feldstein, A. E. Levitin, B. A. Belov, L. I. Gromova, and T. E. Valchuk, Equivalent ionospheric currents above Antarctica during the austral summer, *Antarctic Sci.*, 2, 267, 1990.
- Papitashvili, V. O., B. A. Belov, D. S. Faermark, Ya. I. Feldstein, S. A. Golyshev, L. I. Gromova, and A. E. Levitin, Electric potential patterns in the Northern and Southern polar regions parameterized by the interplanetary magnetic field, *J. Geophys. Res.*, 99, 13,251, 1994.

- Papitashvili, V. O., C. R. Clauer, A. E. Levitin, and B. A. Belov, Relationship between the observed and modeled modulation of the dayside ionospheric convection by the IMF B_y component, *J. Geophys. Res.*, *100*, 7715, 1995.
- Reiff, P. H., R. W. Spiro, and T. W. Hill, Dependence of polar cap potential drop on interplanetary parameters, *J. Geophys. Res.*, *86*, 7639, 1981.
- Reiff, P. H., and J. G. Luhmann, Solar wind control of the polar cap voltage, in *Solar Wind - Magnetosphere Coupling*, edited by Y. Kamide and J. A. Slavin, pp. 453-476, Terra Sci. Publ. Co., Tokyo, 1986.
- Rich, F. J., and M. Hairston, Large-scale convection patterns observed by DMSP, *J. Geophys. Res.*, *99*, 3827, 1994.
- Rich, F. J., and N. C. Maynard, Consequences of using simple analytical functions for the high-latitude convection electric field, *J. Geophys. Res.*, *94*, 3687, 1989.
- Ruohoniemi, J. M., and R. A. Greenwald, Statistical patterns of high-latitude convection obtained from Goose Bay HF radar observations, *J. Geophys. Res.*, *101*, 21,743, 1996.
- Sergeev, V. A., and B. M. Kuznetsov, Quantitative dependence of the polar cap electric field on the IMF B_z -component and solar wind velocity, *Planet. Space Sci.*, *29*, 205, 1981.
- Walker, J. K., and V. O. Papitashvili, Models of polar geomagnetic activity and equivalent currents: Approach and comparison of techniques, in *Solar--Terrestrial Energy Program: The Initial Results from STEP Facilities and Theory Campaigns*, edited by D. N. Baker, V. O. Papitashvili, and M. J. Teague, COSPAR Colloquia Series, Vol.5, pp. 813-818, Pergamon Press, Oxford, 1994.
- Weimer, D. R., Models of high-latitude electric potentials derived with a least error fit of spherical harmonic coefficients, *J. Geophys. Res.*, *100*, 19,595, 1995.

Associate did not participate in the program.

ANALYSIS OF TURBULENT MIXING IN THE STRATOSPHERE AND TROPOSPHERE

Louis F. Rossi
Associate Professor
Department of Mathematical Sciences Engineering

University of Massachusetts Lowell
One University Avenue
Lowell, MA 01854

Final Report for:
Summer Research Program
Phillips Laboratory

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

And

Phillips Laboratory

September 1997

ANALYSIS OF TURBULENT MIXING IN THE STRATOSPHERE AND TROPOSPHERE

Louis F. Rossi
Assistant Professor
Department of Mathematical Sciences
University of Massachusetts Lowell

Abstract

By analyzing the wavelet averages of thermal dissipation in the troposphere measured by the Argus platform, it is possible to determine regions where Kolmogorov's universal equilibrium hypotheses are valid. Not surprisingly, these regions of the flow exhibit spectra obeying the predicted $k^{-5/3}$ power law. In these regions, one also finds p^{th} -order structure functions that similarly obey the predicted $r^{p/3}$ power law. However, the nature and distribution of the regions of the flow wherein the dissipation is not constant remains unexplored, save that power law scalings are inapplicable in many cases. For all that it can accomplish, my treatment is relatively primitive. The underlying physical model is fairly simple and would benefit from further analysis such as multifractal distributions of scaling exponents. Also, more sophisticated signal processing would improve the quality of the data we are trying to analyze. For instance, the cold-wire probe response smooths the data to a resolution of 50 centimeters, though the data acquisition rate is capable of resolving scales down to 2 centimeters. It is a distinct possibility that the high frequency portion of the data can be recovered by deconvolving the raw data with an appropriate filter. This summer, I have been able to support Air Force calculations of the upper atmosphere by studying fundamental issues in turbulent mixing and signal processing.

1 Introduction

With AFOSR support, I have entered into a productive collaboration with Phillips Laboratory Hanscom and Kirtland scientists to analyze aerothermal data in connection with the Airborne Laser (ABL) project, and at the same time, analyze the fundamental properties underlying the turbulent mixing of passive scalars. The underlying objective of this work is to make decisions on whether or not a laser beam can be propagated through the upper atmosphere over distances of hundreds of kilometers in certain theatres of operation. Aerothermal measurements are relatively inexpensive compared to airborne optical experiments. Knowing the statistics of the thermal fluctuations over the path of the laser,

$$E(k) \propto C_T^2 k^{-\alpha} \quad (1)$$

where $E(k)$ is the power spectral density of the thermal fluctuations, C_T^2 is a reflection of the strength of these variations and α is predicted by Kolmogorov, Oboukov and Corrsin to be $-5/3$, it is possible to simulate the beam propagation and make a decision as to whether the beam will scintillate or not. Of course, without proper treatment, the aerothermal data is useless and tells us nothing. As a personal note, it is gratifying to see that the fundamental problem of passive scalar mixing remains at the heart of a critical application such as the ABL project.

2 Some pertinent aspects of atmospheric turbulence and passive scalar mixing

The central issues in this paper are Kolmogorov's universal equilibrium, turbulent mixing, intermittent bursts, integral scales over which averages are taken, and non-stationarity. Kolmogorov's seminal works, often referred to as K41 and K61 [15, 14, 16], are the most substantial and predictive description of turbulence from first principles through the Kármán-Howarth-Monin relation. Overviews of Kolmogorov's assumptions and its implications can be found in and many others [7, 21, 22, 29]. While these ideas only apply to kinetic energy dynamics, velocity increments, power spectral densities and related quantities, Corrsin and Oboukhov independently extended universal equilibrium concepts to thermal fluctuations [6, 23], and this was later confirmed experimentally by Gibson and Schwarz [9]. One must also take into account that while the equations governing the motions of slight thermal variations and passive scalars are linear, they are coupled to the nonlinear Navier-Stokes equations, and even simple solutions can yield chaotic trajectories and effectively mix scalar quantities (see [2] for an overview). In this article, we will be analyzing thermal data in the absence of any information about the kinetic energy of the flow. Furthermore, experimental evidence shows that universal equilibrium, and in particular constant mean dissipation, is violated by intermittent bursts, the statistics of which are an active area of research [5, 20, 24, 26, 27]. Central to the understanding of the statistics underlying any fluid quantity, f , is its average, $\langle f \rangle$. Since atmospheric flows contain many large-scale structures that vary with time, a considerable amount of attention is dedicated to determining meaningful ways to calculate $\langle f \rangle$. In this paper, we study each of these subjects using wavelet analysis.

My analysis focuses on horizontal measurements of thermal fluctuations in the troposphere where the Reynolds numbers are very large. Universal equilibrium has been very successful in describing the homogeneous turbulence in a wide range of laboratory and field experiments. However, in atmospheric flows, there is very strong evidence that there are many distinct regions within the atmosphere, each having distinct turbulent statistics. Many of these flat pancake-like regions or "blinis" co-exist at the same altitude, defying the simpler assumption that the atmosphere can be treated as sphere covered by many concentric layers like an onion, each with its own uniform turbulent statistics. While universal equilibrium exists within the distinct pancake-like regions, the kinetic or thermal dissipations may vary from one region to the next.

Within these atmospheric layers at small scales, the incompressible Navier-Stokes equations apply:

$$\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\nabla p + \frac{1}{\text{Re}} \nabla^2 \mathbf{u}, \quad (2)$$

where $\mathbf{u}(\mathbf{x}, t)$ is the velocity field, $p(\mathbf{x}, t)$ is the pressure, and Re is the Reynolds number, a dimensionless ratio of inertial to viscous forces. Furthermore, we assume that the velocity field is incompressible,

$$\nabla \cdot \mathbf{u} = 0, \quad (3)$$

and that fluctuations in the temperature, $\theta(\mathbf{x}, t)$, are very slight and can be treated as a passive quantity. That is, θ moves and diffuses with the flow,

$$\partial_t \theta + (\mathbf{u} \cdot \nabla) \theta = \frac{1}{\text{Pe}} \nabla^2 \theta, \quad (4)$$

but θ does not affect the fluid velocity as described by Equations (2) and (3). Here, Pe is the Peclet number, a dimensionless ratio of inertial to diffusive forces in the scalar field. The mean temperature plays no role in this analysis because equation (4) is linear and involves only derivatives of θ .

Kolmogorov's universal equilibrium assumptions operate in the limit $\text{Re} \rightarrow \infty$, at small scales, and far from boundaries in isotropic homogeneous flows. A more detailed discussion can be found in [7, 11, 29]. The only aspect of universal equilibrium that we shall focus on here is the hypothesis that *the energy dissipation per unit mass, given by*

$$\varepsilon = \frac{1}{2\text{Re}} \sum_{i,j=1}^3 \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)^2, \quad (5)$$

remains constant. Provided all Kolmogorov's assumptions are met, one can expect a number of consequences. First, for very large Reynolds numbers, there exists a range of scales, called the *inertial range*, where the turbulent flow statistics depend solely on scale (or wavelength) and the (constant) mean dissipation. If the flow is stationary and isotropic, then the correlation function depends only on the distance:

$$\langle \mathbf{u}(\mathbf{x} + \mathbf{r}) \cdot \mathbf{u}(\mathbf{x}) \rangle = B(r), \quad r \equiv |\mathbf{r}|, \quad (6)$$

where $\langle \cdot \rangle$ denotes the ensemble average over all possible solutions. (In truth, our analysis requires only *local* stationarity, as will be seen later. In §7, we will segment the data into

distinct regions wherein equation (6) is valid.) To describe the statistics of the flow in the inertial range, define the *power spectral density*,

$$E(k) \equiv \hat{B}(k), \quad (7)$$

(where $\hat{B}$ is the Fourier transform of B), and the p^{th} -order structure function,

$$S_p(\mathbf{r}) \equiv \langle |\delta \mathbf{u}(\mathbf{r})|^p \rangle, \quad p > 0, \quad (8)$$

where $\delta \mathbf{u}$ is a velocity increment

$$\delta \mathbf{u}(\mathbf{r}) \equiv \mathbf{u}(\mathbf{x} + \mathbf{r}) - \mathbf{u}(\mathbf{x}). \quad (9)$$

Then, a second consequence of Kolmogorov's assumptions is that the power spectral density and the structure function obey the following power laws within the inertial range:

$$E(k) = C k^{-5/3}, \quad (10)$$

$$S_p(\mathbf{r}) = C_p r^{p/3}. \quad (11)$$

As aluded to earlier, the power law behavior in kinetic energy and velocity increments can be extended to passive scalar quantities described by equation (4). In this case, one must assume that the Peclet number is very large and the thermal dissipation,

$$\chi = \frac{1}{2\text{Pe}} \sum_i \left(\frac{\partial \theta}{\partial x_i} \right)^2, \quad (12)$$

remains constant as $\text{Pe} \rightarrow \infty$. As first deduced by Batchelor and elaborated by others, several regimes can exist where Re or Pe can play some role in the dynamics [3, 4, 10, 17]. However, the scales at which we measure thermal variations in the troposphere are well within the inertial-convective range where the flow properties depend only on χ , ε and the wavelength (or scale) of the quantities being described. In this case, the power spectral density will be proportional to $k^{-5/3}$ and the temperature structure functions,

$$T_p(\mathbf{r}) \equiv \langle [\delta \theta(\mathbf{r})]^p \rangle, \quad (13)$$

are predicted to have similar $p/3$ exponents,

$$T_p(\mathbf{r}) = C |\mathbf{r}|^{p/3}. \quad (14)$$

Of course, the theorems of Kolmogorov and others are based on these existence of a universal equilibrium and an inertial-dissipative range, and so their applicability to real measurements of thermal fluctuations is severely limited by our ability to find regions of fluid flows where these conditions exist.

Sadly, the fact is that neither kinetic energy dissipation nor thermal dissipation has been known to remain constant in the inertial or inertial-dissipative range. Indeed, intermittent bursts, characterized by regions of relatively intense dissipation, play a central role in real turbulent flows. The statistical properties of these bursts are the subject of intensive phenomenological efforts, the most successful of which involve log-normal and log-Poisson

statistics for the turbulent dissipation [8, 26]. As a further complication, the scaling exponents for passive scalar quantities have been observed to be quite different from those for velocity structure functions and theoretically have very different properties [1, 18, 19, 28]. Nonetheless, turbulent flows typically contain large regions over which the mean dissipation is roughly constant. Rather than attempt to model the bursts, we shall use wavelets to identify the localized bursts and segment the data into regions where Kolmogorov's universal equilibrium hypotheses are likely to apply.

Underlying these results and applications is the notion of a mean quantity or ensemble average. Since ensemble averages are not generally available, we rely upon spatial or temporal averages over some fixed interval. Wavelets are especially useful when analyzing average quantities because the wavelet scaling function naturally computes the mean of a function. Of course, wavelets do not automatically select the appropriate scale for which a spatial average will be sufficiently close to the true mean. This scale, called an *integral scale*, is related to velocity and thermal autocorrelations.

3 A track shoe summary of fast Haar wavelet transforms

This section is not intended to provide a tutorial in wavelet analysis. (For this, I would recommend Gerry Kaiser's most amicable book [12].) Rather, I will describe in broad strokes Haar wavelets and why they can help us understand turbulent data.

The basic idea in wavelet analysis is to use a localized function, $\psi(x)$, to create a family of functions or wavelets

$$\psi_{s,t}(x) = \frac{1}{\sqrt{s}} \psi\left(\frac{x-t}{s}\right) \quad (15)$$

where $s > 0$. The wavelet has scale, s , and position, t . Convolving with a wavelet is similar to a windowed Fourier transform except that in this case the window is built into *and scales with* the wavelet. Using this wavelet, one can define a continuous wavelet transform (CWT)

$$\tilde{f}(s,t) = \int_{-\infty}^{\infty} \psi_{s,t}(x) f(x) dx \quad (16)$$

Like the Fourier transform, there is an inversion formula, but we shall leave CWTs at this point and move to discrete Haar transforms.

If ψ has reasonable properties, the continuous $(R^+ \times R)$ basis can be discretized $(Z \times Z)$, permitting multiresolution analysis of the signal. Multiresolution analysis consists of a scaling function $\phi(x)$ and associated wavelet $\psi(x)$. This scaling function is a localized bump. The wavelet can be deduced from the scaling function. Together, the scaling function and the wavelets of infinitely decreasing scale can form a family of nested, orthogonal bases.

The scaling function for Haar wavelets is simply the indicator function:

$$\phi(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

Notice that a convolution with ϕ is a moving average. Assuming that a turbulent flow is ergodic, the Haar wavelet average directly corresponds to an ensemble average if the scale is chosen properly.

Detail information is obtained from the Haar wavelet which is a linear combination of translates of the compressed Haar scaling function:

$$\psi(x) = \phi(2x) - \phi(2x - 1) \quad (18)$$

Together, the Haar scaling function and Haar wavelets form a discrete orthonormal basis

$$\psi_{mn} = \left(\frac{1}{\sqrt{2}}\right)^m \psi\left(\frac{x-n}{2^m}\right) \quad (19)$$

with which one can decompose a signal into local scale components.

$$a_m(n) = \int_{-\infty}^{\infty} \phi_{0n}(x) f(x) dx \quad (20)$$

$$d_m(n) = \int_{-\infty}^{\infty} \psi_{mn}(x) f(x) dx \quad (21)$$

Then

$$f(x) = \sum_{n=-\infty}^{\infty} a_0(n) \phi_{0n}(x) + \sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} d_m(n) \psi_{mn}(x) \quad (22)$$

Furthermore, there is a direct correspondence between Fourier spectra and wavelet scale spectra. Since the wavelets are mutually orthogonal,

$$\int_{-\infty}^{\infty} [f(x)]^2 dx = \int_{-\infty}^{\infty} \left[\sum_{n=-\infty}^{\infty} a_0(n) \phi_{0n}(x) + \sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} d_m(n) \psi_{mn}(x) \right]^2 dx \quad (23)$$

$$= \sum_{n=-\infty}^{\infty} a_0(n)^2 + \sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} d_m(n)^2. \quad (24)$$

The energy at scale, $m > 0$ (wavelength 2^{-m}) is proportional to $\langle d_m(n)^2 \rangle$, where the ensemble average is taken over the n index [13]. Thus, when I calculate spectra, I typically use wavelet scale spectra. Further details is contained in a paper I am preparing with Gerry Kaiser and Don Washburn [25].

4 A simple segmentation procedure

This segmentation of the thermal data is based upon finding regions where Kolmogorov's hypotheses are valid because the thermal dissipation is constant in the sense that averages of the dissipation over integral scales remain constant. This approach is based on the following principal assumptions.

Assumption 1 *The temperature field being measured varies along streamlines.*

Trivially, if there are no temperature variations, no mixing can occur. Furthermore, if the temperature field is constant along streamlines for all time (though it is difficult to imagine this outside of a laminar regime) mixing is restricted to diffusive transport and advection will play no role. However, if there are thermal fluctuations along streamlines, advective and diffusion mixing can occur. For a turbulent flow, this assumption is hardly restrictive.

Assumption 2 A sudden increase in kinetic energy dissipation, $\langle \varepsilon \rangle$, will correspond to a similar increase in thermal dissipation, $\langle \chi \rangle$.

This assumption is only stated in one direction because chaotic mixing is possible without an underlying turbulence flow. However, localized bursts of kinetic energy dissipation directly correspond to intense velocity gradients. These gradients generate the strong shears required for advective mixing in any laminar or turbulent flow. Of course, determining these quantities requires an ensemble or spatial average.

5 Aside: integral scales

To come close to calculating a true ensemble average, $\langle \theta \rangle$, we must resort to a temporal average, or a spatial average relying upon Taylor's hypothesis. Without delving further into issues of ergodicity, we shall assume that we can use spatial averages, and focus instead upon the *integral scale* or the scale over which averages converge to the true mean. Knowing the correct scale can determine which wavelet scaling functions will be meaningful when measuring spectra or structure functions.

If we define

$$\langle \theta \rangle_L = \frac{1}{L} \int_{x_0}^{x_0+L} \theta(s) ds, \quad (25)$$

where it is understood that the average $\langle \theta \rangle_L$ can be displaced an arbitrary distance x_0 , the quantity to calculate is the error

$$\epsilon(L) = \langle (\langle \theta \rangle_L - \langle \theta \rangle)^2 \rangle. \quad (26)$$

In this context, $\langle \theta \rangle$ can be defined to be $\lim_{L \rightarrow \infty} \langle \theta \rangle_L$. Without loss of generality, we shall treat $x_0 = 0$ from this point onward. If we designate the zero-mean portion of the function to be

$$\tilde{\theta}(x) = \theta(x) - \langle \theta \rangle, \quad (27)$$

then

$$\epsilon(L) = \langle \langle \tilde{\theta} \rangle_L^2 \rangle \quad (28)$$

$$= \left\langle \frac{1}{L^2} \int_0^L \int_0^L \tilde{\theta}(s_1) \tilde{\theta}(s_2) ds_1 ds_2 \right\rangle \quad (29)$$

$$= \frac{1}{L^2} \int_0^L \int_0^L \langle \tilde{\theta}(s_1) \tilde{\theta}(s_2) \rangle ds_1 ds_2. \quad (30)$$

If the function θ is stationary over $(s_1, s_2) \in [0, L] \times [0, L]$, then

$$\langle \tilde{\theta}(s_1) \tilde{\theta}(s_2) \rangle = \rho(s_1 - s_2) \langle \tilde{\theta}^2 \rangle \quad (31)$$

where $\rho(\Delta)$ is the autocorrelation coefficient and Δ is the separation between the points being correlated. Substituting the autocorrelation coefficient into the previous expression,

changing variables to $\Delta = (s_1 - s_2)/2$ and $D = (s_1 + s_2)/2$, and using the fact that $\rho(\Delta)$ is even, one obtains

$$\epsilon(L) = \frac{\langle \tilde{\theta}^2 \rangle}{L^2} \int_0^L \int_0^L \rho(s_1 - s_2) ds_1 ds_2 \quad (32)$$

$$= \frac{\langle \tilde{\theta}^2 \rangle}{L^2} \int_{-L/2}^{L/2} \int_{-(L-|\Delta|)}^{L-|\Delta|} \rho(\Delta) dD d\Delta \quad (33)$$

$$= \frac{2\langle \tilde{\theta}^2 \rangle}{L} \int_0^{L/2} \left(1 - \frac{\Delta}{L}\right) \rho(\Delta) d\Delta. \quad (34)$$

For large L , the contribution by the $|\Delta|\rho(\Delta)/L$ term is small because $\rho(\Delta)$ decays to zero independent of L . What remains is

$$\epsilon(L) = \frac{2\langle \tilde{\theta}^2 \rangle}{L} \int_0^L \rho(\Delta) d\Delta \approx \frac{2\langle \tilde{\theta}^2 \rangle}{L} \int_0^\infty \rho(\Delta) d\Delta. \quad (35)$$

The integral $\int_0^\infty \rho(\Delta) d\Delta$ has units of length (or time for temporal averages) and is commonly referred to as the *integral scale*, $\mathcal{L}$. Hence, the error in one's estimate of $\langle \theta \rangle$ is approximately

$$\epsilon(L) \approx 2\langle \tilde{\theta}^2 \rangle \frac{\mathcal{L}}{L}, \quad (36)$$

and one can expect $\langle \theta \rangle_L$ to be a good approximation if one averages over several integral scales.

Of course, a given integral scale is only valid in one stationary domain. Furthermore, the integral scale varies with the function one is measuring. For instance, rather than measure the mean temperature, $\langle \theta \rangle_L$, one might be interested in approximating structure functions,

$$T_p(r) \approx \langle [\delta\theta(\mathbf{r})]^p \rangle_L. \quad (37)$$

Also, even if one can calculate the integral scale in terms of correlations, it is not clear how many integral scales one must use for one's calculations. However, we shall see in the next subsection that these scales can also be determined empirically directly from the data.

6 Treatment and limitations of the data

The high-resolution data from the airborne laboratory is filtered electronically into low-pass and high-pass components as it is collected. Since we are only interested in temperature fluctuations away from a mean quantities, we only study the high-pass component. Various data sets are gathered at different altitudes, airspeeds and flight paths. In order to compare different data sets, all time variables are converted into spatial variables using the jets true airspeed. Thus, from one data set to another, the sampling intervals may vary though this variation is slight. As a rule of thumb, the instruments are capable to measuring the temperature every 2 centimeters. However, the cold-wire is much slower than this, and one would only expect to resolve features larger than 20 cm without more sophisticated treatment of the data. Also, when atmosphere turbulence intensities are low, the turbulence generated

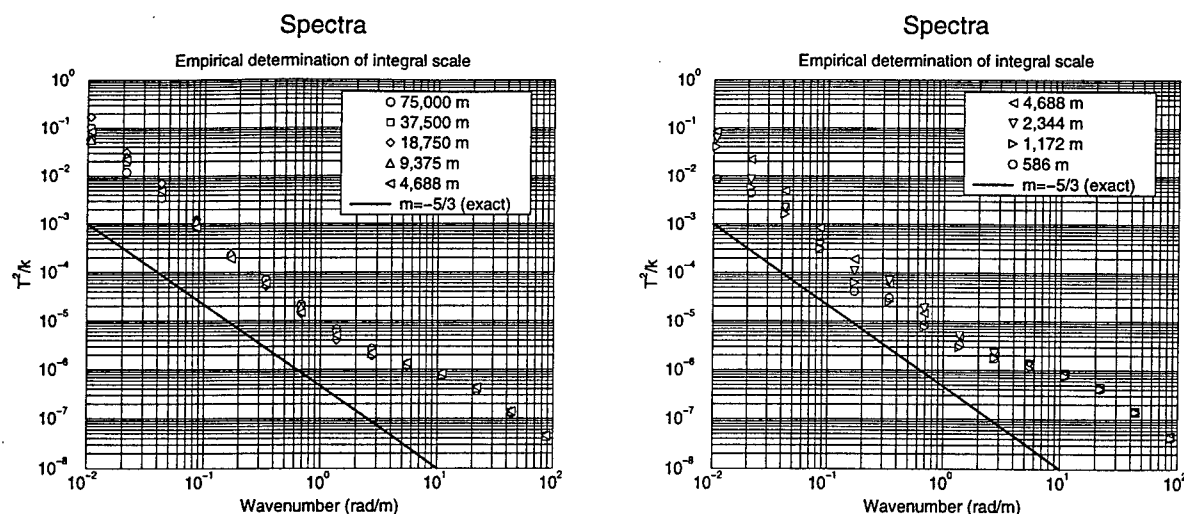


Figure 1: Empirical determination of integral scale by calculating the spectra over varying intervals. By varying the interval over which scale spectra are calculated, it is clear to see what lengths are required for good frequency resolution. These calculations are made over a long segment of near constant dissipation.

by the aircraft also contaminates the data at wavelengths corresponding to about 6 meters. Since the high pass filter rolls off at about 60 meters, we consider our data to be valid for roughly one decade between 6 and 60 meters in low turbulence. If the turbulence is strong, we can expect to resolve inertial ranges from 60 meters down to about 20 cm where probe effects become important.

Since the aircraft only has one probe and can only fly in one direction at the same time, we cannot calculate all three components of the thermal dissipation, but we can calculate $\left(\frac{d\theta}{dx}\right)^2$ using centered differences. If the flow is locally isotropic within the resolution of our measurements (less than one meter!) and we average over a small distance, we rest assured that this quantity is proportional to the thermal dissipation. George Papanicolaou and his group have calculated $\mathcal{L}$ for the 96223 data and found the integral scale to be roughly 2 to 4 km, in agreement with my empirical calculations based on spectral convergence and convergence of the structure functions. In Fig. 1, one sees that the spectrum converges over scales of between 2 and 5 km. Similarly, in Fig. 2, the 12th order structure function converges with similar averaging intervals.

7 Results of segmentation

By examining the wavelet averages of the thermal dissipation data, we find substantial advantages to segmenting the data in regions of uniform dissipation. If one measures spectra and structure functions without regard to dissipation, one quickly finds that the spectra and structure functions do not obey K41 power laws. However, calculating the same functions along regions that have constant dissipation, one recovers Kolmogorov scalings. Furthermore, one can calculate spectra over many disconnected subsets of the signal in instances where the size a single region of constant dissipation is small relative to the integral scale.

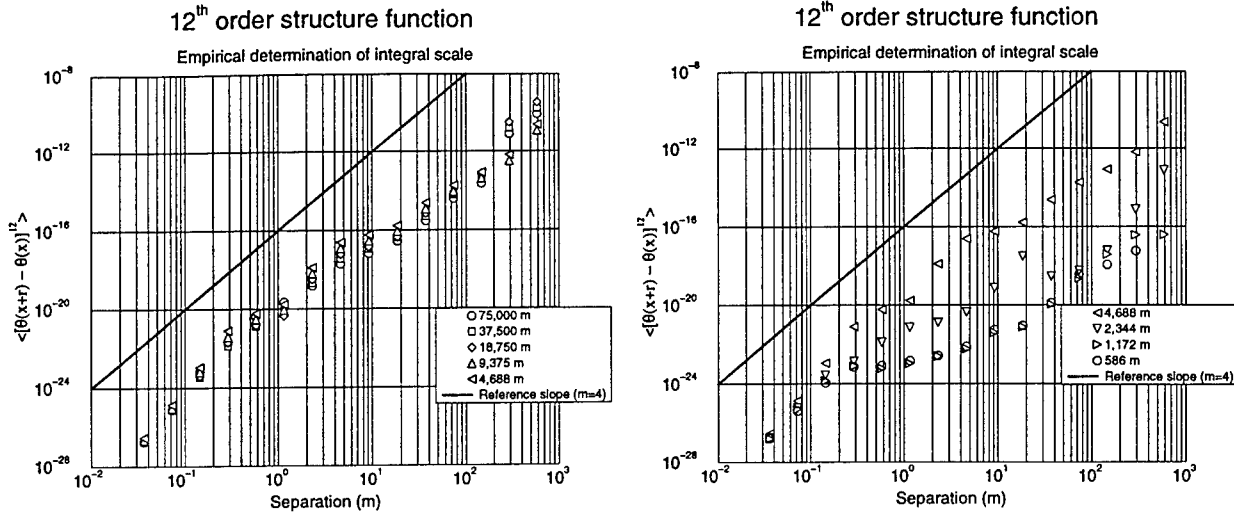


Figure 2: Empirical determination of integral scale by calculating the structure functions over varying intervals. Once again, the integral scale for turbulent mixing can be calculated empirically by varying the interval over which structure functions are averaged. These calculations are made over a long segment of near constant dissipation.

A second and unforeseen advantage to wavelet analysis of the thermal dissipation is the discover of regular patterns buried within the thermal measurements. It should be added at this point that we attempted wavelet analysis of the thermal data as a supplement to an exhaustive Fourier analysis of this data at Phillips Laboratory. Whether or not the source of these patterns are natural or artificial is still under active investigation, but the fact remains that their existence was invisible to standard Fourier analysis.

8 Segmentation based upon dissipation

As described earlier, a signal (dissipation) can be represented as the sum of a Haar wavelet average plus an infinite sum of Haar wavelet differences. One can choose the scale of the averages arbitrarily, and this determined the coarsest scale of the wavelet differences. In the case of thermal dissipation in a turbulent flow, the Haar averages are physically relevant because

$$a_m = \langle \chi \rangle_{2^m}. \quad (38)$$

Thus, we examine wavelet averages primarily even though the orthogonal decomposition of the signal involves the wavelet differences. Often supporting information can also be gathered from the wavelet differences.

Upon viewing wavelet averages and differences of the thermal dissipation in the troposphere, one is immediately impressed with the fact that the dissipation is not constant even at coarse resolutions (see Fig. 3). However, within the 80 kilometer segment, there are some regions wherein the dissipation does appear to be roughly constant. For instance, the first 45 kilometers of the thermal dissipation shown in Fig. 3 exhibits a thermal dissipation somewhere between 4×10^{-3} and 5.1×10^{-3} . Afterward, there are localized regions where the dissipation surges several times higher depending upon the scale over which one is averaging.

Dissipation

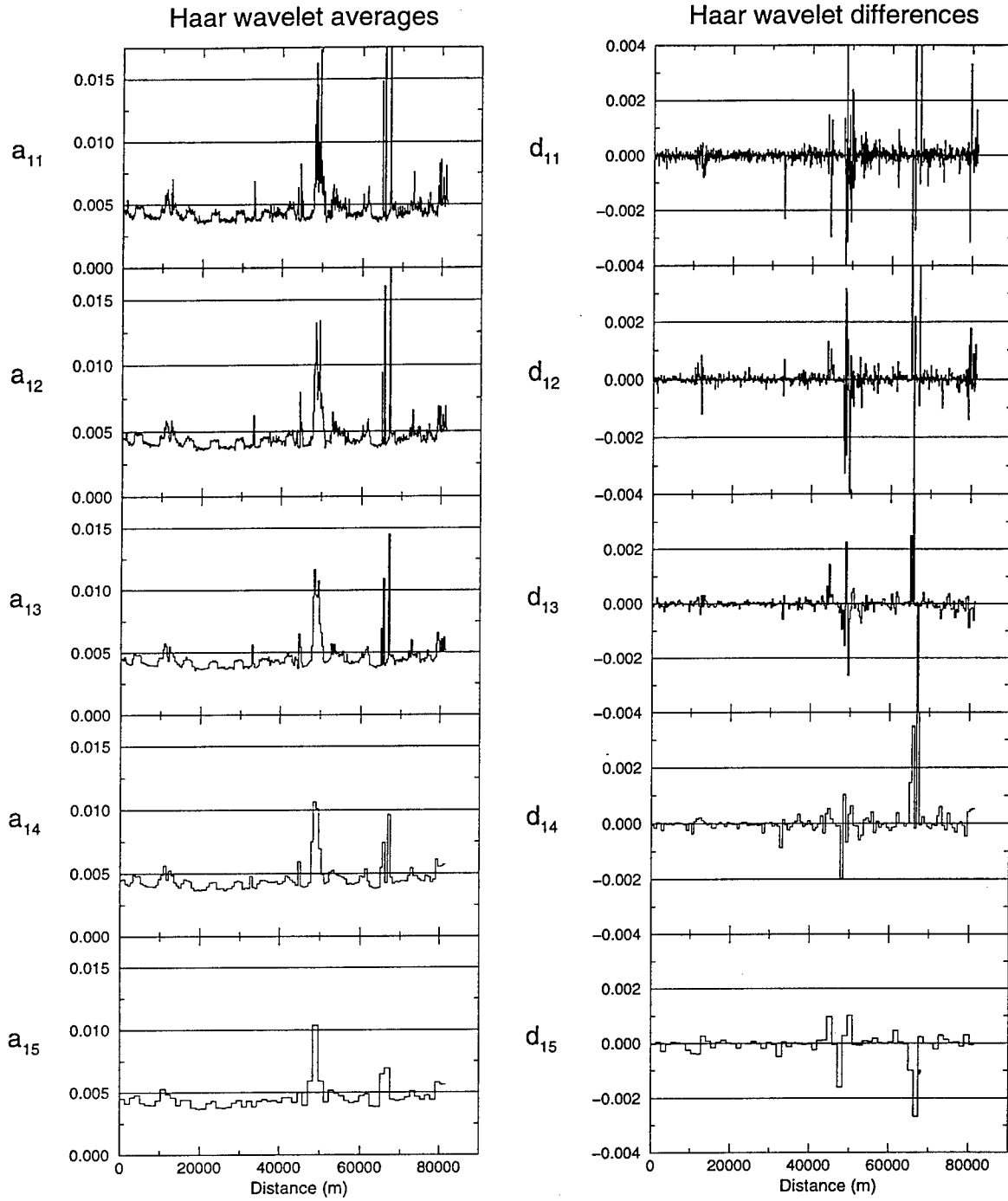


Figure 3: Dilated Haar wavelet averages and differences of $\left(\frac{d\theta}{dx}\right)^2 \propto \chi$ for one set of thermal measurements in the troposphere. Haar wavelet averages can be thought of as running averages of the signal, and the differences can be expressed as differences of averages at the previous (finer) scale.

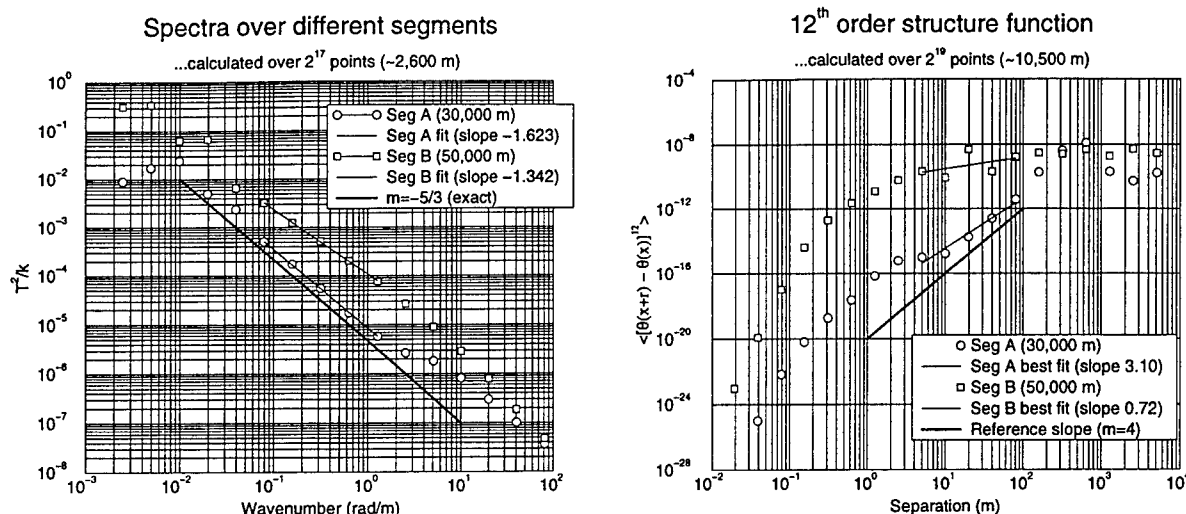


Figure 4: Spectrum (left) and 12th-order structure function (right) at different positions using the signal displayed in Fig. 3. In Segment A (displaced 30,000 m), the dissipation is roughly constant while in Segment B (displaced 50,000 m), the dissipation is changing abruptly. A best fit to the data over the range of values not affected by the low-pass roll-off and the aircraft induced turbulence shows that Kolmogorov spectra is realized in Segment A while it is not in Segment B. The high order structure function on the right shows an even sharper difference between the two segments. Larger windows are required for the high order structure functions because they have longer integral scales.

Another interesting feature of thermal dissipation in the troposphere is that the brief regions of elevated thermal dissipation have substantial structure at small scales while the regions of constant dissipation have relatively little structure aside from expected small scale variations as $2^m \Delta t$ drops below the integral associated with averaging thermal dissipation. (Though 2 kilometers would hardly qualify as “brief” in many atmospheric contexts, we use the term here relative to the vast regions of constant dissipation.) We can segment the data using the wavelet averages of dissipation to tell us where Kolmogorov universality equilibrium hypotheses are valid. Not surprisingly, if one calculates the power spectral density in a region of constant dissipation, one recovers Kolmogorov spectrum. Moreover, one can also recover very high order structure functions in these regions (see Fig. 4). However, segmenting the data near the localized bursts can yield very non-Kolmogorov behavior.

This behavior is explained, not in terms of intermittent bursts, but in terms of stationarity. In this case, the elevated regions of thermal dissipation can occupy several kilometers of space. While they are localized, these regions have a substantial scale. This data is entirely consistent with radar observations of large flat regions in the atmosphere co-existing at the same level. In other words, one explanation for the data is that the aircraft passed through a 2 kilometer region having a mean thermal dissipation different from the surrounding atmosphere. If this were the case, we should be able to recover Kolmogorov statistics when we calculate spectra entirely within the burst. Indeed, as seen in Fig. 5, this is precisely the case. Anomalous behavior in spectra can be explained as a lack of frequency resolution for the dominant part of the signal. For this window, one would expect the spectrum to be

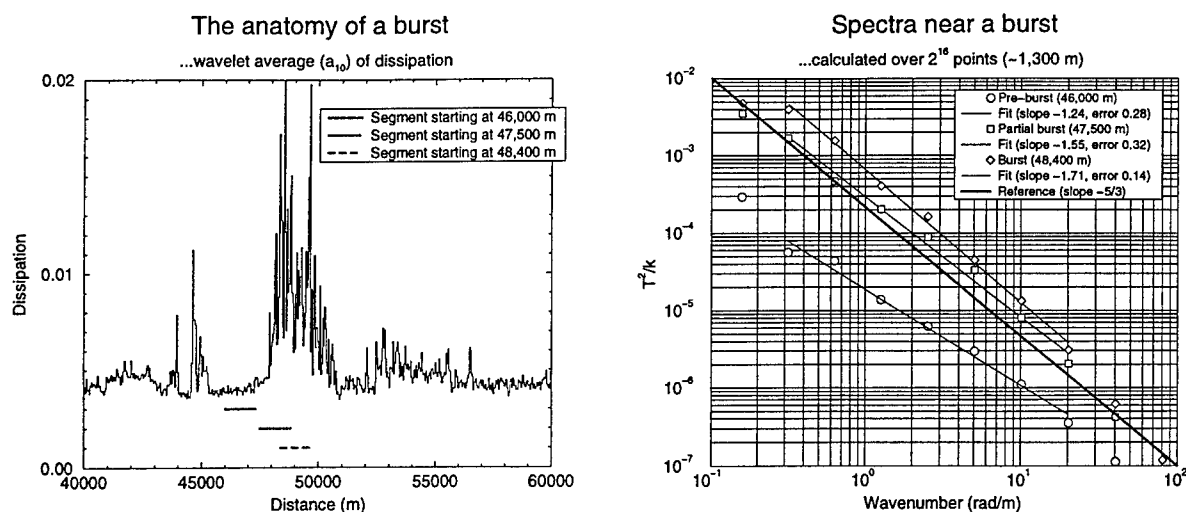


Figure 5: Spectra calculated near a burst. On the left, one can see three segments near a sharp increase in dissipation. The first is in a region of relatively calm dissipation. The last is entirely within the region of high dissipation. The middle segment straddles the two. On the right, we calculate the spectra over these intervals. Notice that the spectrum does not obey a power law in the transition region. Though a best fit is provided, one can see considerable variability in the slope over the region of interest.

the sum of two distinct $Ck^{-5/3}$ spectra. The high dissipation portion would dominate the sum because C is directly related to χ . However, one can also see in Fig. 3 that the high dissipation portions of the signal has relatively small measure, meaning that there is very low frequency resolution. Therefore, one would expect the spectrum and structure functions to have a fair amount of noise in them.

9 Observation of embedded waves

In addition to providing information about thermal dissipation, the wavelet analysis of the thermal dissipation uncovered an unanticipated feature in the data that may be physical or may be instrumental. There are distinct asymmetric square waves in the dissipation.

Re-examining Fig. 3, one will notice a regular pattern in the calm region extending from 13 km up to 30 km, and it would appear that there is a square wave buried in the turbulence signal. This apparent periodicity in $\left(\frac{d\theta}{dx}\right)^2$ means that the roughness of $\theta(x)$ has a regular spatial structure though it says nothing about the magnitudes of the temperature fluctuations themselves. These fluctuations are very minor accounting for less than 1% of the total energy in the data, yet they persist in every broad region of constant dissipation through all the data collected on this particular day (see Fig. 6). There are two reasons why features like this do not emerge with the traditional approaches currently used at Phillips Laboratories. First, they have not been studying dissipation because they are basing their analysis solely on phenomenological reasoning rather than the dynamics of the mixing. I studying fundamental problems behind ABL, I decided to look at the dissipation and found something interesting. The second reason why it would not show up with the standard

Observed waves in thermal dissipation

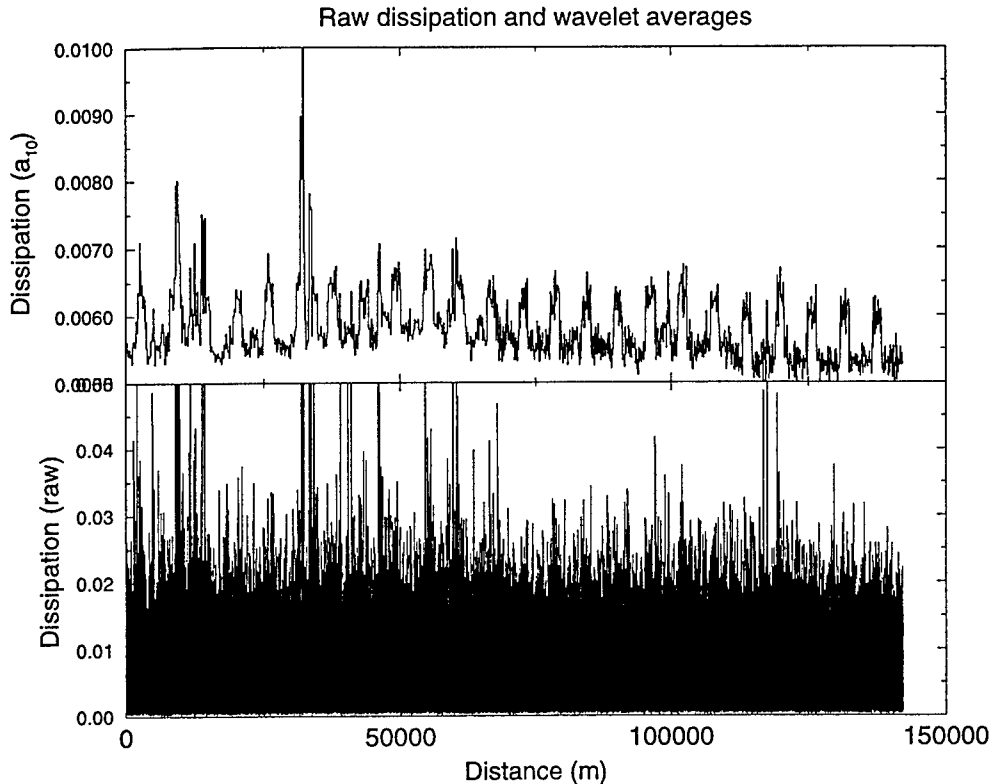


Figure 6: Observed periodicity in thermal dissipation. The lower graph is the raw data, and the upper graph is the wavelet average (running average over 2^{10} points) of the same signal. A suspicious pattern emerges which we are still trying to explain. Aspects like this in the signal are not detected with standard processing techniques. Note: The thickness in the latter half of the wavelet series is an artifact of the plotting package and not the signal itself.

reductions used at Phillips is that they are looking for features in the Fourier domain. While these waves have frequency content, the full signal also contains bursts and large deviations from these waves that would inhibit any spike that appears. Finally, this pattern occurs at wavelength much greater than the frequency cutoffs of the data acquisition circuitry, so naturally, one would not look to find signals at these wavelengths.

Though we lack a full explanation for the periodicity, we have studied it in some detail. The waves are asymmetric in that the peaks are short than the troughs. The spectrum of θ does not vary in and around the peaks and troughs as one might expect. The wavelengths vary with altitude but not direction of flight. (If the waves were physical and moved at a speed different from the mean wind velocity such as leeward waves near mountain ranges, we would expect to observe a Doppler shift.) The variations in wavelength with altitude have eliminated systematic spatial or temporal fluctuations in the aircraft or data acquisition software and hardware. The fact remains that though we cannot explain them (yet), the waves emerged naturally from routine wavelet analysis while this is certainly not the case with Fourier analysis of the same data.

10 Future directions

There are two avenues of exploration on which I can provide immediate support to the aerothermal program, treatment of the raw data and analytic techniques for analyzing the data. Some of the problems with the data collection can be treated post-experimentally and some require instrumental changes. My expertise can only address the first issue directly, but I feel that I can help assess the impact of more costly instrumental changes. By understanding when and where instrumental errors contaminate the data, I can help assess whether or not we really need to correct them. For instance, the "noise floor" only appears in relatively low intensity turbulence, so perhaps it is not important to correct a problem that only exists outside of the regime of interest. Nonetheless, we need to confirm that the noise floor is not concealing information that is of interest in the long term.

Presently, everyone involved with the project treats voltages from the cold-wires as being linearly proportional to the temperature. Knowing from the ABLE ACE report that the probes' response is somewhere near 50 cm (400 Hz), it is clear that the data is substantially oversampled. The "noise floor," which most people involved in the project believe is caused by aircraft turbulence, further corrupts the measurements up to 6 meters in areas of weak turbulence. Finally, no one has investigated the role of dynamic heating in these measurements.

The current program of calculating C_T^2 at Phillips Laboratories Kirtland is based on uniform, systematic analysis of the aerothermal data. This summer, I have shown that wavelet analysis of thermal dissipation can reveal some aspects of the underlying turbulent flow which in turn, provide a degree of confidence to the assumption that a power law exists. The fact remains that power law behavior might not exist everywhere. The data is nonstationary and often subject to strong intermittent bursts. Understanding the properties and structures that are physically relevant in the atmosphere may lead to better techniques for analyzing the raw data.

11 Conclusions

This summer has been a very productive time for me thanks to the Faculty Summer Research Program. In a relatively short time, I have refamiliarized myself with some of the important aspects of turbulent mixing, and at the same time, learned a lot about wavelets and signal processing. During the entire summer, I worked hard at keeping this work focused upon the immediate needs of the ABL program by pursuing problems that my focal point thought were important. As a result, my collaborators and I have developed a new technique for determining stationary regions in turbulent flows and calculating turbulent statistics that can be used for comparisons and validation. Now, I would like to continue this research into the upcoming year, and I hope the AFOSR will continue to support this effort through their Summer Research Extension Program.

12 Acknowledgements

The author would like to thank the AFOSR for their summer support and his collaborators, especially Gerry Kaiser, Bob Beland, Don Washburn, George Jumper and many others at Phillips Laboratories, Hancorn and Kirtland.

References

- [1] R. A. Antonia, E. J. Hopfinger, Y. Gagne, and F. Anselmet. Temperature structure functions in turbulent shear flows. *Phys. Rev. A*, 30(5):2704–2707, 1984.
- [2] H. Aref. Integrable, chaotic and turbulent vortex motion in two-dimensional flows. In *Ann. Rev. Fluid Mech.*, volume 15, pages 345–389. Techbooks, Fairfax, VA, 1983.
- [3] G. K. Batchelor. Small-scale variation of convected quantities like temperature in turbulent fluid. part 1. general discussion and the case of small conductivity. *J. Fluid Mech.*, 5:113–33, 1958.
- [4] G. K. Batchelor, I. D. Howells, and A. A. Townsend. Small-scale variation of convected quantities like temperature in turbulent fluid. part 2. the case of large conductivity. *J. Fluid Mech.*, 5:134–9, 1958.
- [5] O. N. Boratov. On recent intermittency models of turbulence. *Phys. Fluids*, 9(5):1206–8, 1997.
- [6] S. Corrsin. On the spectrum of isotropic temperature fluctuations in an isotropic turbulence. *J. App. Phys.*, 22(4):469–73, 1951.
- [7] U. Frisch. *Turbulence: The Legacy of A.N. Kolmogorov*. Cambridge University Press, 1995.
- [8] U. Frisch, P.-L. Sulem, and M. Nelkin. A simple dynamical model of intermittent fully developed turbulence. *J. Fluid Mech.*, 87:719–736, 1978.
- [9] C. H. Gibson and W. H. Schwarz. The universal equilibrium spectra of turbulent velocity and scalar fields. *J. Fluid Mech.*, 16:368–86, 1963.
- [10] R. J. Hill. Models of the scalar spectrum for turbulent advection. *J. Fluid Mech.*, 88(3):541–562, 1978.
- [11] J. O. Hinze. *Turbulence*. McGraw-Hill, second edition edition, 1975.
- [12] G. Kaiser. *A Friendly Guide to Wavelets*. Birkhauser, Boston, 1994.
- [13] G. G. Katul, J. D. Albertson, C. R. Chu, and M. B. Parlange. Intermittency in atmospheric surface layer turbulence: The orthonormal wavelet representation. In E. Foufoula-Georgiou and P. Kumar, editors, *Wavelets in Geophysics*. Academic Press, Inc., 1994.

- [14] A. N. Kolmogorov. Dissipation of energy in the locally isotropic turbulence. *Dokl. Akad. Nauk SSSR*, 32(1), 1941. Reprinted in *Proc. Roy. Soc. London A* (1991) 434, 15-17. Translated by V. Levin.
- [15] A. N. Kolmogorov. The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers. *Dokl. Akad. Nauk SSSR*, 30(4), 1941. Reprinted in *Proc. Roy. Soc. London A* (1991) 434, 9-13. Translated by V. Levin.
- [16] A. N. Kolmogorov. A refinement of the previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high reynolds numbers. *J. Fluid Mech.*, 13:82-85, 1962.
- [17] R. H. Kraichnan. Small-scale structure of a scalar field convected by turbulence. *Phys. Fluids*, 11(5):945-53, 1968.
- [18] R. H. Kraichnan. Anomalous scaling of a randomly advected passive scalar. *Phys. Rev. Lett.*, 72(7):1016-1019, Feb 1994.
- [19] R. H. Kraichnan, V. Yakhot, and S. Chen. Scaling relations for a randomly advected passive scalar field. *Phys. Rev. Lett.*, 75(2):240-243, Jul 1995.
- [20] E. Kuznetsov, A. C. Newell, and V. E. Zakharov. Intermittency and turbulence. *Phys. Rev. Lett.*, 67(23):3243-3246, Dec 1991.
- [21] A. S. Monin and A. M. Yaglom. *Statistical Fluid Mechanics*, volume 1. The MIT Press, 1965.
- [22] A. S. Monin and A. M. Yaglom. *Statistical Fluid Mechanics*, volume 2. The MIT Press, 1975.
- [23] A. M. Oboukhov. Structure of the temperature field in turbulent flow. *Izv. Akad. Nauk SSSR, Ser. Geogr. i Geofiz.*, 13:58-69, 1949.
- [24] G. Parisi and U. Frisch. Fully developed turbulence and intermittency. In M. Ghil, R. Benzi, and G. Parisi, editors, *Turbulence and predictability in geophysical fluid dynamics and climate dynamics, Proc. Int. School of Physics "E. Fermi"*, pages 84-88, Amsterdam, 1983. North-Holland.
- [25] L. F. Rossi, G. Kaiser, and D. Washburn. Recovery of kolmogorov statistics in thermal mixing in the troposphere: The hazards of real data. *J. Math. Phys.*, In preparation.
- [26] Z.-S. She and E. Leveque. Universal scaling laws in fully developed turbulence. *Phys. Rev. Lett.*, 72(3):336-339, Jan 1994.
- [27] Z.-S. She and E. C. Waymire. Quantized energy cascade and log-Poisson statistics in fully developed turbulence. *Phys. Rev. Lett.*, 74(2):262-265, Jan 1995.
- [28] K. R. Sreenivasan and R. A. Antonia. The phenomenology of small-scale turbulence. In J. L. Lumley, M. V. Dyke, and H. L. Reed, editors, *Annual Review of Fluid Mechanics*, volume 29, pages 435-472. Techbooks, Fairfax, VA, 1997.

[29] H. Tennekes and J. L. Lumley. *A first course in turbulence*. The MIT Press, 1990.

**ATMOPHERIC EFFECTS UPON SUB-ORBITAL BOOST GLIDE
SPACEPLANE TRAJECTORIES**

David P. Stapleton
Associate Professor
Department of Mathematical Statistics

University of Central Oklahoma
100 N. University Drive
Edmond, OK 73034

Final Report for:
Summer Research Program
Phillips Laboratory

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

And

Phillips Laboratory

September 1997

ATMOSPHERIC EFFECTS UPON SUB-ORBITAL BOOST GLIDE SPACEPLANE TRAJECTORIES

David P. Stapleton
Assistant Professor
Department of Mathematics and Statistics
University of Central Oklahoma

Abstract

The Reusable Aero Space Vehicle (RASV) concept is considered in order to analyze the effects of various atmospheric conditions upon a mission to release a payload near to a polar, low earth orbit. After payload release, the spaceplane is intended to circumnavigate the globe and return to launch site by skipping off the atmosphere, while the payload nominally includes a small upper stage that boosts a satellite into a final orbit. In order to evaluate atmospheric effects, Monte Carlo atmospheric temperature, pressure, density, and winds (based on time of the year, recent solar activity, and vehicle geographic position and altitude) are applied to a baseline RASV mission. Results indicate that these around-the-world RASV trajectories are feasible under a broad range of atmospheric conditions. However, for a given set of flight controls, it is found that maximum vehicle altitude, aerodynamic heat rate, and skip amplitude, vary substantially with atmospheric perturbations.

ATMOSPHERIC EFFECTS UPON SUB-ORBITAL BOOST GLIDE SPACEPLANE TRAJECTORIES

David P. Stapleton

Introduction

It is desired to simulate the effects of varying atmospheric conditions (including winds) upon the performance of a spaceplane that attempts to fly an around-the-world mission from within the continental United States, en route releasing a payload into prescribed near-orbit conditions. In order to reduce the influence of the earth's rotation and to more accurately assess the influence of latitudinal climate differences, trajectories that approximately cross the earth's poles are considered.

Flight simulations have been generated by combining existing software to model trajectories with software to simulate random atmospheres. Specifically, the Program to Optimize Simulated Trajectories (POST Version 5.1 - see Brauer, 1990), developed by Martin Marietta Corporation, has been combined with the NASA/MSFC Global Reference Atmospheric Model - 1995 Version (GRAM-95, discussed by Justus et al., 1995). Both programs have been slightly modified to reduce execution time and so that GRAM-95 can be called as a subroutine by the POST program. The specifications of a sub-orbital version of a Reusable Aero Space Vehicle (RASV) as envisioned by Boeing Corp. (see also Froning, 1996, and Johnson, 1992) prescribe the spaceplane model parameters for POST.

The Atmosphere Model

The GRAM-95 computer code was chosen to provide atmospheric data because it is a well tested program that can provide geographically and time varying, correlated, random values of necessary parameters. It computes data in the lower (0-27 km), middle (20-120 km), and upper (90+ km) altitudes via three different models and obtains results at transition altitudes by "fairing" estimates

of the individual models. Temperature, density, pressure, speed of sound, and wind components are the quantities subsequently used in (POST) trajectory calculations.

The middle atmosphere model in GRAM-95 is based on data from the Middle Atmosphere Program (MAP - see Barnett and Corney, 1985a, and 1985b) and provides the most critical data for skipping trajectories to POST from the vehicle altitude and latitude, the elapsed time, and the month of the year. Diurnal effects are not modeled; nor are solar effects, nor gravity waves.

The upper and lower portions of the atmosphere have a smaller influence upon spaceplane performance since little time is spent at these altitudes. It is useful to understand, however, that (i) in the lower atmosphere the Global Upper Air Climatic Atlas (GUACA Version 1.0 - see Ruth, et al., 1993), provides historical data from which random parameters are obtained, and (ii) in the upper atmosphere model, or Jacchia model (see Hickey, 1988a, and 1988b), influences of the day of the month, diurnal variations, and solar effects (as well as altitude, latitude, time elapsed, and month) are taken into account.

The Baseline Trajectory Model (Using the U.S. Standard Atmosphere, 1962)

A baseline trajectory was generated for a flight of the RASV from Edwards Air Force Base, California, around the world and back, by two POST runs using a polar trajectory headed initially due south, through a geographically uniform U.S. Standard Atmosphere, 1962 (with no winds).

In the first run of the baseline case POST was requested to maximize the weight of a payload subject to prescribed conditions at the end of the powered phase of ascent:

- (i) inertial flight path angle must be within 0.001° of 0.0° ,

- (ii) inertial velocity is required to be within 10 ft/sec of 25,814 ft/sec, and
- (iii) altitude must lie within 100 ft of 340,495 ft (about 104 km).

These conditions occurred just prior to an unpowered elliptical orbit (Hohmann) transfer to an apogee altitude at which the payload could be released. The vehicle assumed a horizontal lift-off position, was sled-launched at an azimuth of 180° (due south) with an initial relative velocity of 600 ft/sec from a runway at 2,276 ft altitude, and was powered by two modified space shuttle main engines (SSMEs) for 522.31 seconds thereafter. The engines were assumed to initially deliver 1.01 million lbs of thrust, have an instantaneous specific impulse (ISP) of 449.8 sec and an exit area of 58 ft². When the RASV reached 50,000 ft, its extendable nozzles were moved to a second position so that 1.041 million lbs of thrust was delivered with an ISP of 463.5 seconds and an exit area of 174 ft². The propellant weight decreased from 1,090,700 lbs as shown in Fig. 1 and the vehicle dry weight was 134,300 lbs. (not including a payload of about 25,000 lbs) for a gross takeoff weight of about 1.25 million lbs.

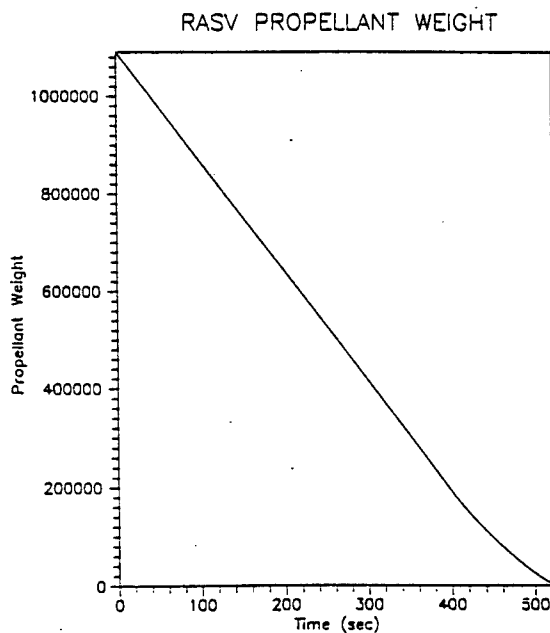


Figure 1. RASV Propellant Weight as a Function of Time

Drag and lift calculations used the coefficients shown in Fig. 2 and Fig. 3 with a reference area of 5670 ft².

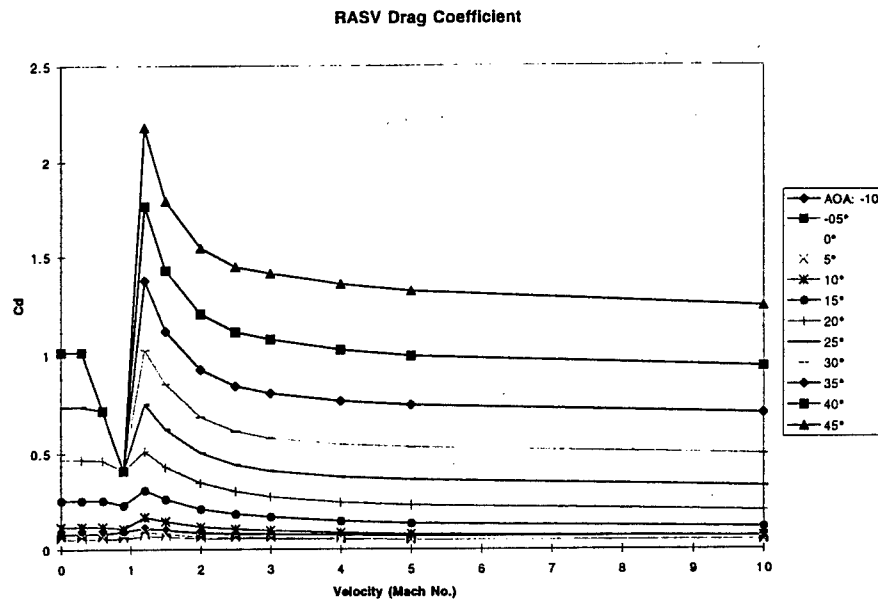


Figure 2. RASV Drag Coefficient (C_D) Verses Mach Number

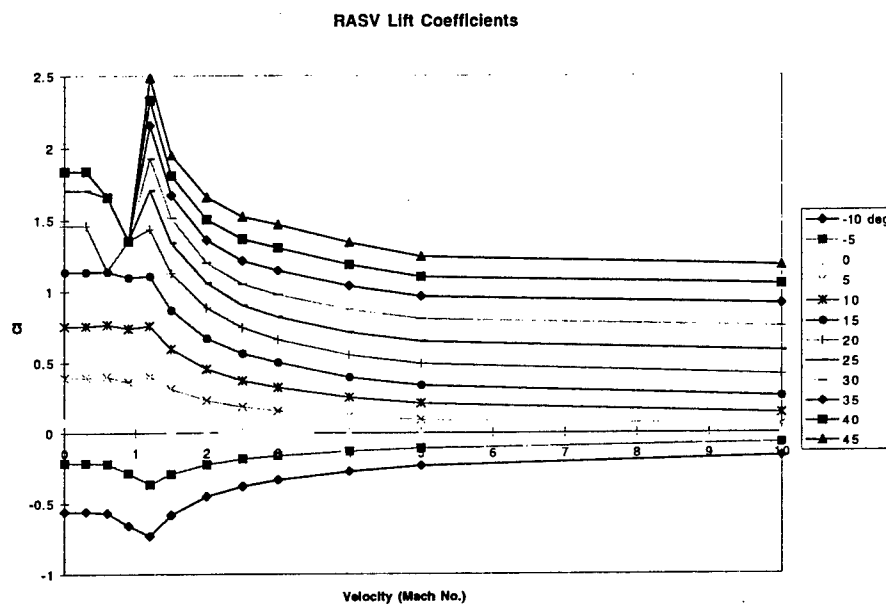


Figure 3. RASV Lift Coefficient (C_L) Verses Mach Number

It was found that a payload of 25,100 lbs could be lifted to the near-orbit conditions if the attack angles shown in Table 1 were employed (with bank angle zero).

Table 1.
Attack and Bank Angles for RASV Ascent

Event	Time (sec)	Attack Angle (deg)
1	0.	3.5685925
2	15.	6.9171659
3	35.	2.4401747
4	75.	1.0975112
5	175.	2.1646243
6	295.	1.684
7	345.	0.81079
8	395.	0.7131
9	445.	-.23185

The attack angles and payload weight were inserted into a second POST computation in which the RASV was required to return to Edwards AFB from the payload release point - but, without releasing its payload (it was assumed that a return with the payload would be more difficult than a return after a release, due to additional weight). The aerodynamic heat rate was constrained to a maximum of 200 BTU/ft²/sec¹. POST was requested to meet these objectives by choosing an angle of attack and a bank angle during each of the phases: (i) from vehicle apogee until 70,000 lbs of lift was generated, and (ii) from this point until the vehicle relative velocity dropped to 5,000 feet per second. From this point the RASV was (somewhat unrealistically) assumed to dive to landing by use of a constant flight path angular rate. A skipping return to Edwards AFB was found to be possible using attack angles of 6.704° and 6.771° with corresponding bank angles of -12.82° and -12.91° for the two phases, respectively. The baseline ground track (subvehicle path) appears in Fig. 4, and the associated skipping profile of altitude verses time is presented in Fig. 5.

¹ Based upon a one foot radius reference sphere.

optimize rasv ascent(polar leo) with heating constraint

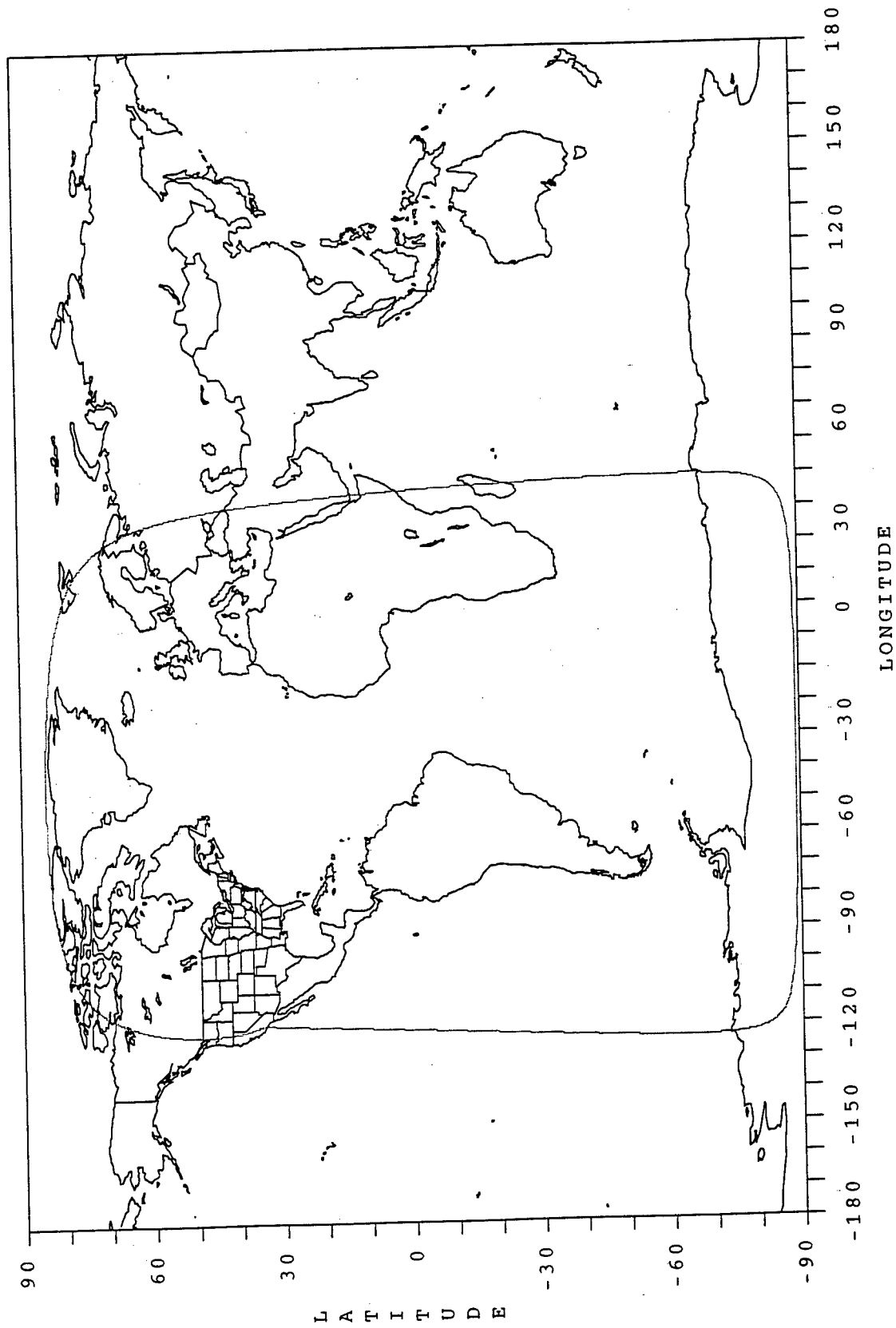


Figure 4. Ground Track for the Baseline RASV Skipping Trajectory

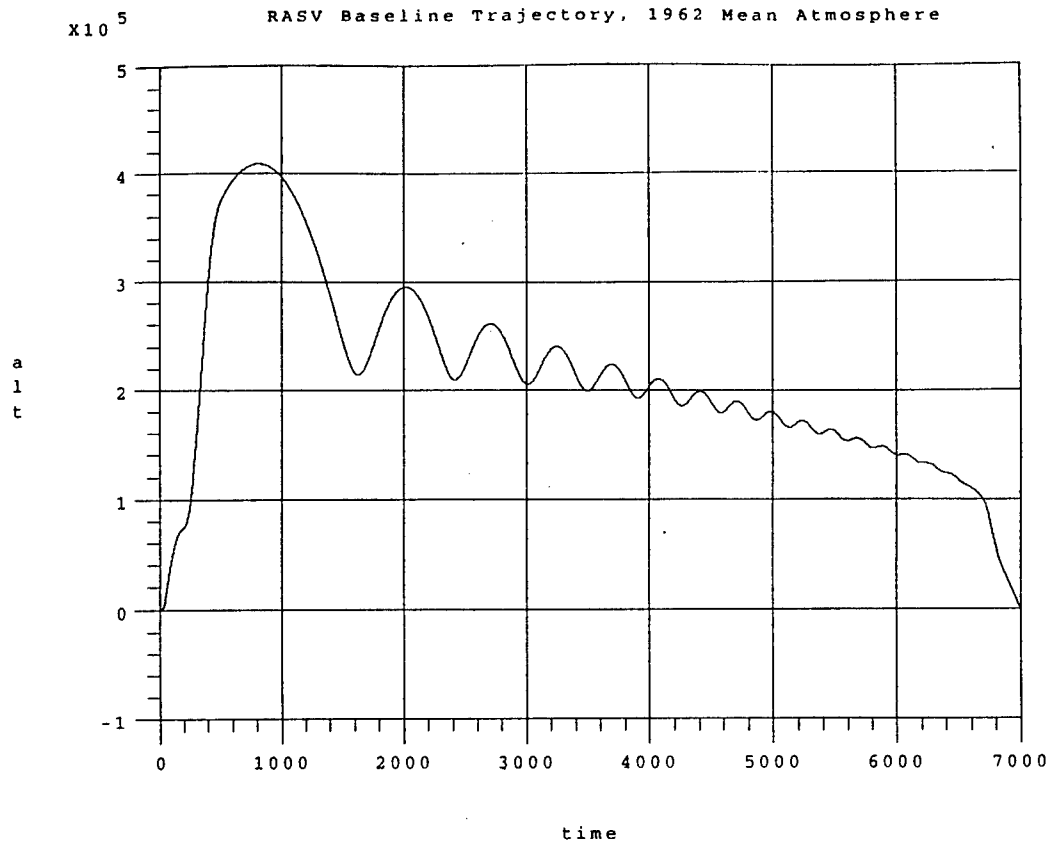


Figure 5. Skipping Profile for the Baseline RASV Trajectory

Monte Carlo Flight Simulations

With the optimized or baseline trajectory for the uniform atmosphere complete, it was possible to execute Monte Carlo POST runs with GRAM-95 generated, globally varying, random atmospheres by implementing the attack angles, bank angles, and payload weight of the baseline trajectory. Twenty such runs for each of January, April, July, and October were performed. Each run was made on a 200 MHz Pentium-class PC, and required approximately 2 minutes using POST's projected gradient optimization mode with zero control vector iterations.

Since GRAM-95 requires that parameters for solar activity be input (only for effects above 90 km, or 295,275 ft) in each run, random solar parameters were generated externally. The 3-hourly geomagnetic index a_p (used in the GRAM model to compute a geomagnetic correlation to the exospheric temperature) was selected from the percent frequency distribution shown in Fig. 6.

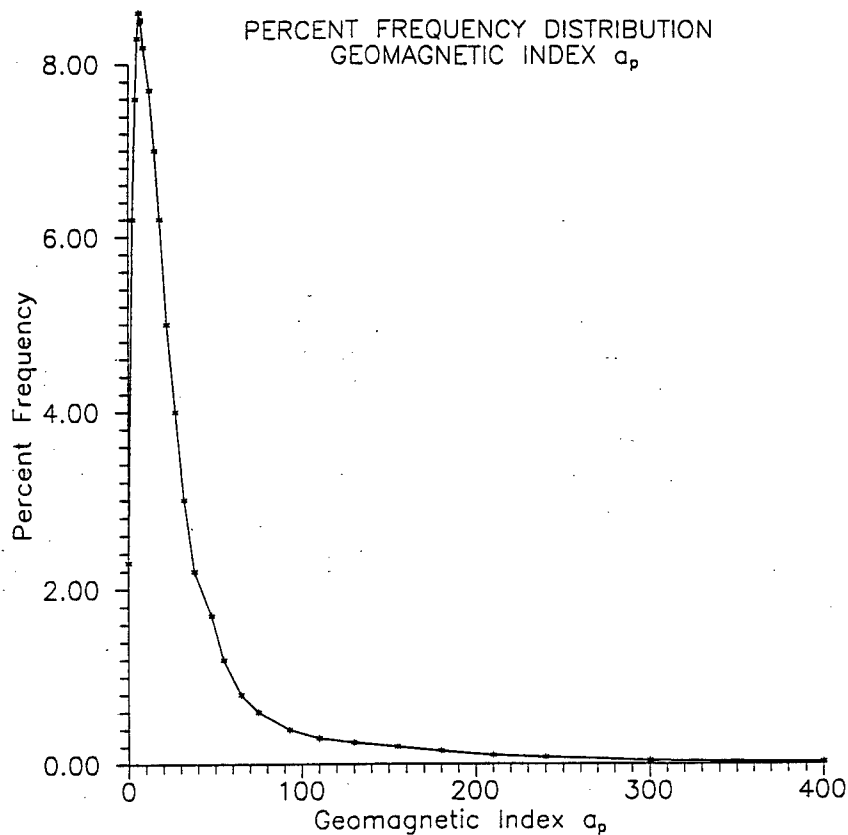


Figure 6. Approximate Historical Percent Frequency of Occurrence for Values of a_p

The daily and mean 162-day values for the solar 10.7-cm radio noise flux were assumed to be equal, with a value selected randomly by use of the uniform distribution $U[40,240]$. The 15th day of the month, and Greenwich mean time (UTC) 1:00 were used for all runs - they only influence GRAM-95 output moderately above 295,276 ft. Unfortunately, diurnal variations at middle and lower altitudes could not be simulated.

One October run failed to meet the 70,000 lbs of lift baseline phase criterion, so phasing was implemented at 65,000 lbs for that run; however, all other runs completed normally. Since Monte Carlo runs were not subjected to the restrictions of the baseline run, but merely used the same payload, bank angles, and attack angles, landing location varied as shown in Figure 7.

MONTE CARLO LANDING SITES FOR RASV SKIPPING TRAJECTORIES

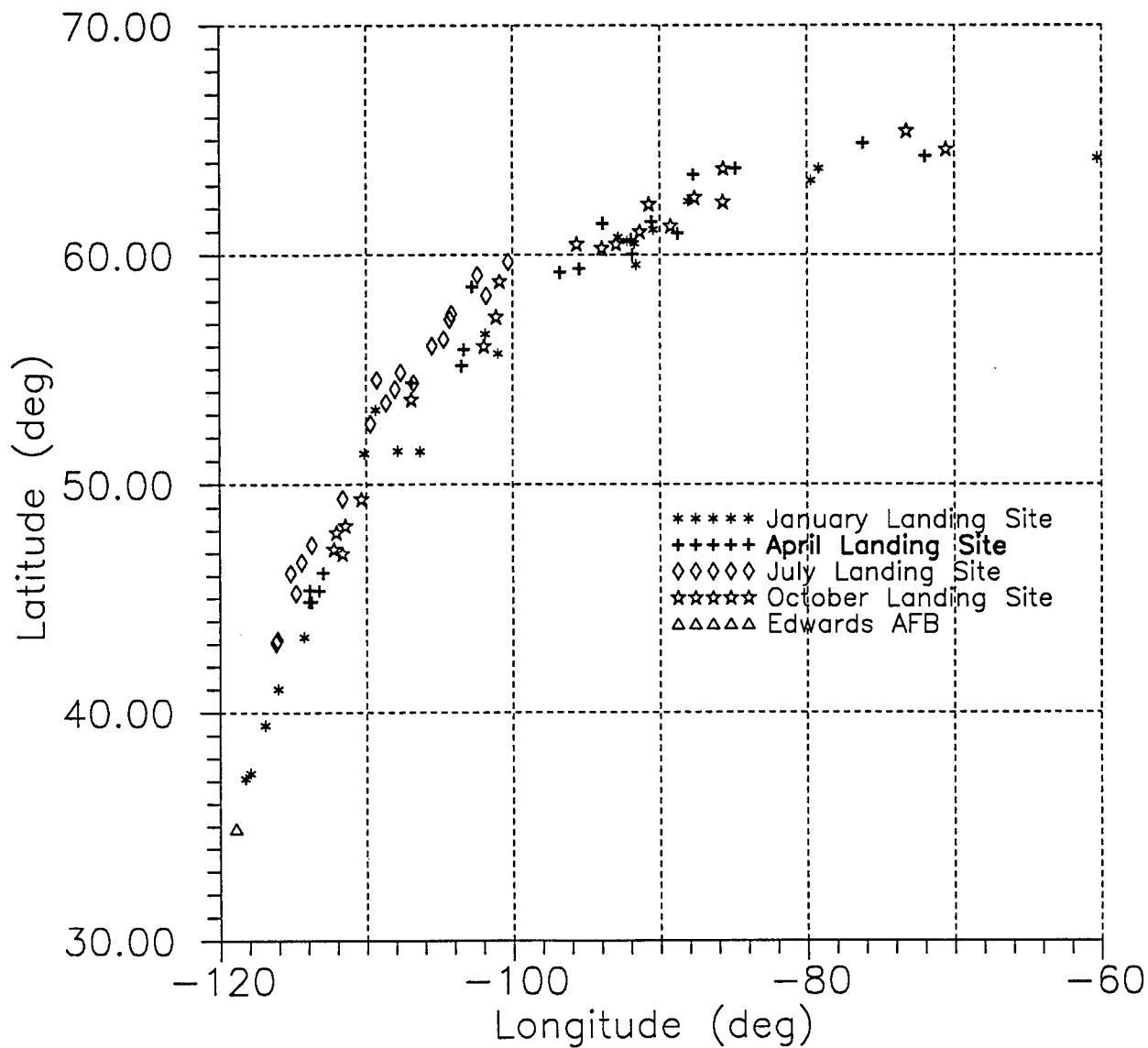


Figure 7. Return Sites for Monte Carlo, Around-the-World, Skipping Trajectories

The peak altitudes of the trajectories averaged only 425,239 ft for January, but averaged 470,713 ft for April, 484,768 ft for October, and 593,805 ft for July. Vehicle peak heat rate was higher for the higher flights - averaging (in Btu/ft²/sec) 238.6 for January, 276.2 for April, 280.0 for April, and 345.5 for July; however, total aerodynamic heat turned out to be lower in the summer. The situation is illustrated in Figures 8, 9, and 10 for the case of an especially high July flight and an especially low January flight - in which possible re-radiation of heat was not taken into account.

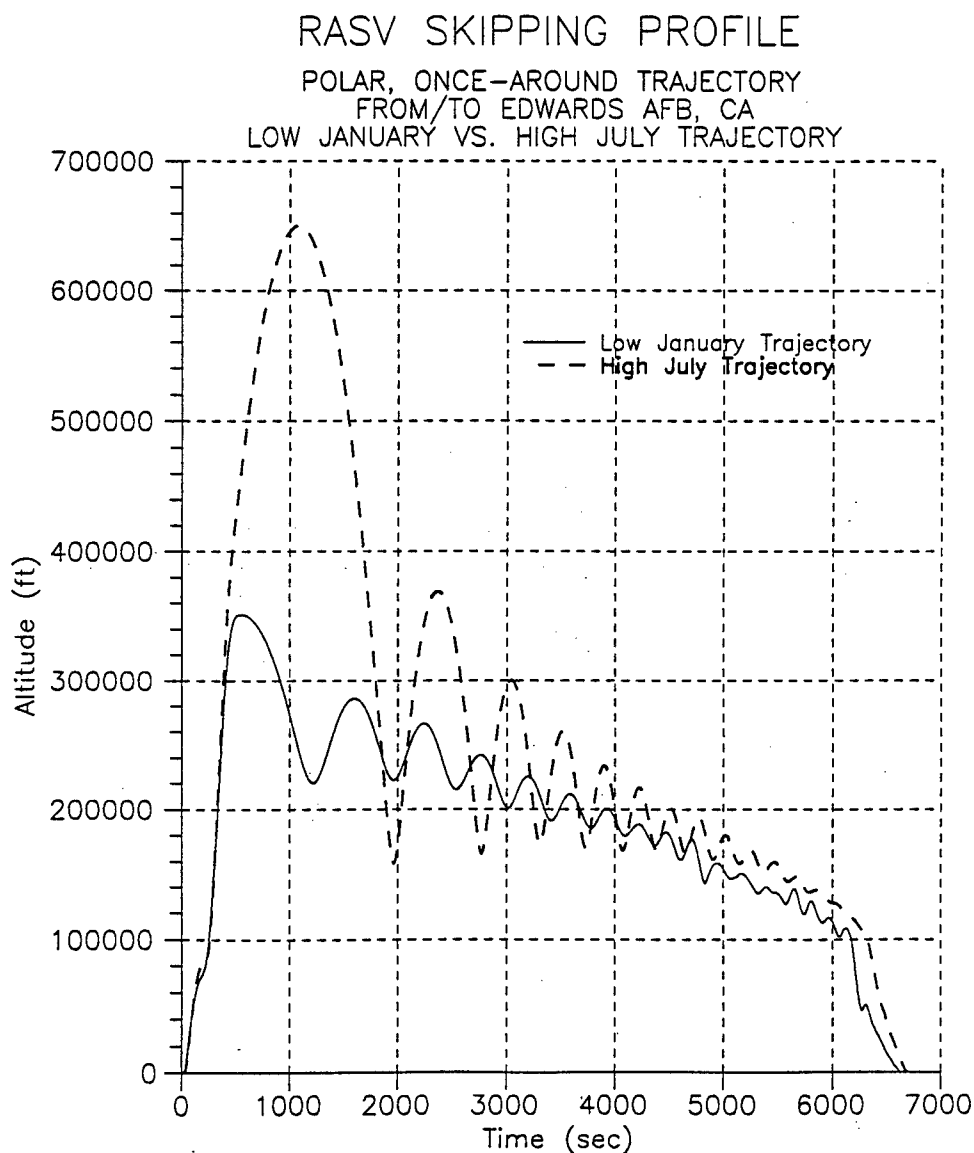


Figure 8. Altitude Profiles for a Low January Flight and a High July Flight

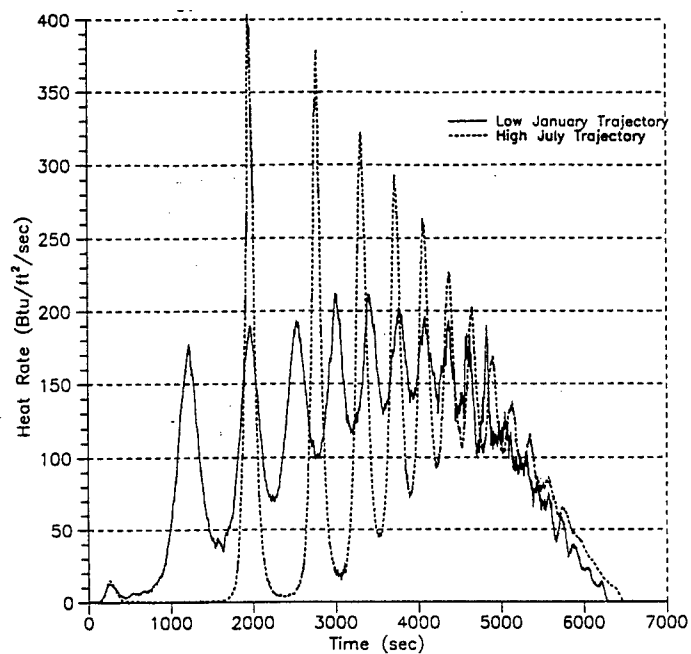


Figure 9. Heat Rate for a Low January Flight and a High July Flight

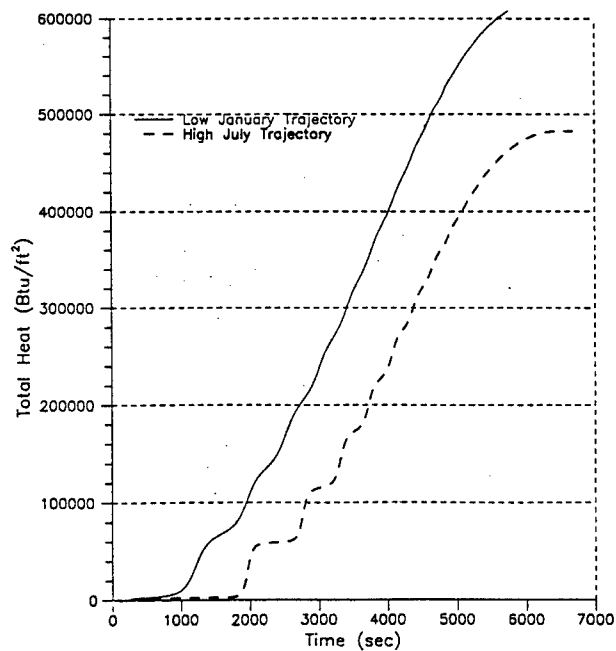


Figure 10. Total Heat for a Low January Flight and a High July Flight

Confidence intervals for peak altitude and peak heat rate were deduced from each set of 20 runs and are given in Tables 2 and 3.

Table 2.
Seasonal 90% Confidence Intervals for Peak RASV Altitude

	January	April	July	October
Lower Altitude Limit (ft)	366,632	403,071	518,158	412,436
Upper Altitude Limit (ft)	489,024	533,342	665,747	546,614

Table 3.
Seasonal 90% Confidence Intervals for Peak RASV Heat Rate

	January	April	July	October
Lower Heat Rate Limit (Btu/ft ² /sec)	213.9	225.7	276.2	230.5
Upper Heat Rate Limit (Btu/ft ² /sec)	279.3	320.9	398.4	330.5

Despite the variability in peak altitude, total energy ($\text{mass} \times \text{gravity} \times \text{height} + \text{mass} \times \text{velocity}^2/2$) was relatively equal among trajectories - i.e. high trajectories had slightly lower velocities at apogee.

Substantial variations were also noted in skipping behavior. Figures 11 and 12 illustrate the wide variation in amplitude of skips late in flight (without a change in attack or bank angle). Note that Figure 11 does not display one of the 80 Monte Carlo runs but a January run without winds - including winds slightly reduced these late amplitudes.

Some additional POST/GRAM-95 runs were made in order to estimate the influence of random winds and of geomagnetic storms. Figure 13 shows the results for a July case in which a Monte Carlo atmosphere was generated (i) with and without the presence of a very extreme geomagnetic

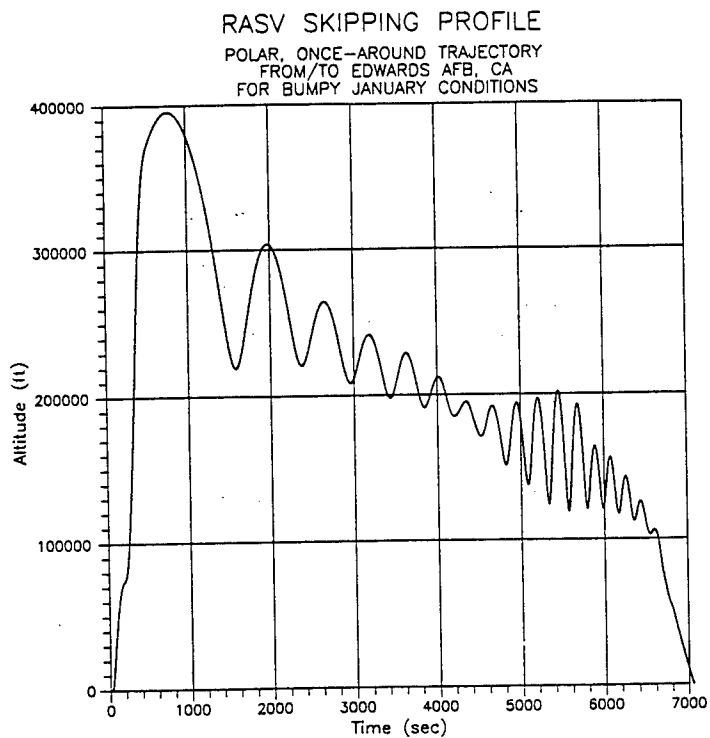


Figure 11. High Steps in a January RASV Profile

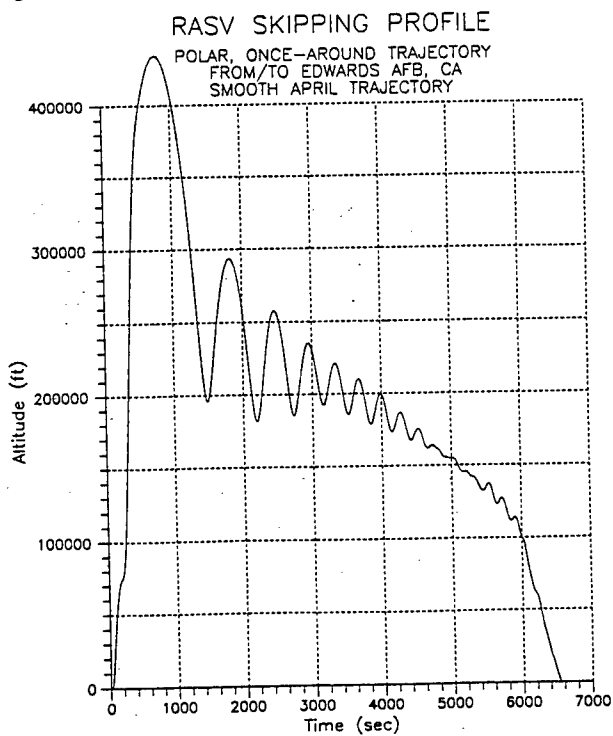


Figure 12. Baby Steps Late in an April RASV Profile

storm ($a_p = 400$, daily f10.7 flux = 230, 162-day average flux = 230), and (ii) with and without random winds. Evidently, winds significantly impact RASV performance but solar effects are minimal; it should be recalled, however, that GRAM-95 fails to apply solar influences below 90 km (295,276 ft).

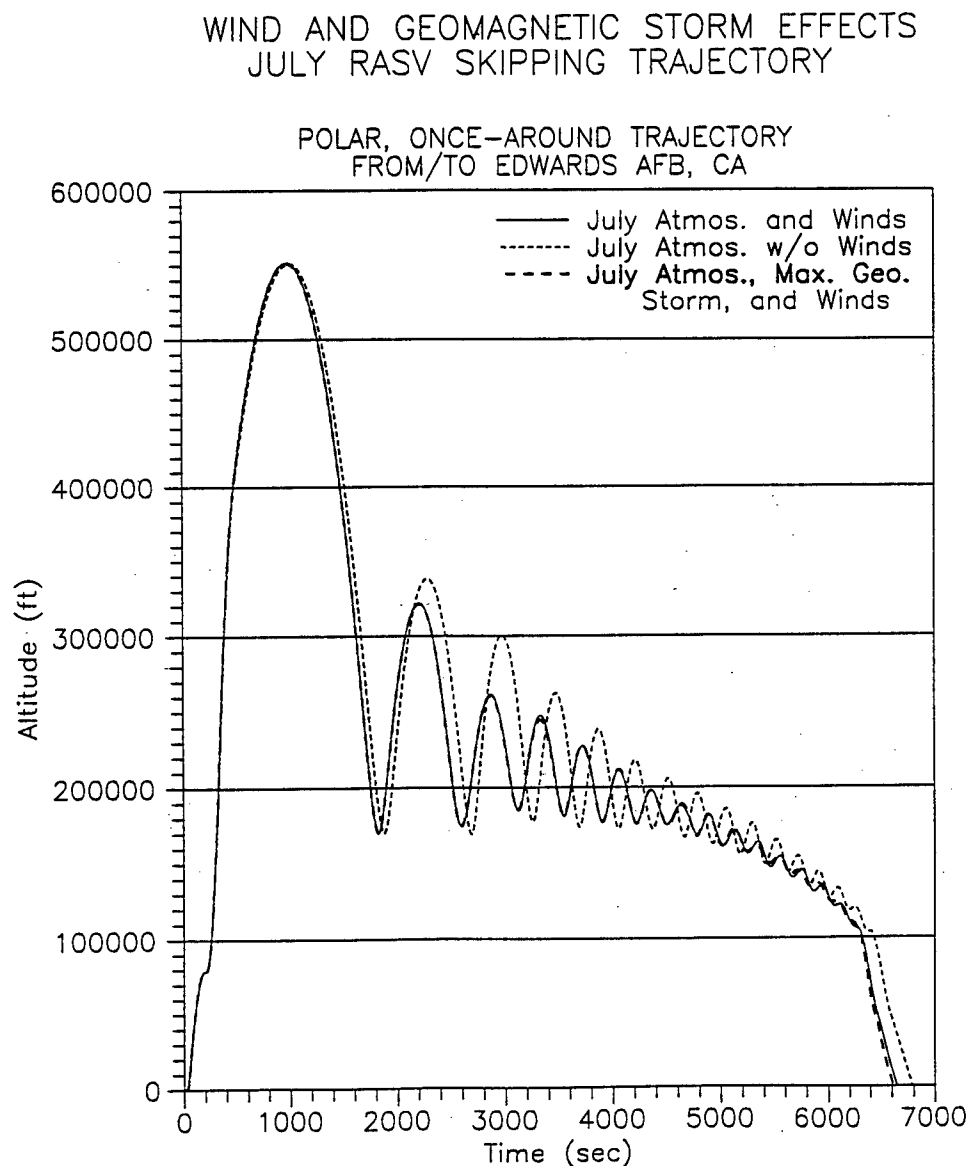


Figure 13. Geomagnetic Storm and Wind Influences Upon a July Trajectory

The changes in trajectory characteristics within a given month were typically less substantial than those between months. Figure 14 displays the variation among altitude profiles within the 20 January simulations. (Recall from Fig. 7 that January provided the largest variety of return sites for the given set of baseline controls.)

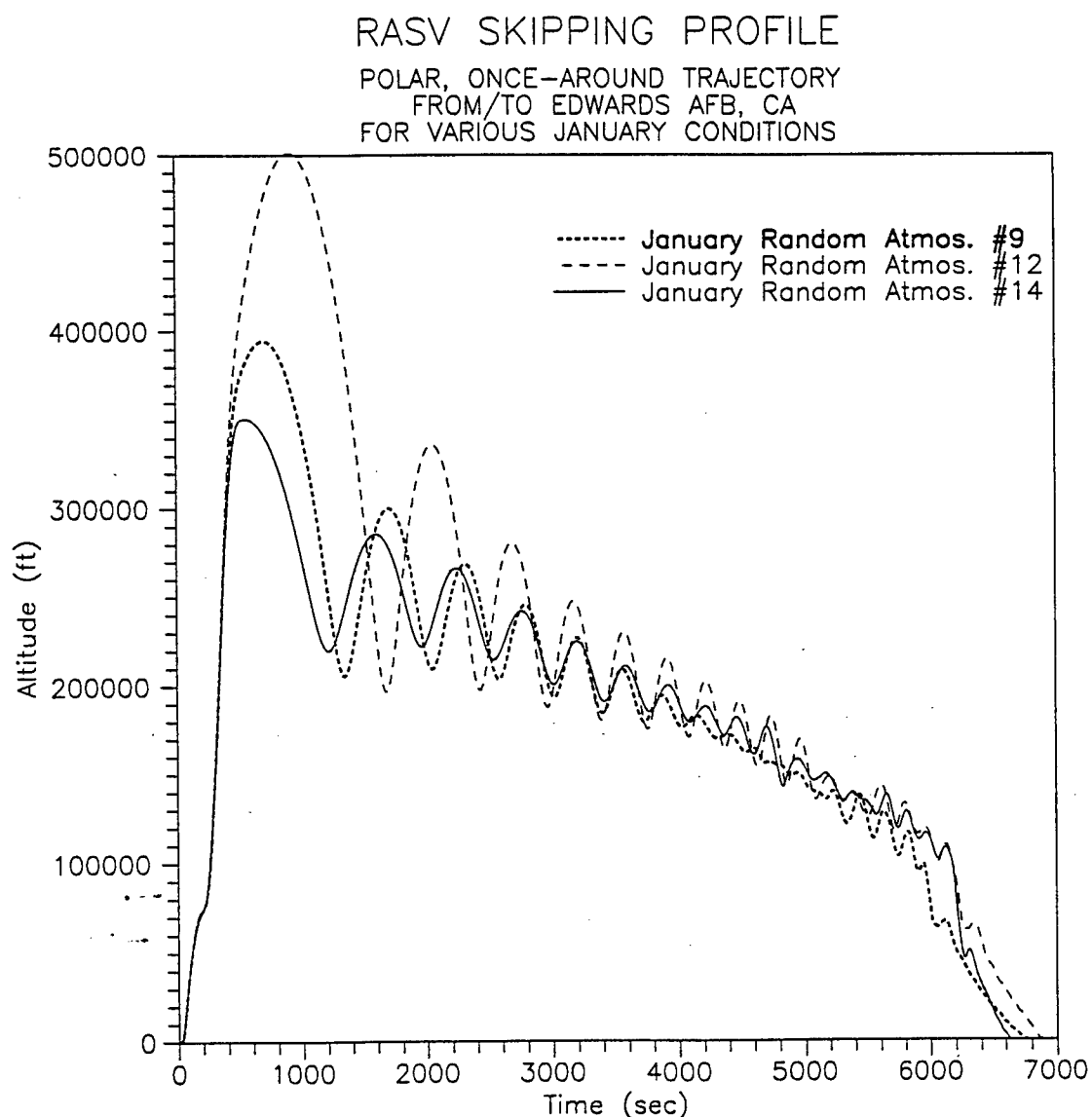


Figure 14. Variation Among January Skipping Profiles

Summary

A sequence of 80 Monte Carlo, geographically varying atmospheres, together with flight parameters obtained from a baseline trajectory for a conceptual Boeing RASV spaceplane has indicated that around-the-world polar, skipping trajectories can be used to lift payloads of about 25,000 lbs into near-orbit conditions in a variety of atmospheric cases. The released payloads would be capable of lifting satellites into low earth orbit by small (e.g. 5,000 ft/sec) upper stages.

Simulations employed a sled-launched, SSME powered vehicle with two-position extendable nozzles and supposed that the payload returned to earth (a worst-case condition) with the spaceplane.

The baseline flight parameters were optimized in order to lift the largest possible payload into the prescribed near-orbit conditions. Some flight characteristics (peak altitude, skip amplitudes, and heat rate) displayed substantial variability due to atmospheric and wind variations when flight controls from the baseline case were applied.

Further work is required in order to determine the nature of larger skipping amplitudes that occurred late in flight within certain atmospheres and to find an ideal skipping motion with controls for maintaining the motion.

References

Barnett, J.J., and Corney, M. (1985a): *Middle Atmosphere Reference Atmosphere Model Derived from Satellite Data*, MAP Handbook 16, K. Labitzke et. al., pp. 47-85.

Barnett, J.J., and Corney, M. (1985b): *Planetary Waves, a Climatological Distribution*, MAP Handbook 16, K. Labitzke et al., pp. 86-137.

Brauer, D.E., et al. (1990): *Program to Optimize Simulated Trajectories (POST)*, Vol. I & II, Martin Marietta Corp., Denver, CO.

Froning, H.D., et al. (1996): *Aerospace Plane Trajectory Optimization for Sub-Orbital Boost Glide Flight*, AIAA 7th International Space Planes and Hypersonic Systems and Technologies Conference, Norfolk, VA, AIAA 96-4519.

Johnson, D.B., et al. (1992): *NASP Derived Vehicles: Not Just to Space!*, 4th International Aerospace Planes Conference, Orlando, FL, AIAA 92-5020

Justus, W.R., et al. (1995): *The NASA/MSFC Global Reference Atmospheric Model - 1995 Version (GRAM-95)*, NASA Technical Memorandum 4715, NASA Marshall Space Flight Center.

Hickey, M.P. (1988a): *The NASA Marshall Engineering Thermosphere Model*, NASA CR-179359.

Hickey, M.P. (1988b): *An Improvement in the Numerical Integration Procedure Used in the NASA/Marshall Engineering Thermosphere Model*, NASA CR-179389.

National Oceanic and Atmospheric Administration, National Aeronautics and Space Administration, United States Air force (1962): *U.S. Standard Atmosphere, 1962*, U.S. Government Printing Office.

Ruth, D.B., et al. (1993): *Global Upper Air Climatic Atlas (GUACA)*, U.S. Navy - Dept. of Commerce, Naval Oceanography Command Detachment Ashville National Climatic Data Center.

**THERMODYNAMIC STABILITY AND OXIDATION BEHAVIOR OF
REFRACTORY (Hf, Ta, Zr) CARBIDE/BORIDE COMPOSITES FOR
ULTRA HIGH TEMPERATURE APPLICATIONS**

**Jenn-Ming Yang
Professor
Department of Materials Science and Engineering**

**University of California, Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90095-1595**

**Final report for:
Summer Faculty Research Program
Phillips Laboratory**

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

and

Phillips Laboratory

September, 1997

THERMODYNAMIC STABILITY AND OXIDATION BEHAVIOR OF REFRACTORY (Hf, Ta, Zr) CARBIDE/BORIDE COMPOSITES FOR ULTRA HIGH TEMPERATURE APPLICATIONS

Jenn-Ming Yang
Professor
Department of Materials Science and Engineering
University of California, Los Angeles

ABSTRACT

The multicomponent thermodynamic stability diagrams for the complex (Hf, Ta, Zr, Si)-C(B)-O, Hf-Ta-C(B)-O and Zr-Si-C(B)-O systems were generated using the proposed linear inequality method, where the stability area of compound is the solution of a set of linear inequalities which is directly obtained from the free energy changes of general chemical reactions. The thermodynamic relationships in the oxide scale and the complicated interfacial reactions in the high temperature oxidation of these systems were analyzed using the generated diagrams. The possible phases formed on the HfC-TaC, HfB₂-TaC, HfB₂-TaB₂, HfC-TaB₂, ZrB₂-SiC, ZrC-SiC composites at high temperatures were predicted. The interactions between oxidation byproducts (CO, CO₂, B₂O₃ gases), carbides (or borides) and porous oxides were analyzed. The possible kinetic rate-controlling processes were discussed based on the considerations of outward and inward gas diffusion model and outward and inward ionic diffusion model. Finally, attempts to improve the oxidation resistance of refractory carbide/boride composites will be discussed.

THERMODYNAMIC STABILITY AND OXIDATION BEHAVIOR OF REFRACTORY (Hf, Ta, Zr) CARBIDE/BORIDE COMPOSITES FOR ULTRA HIGH TEMPERATURE APPLICATIONS

INTRODUCTION

Advanced materials with temperature capability of over 2000 °C are needed for ultra-high temperature structural applications, such as rocket engines and thermal protection systems for space vehicles. The potential ceramic systems that can operate as oxidation-resistant materials with temperature capabilities from 2000 to 2400 °C are refractory oxides, carbides, borides and nitrides as shown in Figure 1. Oxide-based ceramics are often chemically inert but are more brittle than the non-oxide based ceramics. Non-oxide ceramics, on the other hand, are generally not very oxidation resistant when heated to such high temperatures in oxidizing environment, but for those which can form continuously protective oxide films are exceptions. There are relatively few refractory oxides that are stable in an oxidizing atmosphere above 2000 °C. Hafnia (melting point 2900 °C) and zirconia (melting point 2770 °C) have sufficiently high melting temperatures and relatively low vapor pressures. However, they undergo solid phase transformations: from monoclinic to tetragonal structure at 1150 and 1650 and from tetragonal to cubic at 2370 and about 2700 °C for zirconia and hafnia respectively, with a corresponding large volume change. The large volume change would result in the destruction of any large scale component made from these materials, and so in practice the materials must be stabilized with an appropriate additive such as calcium oxide, magnesium oxide, or yttrium oxide.

Refractory carbides and borides of Hf, Zr and Ta etc. are potential candidates for ultra high temperature structural applications since they have melting temperatures considerably higher than their associated oxides, do not undergo any solid phase transformation, and have relatively good thermal shock resistance. However, very limited studies¹⁻⁵ have been conducted to investigate the oxidation behavior of refractory carbides and borides at temperatures above 2000 °C. The oxidation processes of refractory carbides and borides have been shown to be the combined processes of oxygen inward or metal ion outward diffusion and gaseous (or liquid at relatively lower temperatures) byproduct outward diffusion through oxide scale⁶⁻¹⁰. Therefore, the oxidation resistances of carbides and borides are mainly influenced by the formation and escaping of gaseous byproducts (such as CO, CO₂, B₂O₃) during the oxidation processes which are significantly different from those of their metal counterparts. Typically, the oxide scale formed at high temperature consists of at least two distinctive layers: (1) a much less porous inner oxide layer and (2) a porous outer oxide layer¹¹⁻¹². However, Barger et al.¹ indicated that, in the oxidized HfC film, an oxycarbide (HfO_{2-x}C_y) interlayer was found between the outer porous HfO₂ layer and a residual carbide layer with dissolved oxygen in the lattice. The oxide interlayer was found to be a better diffusion barrier for oxygen than either the hafnium oxide or carbide layers. The presence of oxycarbide indicated that the oxygen partial pressure would arrive to a relatively lower value at carbide and oxycarbide interface. Various diffusion models have also been proposed to describe the oxidation behavior of refractory carbide and borides. Holcomb et al.¹³ proposed a counter-current gaseous diffusion model to describe the oxidation behavior of HfC. Barger et al.¹ proposed an oxidation model for HfC based on moving-boundary diffusion theory. Attempts have also been made

to improve the oxidation resistance of refractory carbides and borides with appropriate additives. For example, the addition of SiC can improve the oxidation resistance of both HfB_2 and ZrB_2 ⁵ . However, much more experimental work and theoretical study need to be conducted to better understand the oxidation behavior of the refractory carbides/borides.

The major objective of this research is to construct the multicomponent thermodynamic stability diagrams using the linear inequality methods for predicting the sequence of the layers for gas-solid-solid interfacial reactions and to interpret the oxidation behavior of several refractory carbide/boride composites. The applications of these thermodynamic stability diagrams will be discussed. Furthermore, the possible gaseous and ionic diffusion models will be discussed based on information obtained in the generated thermodynamic stability diagrams.

THERMODYNAMIC MODELS FOR THE GENERATION OF MULTICOMPONENT THERMODYNAMIC STABILITY DIAGRAMS

Since the potential-pH diagrams were first proposed by Pourbaix ¹⁴ , the potential-pH diagrams, two partial pressure diagrams etc. for metal-nonmetal-nonmetal systems have been widely used in many fields of corrosion science and metallurgy etc.. Methods of calculating thermodynamic equilibria can be categorized into two approaches, namely, the stoichiometric and non-stoichiometric ones ¹⁵ . Stoichiometric approach makes use of independent reactions and their equilibrium constants, whereas the non-stoichiometric one is based on linear equations in terms of the chemical potentials of system components to

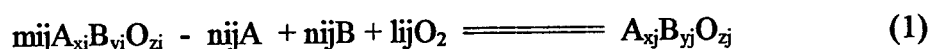
obtain the triple points of the stability areas of three condensed species considered.

Yokokawa et al.¹⁶⁻²² first proposed to use the generalized chemical potential diagrams for metal-metal-nonmetal systems in which $\text{Log}(a_{M1}/a_{M2})$ coordinate was successfully introduced to the generated diagrams by the non-stoichiometric method. The introduction of $\text{Log}(a_{M1}/a_{M2})$ coordinate makes the thermodynamic stability (or phase stability , or chemical potential) diagrams much more powerful than one expected in the thermodynamic stability diagrams which only deal with pure metals. The generation of such thermodynamic stability diagram can also be done by stoichiometric method as outlined in this section.

For a system consisting of N different possible condensed compounds (solid or liquid), the stability area of certain compound I ($1 < I < N$) corresponds to the region in which the free energy changes of $N-1$ chemical (or electrochemical) reactions of compound I to compound $1, \dots, I-1, I+1, \dots, N$, respectively , at certain condition are either greater than or equal to zero. If one of three parameters for the metal-nonmetal-nonmetal systems (i.e. T , pH and electrode potential in the potential- pH diagrams; T and two partial pressures in the partial pressure diagrams and partial pressure- $1/T$ diagrams) is constant, the thermodynamic stability diagrams are two-dimensional. If two of four parameters for metal-metal-nonmetal-nonmetal systems or one of three parameters for metal-metal-nonmetal systems are constant, the multicomponent thermodynamic stability diagrams are also two-dimensional. The geometrical shape of the stability area of compound is a convex polygon. It is known that the stability area of compound can be considered to be the solution of a set of linear inequalities which is directly obtained from the free energy changes of chemical (or electrochemical) reactions.

The multicomponent thermodynamic stability diagrams of metal-metal-nonmetal systems will be selected as an example to describe the construction of a set of linear inequailities. Two additional examples to illustrate how to obtain the $\text{Log}(a_{M1}/a_{M2})$ and $\text{Log}(a_{M3}/a_{M2})$ coordinates in two or three metallic component systems are given in the Appendix.

An A-B-O ternary system is considered here (A and B are metallic elements, O being oxygen; these are the system components). If there are N possible condensed species, then the reactions between condensed species, two metallic components and oxygen can be studied in terms of chemical reactions, which proceed according to the general equation:



where $A_{x_i}B_{y_i}O_{z_i}$ and $A_{x_j}B_{y_j}O_{z_j}$ are the general forms of condensed species in A-B-O ternary systems, x_i, y_i, z_i, x_j, y_j and z_j are the stoichiometric numbers of these two compounds, respectively.

The elemental balances for each of the three elements, i.e. A, B and O, are then used to find n_{ij} , m_{ij} and l_{ij} for above reaction, and following equations can be yielded, respectively:

$$m_{ij} = (x_j + y_j)/(x_i + y_i) \quad (2)$$

$$n_{ij} = (y_j x_i - x_j y_i)/(x_i + y_i) \quad (3)$$

$$l_{ij} = (x_i z_j + y_i z_j - x_j z_i - y_j z_i)/2(x_i + y_i) \quad (4)$$

where $i=1, \dots, N; j=1, \dots, i-1, i+1, \dots, N$.

According to the thermodynamic properties of equilibria and from the general reactions (1), the stability area of substance i ($A_{x_i}B_{y_i}O_{z_i}$) (where i is the certain value $1 < i < N$)

obtained by comparison with other substances j ($A_{xj}B_{yj}O_{zj}$) (where $j=1, \dots, i-1, i+1, \dots, N$) must satisfy to the following conditions:

$$G_{ij} = G_{ij}^0 + RT \ln K_{pij} > 0$$

where $G_{ij}^0 = G_{A_{xj}B_{yj}O_{zj}}^0 - l_{ij}G_{O_2}^0 + n_{ij}G_A^0 - m_{ij}G_B^0 - m_{ij}G_{A_{xi}B_{yi}O_{zi}}^0$, and

$K_{pij} = a_{A_{xj}B_{yj}O_{zj}}/P_{O_2}^{l_{ij}} a_{A_{xi}B_{yi}O_{zi}}^{m_{ij}} (a_A/a_B)^{n_{ij}}$; K_{pij} is the thermodynamic equilibrium constant

for reaction (1), a_A and a_B are the activities of two metallic components, $a_{A_{xi}B_{yi}O_{zi}}$ and $a_{A_{xj}B_{yj}O_{zj}}$ are the activities of two condensed species considered having the stability areas, P_{O_2} is the oxygen partial pressure, respectively.

Suppose that all the activities of condensed species are equal to 1 (i.e. $a_{A_{xi}B_{yi}O_{zi}} = a_{A_{xj}B_{yj}O_{zj}} = 1$), and let:

$$k_{ij} = G_{ij} / 2.303RT$$

$$X_1 = - \text{Log} P_{O_2}$$

$$X_2 = + \text{Log} (a_A / a_B)$$

, then $N-1$ linear inequalities can be written as:

$$k_{ij} + l_{ij} X_1 + n_{ij} X_2 > 0 \quad (5)$$

where i is the certain value ($1 < i < N$); $j=1, \dots, i-1, i+1, \dots, N$.

Let:

$$\begin{array}{lll} C_{ij} = k_{ij} & & C_{ij-1} = k_{ij} \\ A_{ij} = l_{ij} & \text{when } i > j & A_{ij-1} = l_{ij} \\ B_{ij} = n_{ij} & & B_{ij-1} = n_{ij} \end{array} \quad \text{when } i < j$$

, then the $N-1$ linear inequalities (5) can be changed into:

$$A_{ij}X_1 + B_{ij}X_2 > C_{ij} \quad (6)$$

where $j=1, \dots, N-1$.

The set of linear inequalities (6) defines the stability area of condensed compound i ($A_{xi}B_{yi}O_{zi}$). When i changes from 1 to N , the N sets of linear inequalities as (6) can be obtained in the same way.

When there are many compounds to be considered in the system, it is necessary to use computer program to generate the thermodynamic stability diagrams. The computer algorithm for the generation of the phase stability diagrams had been proposed by Wang et al.²³, in which a series of relevant objective functions with minimum numbers were introduced to the set of linear inequalities to form the problems of linear programming. The FORTRAN program had been constructed to produce the phase stability diagrams by the use of Revised Simplex Method that was used to solve the problems of linear programming. Various researchers²⁴⁻²⁹ had also developed the computer algorithms and produced the phase stability diagrams which only deal with pure metals. The generated thermodynamic stability diagrams by the above linear inequality method are equivalent to those generated by Yokokawa et al.'s chemical potential method. The slopes of boundary lines can be directly obtained from the specific general reactions (for example, in equation (1), the slope of boundary line for the stability areas of two possible condensed species is n_{ij}/l_{ij}). The convex polygon stability area of certain condensed substance is directly obtained by the solution of a set of linear inequalities. However, unlike chemical potential method proposed by Yokokawa et al.¹⁶⁻¹⁷ where the fundamental stoichiometric number matrix represents linear equations, the fundamental linear inequalities with at least one coordinate of $\text{Log}(a_{M1}/a_{M2})$ in the linear inequality method are directly obtained from the

free energy changes of the specific general chemical (or electrochemical) reaction equation as shown in (1). Such kinds of specific general chemical (or electrochemical) reaction equations can be obtained in any system which includes at least two metallic components and one non-metallic component.

THERMODYNAMIC DATA

Thermodynamic and physical property data were collected from different sources. However, Thermochemical Properties of Inorganic Substances³⁰ and JANAF Thermodynamic Tables³¹ were used as the primary sources for free-energy values, heat of formation, heat capacity (Cp), and entropy values. These information are listed in Table I.

RESULTS AND DISCUSSION

1. Partial Pressure Diagrams

The generated LogP_{O_2} - LogP_{CO} , LogP_{O_2} - $\text{LogP}_{\text{CO}_2}$, $\text{LogP}_{\text{CO}_2}$ - LogP_{CO} and LogP_{O_2} - $\text{LogP}_{\text{B}_2\text{O}_3}$ diagrams for Hf-C(or B)-O, Zr-C(or B)-O, Ta-C(or B)-O and Si-C-O systems at 2300 K are shown in Figs.2-5, respectively. From the analysis of these thermodynamic stability diagrams, it is known that, in the temperature ranges of 2000-2700 K, hafnium, zirconium, and tantalum oxides can be formed in a wide range of O_2 , CO_2 , CO and B_2O_3 partial pressure. Obviously, hafnium, zirconium and tantalum borides are not stable in oxidizing atmosphere and will directly form oxides. However, hafnium, zirconium and tantalum carbides could be stable in the reducing atmosphere of $\text{LogP}_{\text{O}_2} =$

-10 to -16, and $\text{Log}P_{\text{CO}} = -7$ to -1 (atm.). Thus, hafnium, zirconium and tantalum carbides could be the high temperature oxidation-resistant materials in the very low oxygen and high carbon activity atmospheres.

Thermodynamic analysis also provides an insight into the mechanism of the formation of outer porous oxide layer during oxidation of refractory carbide and boride as reported by Bargerion et al. and Courtright et al. etc.. For example, the thermodynamic stability boundary line of HfC/HfO_2 is obtained from the following two reactions:



There exists another equilibrium reaction in the gas environment:



Table 2 shows the equilibrium CO , CO_2 partial pressures with different oxygen partial pressure at the HfC/HfO_2 interface at 2300 K. From the analysis of table 2, it is known that when the oxygen partial pressure is higher than 10^{-12} atm., the CO equilibrium partial pressure could be extremely high. As a result, if the oxygen partial pressure at carbide/oxide interface remains a relatively high value, the formed oxide layer could be destroyed by the entrapped gas. In this case, the formation of porous oxide layer is necessary to allow the escaping of gas and to maintain the integrity of the oxide scale.

Table 2. The equilibrium partial pressures at HfC/HfO₂ interface at 2300 K

LogP _{O2}	-1	-10	-12	-14	-16	-18
LogP _{CO}	+16.927	+3.427	+0.427	-2.573	-5.573	-8.573
LogP _{CO2}	+18.354	+0.354	-3.646	-7.646	-11.646	-15.646

2. The Log(a_A/a_B) vs LogP_{O2}, Log(a_A/a_B) vs LogP_{CO} (or LogP_{CO2}) and

Log(a_A/a_B) vs LogP_{B2O3} Diagrams

The generated diagrams of Log(a_{Ta}/a_{Hf}) vs LogP_{O2} (LogP_{CO}, LogP_{CO2} and LogP_{B2O3}) for Hf-Ta-C(B)-O systems and Log(a_{Si}/a_{Zr}) vs LogP_{O2} (LogP_{CO}, LogP_{CO2} and LogP_{B2O3}) for Zr-Si-C(B)-O systems are shown in Figs.6-15, respectively. From these diagrams, it is very clear that when oxygen partial pressure remains relatively high values, the stability areas of oxides in the oxide scale depend only on the activities of metallic components. It is well known that the activity gradient of metallic component across the scale is mainly dependent on the diffusion coefficient of metallic ion through the oxide layer. The metallic component which has the highest diffusion coefficient is usually formed as the outermost oxide layer. The metallic component which has the lowest diffusion coefficient is usually formed as the inner oxide layer. Based on the consideration of diffusion effect, the thermodynamic stability diagrams can be directly used to predict the formation of possible phases during the oxidation of HfB₂-TaC, HfC-TaC, HfB₂-TaB₂, HfC-TaB₂, ZrB₂-SiC, and ZrC-SiC composites. In the following, a detailed analysis in the oxidation processes of several refractory carbide/ boride composites will be presented.

2.1 The interfacial reaction and layer sequence in the high temperature oxidation of HfC-TaC composite

The high temperature oxidation of HfC-TaC composite in the temperature range of 1400-2200 °C was studied by Courtringt et al. ² and Patterson et al. ⁴. It was found that the oxide layers that formed on HfC-TaC obeyed the parabolic growth kinetics. However, there was a break in the kinetics around 1800 °C. The oxidation kinetics of HfC-TaC above 1800 °C were in fair agreement with those reported for the Hf-Ta alloys. Nevertheless, below 1800°C, the notable difference between the oxide layer that formed on the carbide and the metallic alloy system was observed. The Hf-Ta alloy formed a dense, tenacious, crack-resistant oxide; whereas, the carbide developed an oxide that was quite porous and prone to cracking. The oxide layer that formed on the carbide was a mixture of HfO_2 and $\text{Ta}_6\text{Hf}_2\text{O}_{19}$ ($>1900^\circ\text{C}$). Around 1800 °C, the outermost oxide layer was found to be Ta_2O_5 . This indicated that the diffusion coefficient of Ta ion through the oxide is greater than that of Hf ion at this condition.

From the generated thermodynamic stability diagram for the HfC-TaC composite as shown in Fig. 10, the sequence of porous oxide layer would be $\text{Ta}_6\text{Hf}_2\text{O}_{19}$ (or Ta_2O_5) (outermost)/ HfO_2 (inner porous oxide layer). In the outer porous oxide layer, because oxygen partial pressure remains relatively high values, the oxide layer sequence is mainly dependent on the diffusion coefficients of metallic ions through the oxide layer. The possible oxides (Ta_2O_5 , $\text{Ta}_6\text{Hf}_2\text{O}_{19}$ and HfO_2) are all thermodynamically stable in different $\text{Log}(a_{\text{Ta}}/a_{\text{Hf}})$ ranges. The influence of oxygen partial pressure (in the range $\text{Log}p_{\text{O}_2} = -9$ to 0) on the stability areas of various possible oxides are almost negligible. The stability

areas of various oxides are mainly dependent on the activities of metallic components. Thus, Ta_2O_5 or $\text{Ta}_6\text{Hf}_2\text{O}_{19}$ is formed as the outermost layer because the activity of Ta ion is much greater than that of Hf ion. HfO_2 is formed as the inner porous oxide layer because HfO_2 is stable even in the lower Hf activity range ($\text{Log}(a_{\text{Hf}}/a_{\text{Ta}}) > -4$) and relatively low oxygen partial pressure. The layer sequence in the less porous inner oxide layer would be $\text{HfO}_2/\text{HfC}_x\text{O}_y + \text{Ta}(\text{TaC})$. In this layer, the stabilities of oxides and carbides and interfacial reactions are strongly influenced by oxygen partial pressure and activities of metallic components. The results obtained from the thermodynamic analysis are in good agreement with the experimental results.

2.2 The interfacial reaction and layer sequence in the high temperature oxidation of ZrB_2 (or ZrC)/SiC

The effect of an SiC addition on the oxidation of ZrB_2 up to 1500°C was studied by Tripp et al.³². The experimental results showed that in the temperature range of $1300 - 1500^\circ\text{C}$, SiC additions to ZrB_2 improve its oxidation resistance. The SiO_2 -rich glassy layer was found to be the outermost oxide layer, whereas ZrO_2 -rich oxide layer was the inner oxide layer. SiC could coexist with ZrO_2 in the inner oxide layer.

The generated thermodynamic stability diagrams for SiC/ ZrB_2 at 1700 K at different oxygen partial pressures are shown in Figs. 16-17, respectively. From the thermodynamic stability diagrams, it is known that the complex oxide SiZrO_4 has the stability area and might be formed during oxidation process. However, the complex oxide SiZrO_4 is not

stable when $T > 1680\text{ }^{\circ}\text{C}$. At higher temperature, the oxides which can be formed on the surface are ZrO_2 and SiO_2 (s,l). Thus, the sequence of the multilayer oxide scale at higher temperatures will be different from that at lower temperatures. The layer sequence in the temperature range of $1300 - 1500\text{ }^{\circ}\text{C}$ would be SiO_2 (outermost)/ $\text{SiZrO}_4 + \text{SiO}_2/\text{SiZrO}_4 + \text{ZrO}_2/\text{ZrO}_2 + \text{SiC}/\text{ZrB}_2 + \text{SiC}$. The sequence of the multilayer oxide scale in the temperature range of $1700 - 2200\text{ }^{\circ}\text{C}$ would be $\text{SiO}_2(\text{l})$ (outermost)/ $\text{SiO}_2(\text{l}) + \text{ZrO}_2/\text{ZrO}_2 + \text{SiC}/\text{ZrB}_2 + \text{SiC}$. The oxidation behavior of ZrB_2/SiC composite in the temperature range of $1800\text{--}2400\text{ }^{\circ}\text{C}$ had been studied by Bull et al. ⁵. The experimental results confirmed that silicon oxide was formed as the outermost layer, while zirconium oxide was formed as inner oxide layer. Silicon carbide coexisted with zirconium oxide in the internal oxidation region.

It is well known that the formation of liquid glass (such as $\text{B}_2\text{O}_3(\text{l})$ in the temperature range $500 - 1000\text{ }^{\circ}\text{C}$ and $\text{SiO}_2(\text{l})$ when $T > 1683\text{ }^{\circ}\text{C}$) can separate two totally different gas environments (one with high oxygen partial pressure and another one with very lower oxygen partial pressure). The formation of liquid SiO_2 ($T > 1700\text{ }^{\circ}\text{C}$) might serve as a diffusion barrier. However, at higher temperatures, the volatilization of $\text{SiO}(\text{g})$ has significant influence on the stability of liquid SiO_2 . The effect of SiC additive on the oxidation resistance of ZrB_2 is greatly limited to the formation of $\text{SiO}(\text{g})$. The thermodynamic boundary line of SiC/SiO_2 is obtained from the following two reactions:



There exists a volatilization equilibrium between SiO(g) and SiO₂:



Table 3 shows the equilibrium partial pressures at SiC/SiO₂(l) and SiO(g)/SiO₂(l) interfaces at 2000, 2300 and 2500 K, respectively. From the analysis of table 3, the equilibrium SiO(g) partial pressure decreases with increasing oxygen partial pressure and increases with increasing temperature. Thus, at higher temperatures (T > 2500 K), the volatilization of SiO(g) becomes the dominant process through reaction (10) to reaction (12). In this case, SiO₂(l) formed on the surface will volatilize rapidly and the glassy liquid film can not serve as an effective diffusion barrier. However, at lower temperature (T < 2300 K), glassy SiO₂ (l) film is stable and can be used as the diffusion barrier. The existence of such diffusion barrier would greatly improve the oxidation resistance of carbide or boride.

Table 3 The equilibrium partial pressures of SiO(g), CO and CO₂ with different P_{O2}

	T = 2000 K			T = 2300 K			T = 2500 K		
LogP _{O2}	-1	-10	-14	-1	-10	-14	-1	-10	-14
LogP _{CO}	+19.099	+5.599	-0.401	+15.932	+2.432	-3.568	+14.248	+0.748	-5.252
LogP _{CO2}	+21.471	+3.471	-4.529	+17.359	-0.641	-8.641	+15.174	-2.826	-10.826
LogP _{SiO}	-7.195	-2.695	-0.695	-4.511	-0.011	+1.989	-3.406	+1.094	+3.094

CONCLUSION

High temperature oxidation of refractory carbide or boride is complicated by the formation of gaseous oxidation byproducts. In order to understand the oxidation mechanism, both outward diffusion (or escaping) of metallic ions and formed gases and inward diffusion of oxygen should be considered based on thermodynamics and kinetics of various possible interfacial reactions. The thermodynamic stability diagrams which include at least two metallic components can clearly represent the stability area of complex oxide in certain $\text{Log}(a_{M1}/a_{M2})$ range. The possible oxides which are stable and can be used as the oxidation-resistant materials in the temperature range of 2000 - 2500 °C are HfO_2 and ZrO_2 . However, the high oxygen anionic mobilities and porous, defective structures of these two oxides formed during the oxidation processes of their carbides or borides indicate that hafnium and zirconium carbides or borides alone can not be used as high temperature oxidation-resistant materials for extended period of time. In order to improve the oxidation resistance of hafnium and zirconium carbides or borides, attempts should be directed toward incorporating of appropriate additives such as SiC, TiC, etc. Al_2O_3 may be another possible additive. The incorporation of these additives to the base hafnium and zirconium carbides or borides may lead to the formation of solid (most likely complex oxide) or viscous liquid diffusion barrier. Liquid SiO_2 film can be used as the diffusion barrier when $T < 2300$ K. Liquid Al_2O_3 and TiO_x films may be used as the diffusion barrier when $T < 2700$ K. Furthermore, liquid Al_2O_3 and TiO_x may react with HfO_2 or ZrO_2 to form various solid complex oxides in the temperature range of 2200 -

2500 K. These complex oxide layers may also serve as diffusion barriers to further improve the oxidation resistance. Experimental verification is in progress to elucidate the role of additives on the oxidation behavior of refractory carbide/boride composites.

APPENDIX

The approach to obtain the $\text{Log}(a_{M1}/a_{M2})$ and $\text{Log}(a_{M3}/a_{M2})$ coordinates in two or three metallic component systems is illustrated here.

For an A-B-O-S system (A and B are metallic elements, O and S are non-metallic elements, respectively; these are the system components), the specific general reaction equation can be written as following:



where m_{ij} , n_{ij} , l_{ij} and k_{ij} are directly defined by stoichiometric numbers in the two condensed substances i and j. The following four equations are obtained based on the consideration of elemental balances:

$$m_{ij} = (w_j + x_j)/(w_i + x_i) \quad (14)$$

$$n_{ij} = (w_{ix_j} - w_{jx_i})/(w_i + x_i) \quad (15)$$

$$l_{ij} = (y_jw_i + y_jx_i - y_iw_i - y_ix_i)/(w_i + x_i) \quad (16)$$

$$k_{ij} = (z_jw_i + z_jx_i - z_iw_i - z_ix_i)/(w_i + x_i) \quad (17)$$

The $\text{Log}(a_A/a_B)$ coordinate is then obtained directly from the equilibrium constant K_{pij} for the reaction equation (13).

For an A-B-C-O system (A, B and C are metallic elements, O is nonmetallic element, respectively; these are the system components). The specific general reactions can be written as following:



where m_{ij} , n_{ij} , l_{ij} and k_{ij} are given by following equations:

$$m_{ij} = (w_j + x_j + y_j)/(w_i + x_i + y_i) \quad (19)$$

$$n_{ij} = (w_{ixj} + w_{iyj} - w_{jxi} - w_{jyi})/(w_i + x_i + y_i) \quad (20)$$

$$l_{ij} = (w_{jxi} + y_{jxi} - w_{ixj} - y_{ixj})/(w_i + x_i + y_i) \quad (21)$$

$$k_{ij} = (z_{jwi} + z_{jxi} + z_{jyi} - z_{iwj} - z_{ixj} - z_{iyj})/(w_i + x_i + y_i) \quad (22)$$

The $\text{Log}(a_A/a_C)$ and $\text{Log}(a_B/a_C)$ coordinates are then obtained directly from the equilibrium constant K_{pij} for the reaction equation (18).