UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1993

SUMMER RESEARCH PROGRAM FINAL REPORTS

VOLUME 9

ROME LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Col. Hal Rhoades

Program Manager, RDL
Scott Licoscos

Program Administrator, RDL
Gwendolyn Smith

Program Administrator, RDL
Johnetta Thompson

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

Decmeber 1993

Master Index For Graduate Students

Ahmad, Imad
BS
Electrical Engineering
Portland State University
Portland, OR   97207-0000

Field: Electrical Engineering
Laboratory: PL/WS

Vol-Page No: 8-19

Anselmo, Andrew
MS
Mechanical Engineering
Columbia University
New York, NY   10027-0000

Field: Mechanical Engineering
Laboratory: RL/ER

Vol-Page No: 9- 3

Bapty, Theodore
MS
Box 1649 Station B
Vanderbilt University
Nashville, TN   37235-0000

Field: Electrical Engineering
Laboratory: AEDC/

Vol-Page No: 11- 1

Barber, Brian
BS
Electrical Engineering
Washington University
St. Louis, MO   63130-0000

Field: Electrical Engineering
Laboratory: RL/OC

Vol-Page No: 9-10

Bard, David
MS
Mechanical Engineering
Columbia University
New York, NY   10027-0000

Field: Applied Mechanics
Laboratory: PL/SX

Vol-Page No: 8-14

Barnett, William
BS
CLOUD & Aerosol Science L
University of Missouri
Rolla, MO   65401-0000

Field: Chemistry
Laboratory: PL/LI

Vol-Page No: 8- 6

Bartell, Craig
BS
Chemistry
Wright State University
Fairborn, OH   45435-0000

Field: Chemistry
Laboratory: WL/PO

Vol-Page No: 10-26

Bartelt, Timothy
MS
Physics
Clarkson University
Potsdam, NY   13676-0000

Field: Physics
Laboratory: RL/ER

Vol-Page No: 9- 4

Brocklehurst, William
BS
Aerospace Engineering
University of Cincinnati
Cincinnati, OH 45221-0000

Field: Aerospace Engineering
Laboratory: PL/RK

Vol-Page No: 8-10

Burns, Paul
BS
Electrical Engineering
Auburn University
Auburn, AL 36849-0000

Field: Electrical Engineering
Laboratory: WL/MN

Vol-Page No: 10-21

Carlen, Edwin
BS
Electrical Engineering
Oakland University
Rochester, MI 48309-4401

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 10- 1

Carter, Charity
MS
Electrical Engineering
Stevens Institute of Tech.
Hoboken, NJ   7030-0000

Field: Interdisciplinary Engr.
Laboratory: RL/IR

Vol-Page No: 9- 8

Caslin, Barry
BS
Computer Science/Engineer
Wright State University
Dayton, OH 45435-0000

Field: Computer Science
Laboratory: WL/ML

Vol-Page No: 10-14

Charley, David
BS
Electrical Engineering
University of Cincinnati
Cincinnati, OH 45221-0030

Field: Computer Science
Laboratory: WL/AA

Vol-Page No: 10- 2

Cisneros, John
BS
Psychology
California State University
Los Angeles, CA 90032-0000

Field: Psychology
Laboratory: RL/XP

Vol-Page No: 7- 8

Craig, Ralph
BS
Mathematics
North Carolina State Univ.
Raleigh, NC 27695-8205

Field: Mathematics
Laboratory: PL/VT

Vol-Page No: 8-15

GSRP Participant Data

Curtis, Joseph
MS
Zoology
University of California
Davis, CA 95616-0000

Field: Biomedicine
Laboratory: AL/OE

Vol-Page No: 7-25

Darnell, Julie
BS
Mechanical Engineering
University of Tennessee
Tullahoma, TN 37388-0000

Field: Mechanical Engineering
Laboratory: AEDC/

Vol-Page No: 11- 2

Davis, Wyatt
BS
Mechanical Engineering
Washington State University
Pullman, WA 99164-2920

Field: Mechanical Engineering
Laboratory: FJSRL/

Vol-Page No: 11-10

DeVilbiss, Alan
BS
Physics
University of Colorado
Colorado Springs, CO 80901-7150

Field: Physics
Laboratory: FJSRL/

Vol-Page No: 11-11

Detwiler, Duane
BS
Aeronautical Engineering
Ohio State University
Columbus, OH 43235-0000

Field: Aerospace Engineering
Laboratory: WL/FI

Vol-Page No: 10- 7

DiPietro, Jr., Anthony
BS
Aerospace Engineering
Virginia Polytechnic Institute
Blacksburg, VA 24060-0000

Field: Aerospace Engineering
Laboratory: AEDC/

Vol-Page No: 11- 3

Dobransky, Mary
MS
Systems Science
Binghamton University
Binghamton, NY 13902-6000

Field: Advanced Technology
Laboratory: RL/XP

Vol-Page No: 9-15

Doss, Ellen
BS
Biophysics
University of Scranton
Scranton, PA 18505-0898

Field: Biophysics
Laboratory: AL/AO

Vol-Page No: 7- 1

GSRP Participant Data

DuBois, Travis
BS
Mechanical Engineering
Old Dominion University
Norfolk, VA 23529-0247

Field: Mechanical Engineering
Laboratory: FJSRL/

Vol-Page No: 11-12

Farquhar, John
MS
Instructional Technology
University of Georgia
Athens, GA 30602-0000

Field: Instructional Technology
Laboratory: AL/HR

Vol-Page No: 7-17

Fischer, Verlyn
BS
Materials Science & Engr.
University of Florida
Gainesville, FL 32611-0000

Field: Physics
Laboratory: WL/ML

Vol-Page No: 10-15

Foster, Robert
BS
Mechanical Engineering
University of Iowa
Iowa City, IA 52240-0000

Field: Mechanical Engineering
Laboratory: WL/PO

Vol-Page No: 10-27

Frank, Jonathan
MS
Mechanical Engineering
Yale University
New Haven, CT 6520-0000

Field: Mechanical Engineering
Laboratory: WL/PO

Vol-Page No: 10-28

Fuller, Joan
BS
Chemistry
University of Alabama
Tuscaloosa, AL   -   0

Field: Chemistry
Laboratory: FJSRL/

Vol-Page No: 11-13

Gavora, Mark
MS
Instructional Systems
Florida State University
Tallahassee, FL 32304-3551

Field: Instructional Systems
Laboratory: AL/HR

Vol-Page No: 7-18

Geierman, Robert
BS
Mechanical Engineering
University of Tennessee
Tullahoma, TN 37388-0000

Field: Mechanical Engineering
Laboratory: AEDC/

Vol-Page No: 11- 4

GSRP Participant Data

Gluck, Kevin
BS
Psychology
Trinity University
San Antonio, TX 78212-0000

Field: Psychology
Laboratory: AL/HR

Vol-Page No: 7-19

Gottlob, Lawrence
MS
Psychology
Arizona State University
Tempe, AZ 85287-0000

Field: Psychology
Laboratory: AL/HR

Vol-Page No: 7-20

Griffin, Steven
BS
Engineering
University of Texas
San Antonio, TX 78249-0665

Field: Mechanical Engineering
Laboratory: PL/VT

Vol-Page No: 8-16

Grosskopf, Kevin
BS
FAC 101
University of Florida
Gainesville, FL 32611-0000

Field: School of Bldg. Construct
Laboratory: WL/FI

Vol-Page No: 10- 8

Harn, Patricia
BS
Technical Communication
University of Washington
Seattle, WA 98195-0000

Field: Communications
Laboratory: AL/HR

Vol-Page No: 7-21

Hickman, Mary
BS
Physics and Astronomy
University of Wyoming
Laramie, WY 82071-0000

Field: Astrophysics
Laboratory: PL/LI

Vol-Page No: 8- 7

Hovey III, Leland
BS
Computer Science
SUNY Institute of Technology
Utica, NY 13504-3050

Field: Mathematics
Laboratory: RL/XP

Vol-Page No: 9-16

Jain, Sanjay
BS

Northeastern Ohio University
Rootstown, OH 44272-0000

Field: Natural Sciences
Laboratory: WHMC/

Vol-Page No: 11-16

Jefferson, Ellen
BS
Biology
Trinity University
San Antonio, TX 78212-0000

Field: Biology
Laboratory: AL/OE

Vol-Page No: 7-26

Jenny, Jason
BS
Materials Science
Carnegie-Mellon University
Pittsburgh, PA 15213-0000

Field: Materials Science
Laboratory: WL/ML

Vol-Page No: 10-16

Jolly, Mohanjit
BS
Aeronautics/Astronautics
MIT
Cambridge, MA   2139-0000

Field: Aeronautics/Astronautics
Laboratory: PL/RK

Vol-Page No: 8-11

Jones, Claud
BS
Electrical & Computer Eng
University of SW Louisiana
Lafayette, LA 70504-3890

Field: Petroleum Engineering
Laboratory: RL/C3

Vol-Page No: 9- 1

Kelly, John
BS
Electrical Engineering
University of Florida
Gainesville, FL 32601-0000

Field: Electrical Engineering
Laboratory: PL/LI

Vol-Page No: 8- 8

Kimball, Lucia
MS
Mathematics
Worcester Polytechnic Inst.
Worcester, MA   1609-0000

Field: Applied Mathematics
Laboratory: PL/GP

Vol-Page No: 8- 1

Kundich, Robert
MS
899 Madison Ave., Ste 801
Univ. Tennessee-Memphis
Memphis, TN 38163-0000

Field: Biomedical Engineering
Laboratory: AL/CF

Vol-Page No: 7- 9

Kvasnak, William
MS
Mechanical Engineering
Clarkson University
Potsdam, NY 13699-5725

Field: Mechanical Engineering
Laboratory: AEDC/

Vol-Page No: 11- 5

Lair, John   Field: Industrial Design
BS   Laboratory: WL/FI
Civil Engineering
University of New Orleans   Vol-Page No: 10- 9
New Orleans, LA 70148-0000

Lee, Daniel   Field: Electrical Engineering
BS   Laboratory: RL/ER
Electrical Engineering
Southern Illinois University   Vol-Page No: 9- 5
Carbondale, IL 62901-0000

Leiweke, Robert   Field: Aerospace Engineering
BS   Laboratory: PL/WS
Aeronautical Engineering
Ohio State University   Vol-Page No: 8-20
Columbus, OH 43210-0000

Litvin, Kerry   Field: Electrical Engineering
MS   Laboratory: RL/OC
Electrical Engineering
Cornell University   Vol-Page No: 9-11
Ithaca, NY 14853-0000

Luker, Stephen   Field: Marine Science
MS   Laboratory: PL/GP
Biological Science
University of Alabama   Vol-Page No: 8- 2
Tuscaloosa, AL 35487-0344

Masterson, Gina   Field: Industrial Engineering
BS   Laboratory: AL/OE
Industrial Engineering
Auburn University   Vol-Page No: 7-27
Auburn, AL   -   0

McMurtry, J.   Field: Mechanical Engineering
MS   Laboratory: WL/MN
Mechanical Engineering
Louisiana Tech University   Vol-Page No: 10-22
Ruston, LA 71272-0046

Menendez-Barreto, Melani   Field: Physics
MS   Laboratory: PL/GP
Physics
University of Puerto Rico   Vol-Page No: 8- 3
Mayaguez, PR   681-0000

GSRP Participant Data

Messerschmitt, Jane
MS
Electrical Engineering
Polytechnic University
Farmingdale, NY   11735-0000

Field:
Laboratory: PL/WS

Vol-Page No: 8-21

Mills, Jeffrey
MS
Cheistry
Indiana University
Bloomington, IN   47405-0000

Field: Chemistry
Laboratory: PL/RK

Vol-Page No: 8-12

Monka, Gary
BS
Chemistry
University of Scranton
Scranton, PA   18510-0000

Field: Chemistry
Laboratory: AL/OE

Vol-Page No: 7-28

Moore, Michael
MS
Electrical Engineering
Vanderbilt University
Nashville, TN   37235-0000

Field: Electrical Engineering
Laboratory: AEDC/

Vol-Page No: 11- 6

Moriarty, Daniel
MS
Nuclear Engineering
MIT
Cambridge, MA   2139-0000

Field: Nuclear Engineering
Laboratory: PL/GP

Vol-Page No: 8- 4

Nagaraja, Chandra
BS
Mechanical Engineering
University of Dayton
Dayton, OH   45469-0000

Field: Mechanical Engineering
Laboratory: WL/PO

Vol-Page No: 10-29

Naghski, David
MS
Electrical Engineering
University of Cincinnati
Cincinnati, OH   45221-0000

Field: Electrical Engineering
Laboratory: WL/ML

Vol-Page No: 10-17

Neaffer, Ronald
BS
Polymer Science
University of Akron
Akron, OH   44325-3909

Field: Physics
Laboratory: WL/ML

Vol-Page No: 10-18

GSRP Participant Data

Newell, Tim
MS
Physics
University of North Texas.
Denton, TX 76203-0000

Field: Physics
Laboratory: PL/LI

Vol-Page No: 8- 9

Nguyen, Sonny
BS
Electrical Engineering
Wright University
Dayton, OH 45401-0000

Field: Electrical Engineering
Laboratory: WL/PO

Vol-Page No: 10-30

Nuteson, Todd
BS
Electrical Engineering
Wright State University
Dayton, OH 45435-0000

Field: Electrical Engineering
Laboratory: WL/EL

Vol-Page No: 10- 4

Olson, Michael
BS
Electrical Engineering
Florida State University
Tallahassee, FL 32316-0000

Field: Electrical Engineering
Laboratory: WL/MN

Vol-Page No: 10-23

Panek, Heather
MS
Biology
University of Scranton
Scranton, PA 18510-0000

Field: Biochemistry
Laboratory: AL/AO

Vol-Page No: 7- 3

Petroziello, Joseph
BS
Biology
University of Scranton
Scranton, PA 18510-0000

Field: Biology
Laboratory: AL/AO

Vol-Page No: 7- 4

Pobst, Jeffrey
BS
Aerospace Engineering
University of Southern Calif.
Los Angeles, CA 90089-1191

Field: Aerospace Engineering
Laboratory: PL/RK

Vol-Page No: 8-13

Povich, Clinton
BS
Aerospace Engineering
University of Kansas
Lawrence, KS 66045-0000

Field: Aerospace Engineering
Laboratory: FJSRL/

Vol-Page No: 11-14

GSRP Participant Data

Radomsky, Leon
MS
Chemical Engineering
Columbia Univeristy
New York, NY   10027-0000

Field: Materials Science
Laboratory: WL/EL

Vol-Page No: 10- 5

Reed, Jason
BS
Electrical Engineering
Cornell University
Ithaca, NY   14853-0000

Field: Physics
Laboratory: RL/OC

Vol-Page No: 9-12

Reichmeyer, Francis
MS
Electlrical Engineering
Syracuse University
Syracuse, NY   13244-1240

Field: Electrical Engineering
Laboratory: RL/C3

Vol-Page No: 9- 2

Reiley, Daniel
MS
Physics
University of Alabama
Huntsville, AL   35899-0000

Field: Physics
Laboratory: WL/MN

Vol-Page No: 10-24

Reister, Craig
BS
Electrical Engineering
North Dakota State University
Fargo, ND   58105-0000

Field: Electrical Engineering
Laboratory: AL/AO

Vol-Page No: 7- 5

Reuster, Daniel
MS
Electrical Engineering
University of Dayton
Dayton, OH   45469-0000

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 10- 3

Riesinger, Joyce
BS
Ecology/Evolutionary Biol
University of West Florida
Pensacola, FL   32514-0000

Field: Marine Biology
Laboratory: AL/EQ

Vol-Page No: 7-14

Ryan, Arthur
BS
Psychology
Wright State University
Dayton, OH   45435-0001

Field: Psychology
Laboratory: AL/CF

Vol-Page No: 7-10

GSRP Participant Data

Samad, Mohammed
BS
Mechanical Engineering
University of New Orleans
New Orleans, LA 70148-0000

Field: Mechanical Engineering
Laboratory: WL/FI

Vol-Page No: 10-10

Schroeder, Mark
BS
Electrical Engineering
North Dakota State University
Fargo, ND 58105-0000

Field: Electrical Engineering
Laboratory: AL/AO

Vol-Page No: 7- 6

Schwing, Ronald
BS
Mechanical Engineering
University of Iowa
Iowa City, IA 52242-0000

Field: Mechanical Engineering
Laboratory: WL/PO

Vol-Page No: 10-31

Seifert, Jason
BS
Psychology
University of Dayton
Dayton, OH 45410-0000

Field: Psychology
Laboratory: AL/HR

Vol-Page No: 7-22

Seydel III, Robert
BS
Aerospace Engineering
University of Missouri
Rolla, MO 65401-0000

Field: Aerospace Engineering
Laboratory: WL/FI

Vol-Page No: 10-11

Shahidi, Anoosh
MS
Intelligent Systems Studi
University of Pittsburgh
Pittsburgh, PA 15260-0000

Field: Intelligent Systems
Laboratory: AL/HR

Vol-Page No: 7-23

Shahrokhi, Kimball
BS
Mechanical Engineering
Vanderbilt University
Nashville, TN 37232-0000

Field: Mechanical Engineering
Laboratory: AEDC/

Vol-Page No: 11- 7

Slater III, Robert
MS
Mechanical Engineering
University of Cincinnati
Cincinnati, OH 45221-0000

Field: Engineering
Laboratory: WL/FI

Vol-Page No: 10-12

Socci, Ed
MS
Materials Science & Engr
University of Virginia
Charlottesville, VA 22903-2442

Field:
Laboratory: WL/ML

Vol-Page No: 10-19

Sodoy, Maureen
BS
Biology
Trinity University
San Antonio, TX 78212-0000

Field: Biology
Laboratory: AL/OE

Vol-Page No: 7-29

Sorensen, Bryant
MS
Electrical Engineering
University of Tennessee
Tullahoma, TN 37388-0000

Field: Mathematics
Laboratory: AEDC/

Vol-Page No: 11- 8

Spencer, Robert
BS
Electrical Engineering
Cornell University
Ithaca, NY 14853-0000

Field: Electrical Engineering
Laboratory: RL/OC

Vol-Page No: 9-13

Stadler, Bethanie
BS
Materials Science
Massachusetts Inst. of Tech.
Cambridge, MA 2139-0000

Field: Materials Science
Laboratory: RL/ER

Vol-Page No: 9- 6

Stansbery, Donald
BS
Aerospace Engineering
University of Missouri
Rolla, MO 65401-0000

Field: Aerospace
Laboratory: WL/MN

Vol-Page No: 10-25

Starchville, Jr., Thomas
MS
Aerospace Engineering
Pennsylvania State University
University Park, PA 16802-0000

Field: Aerospace Engineering
Laboratory: WL/FI

Vol-Page No: 10-13

Starks, Michael
BS
Electrical Engineering
Boston University
Boston, MA 2215-0000

Field: Electrical Engineering
Laboratory: PL/GP

Vol-Page No: 8- 5

GSRP Participant Data

Stauffer, Joseph
MS
Management
University of Iowa
Iowa City, IA 52242-1323

Field: Management
Laboratory: AL/HR

Vol-Page No: 7-24

Stenger, Vincent
BS
Electrical Engineering
University of Cincinnati
Cincinnati, OH 45221-0030

Field: Electrical Engineering
Laboratory: WL/EL

Vol-Page No: 10- 6

Stohs, Jonathan
BS
Physics
University of New Mexico
Albuquerque, NM 87131-0000

Field: Math
Laboratory: PL/VT

Vol-Page No: 8-17

Stroman, Ronald
BS
Pharmacy
University of South Carolina
Columbia, SC 29208-0000

Field: Pharmacy
Laboratory: AL/AO

Vol-Page No: 7- 7

Sullins, Tashia
BS
Chemistry
University of Georgia
Athens, GA 30605-0000

Field: Chemistry
Laboratory: AL/EQ

Vol-Page No: 7-15

Sutcliffe, Ronald
BS
Chemistry
Southwest Texas State
San Marcos, TX 78666-0000

Field: Chemistry
Laboratory: AL/OE

Vol-Page No: 7-30

Sweeney, Walter
MS
Electrical Engineering
University of Maine
Orono, ME   4469-0000

Field: Electrical Engineering
Laboratory: RL/IR

Vol-Page No: 9- 9

Swindal, J.
MS
Applied Physice
Yale University
New Haven, CT   6520-0000

Field: Applied Physics
Laboratory: WL/PO

Vol-Page No: 10-32

GSRP Participant Data

Tascillo, Mark
MS
Electrical Engineering
Binghamton University
Binghamton, NY 13902-0000

Field: Electrical Engineering
Laboratory: RL/OC

Vol-Page No: 9-14

Thomas, Stuart
BS
Chemical Engineering
North Carolina State Univ.
Raleigh, NC 27695-0000

Field: Chemical Engineering
Laboratory: AL/EQ

Vol-Page No: 7-16

Tipton, Kevin
MS
Nutrition and Food Scienc
Auburn University
Auburn, AL 36849-5605

Field: Zoology
Laboratory: AL/CF

Vol-Page No: 7-11

Tornow, Carina
BS
Quantitive Psychology
Ohio State University
Columbus, OH 43210-0000

Field: Psychology
Laboratory: AL/CF

Vol-Page No: 7-12

Trick, Kimberly
MS
Chemical Engineering
University of Dayton
Dayton, OH 45401-0240

Field: Chemical Engineering
Laboratory: WL/ML

Vol-Page No: 10-20

Weaver, Michael
MS
Aerospace Engineering
Georgia Institute of Technolog
Atlanta, GA 30332-0150

Field: Aerospace Engineering
Laboratory: AEDC/

Vol-Page No: 11- 9

Weitzman, Peter
MS
Thayer School of Engineer
Dartmouth College
Hanover, NH   3755-0000

Field: Electrical Engineering
Laboratory: FJSRL/

Vol-Page No: 11-15

Whitmore, Christine
BS
Mechanical Engineering
Washington University
St. Louis, MO   -   0

Field: Mechanical Engineering
Laboratory: AL/CF

Vol-Page No: 7-13

GSRP Participant Data

Widener, Charles
MS
Electrical Engineering
Syracuse University
Syracuse, NY 13244-0000

Field: Electrical Engineering
Laboratory: PL/VT

Vol-Page No: 8-18

Wierman, Mark
MA
Systems Science
Binghamton University
Binghamton, NY 13902-6000

Field: Mathematics
Laboratory: RL/XP

Vol-Page No: 9-17

Willemsen, Balam
MS
Physics
Northeastern University
Boston, MA   2115-0000

Field: Physics
Laboratory: RL/ER

Vol-Page No: 9- 7

Wolfe, Randall
MS
Biopsychology
University of Georgia
Athens, GA 30612-0000

Field: General Psychology
Laboratory: AL/OE

Vol-Page No: 7-31

Worthy, Mark
BS
Environmental Engineering
University of Alabama
Huntsville, AL 35899-1750

Field: Electrical Engineering
Laboratory: PL/WS

Vol-Page No: 8-22

A FAST ATM ROUTING ALGORITHM
FOR THE DYNAMIC THEATER ENVIRONMENT

Claud K. Jones

and

Robert R. Henry
Professor

Department of Electrical & Computer Engineering

University of Southwestern Louisiana
P.O. Box 43890
Lafayette, LA 70504-3890

A FAST ATM ROUTING ALGORITHM FOR THE DYNAMIC THEATER ENVIRONMENT

Claud K. Jones

and

Robert R. Henry
Professor

Department of Electrical & Computer Engineering
University of Southwestern Louisiana

ABSTRACT

Traditional tactical and theater military communication networks are characterized by relatively low bandwidth links. The environment is dynamic in the sense that the links are subject to jamming and the nodes to destruction by the enemy. Modern and future military equipment and tactics require the use of wideband links to exchange bandwidth-intensive information such as video and images. However, current and proposed wideband networks such as ATM have been designed for peacetime, i.e. well-behaved operation. The research described herein proposes and develops an original Fast ATM Routing (FAR) protocol which adapts wideband ATM networks for operation in the dynamic theater environment. It is shown that the FAR protocol provides robust performance in such an environment.

A FAST ATM ROUTING ALGORITHM
FOR THE DYNAMIC THEATER ENVIRONMENT

Claud K. Jones and Robert R. Henry

INTRODUCTION

The research performed this summer by the authors consists of the development and evaluation of a protocol to suitably adapt the Asynchronous Transfer Mode (ATM) implementation of integrated broadband networking to the requirements of a Theater Environment Network (TENet). This research area was previously studied [1] to identify ways in which the advantages of highly adaptable dynamic routing protocols could be combined with the broadband capacity of ATM. This report is divided into four parts. The first describes the relationship between this research effort and other research efforts now being conducted or recently completed at Rome Labs. The second describes the protocol developed by the authors this summer. The third presents the results of a deterministic performance analysis of the protocol, and the fourth describes and presents the results of a detailed simulation of the protocol using the OPNET simulator.

RELATIONSHIP TO CURRENT PROJECTS

Current Projects

The research performed by the authors this summer was chosen to complement and support current research efforts undertaken by Rome Labs. One such project is the Evaluation and Development of Multimedia Networks Under Dynamic Stress (EDMUNDS) project [2]. It is an ongoing effort to develop Secure Tactical Internet Protocols (STIP)

that respond well to a stressed tactical environment. Such an environment is characterized by communication links subject to jamming and nodes subject to destruction by the enemy. The assumption is that the links have relatively low bandwidth, and that there is sufficient time to perform sophisticated processing to determine optimal datagram routing.

The Secure Survivable Communication Network (SSCN) project [3] was initiated in an effort to utilize the emerging Broadband Integrated Services Digital Network (B-ISDN) for military applications. This network relies on the Asynchronous Transfer Mode (ATM) to provide rapid multiplexing and routing of data over wide bandwidth links. Due to high data rates there is relatively little time to determine the optimal route of each packet (cell). Therefore connection oriented services have been selected to provide routing that is fixed for the call duration. This is in direct contrast with the STIP routing protocol.

Two other projects, the Media Resource Controller (MRC) [4] and the Multimedia Communication Capability (M2C2) [5], are similar to the EDMUNDS project in the sense that the routing is adaptive and the links are low bandwidth. However, these projects have actually been implemented and demonstrated in the laboratory. Another laboratory oriented project is the Advanced Multi-Media Information Distribution System (AMIDS) [6] in which a distributed Tactical Air Control Center (TACC) is interconnected by fixed wideband links.

ATM and AMIDS have high capacity links in contrast with the lower capacity MRC and STIP links. Each may be classified into one of two groups based on link capacity and the ability to adapt to a stressed environment as follows:

- ATM/AMIDS: high capacity, low adaptability
- STIP/MRC/M2C2: low capacity, high adaptability.

Interconnection between these two dissimilar groups requires careful consideration so that performance does not revert to the least common denominator, i.e. low capacity and low
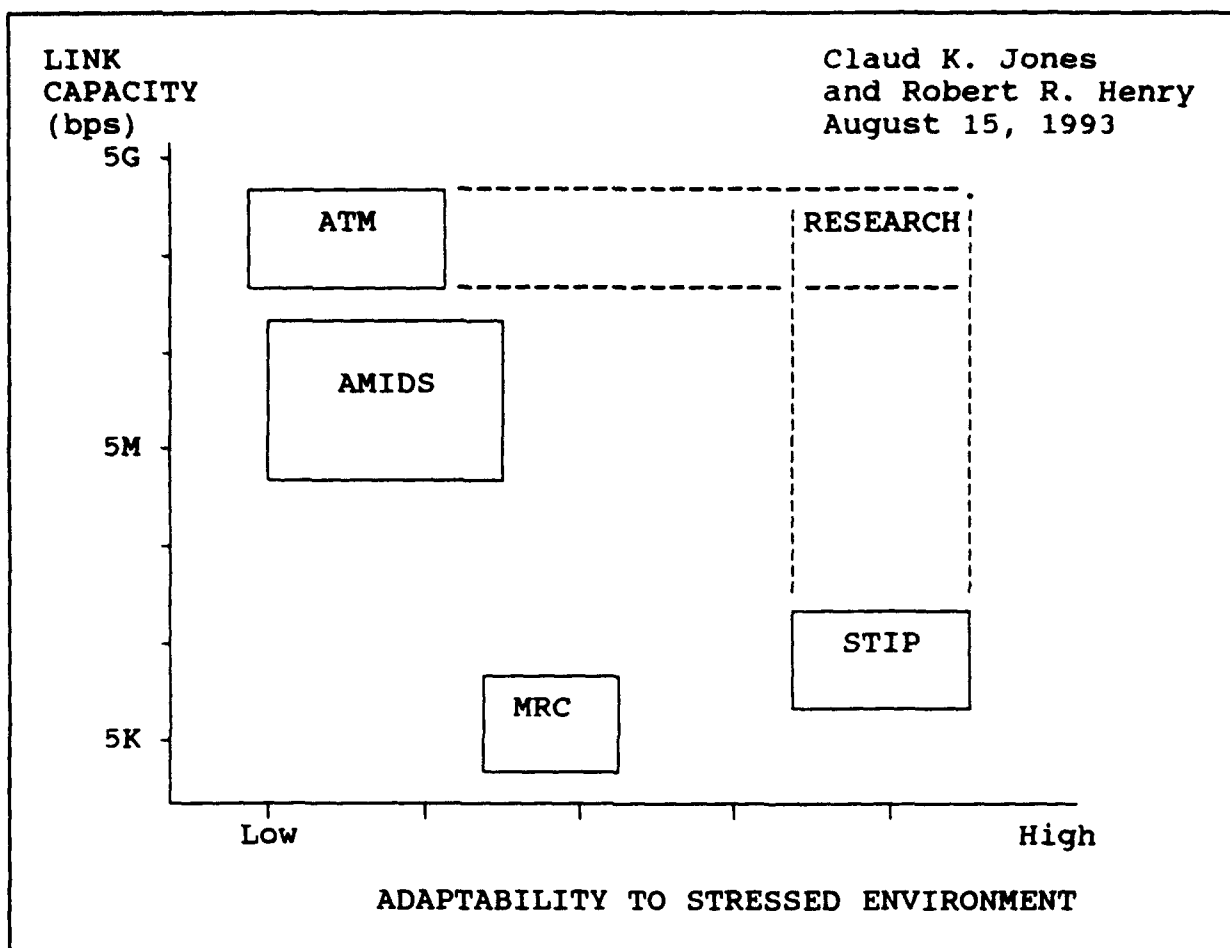
Claud K. Jones
and Robert R. Henry
August 15, 1993

5G

ATM

RESEARCH

AMIDS

5M

STIP

MRC

5K

Low   High

ADAPTABILITY TO STRESSED ENVIRONMENT

Figure 1. Network Protocol Dynamics.

adaptability.

Figure 1 illustrates the networks in these two groups in more detail. As can be seen ATM represents the highest in link capacity, while STIP represents a high degree of adaptability to a stressed environment.

This Research

An underlying concept of the global grid concept is that the National Command Authority (NCA) will be able to communicate directly with personnel in the tactical environment. In order that graphics and video information be delivered in a timely manner, high capacity links are essential. ATM is a candidate for the wide-area part of this network

since it is wideband, is being developed commercially, and is being standardized in international forums. However, since ATM is designed for peacetime (well-behaved) use, adaptive protocols, such as those incorporated into STIP, must also be incorporated into ATM-based networks to provide for survivability in the tactical area. This report documents research by the authors that formulates a protocol which combines the advantages of both the ATM and adaptive (i.e. STIP-like) protocols into a network suitable for the military environment. This is the "RESEARCH" area indicated in Figure 1.

THE FAST ATM REROUTING PROTOCOL

Commercial broadband networks are expected to use very reliable, extremely low noise optical fiber links. Given this assumption, error recovery may feasibly be removed from the internal operation of the network to the network edges. Military networks operating in a theater environment cannot make this assumption. The US military may be called upon to operate in areas far removed from modern networks and so will continue to rely upon links, characterized by high bit error rates, for its communication requirements. Military communication networks are subject as well to the fluid nature of combat as exemplified by frequent network reconfiguration and hostile action.

The thrust of this research is to find ways in which broadband networks based on ATM may be adapted to the dynamic theater environment. The authors assert that fault recovery should not be removed to the network edges in the theater environment but must remain internal to the network. It will be shown that the Fast ATM Rerouting (FAR) protocol proposed by the authors reduces the disruption to the network users due to failure. It does so by automatically rerouting to an alternate virtual path connection (VPC) route in a way that is functionally compliant with ATM standards.

ALGORITHM DESCRIPTION

The purpose of FAR is to overcome some of the problems associated with high speed networks when disruptions in the network occur. Unlike today's relatively low speed networks, propagation delay dominates in broadband networks. This latency means that a very large amount of data (on the order of 10^6 bits per 1000 km, or about 1500 ATM cells) may be in transit from source to destination. In-theater networks must be prepared for frequent disruptions, or rapid and unpredictable changes in their state, resulting in large losses of data. FAR is a step towards minimizing the potential loss of data by allowing the network to sense and react as quickly as possible to these disruptions. This paradigm places in each node, or strategically selected nodes, the ability to assess the quality of its communication with its neighbors and to automatically reconfigure appropriate VPCs.

The FAR Paradigm

When a node determines that its ability to communicate with a neighbor, i.e. the quality of service (QOS), has degraded below some threshold (a "fault" occurs), it will take prompt action to activate an alternate path. The QOS could involve parameters such as bandwidth, latency, security, etc.

The method proposed here maintains alternate standby paths (either hot standby or dormant) that can be activated and used very quickly when a fault is observed. Figure 2 illustrates one such primary VPC with one, pre-defined alternate path branching from node I. Other alternate paths may exist in addition to this one but are not shown in order to simplify the diagram. This representative VPC is also shown in isolation from other VPCs
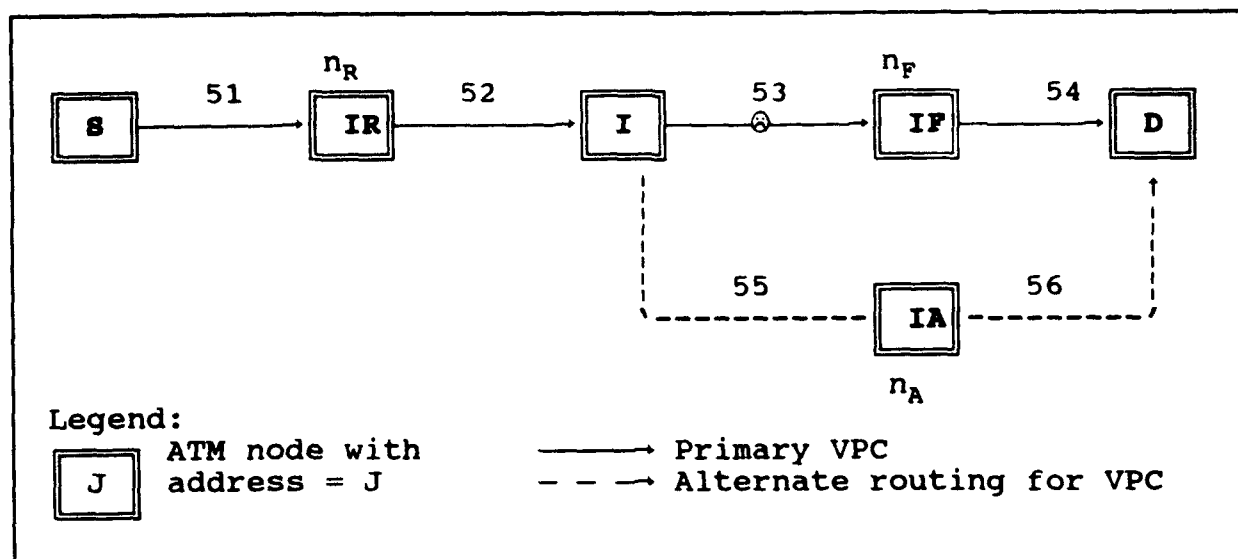
Figure 2. Primary VPC and alternate for node I.

traversing the same nodes. It is defined by source node S, destination node D, intermediate node I , n_R upstream (relative to node I) intermediate nodes IR, n_F downstream intermediate nodes IF, and n_A alternate path intermediate nodes IA. It is also assumed that reverse VPCs exist from node D to node S along the primary and alternate paths.

The typical intermediate node I has detected a fault immediately downstream of itself. In response, node I takes the necessary steps to reestablish the VPC to D along the n_A alternate set of nodes IA. The new route continues to use the n_R IR nodes upstream from the failure, but bypasses the n_F downstream forward path nodes IF downstream of the failure.

In addition to the rerouting mechanism to be described in this paper, it is assumed that there are two other processes running asynchronously in the background at each node. One is the fault detection/reporting function mentioned in the previous paragraph. The second provides the alternate route to be used when a link failure is detected. Such routes may be updated relatively slowly compared to the time in which the node is required to switch to the alternate route. The selection of the alternate routes is an area for future

study and is beyond the scope of this paper.

An advantage of FAR is that most resources are consumed only when needed, with VPI resources being consumed to maintain alternate paths. Bandwidth remains available to other, active VPCs. FAR is also distributed with each node taking action based on OAM (Operation and Maintenance) cells exchanged with its neighbors allowing it to recover from a fault in any link or node. A two-way handshake is provided from the intermediate node I to D and back. However, node I may opt for an early reroute immediately upon detection of the fault before confirmation is received from the handshake as to the availability of resources on the alternate path.

Assumptions for Protocol Implementation

The goal of the research performed in the summer of 1993 at Rome Labs was to fully implement and test a portion of the FAR protocol. The limitation of time dictated that we adopt simplifying assumptions to enable us to reach this goal. First only one alternate path branching at the node just upstream of the fault is assumed. A hard, yes/no decision is made on QOS parameters, ie there is an abrupt transition from acceptable to non-acceptable QOS. Only one try is made at rerouting with the source being notified of the attempt and the status of the attempt. No timeouts are provided for lost OAM cells. A two-way handshake is used from the node initiating recovery to the destination node. No handshake is used upstream from the initiating node but notification of action taken is sent. We stress that these assumptions are not inherent to the FAR protocol but were made only in the interest of reducing the time to achieve measurable results from the protocol.

1-9

OAM Cells

Node I could assess a fault through the use of F1, F2 and F3 (OAM) cells as defined by the Consultative Committee on Telephony and Telegraphy (CCITT) [7]. Further definition by the CCITT and other standards bodies is required before the suitability of using these OAM cells can be ascertained. Therefore, we have formulated generic type 0, 1, 2 and 3 cells that perform specific functions required of the FAR. Type 0 cells detect and report fault conditions between ATM devices. Cell types 1, 2 and 3 are used to recover from the fault condition. The functions of each cell type are described below:

- Type 0   Detects faults and reports on the status of the VPC between nodes I and IF.

- Type 1   Sent to node IA along the alternate route (downstream of I and bypassing the fault) and forwarded until D is reached or the alternate path also fails. It changes the alternate path to primary route status for the VPC in question.

- Type 2   Sent to node S to inform it of the VPC status, "rerouting", "rerouted", or "alt_path_fail".

- Type 3   Sent to node I by node *D* upon receipt of a type 1 cell or by one of the intermediate alternate path nodes IA if the alternate path can not be activated for some reason. It informs I if the alternate path was successfully activated.

Protocol Description

The FAR protocol can be viewed as a Finite State Machine (FSM) residing at or contained within an ATM node. Seven high level states are presently specified as shown in Figure 3. Lower level states may be contained within each high level state.
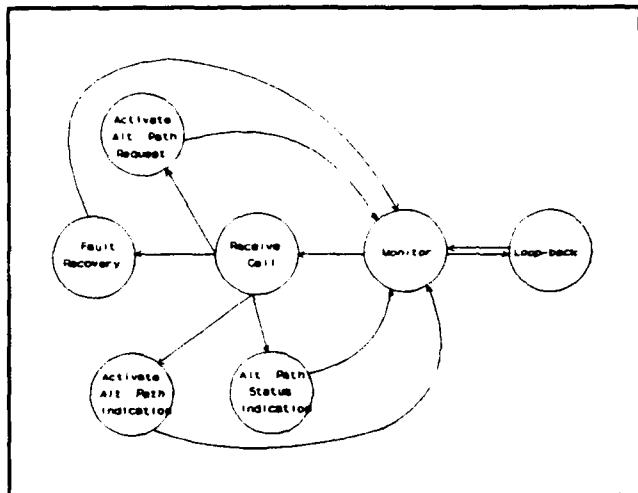
Figure 3. Current FSM for the FAR Protocol.

The loop-back state consists of the exchange of type 0 cells with each neighboring node. Faults are indicated by status conditions returned from downstream or by a timeout. The exact operation in this state is not yet defined pending further definition of standards. Only faults immediately downstream of the subject node are currently detected.
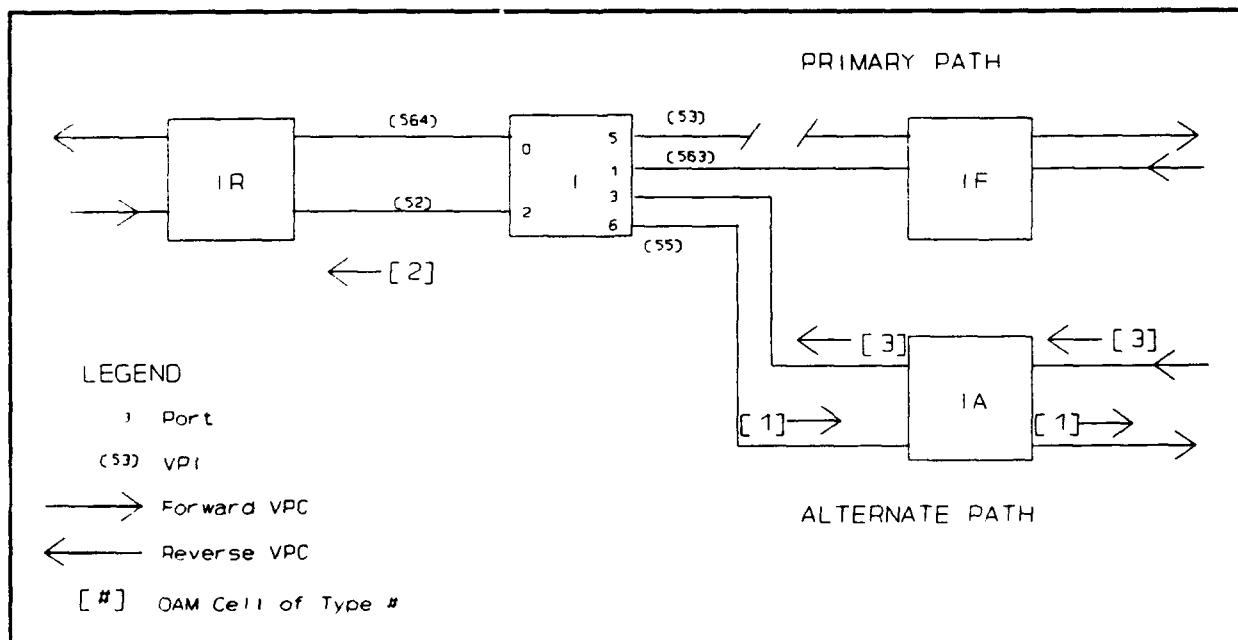
Figure 4. VPC schematic and OAM cell activity.

Figure 4 illustrates the VPCs in the vicinity of node I and the OAM cell traffic generated as a result of the fault. The outgoing port and VPI to route cells along the alternate path are contained in the alternate path output port/VPI table entry for that VPC as shown in Figure 5. In the fault recovery state the FSM first checks to see if adequate resources, (ie bandwidth) is available to begin the activation of the alternate path. If so, a type 1 cell is sent to the next downstream node IA, which then forwards the cell to the next node IA and so on until node D is reached. A type 2 cell is sent directly to the source node S indicating that the network is attempting to reroute the VPC.

VPC Index	CURRENT VPC			ALT PATH VPC		VPC PARAMETERS		Rev. VPC
	IN	OUT	BW	OUT	BW	Src_addr	Dest_addr	OUT
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
I(f)	2/52	5/53	Bc	6/55	Ba	S	D	0/564
I(r)	1/563	0/564	Bc	--	--	D	S	5/53

Legend:
#/###   port/VPI entry for a VPC
I(f)   forward VPC index
I(r)   reverse VPC index
Bc   Bandwidth parameters set for the current VPC
Ba   Bandwidth parameters acceptable if alternate
          path is activated
Src_addr Address of node connected to source
Dest_addr Address of node connected to destination

Figure 5. Expanded VPC table.

If the resources are not available, a type 2 cell only is sent, again to node S, indicating the failure of both the primary and alternate paths. In the representative situation shown in Figure 2, node I has entered the fault recovery state, transmitted a type 1 and type 2 cell and has returned to the monitor state. At a later time a type 3 cell will

return to node I from node D or one of the nodes IA causing it to enter the alternate path indication state. Node S enters the alternate path status indication state on receiving the type 2 cell. The actions that may be taken by the source could include temporarily suspending transmission, preparing for retransmission of possibly lost data, or any other number of actions.

An "early reroute" option is available in the fault recovery state to reroute the VPC immediately after the type 1 cell is sent. Doing so has the advantage of potentially reducing the number of cells buffered or lost. The disadvantage is that if resources are not available downstream congestion problems may occur for other VPCs traversing the same nodes. Evaluating the ramifications of this option is beyond the scope of this paper and is for future research. By contrast the "late reroute" option waits on the acknowledgement from the destination before rerouting.

The activate alternate path request state is entered upon receipt of a type 1 OAM cell. The type 1 cell arrives over the user VPC but is distinguished by a reserved VCI or cell type indicator or both [8]. The receiving node activates the resources requested by the type 1 cell if possible and checks the destination address contained in the data field of the cell to see if it is the end node for that VPC. If it is the destination node and successfully allocated resources to the VPC, it issues a type 3 cell along the reverse VPC defined in Figure 5. If is not the destination node, it sends the cell along the next link in the alternate, forward VPC. If it can not allocate the requested resources to the alternate VPC, it will issue a type 3 cell along the reverse, alternate VPC with an indication that the alternate path has also failed. Figure 4 shows node IA entering the activate alternate path request state, successfully allocating bandwidth, and sending the type 1 cell along the alternate, forward VPC.

When a node receives a type 3 cell, it enters the activate alternate path indication

1-13

state. Figure 4 shows node IA receiving a type 3 cell from downstream, which would occur at some time after forwarding the original type 1 cell on to node D, and forwarding the cell to node I. If the data field of the cell contains an indicator that the alternate, forward VPC was successful node I immediately sends a second type 2 cell to node S indicating that the alternate, forward path was successful. If any node receives a type 3 cell with an indication that the alternate VPC failed, it immediately deallocates the alternate VPC and sends the cell back along the reverse alternate VPC. The type 3 cell contains the destination address of node I so that node I will know the status of the alternate path and will not forward the cell.

DETERMINISTIC PERFORMANCE ANALYSIS

Parametric equations of performance of the FAR protocol have been derived for the representative environment depicted in Figure 2. Two performance metrics are determined, the time to affect early reroute and the time to affect late reroute. A discussion of the resulting equations is beyond the scope of this paper. For this analysis, a VPC 5000 km long was assumed consisting of six, equally spaced ATM nodes. Node one is the source end-point, node 6 is the destination end-point. The VPC fails 3500 km from node 1 placing the fault 500 km downstream of node 4, or half-way between nodes 4 and 5. This makes node 4 responsible for recovery. A reverse direction VPC is assumed between node 1 and node 6. An alternate path consists of node 7 being equidistant between nodes 4 and 6. All links are equidistant at 1000 km. An estimate is made of the number of clock cycles required to perform the protocol operations, and the propagation speed is assumed to be 0.8c. All links assume OC-3 capacity yielding 148.608 Mbps and a cell transmission delay of 2.853 μsec. Propagation delay between switches is 4.167 msec.

Table I. Selected results of deterministic analysis.

VPC Length	# Nodes	Fault Location	Xmt Speed	Clock Speed	Early reroute	Late reroute	ETE early Reroute	ETE later Reroute
km		km	Mbps	MHz	msec	msec	msec	msec
5000	6	3500	148	20	9	25	42	84
5000	6	3500	148	1	9	25	42	84
5000	6	3500	10	10	9	26	43	86
5000	6	0	148	10	9	50	42	84
5000	0	3500	148	10	21	42	42	84
5	6	3.5	148	10	.02	.04	.07	.14
77000	3	53900	148	10	350	620	620	1290
5000	20	3500	148	10	2	16	42	84

The results of the FAR analysis are compared with the end-to-end (ete) protocol that relies on the end-points only to recover. The nominal case assumed two forward VPCs, both 5000 km long with one designated as the alternate path. Various parameters were varied and the results are shown in Table I. In short, the FAR protocol (internal rerouting) performs significantly better than rerouting performed at the network edges. The protocol is insensitive to processing time unless the nodes are very close together as in a LAN configuration.

For VPC lengths ranging from 5 km to 77000, the FAR protocol performs two to three times faster than end-to-end recovery. The distance from a node to a fault is one of the prime determinants in time to reroute. The closer the node spacing, the faster the fault

will be detected. The end-to-end and the FAR protocols converge when a VPC consists only of the two endpoints. Conversely, for a VPC traversing several nodes, the FAR method performs proportionally faster.

Another trend is for the FAR protocol to perform much better if the fault occurs near the downstream end of the VPC rather that the transmitting end. However, if the early reroute option is used, fault location has no effect on performance. Transmission speed was decreased to 10 Mbps having little effect on time to reroute with the exception of short VPCs (LANs). At speeds of 10 Mbps, performance for the FAR protocol remained proportionally faster. All methods had a sharp increase in reroute time due to transmission delay becoming significant relative to propagation delay.

OPNET ANALYSIS
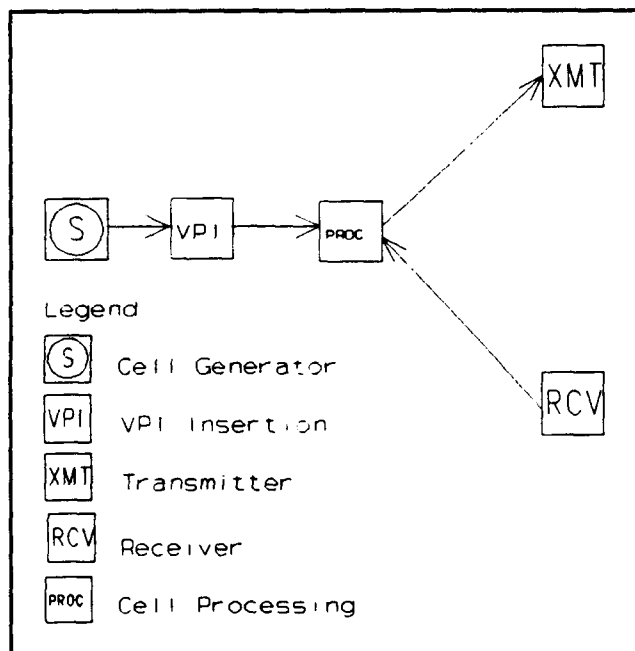
Figure 6. Diagram of OPNET model transceiver.

An OPNET simulation of the FAR protocol was implemented to confirm the correctness of the protocol, and to validate the results of the deterministic performance analysis. OPNET is an event driven simulator which provides general purpose routines for creating and destroying packets, determination of time delays, as well as for user defined functions code written in proto-C. Protocols in OPNET are modeled by a FSM, with Figure 3 showing one developed for the FAR protocol. FSMs are associated with

1-16

processes which are linked to each other via *streams* into the next highest level in the OPNET hierarchy, *nodes*. Figure 6 and Figure 7 show the FAR transceiver and ATM switch *nodes*. An OPNET function named op_pk_send(*index*) is available to send a packet over a stream with an identifier of *index*. OPNET is responsible for delivering the packet to the FSM (process) at the receiving end of the stream. The pre-defined processes queues, point-to-point transmitters and receivers, and sources are utilized in this simulation.
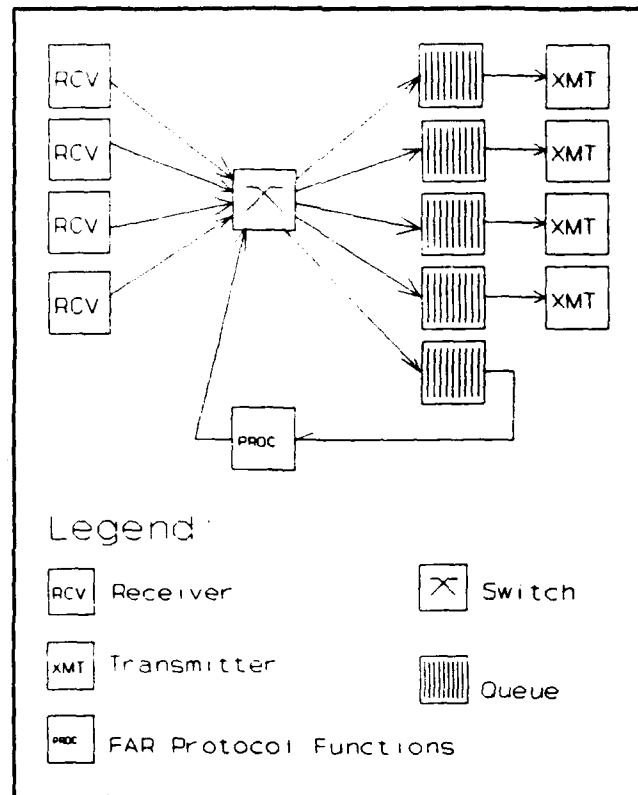
Figure 7. OPNET model of ATM node.

Figure 8 shows the *subnet* of transceiver and ATM node type *nodes* used in the FAR model configuration. The links are representations of physical links for which propagation delay and bit error rate (BER) can be specified. Propagation delay was set at 4.1667 μsec (for a 1000 km link) and BER was not used. Three cases were run, 1 km internodal spacing, 1000 km internodal spacing and 38500 km internodal spacing (3 nodes). These cases emulate a LAN, WAN and satellite system respectively. Delay to reroute once notified of a fault was measured and is in excellent agreement with the analytical results except for the LAN case. This is because the OPNET simulator handles each cell several more times than was assumed in the deterministic analysis. Thus the difference is not noticeable unless propagation delay is very small as in the LAN case. The differences in the two analyses corresponds to the difference in the number of times each cell is handled in the two analyses. Each time a cell is handled adds 2.85 μsec to its total delay and each ATM node in the model handles the cell up to six times (receive, transmit, and 4 queueings) while the deterministic analysis assumed only

one handling.
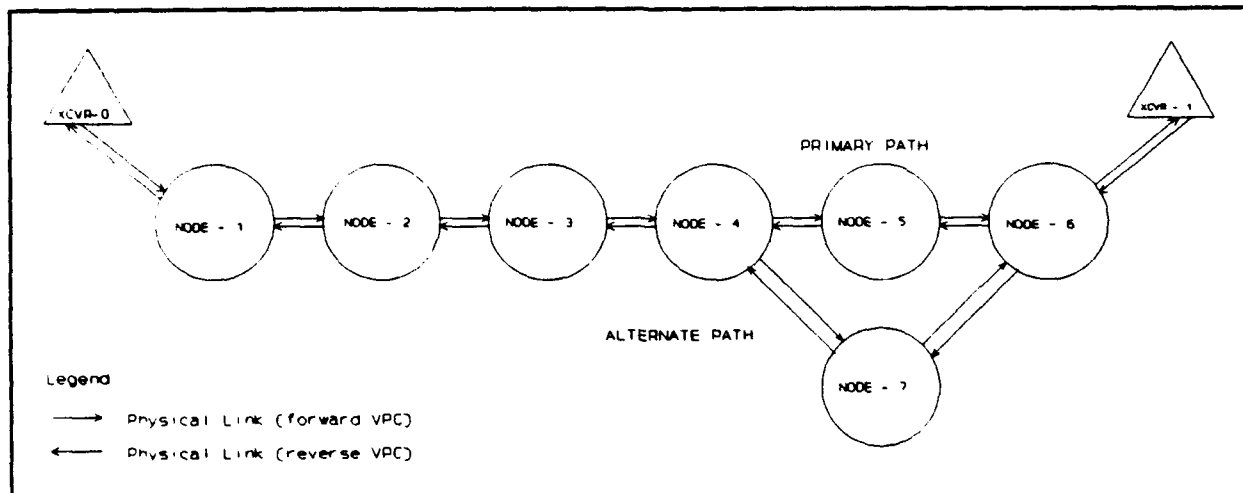
Figure 8. OPNET model of the VPC.

CONCLUSIONS

The authors have developed an original, ATM compliant rerouting protocol, FAR, the Fast ATM Rerouting protocol. It exhibits robust performance over a wide range of parameters such as processing speed, transmission speed, node spacing and VPC length. The protocol has been implemented and verified using the OPNET simulator and compliments and ties together other RL/C3D research efforts.

REFERENCES

1. R.R. Henry, "Wideband ATM Networks with Adaptive Routing for the Dynamic Theater Environment," Final Report AFOSR Summer Research Program, Rome Laboratory, August 1992, page 6-10.

2. "Evaluation and Development of Multimedia Networks in Dynamic Stress," Semiannual Project Report, SRI contract with Rome Laboratory, April 5-9, 1992.

3. "Secure Survivable Communication Network," Users Meeting Report, GTE Government Systems contract with Rome Laboratory, July 15-16, 1992.

4. M.T. Rafter and W.C. Walker, "A Distribution Multiple Media Network Architecture," MILCOM '89 Conference Record, October 15-18, 1989. pp.1-6.

5. "An Overview of the Multimedia Communications Capability Program," Rome Laboratory Telecommunications Division internal report.

6. D.E. Krzysiak, "Advanced Multi-Media Information Distribution System," MILCOM '90 Conference Record, 1990, pp.35.7.1 - 53.7.3.

7. CCITT Recommendation I.610, "OAM Principles of the B-ISDN Access," pp. 12, 14.

8. "ATM User-Network Interface Specification," Version 2.0, The ATM Forum, June 1, 1992, pp 61-63.

EVALUATION OF NETWORK TOPOLOGY IN A DISTRIBUTED PROCESSING ENVIRONMENT

Kamal Jabbour

Associate Professor

Department of Electrical and Computer Engineering

Francis X. Reichmeyer

Graduate Student

Department of Electrical and Computer Engineering

Syracuse University

121 Link Hall

Syracuse, NY 13244-1240

Final Report for:

Summer Faculty Research Program

Rome Laboratory C3AB

August 1993

EVALUATION OF NETWORK TOPOLOGY IN A DISTRIBUTED PROCESSING ENVIRONMENT

Kamal Jabbour
Associate Professor

Francis X. Reichmeyer
Graduate Student

Department of Electrical and Computer Engineering
Syracuse University

Abstract

The impact of network topology on the performance of a distributed processing environment was evaluated through experimentation and simulation. Experimental runs of the JDL experiment on Cronus were used to collect traffic data for use in determining the limitations of Ethernet-based networks. Discrete-event simulation and analytical modeling were used to estimate the performance of heavily-loaded Ethernet and FDDI networks.

EVALUATION OF NETWORK TOPOLOGY IN A DISTRIBUTED PROCESSING ENVIRONMENT

Kamal Jabbour and Francis X. Reichmeyer

1. INTRODUCTION

The attractiveness of distributed computing on networks of heterogeneous workstations has created an interesting class of problems. The communications networks and protocols used in such applications were originally designed with a different usage in mind, primarily reliable packet transport over unreliable and low-capacity links. Distributed operating systems, such as BBN's Cronus, use standard technologies like TCP/IP over Ethernet, and as a result suffer from their severe limitations.

In this research we looked at distributed processing over heterogeneous systems and simulated the expected benefit from using new technologies. In particular, we evaluated the performance of the JDL (Joint Director of Labs) experiment over a group of Ethernet networks, characterized its data traffic, and projected the potential improvement from using FDDI networks.

This report starts with a brief introduction to the JDL experiment and the Cronus distributed operating system, followed by an outline of the data collection strategy and a characterization of the JDL traffic. The results are generalized to provide a generic characterization of the packet traffic for typical distributed computing applications. A computer simulation of the performance of an Ethernet-based system under increasingly heavy load is presented next, followed by an analytical model of an FDDI-based system. The report concludes by identifying unanswered questions and topics for future

research, dealing both with the chosen topology and the transport protocol.

2. CRONUS AND THE JDL EXPERIMENT

In this section we introduce briefly the JDL Experiment and the distributed operating system environment, Cronus, on which it runs. The JDL Experiment was used as a sample distributed processing application for the purpose of characterizing network traffic, and subsequently extrapolating such traffic characteristics to other classes of distributed applications.

2.1 Cronus

Cronus is a software environment for the development and operation of distributed computing applications [1]. Cronus was developed by BBN Systems and Technologies of Cambridge, MA, under contract with Rome Air Development Center (Rome Laboratory) in the early 1980s. The Cronus environment runs at the user level, above the operating system, of several different host hardware and operating system combinations. It uses an object-oriented architecture within which every system resource and application-specific resource is viewed as a typed object under the control of an object manager. An object manager implements all the operations that correspond to a particular object type.

In the Cronus environment, client and object managers need not reside on the same machine, and operations can be invoked across machine boundaries. A client process does not need to know the specific locations of objects and managers, as Cronus automatically routes operation invocations to the appropriate manager.

2.2 The JDL Experiment

The Tri-Service Distributed Technology Experiment was initiated as an

unclassified activity by the Distributed Processing Subcommittee of the Joint Directors of Laboratories (JDL) [2]. The concept of the experiment centers around a low-cost distributed computer system application at three sites: Naval Command, Control, Ocean Systems Center RDT&E Division (NRaD), San Diego, CA; Rome Laboratory (RL), Rome, NY; and Communications and Electronics Command (CECOM), Fort Monmouth, NJ. The distributed computing environment Cronus was chosen for the experiment.

The JDL Experiment consists primarily of eight managers, exchanging data in a packet switched mode through Cronus by using UDP and TCP/IP. The operations of these managers is beyond the scope of this report, since our interest lies in characterizing the data traffic generated by the experiment.

The JDL experiment can be run in a one-site (RL), two-site (RL and CECOM), or three-site (RL, NRaD and CECOM) configuration. Computers at these cites are primarily workstations connected locally with Ethernet LANs. The LANs are in turn interconnected through T-1 links over terrestrial wideband links.

3. TRAFFIC CHARACTERIZATION

We conducted local (RL-only) JDL runs separately under light network traffic and under relatively heavy network traffic, as well as multi-site runs under normal prevailing traffic. Etherfind, a network monitoring software provided with the SUN OS, was used to capture the packet headers on the local network for use in traffic characterization. A major limitation of the software/hardware is that workstations keep time to a resolution of 10 msec, thereby preventing more precise measurement of delay.

3.1 Packet Size Distribution

A single-site JDL run generated about 24,137 packets exchanged between the various computers. The majority (23589 or 98%) of these packets used TCP, while the remaining 548 (2%) used UDP. These proportions changed only slightly (94% vs 6%) in multi-site runs. These packets represent collectively fewer than 8,000 transactions. By using TCP, and its inherent inefficiency, each transaction requires a minimum of 3 packets: request, reply, acknowledgement.

A histogram of Ethernet packet size distribution shows a basic tri-modal distribution at 60 bytes (7675 packets) 213 bytes (6971) and 239 bytes (6971) (Fig 4-1). The largest packet is 619 bytes, well below Ethernet's packet size limit. Thus we have observed no fragmentation.

It may be useful at this point to discuss TCP packet structure to put the above packet sizes in perspective. A typical TCP packet consists of a 14-byte Ethernet header, a 20-byte IP header, a 20-byte TCP header, and up to 6 bytes of padding to achieve a 60-byte packet. Thus a 1-byte acknowledgement carries an additional 59 bytes of overhead in a 60-byte packet. Similarly the 239-byte packets consist of 185 bytes of data and 54 bytes of overhead.

In analyzing JDL traffic, we looked at two sample transactions: one dealing with the Timer Manager (running on Mensa), and the second with the Target Filter Manager (running on Orion).

The Timer Manager accounts for the bulk of the network traffic in JDL, namely 20,904 out of 24,137 packets. A typical timing transaction consists of a request (213-byte packet), a reply (239-byte packet) after about 20 ms, and an acknowledgement (60-byte packet):

Time	bytes	source	destination	srcport	dstport
46.72	239	TCP janus.rl.af.mil	mensa.rl.af.mil	1433	2611
46.74	213	TCP mensa.rl.af.mil	janus.rl.af.mil	2611	1433
46.92	60	TCP janus.rl.af.mil	mensa.rl.af.mil	1433	2611

The Target Filter Manager responds to requests for data and accesses the target database. A typical local transaction consists of a request (387 bytes), an ack (60 bytes) after about 80ms, a reply (219 bytes) after about 2.52 sec, and a final ack (60 bytes) after about 80 ms:

Time	bytes	source	destination	srcport	dstport
692.20	387	TCP phobos.rl.af.mi	orion.rl.af.mil	1154	2611
692.28	60	TCP orion.rl.af.mil	phobos.rl.af.mi	2611	1154
694.80	219	TCP orion.rl.af.mil	phobos.rl.af.mi	2611	1154
694.88	60	TCP phobos.rl.af.mi	orion.rl.af.mil	1154	2611

In general, we can characterize JDL traffic as consisting of a series of transactions, each consisting of a request for service, followed by an optional acknowledgement, a reply, and a final acknowledgement. Acks are 60 bytes long, while data packets range from 76 to 619 bytes, with a predominance in the 200-400 byte range.

3.2 Time Domain Characterization

In the time domain, ack packets are generated within 10-100 ms (Fig 4-3), while manager service times range from a few ms to several seconds (Fig 4-4). Service times consist of two components: communication time and processing time. Under very light load conditions, communication times are primarily data transmission times, of the order of 100-500 microseconds. Since a single-site JDL run requires the transmission of about 24,000 packets over a period of an hour, or

less than 7 packets per second, such a run generates a very light network load and has a negligible effect on network performance.

3.3 Additional Data

In addition to the JDL Experiment described above, experimental runs of two other distributed applications were conducted. The Australian Experiment [3] was run at the DISE facility at Rome Lab, and an ISIS run was performed at the Software Engineering Lab at Syracuse University (Fig 4-2). These applications featured a multimedia component and the corresponding large packets for video screen updates. Whereas the JDL Experiment had primarily small packets, multimedia applications require the transfer a large files, resulting in packet fragmentation.

3.4 Background Traffic

As discussed in the previous sections, JDL traffic alone offers a very light load to an Ethernet. Therefore the performance of a JDL run, or any distributed application for that matter, is directly affected by the background traffic on the network. Therefore we have tried to characterize experimentally the background traffic by collecting a significant amount of network traffic data.

Based on the data collected, we have characterized the packet arrival process with an exponential distribution (Fig 4-5). Further, we have observed that the packet size distribution contained a significant peak (about half the packets) at 60 bytes, a small peak at 1514 bytes (fragments at maximum Ethernet packet size), and an otherwise random distribution between 60 and 566, decaying exponentially beyond that (Fig 4-6).

An analysis of the traffic patterns on both the Syracuse University-SELab and

2-8

Rome Lab-DISE networks suggests that the majority of the traffic is generated by remote login and telnet users. In a remote login session, each keystroke generates 3 TCP packets, each consisting of 60 bytes (1 packet contains the key stroke, a return packet carries the echo back, and a third ack packet completes the 3-way handshake). At the other extreme, Network File Systems (NFS) traffic consists primarily of large files fragmented into packets of 1500 bytes (unless a lower maximum packet size is set at internet gateways).

4. GENERIC TRAFFIC CHARACTERIZATION

The distributed processing applications described earlier allow us to develop a generic model for characterizing network traffic. Such a generic model will be application-independent and allows a better understanding of the impact of the communications environment. We will assume a packet-switched environment where all communication between processes uses packets (datagrams).

4.1 Basic Transaction

The generic model assumes that a distributed processing application consists of a group of service managers located on various computers. A request for service can originate from any host on the network, and may be preceded by an initial broadcast to locate the server. We will define a transaction as a complete set of packet transmissions consisting of a request for service, an optional acknowledgement, a reply packet, and a final acknowledgement. Acknowledgements are required to complete the 3-way handshake necessary in an unreliable network environment (such as using TCP/IP). The assumption of a reliable (low error rate, eg: optical) network allows the use of a more efficient protocol.

4.2 Packet Generation

We assume that the packet generation intervals are exponentially distributed. In an ideally active network supporting several simultaneous distributed processing applications, we assume that all nodes act as clients as well as servers, therefore all nodes have similar packet generation characteristics. The value of the mean interval determines the loading level of the network. For example, a JDL Experiment generates about 7 packets per second on average.

Similarly, background traffic (generated by non-distributed processing users) can be modeled with an exponential interarrival distribution with a variable mean that determines the background load on the network.

4.3 Packet Size Distribution

The distributed processing applications that we examined rely predominantly on packets of a few hundred bytes (200-600 bytes) for their communication. A handshake packet requires a minimum of 1 byte, burdened by up to 60 bytes of header information (addressing, sequencing, error detection...). Images, high-resolution graphics, and multimedia data require large packets of several kbytes (typically 4-20 kbytes).

To simplify our model, we define packet sizes excluding transport layer headers and independent of network limitations (Ethernet limit is 1,500 bytes, FDDI limit is 4,500 bytes). We assume that distributed processing traffic consists of 3 packet sizes, namely 60 bytes (handshake packets), 300 bytes (requests and replies), and 9,000 bytes (for graphics). The relative distribution of packets among these three sizes is left as a model variable, allowing the specification of a range of applications.

4.4 Packet Transmission Time

Packet transmission time is a primary variable of interest in modeling the communication characteristics of a distributed processing environment. It depends primarily on the type of local area network (Ethernet, FDDI, ATM), the internetworking topology for multi-site applications, and the transport protocol used.

Processing Time

The final component of the generic traffic model is the remote service time, or manager processing time at a remote site. This is the time it takes a service manager to complete the necessary processing and generate a reply to a request for service. In the above applications, this processing time ranged from sub-millisecond (request for timing information) to several seconds. We assume that the processing time is exponentially distributed, with the mean determining the computational intensity of a given application.

5. COMPUTER SIMULATION OF AN ETHERNET BASED SYSTEM

Currently, the JDL experiment runs on three Ethernet LANs, one at each of the labs involved in the project, as described above. In this section we describe the simulation of the local (RL only) JDL experiment. The simulation was performed to observe the degradation of Ethernet under heavy loads. Since it is not possible to conduct multiple JDL runs simultaneously on the network, and since one JDL experiment by itself does not generate enough network traffic to cause a degradation of the LAN, simulation provides insight that cannot be gained through experimentation alone.

The simulation program is written in C using the SMPL library of subroutines for discrete-event and random variate simulation [4]. We consider an Ethernet LAN of 32 workstations connected with a total cable length of 2.5 kilometers. Each station is considered to be identical in terms of traffic generation. Therefore, LAN access is modeled as a single queue of size 32, which accepts generated packets and releases them on a first in first out (FIFO) basis according to the Carrier Sense Multiple Access with Collision Detection (CSMA/CD) protocol of Ethernet. If the queue is full when a new packet is generated, that packet is discarded. Throughput and average delay are measured as a function of the offered load to the network.

The Ethernet LAN is modeled as an open system, where the total offered load G to the network is the product of the arrival rate L of new packets (for all stations) and the mean transmission time T for a packet, G=LT. The packet length distribution determines the mean transmission time T and the packet interarrival times are obtained from an exponential distribution with mean 1/L. Both packet length distribution and G are supplied as inputs to the simulation by the user.

We considered several different packet length distributions corresponding to different distributed computing applications including JDL, the Australian experiment, and an ISIS run, all three with no background traffic. That is, the only packets on the LAN are those generated by the experiments. For JDL, the packet distribution is as follows:
60% @ 60 bytes, 30 % @ 300 bytes, and 10 % @ 1500 bytes.
The Australian experiment is more computing intensive, having a distribution of:
60% @ 1500 bytes, 30% @ 300 bytes, and 10 % @ 60 bytes.
The 1500-byte packets represent the fragments of the larger packets (mostly 8000 bytes) produced by this application. The ISIS run, simulating an application

2-12

using the ISIS distributed operating system, has the distribution:

57% @ 120 bytes, 32% @ 1500 bytes, and 11% @ 60 bytes.

We also simulated the JDL and Australian experiments with background traffic. In all cases, the interarrival times were computed from offered load and transmission time for the mean packet size (L=G/T).

The attached figures show plots of throughput and delay versus offered load (Fig 4-7, 4-8). Average packet delay is measured from the time a packet is generated to the time it is received. Throughput is defined as the part of the channel utilization used for successful data transmission, based on the total packet length, including the preamble (ethernet header) and data field.

6. PERFORMANCE ANALYSIS OF AN FDDI BASED SYSTEM

An FDDI network was simulated analytically with the same traffic characteristics used above for Ethernet. We assumed a cable length of 2.5 km, 32 stations spaced equally around the ring, and ring access based on the MAC layer. The Target Token Rotation Time was assumed small enough to permit the transmission of only one packet per token acquisition. Finally, large data packets were fragmented to 4500 bytes as opposed to 1500 bytes in the case of Ethernet.

An analytical model of the FDDI ring was developed to estimate network throughput and delay in a simulated distributed processing environment. We assumed that several applications were running simultaneously, with every station hosting a number of managers. Therefore we assumed a homogeneous ring, as far as traffic patterns are concerned [5].

Throughput is shown in Fig 4-9. Offered load refers to the fraction of total network capacity. Since FDDI provide a deterministic data transmission medium,

2-13

representing throughput in number of packets per second provides more useful information than presenting it as a fraction of offered load (the latter relationship is linear until maximum load is reached, then flattens out).

Average packet delay is shown in Fig 4-10, and included both the queueing delay inside individual nodes and the transmission delay. The wait queue inside the node was modeled as a M/D/1 queue, assuming a Markov arrival rate, and a deterministic service rate.

Comparing the performance of Ethernet and FDDI shows that FDDI's degradation under heavy loads is more graceful than that of Ethernet. In addition to the obvious order of magnitude in transmission speed (100 Mbps for FDDI vs 10 Mbps for Ethernet), the deterioration in Ethernet's performance is caused by collision, whereas FDDI's deterioration comes from the queueing delays within each node.

At low utilization, such as 10% of offered load, it is not surprising that Ethernet's performance is as good, if not better, than FDDI's. With an estimated 10-bit delay in each FDDI station, it takes a long time for a token to rotate idly around a ring.

7. FUTURE EFFORT

Our work has shown that current technology is well-suited for distributed applications of the class of the JDL or Australian experiment, which offer a small load to an Ethernet. However, multimedia intensive applications generating traffic several orders of magnitude larger than JDL cannot run on existing networks, and necessitate new topologies and transport protocols. In addition, the impact of the network architecture on the traffic characteristics requires

further study. The use of an FDDI network, or an FDDI-ATM interconnection, with the resulting deterministic delay characteristics, will reduce significantly the number of smaller packets and result in a larger number of large packets.

In future effort we propose to examine the performance of two optical networks. The first network uses multiple FDDI rings interconnected with high-speed point-to-point optical links through a single central ATM switch. The second network uses a hierarchical ATM architecture where the computers are connected to local ATM switches which are in turn connected to a central ATM switch. We propose to compare the delay and throughput achieved by these configurations and determine their suitability for distributed processing.

REFERENCES

[1] "Cronus User's Reference Manual", BBN Systems and Technologies, 1991.

[2] "Joint Directors of Laboratories (JDL) Tri-Service Distributed Technology Experiment", Maintenance Manual, 30 Sep 1992.

[3] A.M. Newton et al, "1991 DISE Summary Report", RL-TR-92-302 In-House Report, Rome Laboratory, Griffiss Air Force Base, NY, Nov 1992.

[4] M.H. MacDougall, Simulating Computer Systems. The MIT Press,1987.

[5] Salma Abu Ayyash, Francis Reichmeyer, Salim Hariri, Kamal Jabbour, "Distributed Computing on FDDI-Based Network", Proceedings of the 35th Midwest Symposium on Circuits and Systems, Washington DC, August 9-12, 1992, pp. 1197-1200.

Packet Size Distribution of an ISIS Run

Number of Packets

Bytes x 10¹

N (K)

Fig 4-2

Packet Size Distribution of a JDL Run

Number of Packets x 10³

Bytes

Fig 4-1

2-16
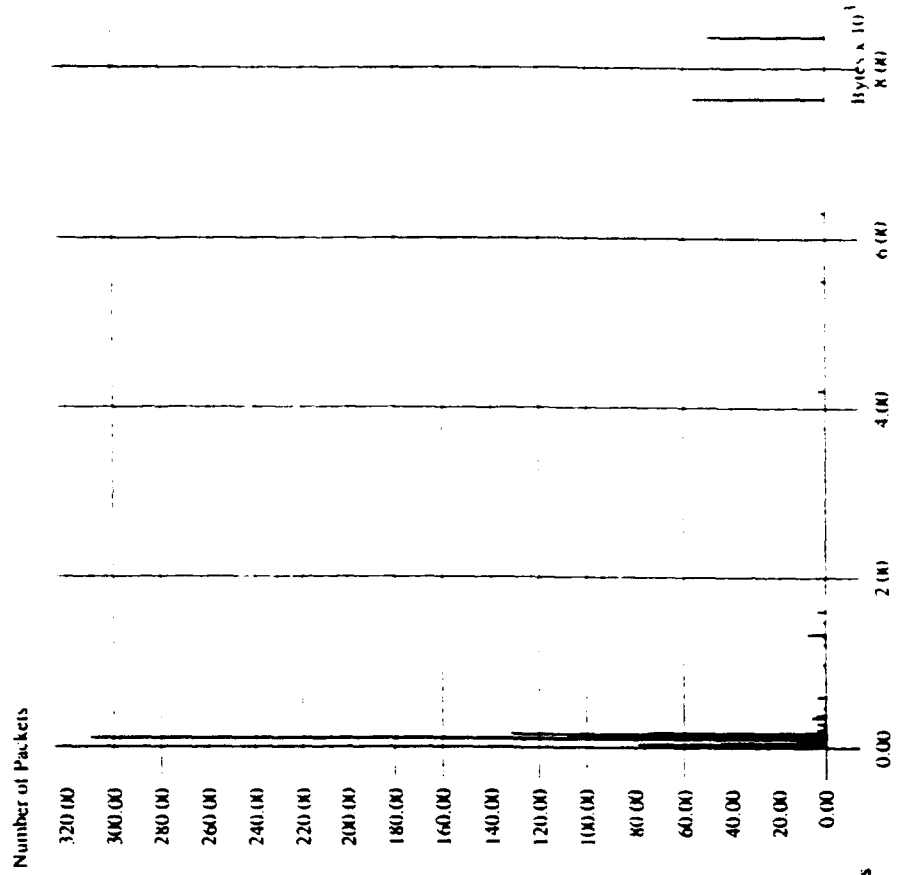
ACK Service Time Distribution in JDL Run

Number of Packets

130.00
120.00
110.00
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00

0.00   50.00   100.00   150.00   200.00

Time x 10 msec
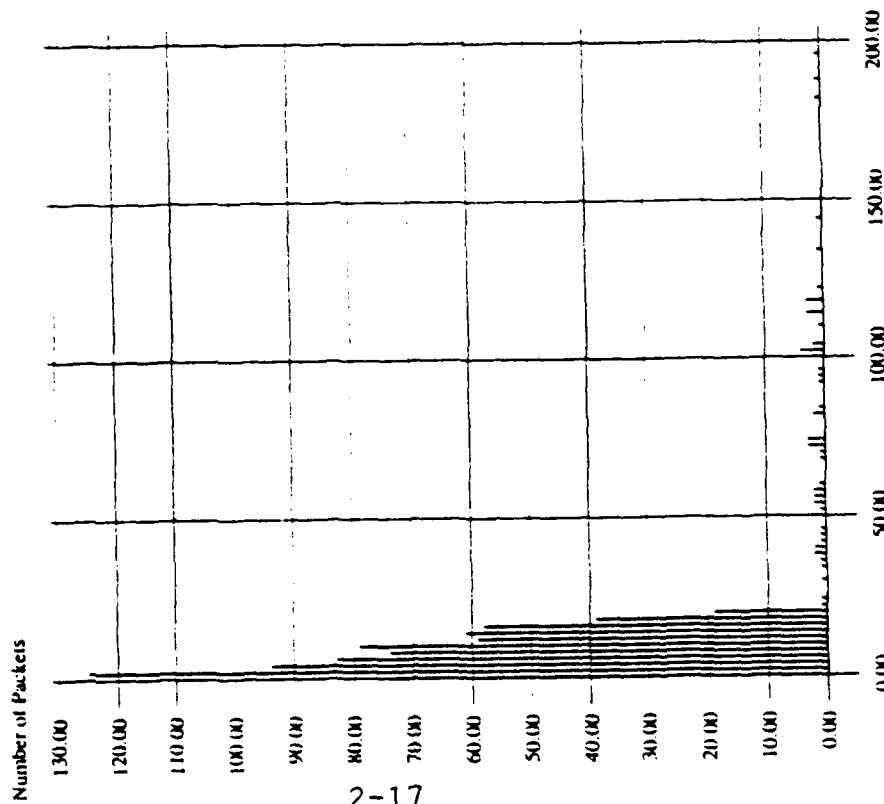
Fig 4-3

Target Filter Manager Service Time

Number of Packets

8.00
7.50
7.00
6.50
6.00
5.50
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00

230.00   240.00   260.00   280.00   300.00

Time x 10 msec

Fig 4-4

2-17

Background Traffic Packet Size Distribution

Number of Packets x 10³

Fig 4-6

Bytes x 10³

Background Traffic Packet Transmission Interval Distribution

Number of Packets x 10³

Fig 4-5

Time x 10 msec

2-18
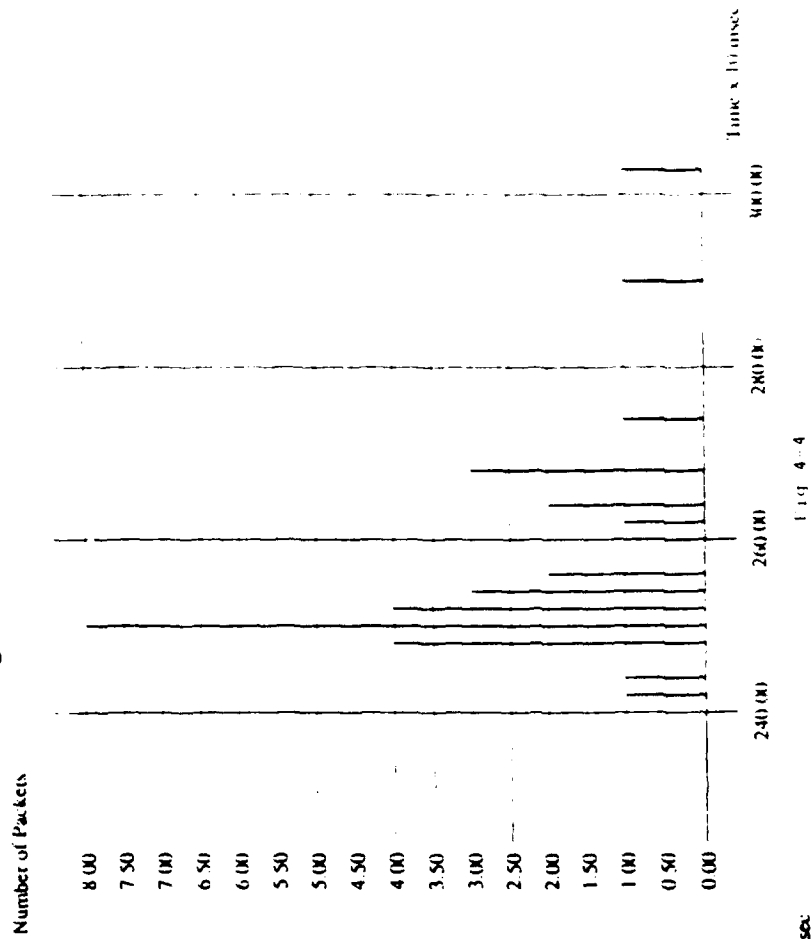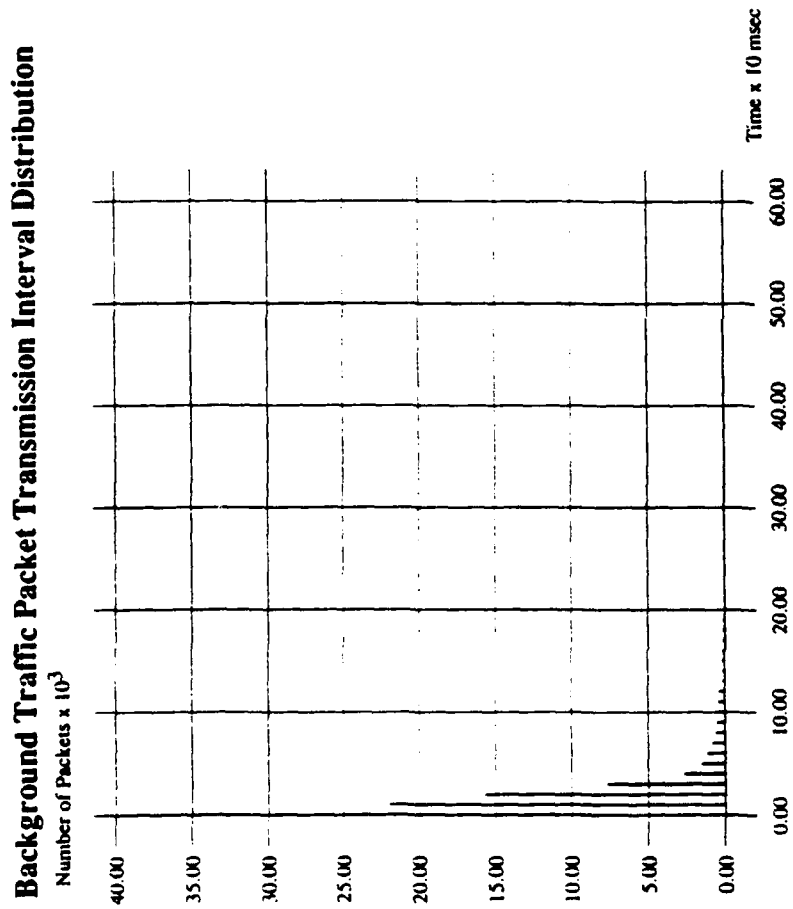
Ethernet Simulation
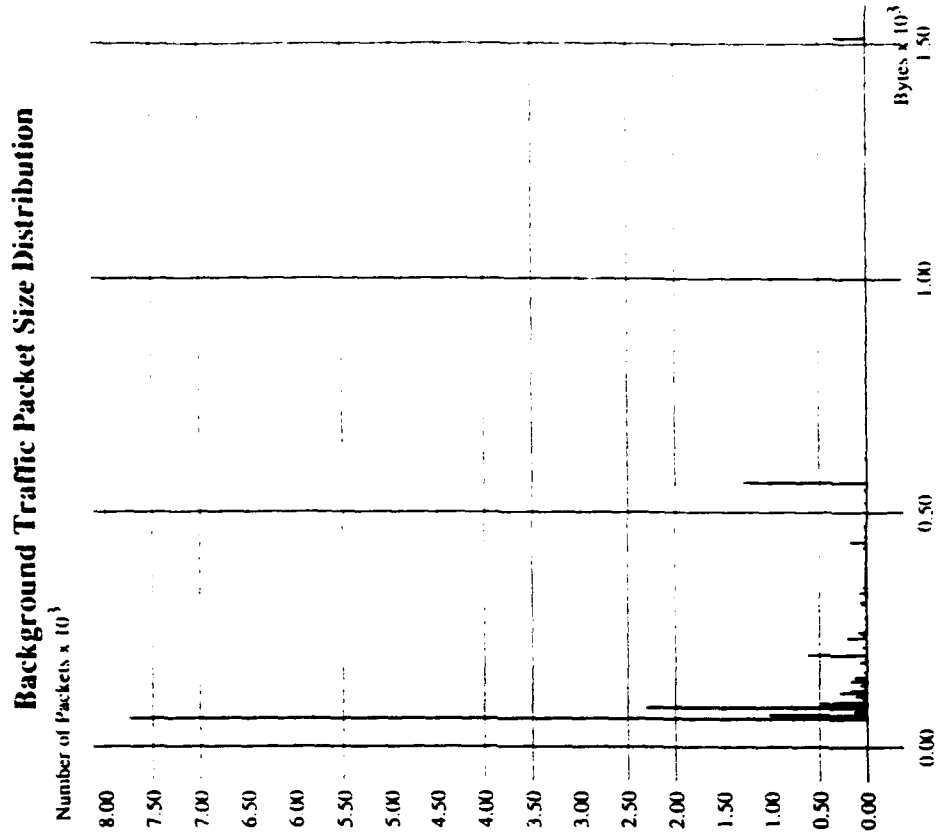
Fig 4-8

Ethernet Simulation

Fig 4-7

2-19
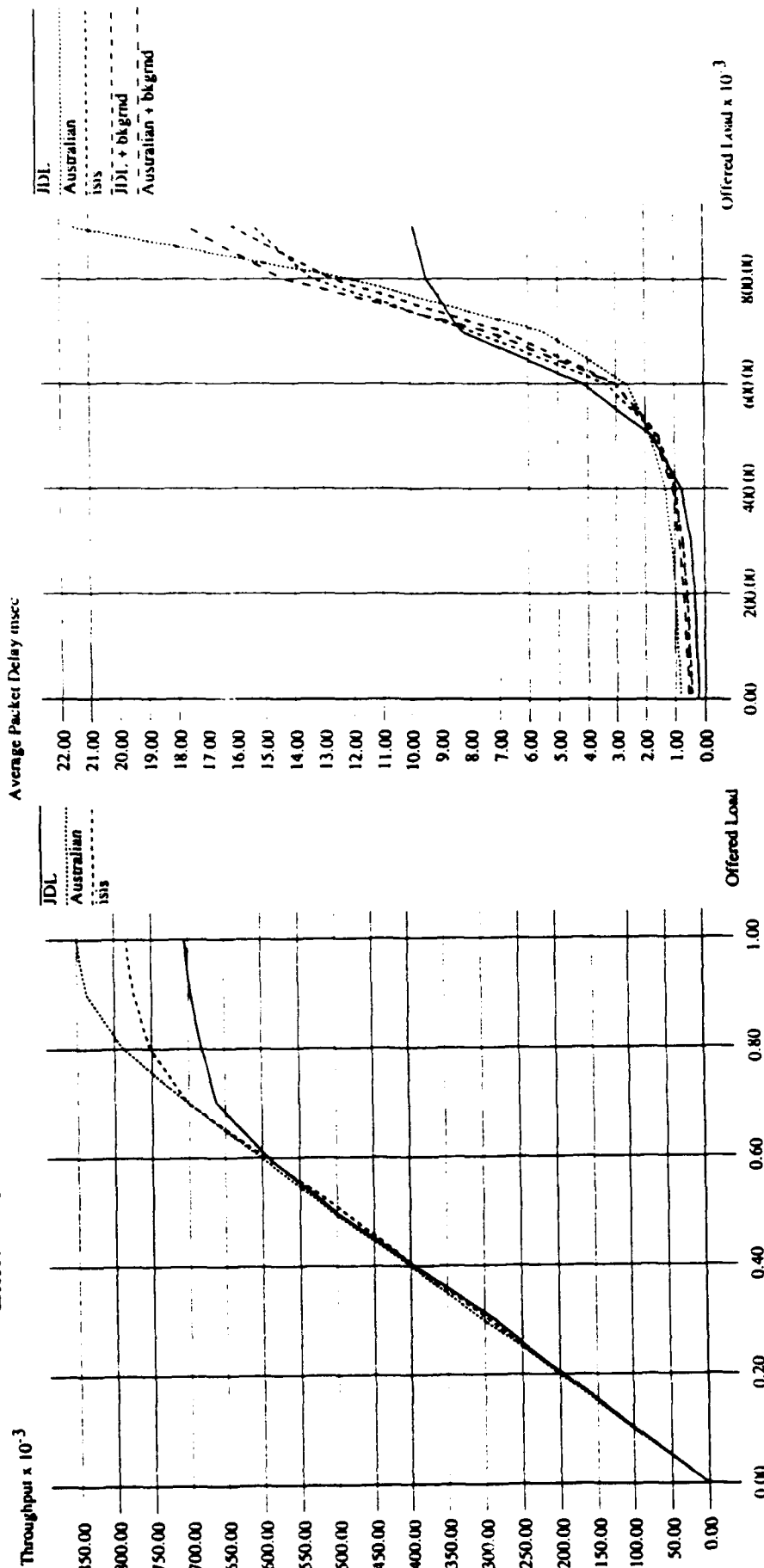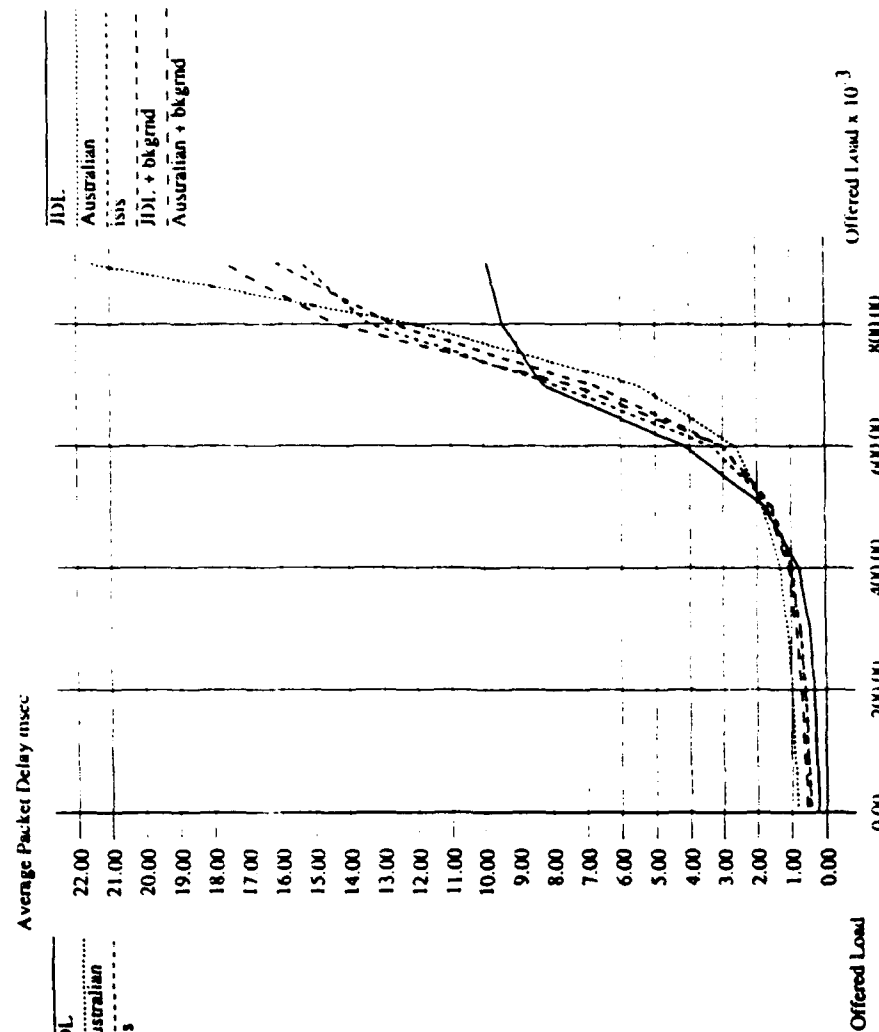
FDDI Simulation

Fig 4-10

FDDI Simulation

Fig 4-9
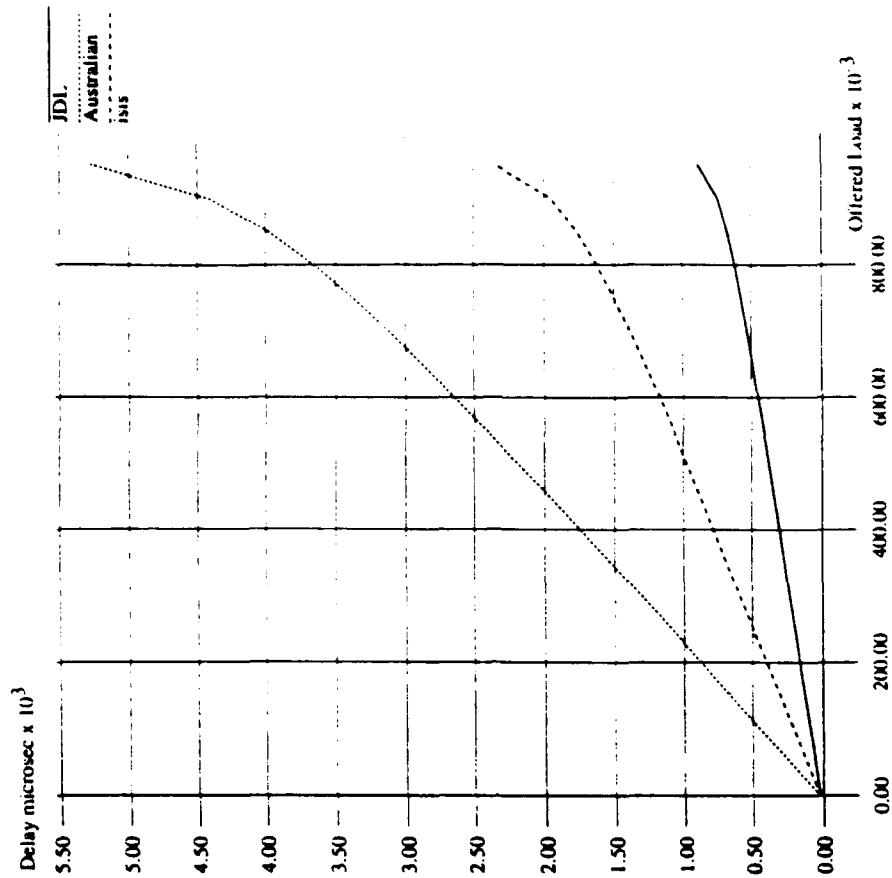
AN EXPERIMENTAL AND NUMERICAL STUDY OF THERMAL CONDITIONS IN A HIGH PRESSURE SYSTEM FOR INDIUM PHOSPHIDE CRYSTAL GROWTH

Vishwanath Prasad, Professor
Department of Mechanical Engineering
State University of New York, Stony Brook, NY 11794-2300

and

Andrew P. Anselmo[1], Doctoral Candidate
Department of Mechanical Engineering
Columbia University, New York, NY 10027

---

1. Present Address: Department of Mechanical Engineering
        State University of New York, Stony Brook, NY 11794-2300

AN EXPERIMENTAL AND NUMERICAL STUDY OF THERMAL CONDITIONS IN A HIGH PRESSURE SYSTEM FOR INDIUM PHOSPHIDE CRYSTAL GROWTH

Vishwanath Prasad
Department of Mechanical Engineering
State University of New York, Stony Brook, NY
and
Andrew P. Anselmo
Department of Mechanical Engineering
Columbia University, New York

Abstract

For one-step, in-situ synthesis of phosphorus vapor and indium melt, and liquid-encapsulated Czochralski growth of InP crystals to succeed and produce single crystals of uniform quality and at lesser cost, it is important to develop a basic understanding of the mechanics of heat transfer and gas flow in a high pressure crystal growth (HPCG) furnace. A series of experiments performed recently in order to characterize the temperature conditions at the graphite susceptor and crucible walls is reported here. These results demonstrate the variations in temperature profiles with the power input as well as with the crucible's vertical location. A computer model to study the gas flow in an HPCG furnace has also been developed. Numerical computations are performed to examine the gas flow and temperature fields with and without the presence of phosphorus injector inside the furnace. Complex recirculatory flows which become oscillatory at high Grashof numbers, are produced by the buoyancy forces. The temperature variations are generally very strong in the vicinity of the encapsulant surface, and weak in the outer region which agree qualitatively with the temperature measurements.

AN EXPERIMENTAL AND NUMERICAL STUDY OF THERMAL CONDITIONS IN A HIGH PRESSURE SYSTEM FOR INDIUM PHOSPHIDE CRYSTAL GROWTH

V. Prasad and A. P. Anselmo

INTRODUCTION

In order to grow single crystals of III-V compounds such as gallium arsenide, indium phosphide and so on, a high pressure crystal growth (HPCG) system is used. The mechanisms of energy transport in such furnaces are very different from that in the low pressure Czochralski puller for silicon crystals. While in a low pressure system, the energy is transferred to the melt and from the melt and crystal, primarily by conduction and radiation, gas convection plays an important role together with other modes of heat transfer in determining the thermal field within an HPCG puller. The buoyancy-driven convective flows in a high pressure chamber are turbulent and oscillatory, and the recirculating gas flow pattern is a strong function of the geometric configuration, temperatures of the r.f. coil, susceptor, melt, crystal and puller walls, and the gas properties. In addition, the convective heat transfer in an HPCG system is strongly coupled with the conduction in various components of this system and radiation exchange between the surfaces which can see each other.

The heat transfer mechanism is further complicated by the process employed to grow the single crystal. Motivated by the high cost of indium phosphide crystals grown by a two-step method - the synthesis of polycrystal by the horizontal Bridgman method and single crystal growth by the liquid-encapsulated Czochralski (LEC) method, a one-step operation has been proposed and investigated by several researchers [1-11]. It is basically an LEC method for pulling a single crystal from an "in-situ" synthesized melt of indium phosphide. Recently, magnetically stabilized liquid-encapsulated Czochralski (MLEC) and Kyropoulos (MLEK) growth of InP single crystals have also been reported [12,13].

For synthesis, the heat radiated by the melt is absorbed by the solid phosphorus for melting and/or sublimation. The convective heat transfer also helps. The temperature of the phosphorus can therefore be controlled by changing the vertical location of the injector. The ampoule is placed directly above the melt and the vapor is transferred to indium melt by a quartz tube whose length is another key variable in the "in situ" synthesis. The heat transfer phenomena during the in-situ synthesis is therefore highly transient. The movement of P-injector changes the temperature field significantly by modifying the radiation exchange between various surfaces and by changing the convective gas flow rate and its structure. The

energy generated into the melt due to the reaction of In and P, and the volume increase (ρ_{In} = 7.31 and ρ_{InP} =4.787) also affect the heat transfer and gas velocity.

For one-step, in-situ synthesis and growth process to succeed and produce InP crystals of uniform quality and at lesser costs, it is important to understand the mechanics of heat transfer and gas flow in an HPCG furnace. This can be achieved only by a well-coordinated experimental and numerical research program. During this summer, we developed a model to analyze the gas flow inside the HPCG furnace, and have performed simulations for a number of selected cases. The computed flow fields support the observed flow phenomena in the furnace [14,15]. Also, whenever a model is used to simulate the crystal growth process, a selection of suitable boundary conditions poses the biggest problem. Special experiments were therefore designed and conducted to determine the boundary conditions on the crucible wall as well as the relationship between the power input and the wall temperature variations as functions of the crucible's location.

DESCRIPTION OF THE HPCG FURNACE

The HPCG furnace (Fig. 1) used for ongoing research at Rome Laboratory uses a 96 mm i.d. quartz crucible or a PBN cup of 2 mm thick wall, placed in a 113 mm o.d. graphite susceptor. The susceptor is insulated from the r.f. coil by "Fiberfrax Lo-Con Felt" insulation placed in an annular space formed by two quartz cylinders. An opaque glass plate of 280 mm diameter separates the bottom portion of the furnace from its upper region. However, the gas can flow downward through a 2 mm gap between this plate and cold puller wall. The plate is placed on top of the quartz cylinders, 305 mm from the bottom surface of the puller, while the mean height of the upper tube of the r.f. coil is 286 mm. The graphite susceptor with crucible can be moved up or down, and can be rotated by a shaft. The shaft holds a boron nitride (BN) hollow cylinder which, in turn, supports the susceptor. A tungsten sheathed W26Re-W.5Re thermocouple probe passes through the shaft and measures temperature at the crucible bottom.

A 75 mm o.d., 100 mm long quartz ampoule is used for phosphorus storage and its vapor injection into the indium melt. This injector has an off-center tube of 12 mm i.d. for the vapor outflow. The injector can be hung from the top with the help of a 12 mm i.d. hollow tube through which a cooled seed shaft can pass. The injector can be moved up or down independent of the seed shaft. The gap (about 2 mm) between the hollow, support tube and the seed shaft can allow the cold gases to flow downward along the seed shaft. To insulate the phosphorus injector from heat losses, 8 mm Lo-Con felt insulation is placed between the injector and an inverted quartz cup (Fig. 1).
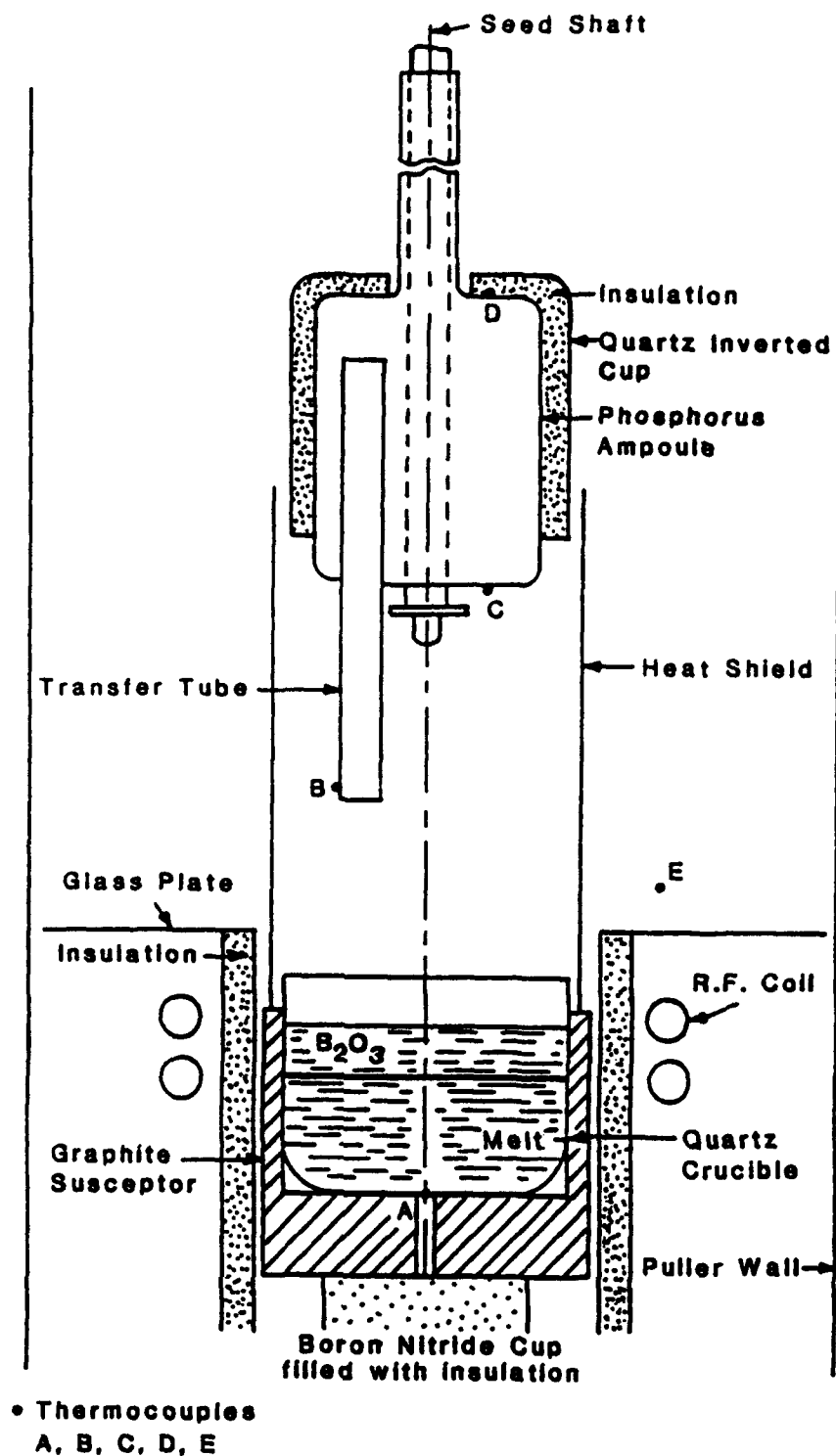
Figure 1. A schematic of the high pressure crystal growth furnace used for the on-going research on in-situ synthesis and MLEK growth of InP single crystals at Rome Laboratory.

A quartz cylinder of 103 mm i.d. is used as a heat shield. However, the heat shield cannot completely eliminate the buoyancy-induced flows - gas flows downward along the seed shaft and upward along the quartz shield - it can reduce the gas velocity significantly. The heat shield rests on the graphite susceptor and therefore can rotate with the crucible. Further details of the HPCG furnace may be found elsewhere [14,15].

There are two viewing ports in this HPCG puller, one is used for video monitoring and photography while the other helps in visual observation. Since the system is run at a high temperature (above 1063°C) and a high pressure (500-600 psi) at which the density of nitrogen is about six to fifteen times larger than at the atmospheric condition, it is possible to see the direction of flow whenever the gas flows over a step. Light reflections also seem to help in this visualization. Although only a small portion of the growth chamber is visible, it is. possible to draw flow patterns qualitatively from the observed gas flow behavior [14,15].

EXPERIMENTAL WORK

A series of experiments to study the thermal conditions in this HPCG furnace was conducted earlier and has already been reported in the literature [15]. The purpose of the present experiment is to measure temperature variations on the crucible wall and the outer surface of the susceptor, and to determine their relationship with the power input as well as with the crucible's location. Several thermocouples were therefore placed outside the susceptor and inside the crucible as shown in Fig. 2. This was the first time that 11 thermocouples were used to record temperature distributions in this system. A conex fitting was used to carry eight thermocouple extension wires. However, this was not a simple job, because placing the thermocouples at right locations and holding them there during the experiments are not easy tasks. Inner thermocouples were placed inside thin quartz tubes attached to the crucible wall, to prevent them from damage from the melt as well as to hold them in place. In order not to alter the susceptor's characteristics, outer thermocouples were placed on the outside of the inner quartz cylinder. We also connected a personal computer to the existing Omega strip chart recorder to monitor and record temperature continuously. Software were developed to collect the temperature information and store them for further analysis. We were therefore able to add some permanent features to the existing system which can help in future experiments.

First, the crucible was held in its central location (@ 2 cm) and temperatures were recorded for three different power inputs. Next, the power was kept fixed and the crucible was moved up or down to study the effect of its location. Figure 3 presents the crucible's location with respect to the r. f. coil. Sufficient time was allowed after every change in either power or location, for the system to reach (quasi) steady-state. From the temperature data collected, it is
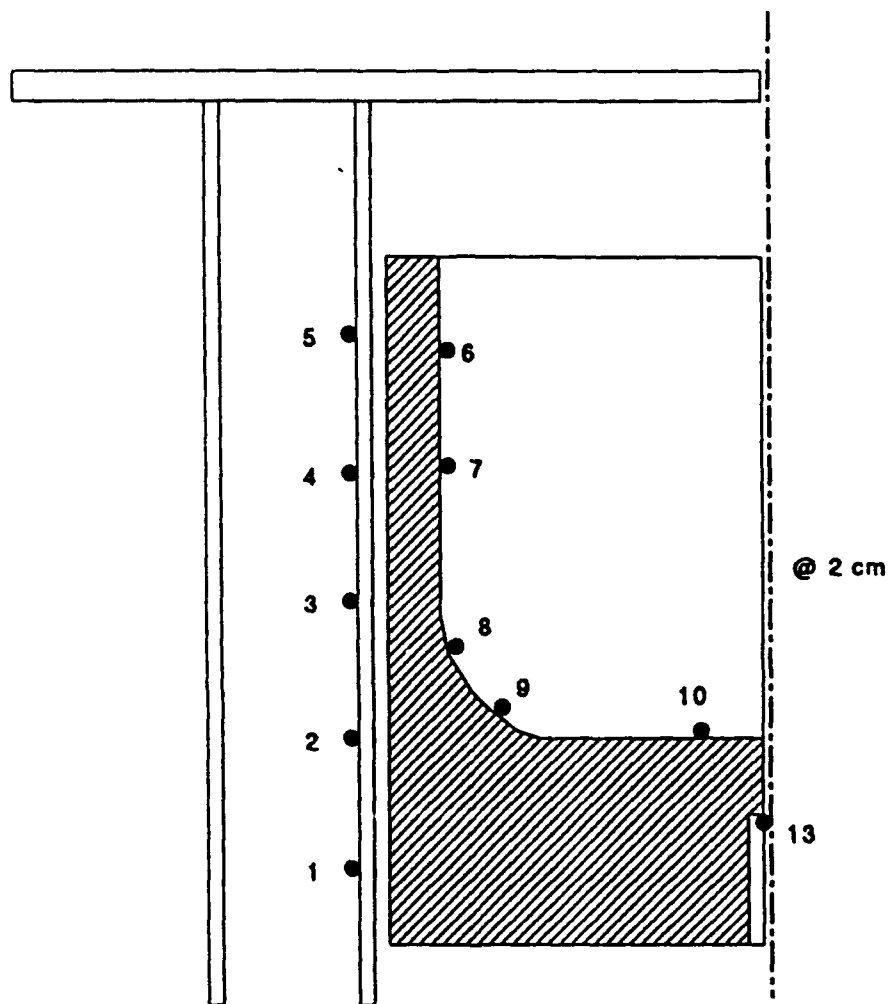
Figure 2. Thermocouple locations for temperature measurements on the inside of the crucible and outside of the graphite susceptor.
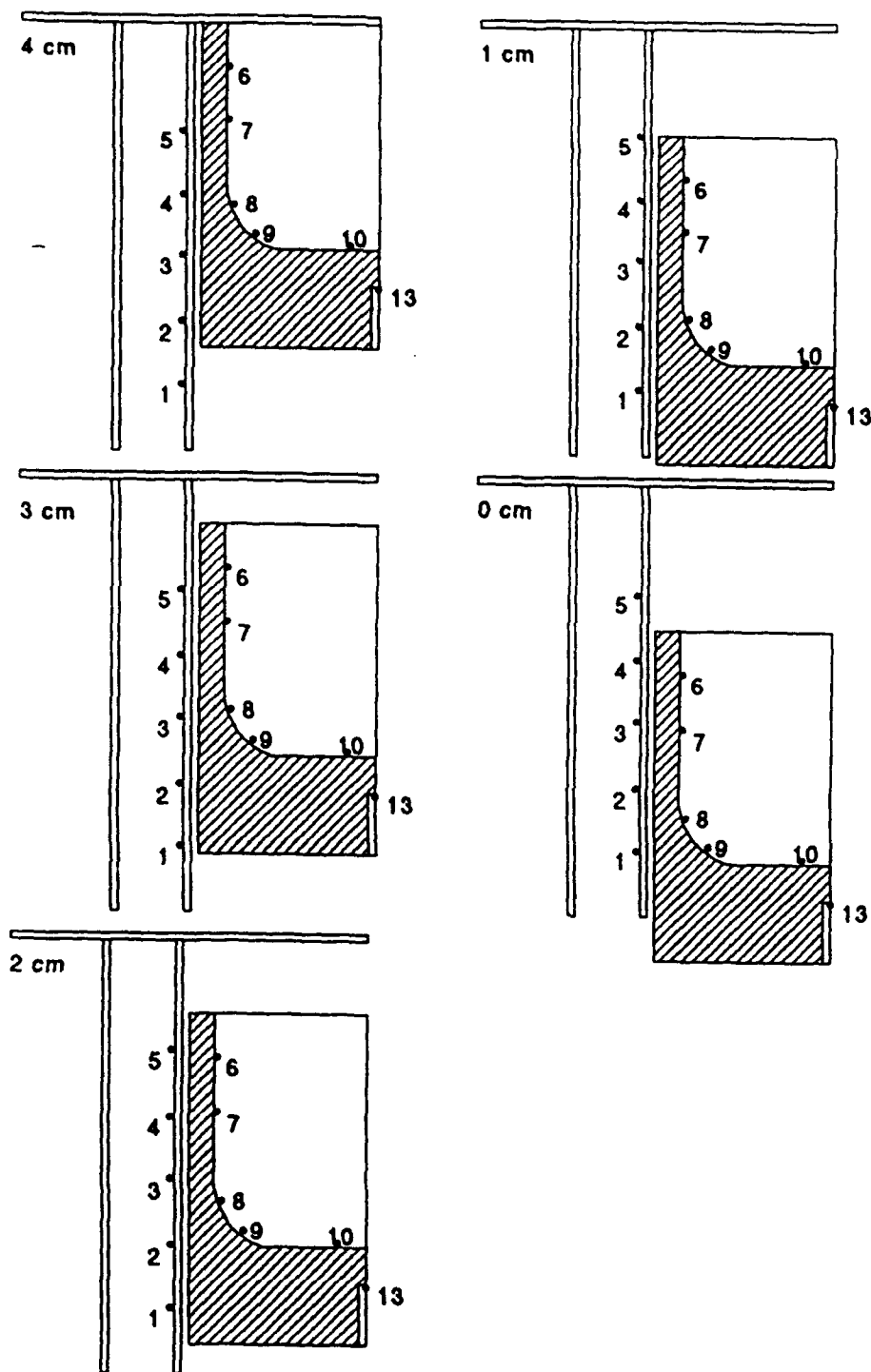
Figure 3. Crucible locations with respect to the upper quartz plate (Fig. 1).

possible to study the transients inside the HPCG furnace. However, this report will present temperatures only under steady conditions (Figs. 4 and 5).

We also designed and conducted two other sets of experiments. The objective of the first experiment was to study the effects of a graphite cover placed on top of the phosphorus injector. This was observed to suppress the temperature fluctuations [14,15] on the injector and temperature rise at the bottom and top of the injector was very smooth. However, the rate of phosphorus vaporization was slow which did not help the synthesis process. The second set of experiments was conducted to examine the effect of a reduction in radiation loss from the melt. The space between the furnace wall and the heat shield (146 mm dia. in this case) was filled with an insulation, placed between two concentric cylinders. This reduced the radiation loss significantly, and a larger temperature rise was observed at lower powers. A detailed analysis of these results will be presented later.

Figures 4(a) and 4(b) present the temperature rise with power. As can be seen the temperature increases monotonically with the power, and the nature of the curve does not change. Indeed the curves are almost parallel and support our earlier observation [14,15] that the melt temperature changes almost linearly with power. In Fig. 4(a), the thermocouple # 2 reads the largest temperature which is slightly higher than that at 1. The temperature, however, drops with the height, moving up from # 2 to # 5. This can be easily explained by considering the heat loss from the susceptor. Thermocouple #5 is in a position where the susceptor loses heat directly to the furnace walls and other components by radiation whereas the energy in the middle portion of the susceptor is transferred to the melt only by conduction. The temperature in the lower portion of the susceptor is smaller because of heat loss to the bottom portion of the furnace. Figure 4(b) confirms this behavior by demonstrating that the temperature is lowest for # 7 and largest for # 8. (Unfortunately, thermocouples 6 and 10 stopped working soon after the experiment was started.)

Figure 5 demonstrates the effects of the crucible's location (central location @2cm). It is observed that the temperature of the susceptor generally increases as it is moved up (Fig. 5a). The only exception is the bottom, thermocouple #1. Figure 5(a) further shows that the temperature at the top of the susceptor is much lower than that at its bottom when the susceptor is at its bottom most location, a temperature drop of 600°C. This temperature variation decreases as the susceptor is moved up, less than 200°C at the topmost location. This does not mean that the topmost location is the most desirable one. As shown in Fig. 5(b) by thermocouple # 7, the melt temperature may be lower when the susceptor is at its top location. It seems that a central location somewhere between 2 and 3 cm is better for the melt.
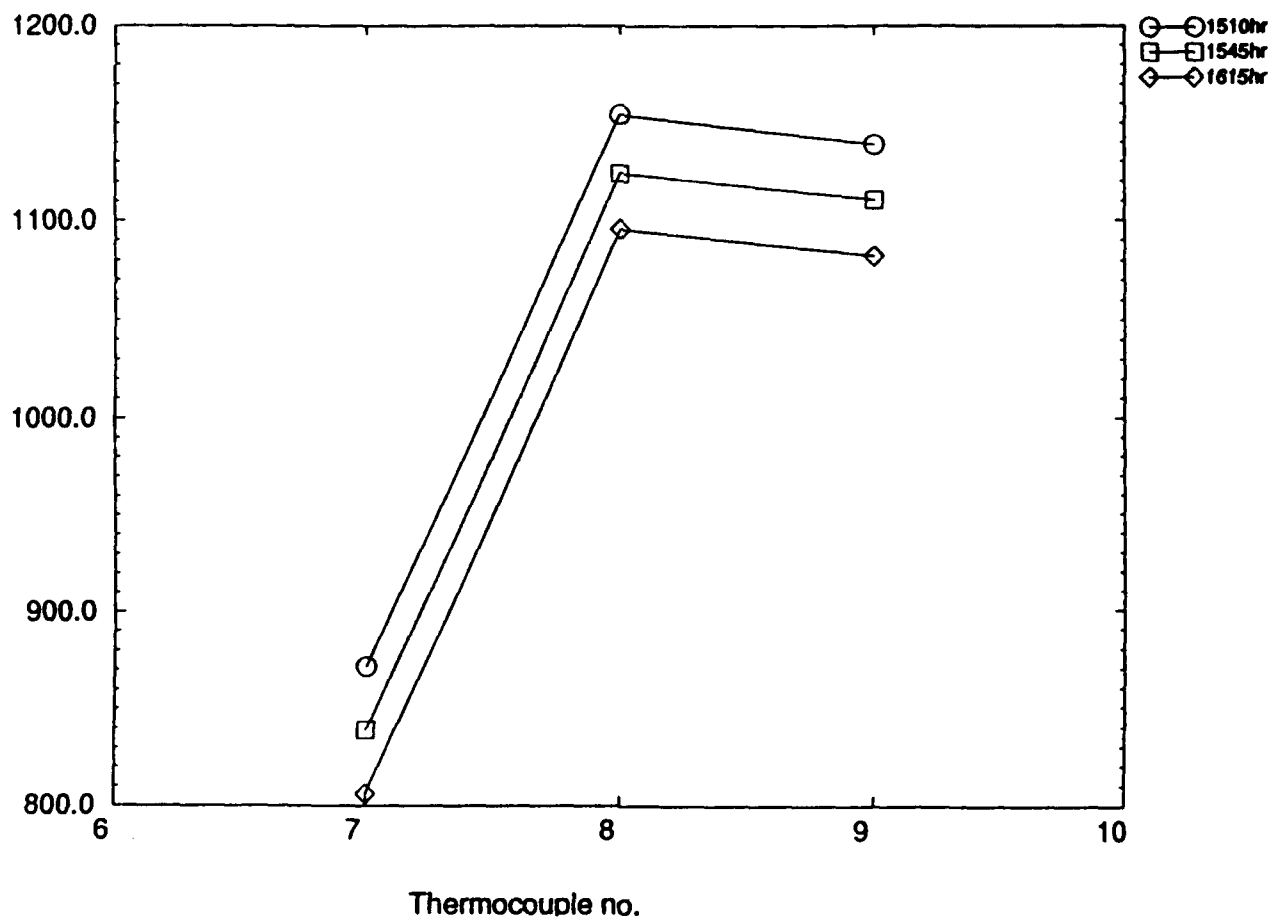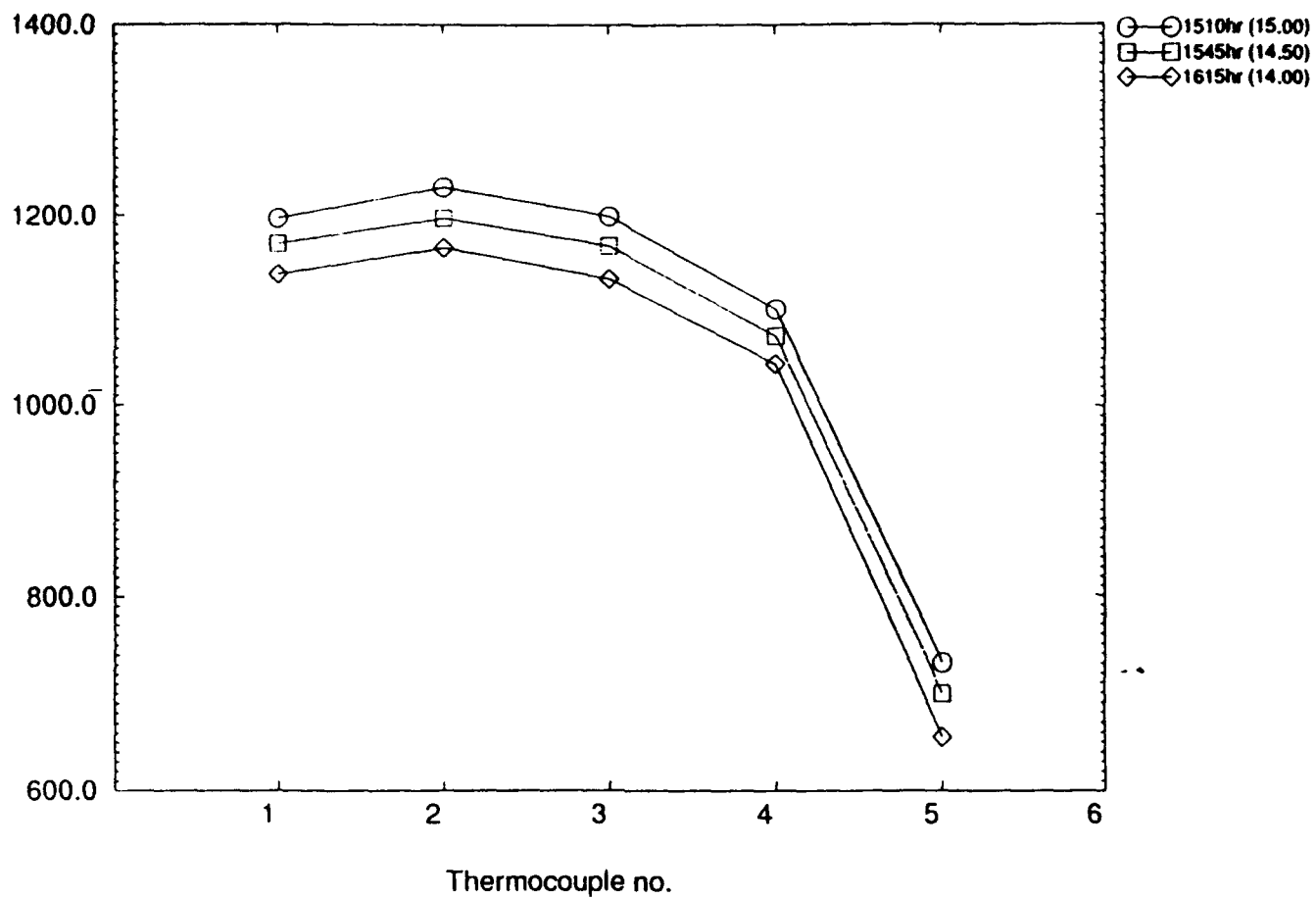
Figure 4. Temperature variation as a function of the power input. The power is expressed
here in terms of the applied voltage. (a) temperature on the outside of the
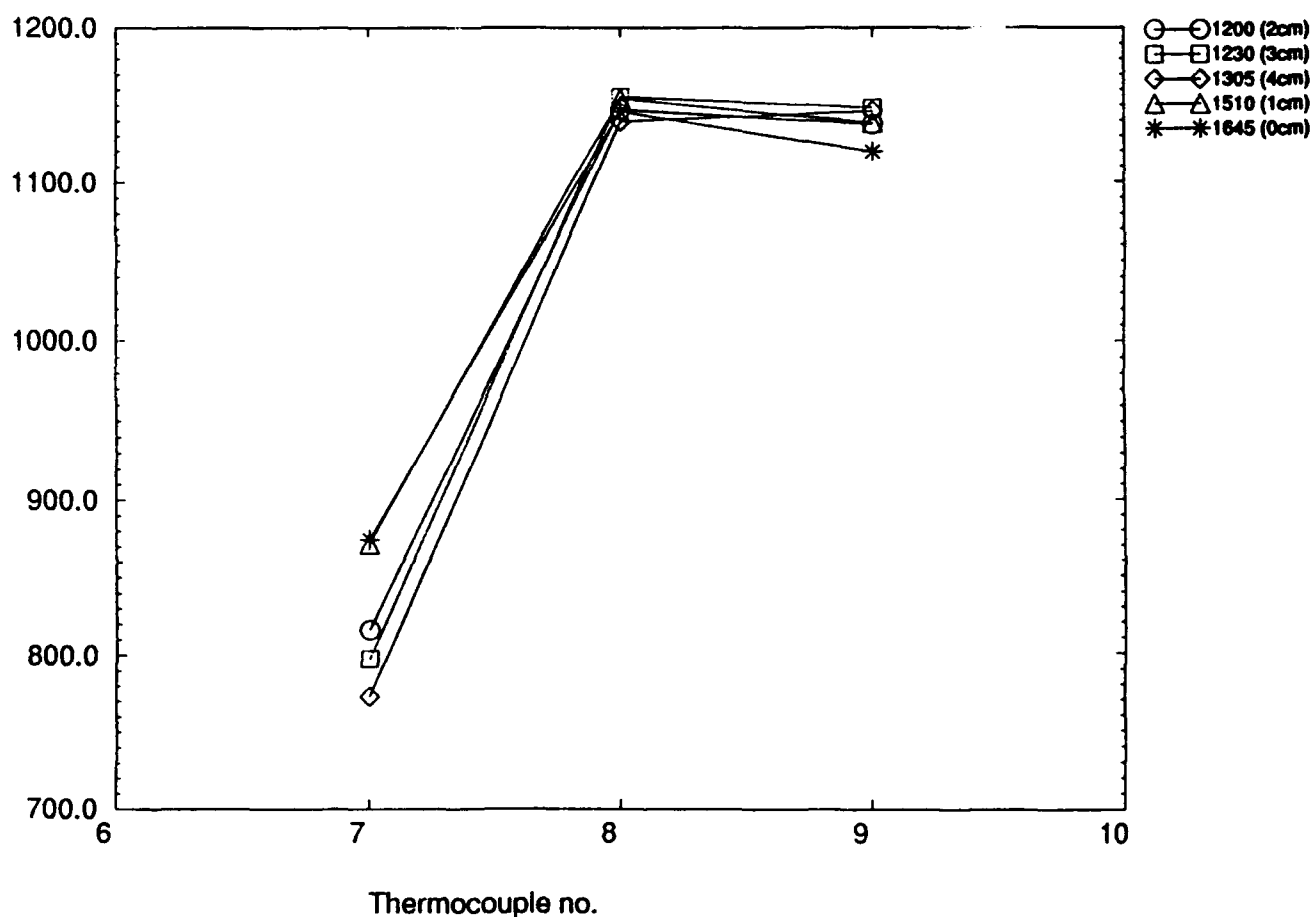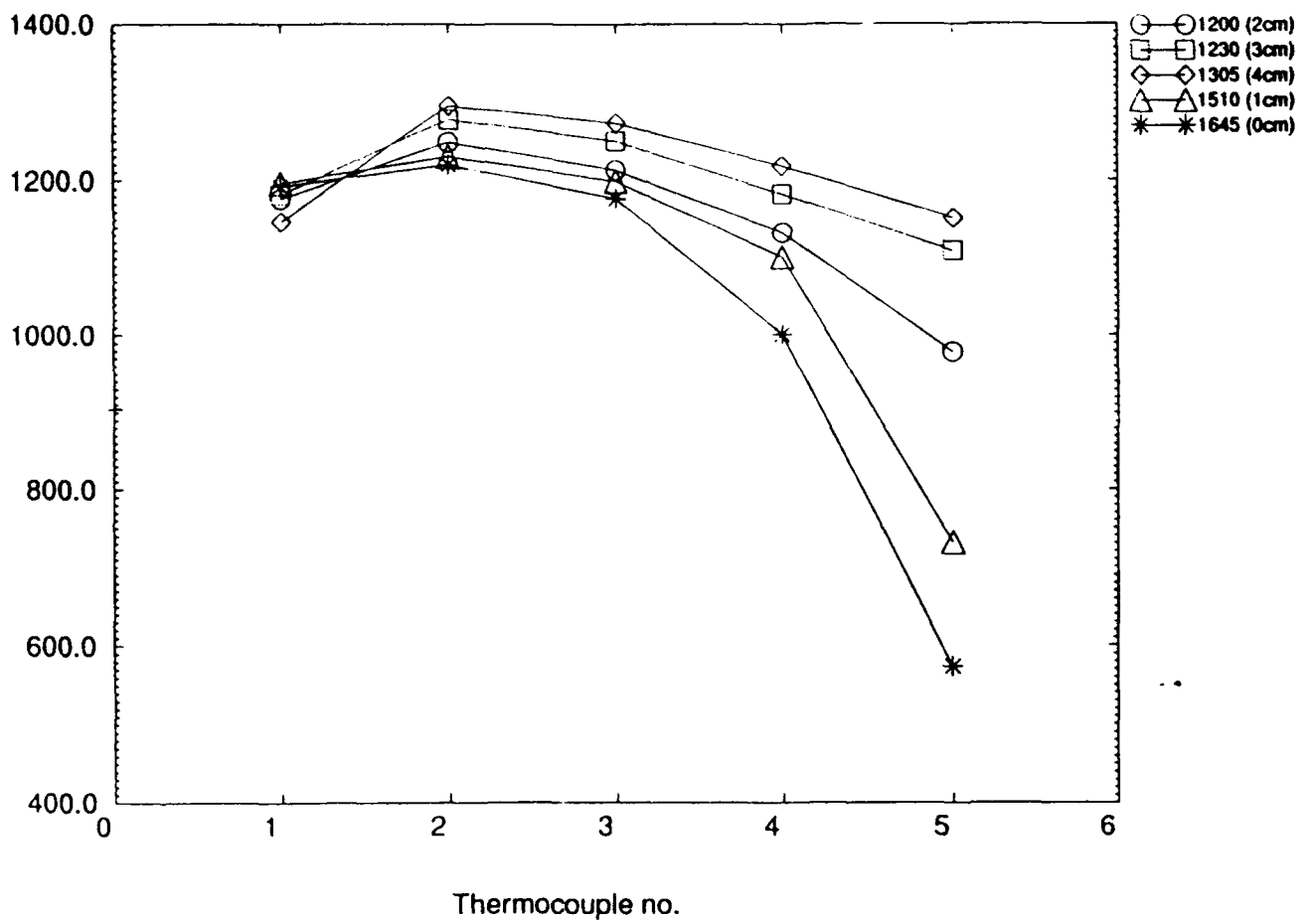graphite susceptor and (b) temperature on the inside of the crucible wall.

Figure 5. Temperature variation as a function of the crucible location. (a) temperature on the outside of the graphite susceptor and (b) temperature on the inside of the crucible wall.

3-11

NUMERICAL STUDY OF GAS FLOW IN AN HPCG FURNACE

As reported earlier [14,15], strong convective flows are induced in an HPCG furnace. The flow is oscillatory and turbulent, and has a complex pattern. A numerical research program was therefore initiated to study the gas flow phenomena in a high pressure system. **Governing Equations.** The model used here to analyze the gas flow assumes that the flow is incompressible and laminar (turbulent flows may be considered in the future), the Boussinesq approximation is applicable, and the viscous dissipation is negligible. The gas flow is considered to be axisymmetric, and therefore, only two dimensional calculations are performed. With these assumptions, the non-dimensional governing equations can be written as follows:

Continuity:

$$\frac{1}{r}\frac{\partial}{\partial r}(ru) + \frac{\partial w}{\partial z} = 0 \tag{1}$$

R-Momentum:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial r} + w\frac{\partial u}{\partial z} = -\frac{\partial p}{\partial r} + \frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial u}{\partial r}) + \frac{\partial^2 u}{\partial z^2} - \frac{u}{r^2} \tag{2}$$

Z-Momentum:

$$\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial r} + w\frac{\partial w}{\partial z} = -\frac{\partial p}{\partial z} + \frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial w}{\partial r}) + \frac{\partial^2 w}{\partial z^2} - Gr\Theta \tag{3}$$

Energy:

$$\frac{\partial \Theta}{\partial t} + u\frac{\partial \Theta}{\partial r} + w\frac{\partial \Theta}{\partial z} = \frac{1}{Pr}\left(\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial \Theta}{\partial r}) + \frac{\partial^2 \Theta}{\partial z^2}\right) \tag{4}$$

which employ the following scales:

length: b, time: b^2/α , velocity: ν/b, and pressure: $\rho\nu^2/b^2$.

Here, the dimensionless velocities in radial (r) and axial (z) directions are represented by u and w, respectively, while p and t denotes the pressure and time. The dimensionless temperature, Θ, is defined as $(T - T_c)/(T_h - T_c)$, where T_h and T_c are the encapsulant surface and wall temperatures, respectively. Here, ν, α, and β are kinematic viscosity, thermal diffusivity and isobaric coefficient of thermal expansion, respectively. The crucible radius, b, has been chosen as the primary length dimension. The governing parameters Gr and Pr are defined as:

Grashof number, $Gr = g\beta b^3(T_h - T_c)/\nu^2$ \hfill (5)

Prandtl number, $Pr = \nu/\alpha$ \hfill (6)

The flow boundary conditions are relatively well-defined since the gas occupies a specifically prescribed control volume with rigid boundaries except the encapsulant surface. For present calculations, the encapsulant surface has also been assumed to be rigid. No-slip conditions are therefore employed on all boundaries. However, the thermal boundary conditions are not as accurately known, since precise conditions within the real systems are not known. For present calculations, we consider $\Theta = 1$ at the encapsulant surface while the furnace walls are assumed to be at $\Theta = 0$. More realistic boundary conditions based on temperature measurements may be employed in future.

In this numerical algorithm, the solid region within the computational domain is introduced by imposing an extremely large artificial viscosity in that region so that no flow can take place in this area. Heat conduction is allowed to take place through the partitions walls/heat shields. This does not pose any difficulty since the energy equation (4) automatically reduces to the conduction equation when the velocities are zero. Note that the thermophysical properties can be allowed to vary from grid to grid [16,17].

Numerical Scheme. The governing equations (1)-(4) are discretized using the control volume integral approach and a staggered grid. It is Patankar's SIMPLER method which is used to solve the equations [18]. In this scheme, each term is integrated over its volume element. In a similar manner, the continuity equation is integrated over the same control element, then multiplied by the dependent variable of the equation being discretized. The discretized continuity equation is eventually subtracted from the discretized momentum or energy equation. This manipulation enables the flow field to always satisfy mass conservation. Two other equations, the pressure and pressure-correction equations are obtained by substituting the discretized momentum equations into the continuity equation. The latter has the task of correcting the velocities during iterations, while the former improves the pressure field. The SIMPLER method is based on a semi-implicit procedure. In the present algorithm, the convective fluxes are approximated by the power law scheme [16-18] which is one of the most robust algorithms. It approximates the central difference when velocities are small, and gradually mimics the upwind difference when velocities becomes large.

The discretized equations are underrelaxed and solved iteratively, within each time step to resolve the coupling and nonlinearity of the flow and energy equations, using a line-by-line tri-diagonal matrix algorithm. Sweeps are performed in both forward and backward directions so that the effects in one part of the computational domain may be felt throughout the entire domain without bias. Convergence at each time step is declared when the mass source and the residuals of the discretized equations are less than a specified absolute relative tolerance. Strict convergence within each time interval is crucial for the accuracy of the final solutions, particularly when the flow is oscillatory. Since the temperature gradients are generally strong

in the vicinity of the furnace walls, the quartz shields and the encapsulant surface, a suitable non-uniform grid distribution is used for the present computations. Only limited trials runs were carried out to optimize the grid distribution, convergence criteria and time step since our objective here is to obtain some qualitative information first and see how do they compare with the observed flows.

Numerical Results. Figures 6 and 7 present the flow and temperature fields in half of the cylindrical enclosure with a partition for Gr $= 10^6$ and 5×10^7, respectively. For these calculations, the gas properties are assumed to be constant. As can be seen in Fig. 6a, there exist two main recirculating cells, one on top of the encapsulant (rotating counter-clockwise) and the other on the outside of the partition wall extending to the top (rotating clockwise). The mechanism of formation of these cells is straight-forward. The gas is heated by the encapsualnt surface and moves up along the quartz shield since the cold gas flows downward along the seed rod which is at a lower temperature than the B_2O_3 surface (Fig. 1). On the outside of the quartz shield, the hot gas rises along its surface and flows downward along the cold furnace wall. This is a basic flow pattern which is induced by the thermal conditions in this HPCG furnace and agrees with the visualization reported earlier [14,15].

Figure 6(b) also shows that a large temperature gradient exists near the B_2O_3 surface and even temperature inversion is possible. The temperature variations in the region outside of the quartz shield and in the upper portion of the furnace are weak. However, a strong flow is induced in this region primarily because of the shear effects produced by the inner cell, and a strong buoyancy force, since the length scale is very large. Note that the Grashof number is proportional to the third power of the length (equation 5).

Figure 7 (Gr $= 5 \times 10^7$) demonstrates the effect of an increase in the buoyancy force. Although the primary flow pattern remains the same, the convective flow oscillates at this Grashof number (Fig. 7a). Several secondary cells whose shapes and sizes change with time are produced and the flow becomes extremely complex. This supports our earlier observation that the flow in an HPCG furnace is oscillatory [14,15]. Note that this Gr is much lower than the real value encountered in an HPCG furnace, and at higher Grashof numbers, the flow can certainly become chaotic and turbulent. Figure 7b further shows that with an increase in Grashof number, the plume effect has been extended much beyond the vicinity of the B_2O_3 surface producing large temperature gradients in other regions (compare with Fig. 6b).

In the above calculations, the presence of the phosphorus injector inside the furnace was not considered, as well as the thermophysical properties of the nitrogen gas were assumed to be constant. A set of calculations was therefore performed to examine the effects of these parameters. The phosphorus injector was placed at its topmost location which simulates the
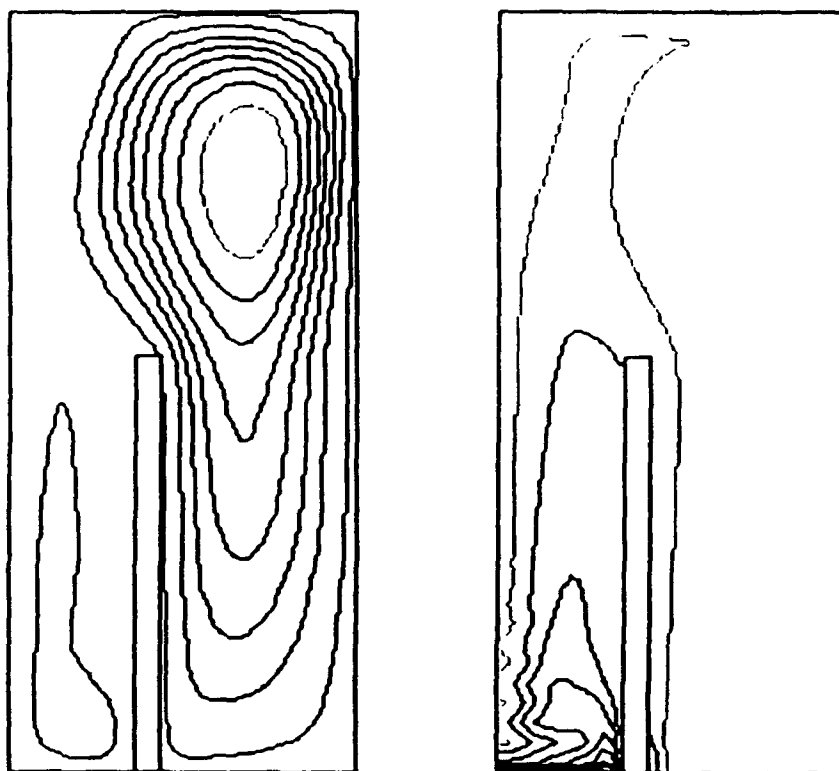
Figure 6. Predicted temperature and flow fields in right half of the enclosure for Gr $= 10^6$, the effect of a partition wall, (a) flow field and (b) temperature distribution.
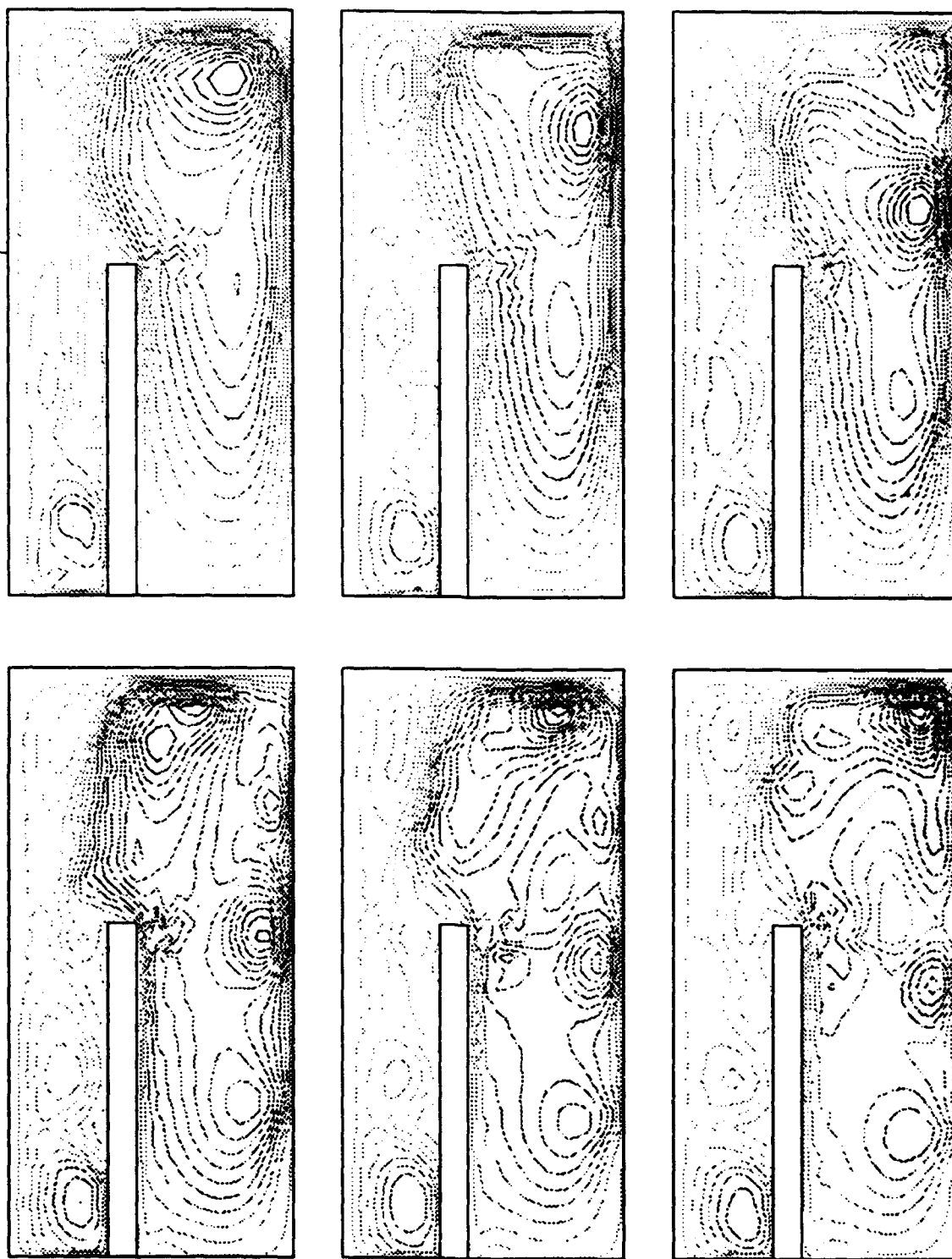
Figure 7a. Predicted flow fields in right half of the enclosure for Gr = 5×10^7, the effect of a partition wall.
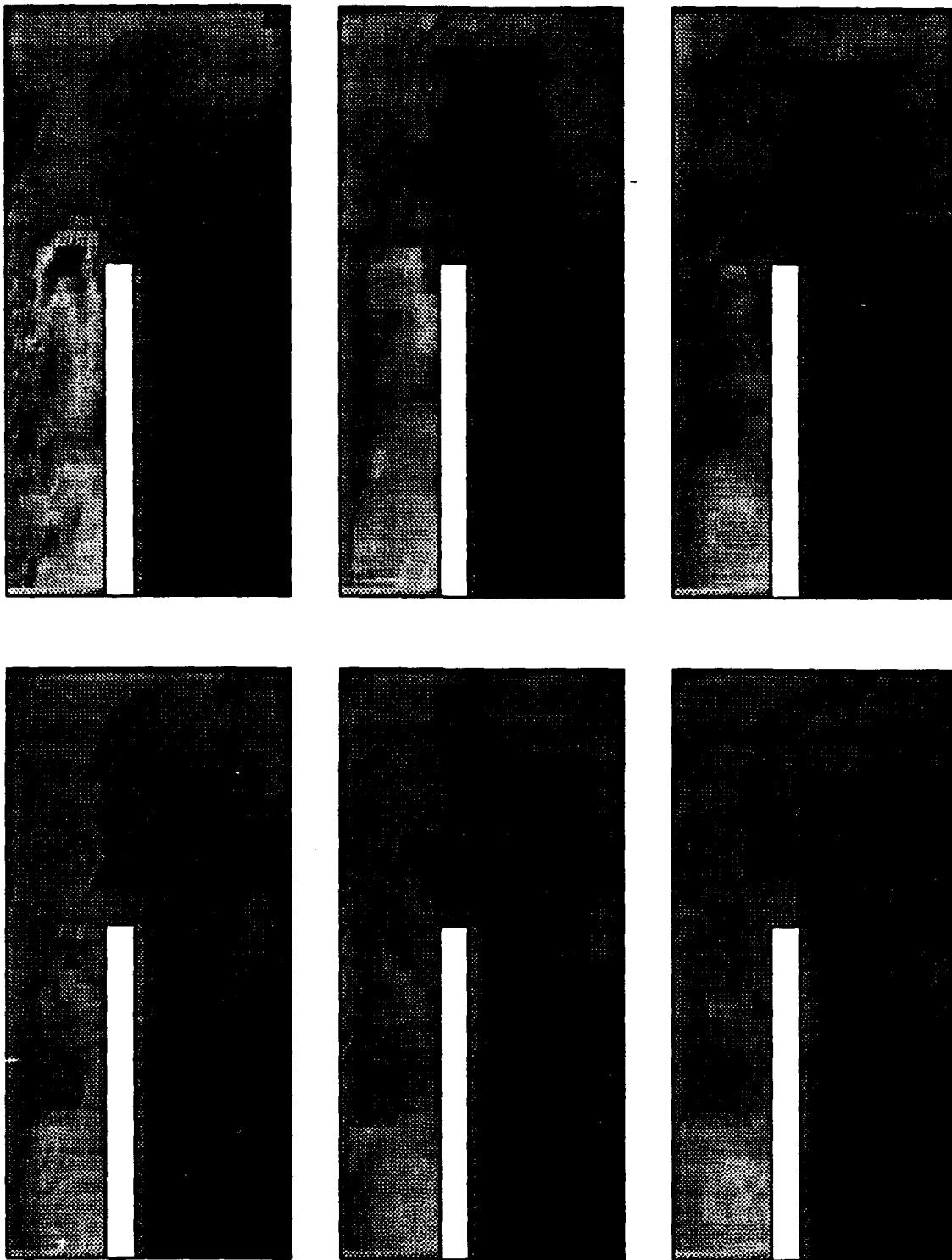
Figure 7b. Predicted temperature field in right half of the enclo   or Gr = 5x10^7, the effect of a partition wall.

crystal growth conditions, and the gas properties were allowed to vary with temperature. These results are presented in Figs. 8-9 for Gr = 10^6, 5×10^7 and 5×10^8, respectively.

In comparison to Fig. 6a, it is now observed that several cells are produced in the melt region at Gr = 10^6 (Fig. 8a), one on top of the other (underneath the P-injector). This is primarily because the phosphorus injector restricts the flow passage, and the gas can go out only through a small passage between the injector and the heat shield. The flow in the outer region has also become weaker and does not extend to the top portion of the furnace as much as it did earlier. As a result the temperature gradient is also localized in the melt region (Fig. 8b).

An increase in Gr again complicates the flow, and two large cells are produced in the outer region (Fig. 9a). These two convective cells have resulted from the hot gases coming out through the gap between the heat shield and the injector, and getting cooled on the furnace side wall and top as well as by the seed rod. The top surface and seed rod provide negative buoyancy for the upper cell while the side wall produces a second cell rotating in the clockwise direction. Several secondary cells are produced when the Grashof number is high although the basic pattern of the flow remains the same. The flow at Gr = 5×10^8 is oscillatory (not shown here). It seems that the flow restriction has reduced the buoyancy effect in the outer region thereby suppressing the oscillatory behavior to some extent. Figure 9(b) demonstrates the complexity of the temperature fields at high Grashof numbers. The temperature fields reported in Figs. 6(b)-9(b) qualitatively agree with the temperature measurements reported earlier.

CONCLUDING REMARKS

In order to develop a successful process for one-step, in-situ synthesis of phosphorus vapor and indium melt and growth of InP crystals of large sizes, uniform properties and at lesser cost, it is extremely important to understand the mechanism of heat transfer and gas flow in a high pressure crystal growth system. A series of experiments and numerical computations have been performed to study the thermal conditions in an HPCG furnace. Temperature distributions have been measured for the first time on the outer surface of the graphite susceptor and the inner surface of the crucible. The effects of variation in power supply and the vertical location of the crucible have also been examined. The numerical results demonstrate interesting flow patterns and show that the flow becomes oscillatory at high Grashof number.

Even though these experiments and computations provide a much better understanding of the HPCG furnace for one-step, in-situ synthesis and growth of InP crystals, they are by no means complete. Further experiments and simulations are needed to characterize the heat
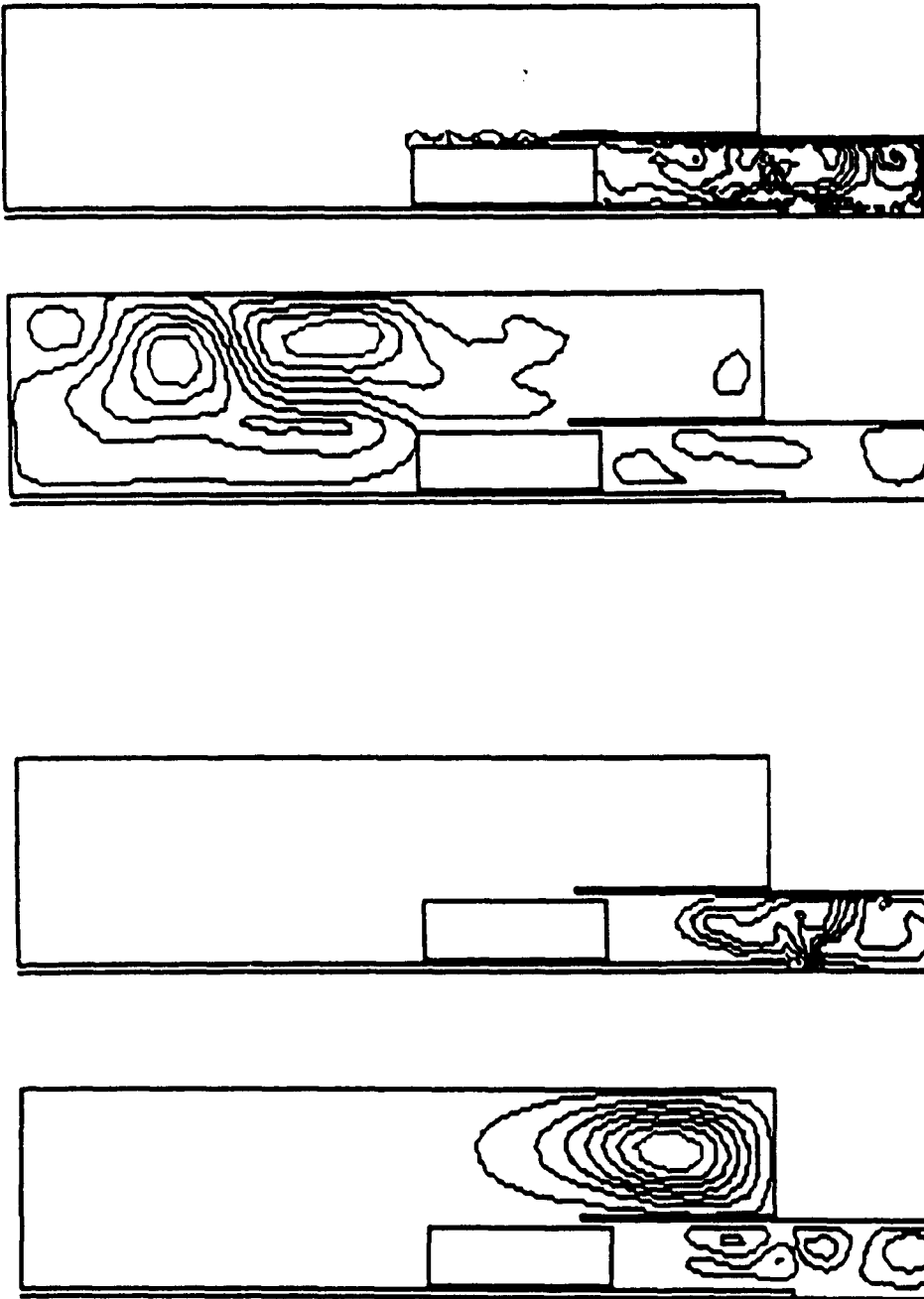
Figure 9. Predicted temperature and flow fields in right half of the enclosure for Gr = 5x10⁷, the effect of a partition wall and the phosphorus injector, (a) flow field and (b) temperature distribution.
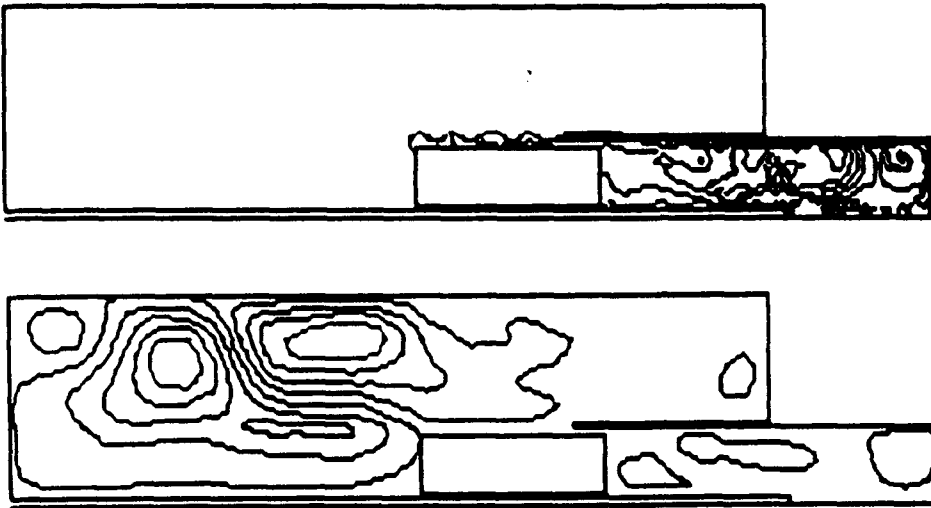
Figure 8. Predicted temperature and flow fields in right half of the enclosure for Gr = 10⁶, the effect of a partition wall and the phosphorus injector, (a) flow field and (b) temperature distribution.

transfer and gas flow under the growth conditions. A simultaneous theoretical/numerical and experimental research program will further help in determining the optimized conditions for both the synthesis and the growth, and in suggesting the modifications in the existing system in order to achieve these conditions.

Acknowledgements

REFERENCES

1. J. P. Farges, *J. Crystal Growth*, 59, 665-668 (1982).
2. J. P. Farges, *2nd NATO Workshop on Materials Aspects of InP*, 1, 9.1-6 (1983).
3. G. A. Antypass, *2nd NATO Workshop on Material Aspects of InP*, 1, 8.1-9 (1983).
4. S. B. Hyder and C. J. Holloway, Jr., *J. Electron, Mater.*, 12, 575-585 (1983).
5. Y. Sasaki, J. Nakagawa and K. Kurata, *46th Aut. Meet., Jpn. Soc. Appl. Phys.* (1985).
6. S. B. Hyder and G. A. Antypass, *3rd NATO Workshop on Material Aspects of Indium Phosphide*, 1 (1986).
7. D. J. Dowling, R. A. Brunton, D. A. E. Crouch, A. J. Thompson and J. E. Wardill, *J. Crystal Growth*, 87, 37-41 (1988).
8. D. F. Bliss, R. M. Hilton and J. A. Adamski, *4th Int. Conf. on Indium Phosphide and Related Materials*, 262-265 (1992).
9. T. Inada, T. Fujii, M. Eguchi and T. Fukuda, *J. Crystal Growth*, 82, 561-565 (1987).
10. T. Inada, T. Fujii, M. Eguchi and T. Fukuda, *Appl. Phys. Lett.*, 50, 86-88 (1987).
11. K. Kohiro, K. Kainosho, H. Shimakura, T. Fukui and O. Oda, *4th Int. Conf. on Indium Phosphide and Related Materials*, B.5, 35-38 (1992).
12. S. Bachowski, D. F. Bliss, B. Ahern, R. M. Hilton, J. Adamski and D. J. Carlson, *2nd Int. Conf. on Indium Phosphide and Related Materials*, 30-34 (1990).
13. D. F. Bliss, J. A. Adamski, W. M. Higgins, V. Prasad and F. X. Zach, *Int. Conf. on Indium Phosphide and Related Materials, Paris* (1993).
14. V. Prasad, Thermal Characterization of In-situ Synthesis for LEC/MLEK Growth of InP Single Crystals, Final Report for Summer Research Program 1992, AFOSR.
15. V. Prasad, D. F. Bliss and J. A. Adamski, *J. Crystal Growth* (submitted).
16. A. Anselmo, V. Prasad, J. Koziol and K. P. Gupta, *J. Crystal Growth*, 131, 247-264 (1993).
17. A. Anselmo, V. Prasad, J. Koziol and K. P. Gupta, *J. Crystal Growth*, in press (1993).
18. S. V. Patankar, *Numerical Heat Transfer and Fluid Flow*, Hemisphere Publ. (1980).

A MOLECULAR DYNAMICS SIMULATION OF ELECTROMIGRATION

H.F. Helbig
Professor
Department of Physics

T. Bartelt
Graduate Student
Department of Physics

Clarkson University
Potsdam. NY 13699

A MOLECULAR DYNAMICS SIMULATION OF ELECTROMIGRATION

H.F. Helbig
Professor
Department of Physics
Clarkson University

T. Bartelt
Graduate Student
Department of Physics
Clarkson University

Abstract

A flexible model of polycrystalline metallization on integrated circuits was developed, and the construction of a computer simulation to study electromigration in such structures has begun. This work was part of an ongoing effort to determine whether microcircuit failure due to electromigration can be reduced or eliminated. Experiments have revealed that void formations in aluminum interconnections on integrated circuits result from current-induced transport of metal atoms [1]. A computer model is being developed to simulate this effect. The model will permit variation of boundary conditions, including grain size and orientation, thermal effects, crystallite composition, as well as interatomic potentials. The roles of grain boundaries and interstitial atoms are of particular interest since these are parameters that can be controlled, to some degree, during fabrication. The computer code is not complete; specific results are not reported. However, progress was made in the development of the model, and significant portions of the code have been written. A commitment to completion endures regardless of further support.

A MOLECULAR DYNAMICS SIMULATION OF ELECTROMIGRATION

H.F. Helbig

T.Bartelt

Introduction

Microcircuit failure due to current-driven changes in the interconnections (traces) between circuit elements is a serious obstacle to reliable performance. Such failures are due to the transport of metal atoms away from their original, more or less uniform, distribution along the trace; a process called electromigration. Operating temperature, thermal gradients, mechanical stress, current density, grain size and structure, and defects (particularly vacancies and impurities) affect failure rate. Experimental studies to determine the effect of each of these parameters are challenging because of the difficulty of sample preparation and the destructive nature of both the experiment and measurement techniques. A realistic, three-dimensional, polycrystalline model that permits control over individual parameters can be pivotal to the understanding of this phenomenon.

If one represents the interactions between atoms by empirical potential functions (rather than doing *ab initio*, quantum mechanical, many body calculations), it is currently possible to handle the motions of thousands of atoms, in reasonable amounts of computing time, without resort to super-computers. Once the concepts described here are proven on commonly available machines (386/486 PCs and workstations), it will undoubtedly be desirable to move the code to one of the national super-computer facilities to study larger structures. The simulation code is being written in the C programming language so that it will be easy to port to a wide range of computers.

The primary interface to the program is a text file containing information about initial condition, including crystal structure and composition. The output will be positions and velocities of each of the atoms, directed to files at predetermined intervals. Additional modules will be used to analyze the output files to determine such things as diffusion coefficients and failure rates. A graphical interface would also be useful, and is being considered. Linking a graphical interface directly to the main program, however, would severely restrict portability.

The principal objective of this project is to study the migration of metal atoms in a polycrystalline conductor as they react to thermal and "electron wind" processes. That objective is to be accomplished in four steps proceeding, generally in the following sequence.

1. Completion of the simulation program, MDEM

2. Preparation of test sample configurations

3. Collection of simulation data

4. Comparison of simulation results and experimental data

The sections to follow describe these steps in detail.

Step 1. describes the simulation program by subheadings corresponding to the names of the major C-code modules from which MDEM is constructed. The functions that calculate the various forces that act on the atoms represent choices about the fundamental physical mechanisms believed to control atomic migration in a conductor. For that reason those functions contain more extensive discussion than those that attend mainly to bookkeeping chores.

Step 2. demands special comment as it includes an effort complementary to the simulation as originally conceived. In particular, we propose to simulate the deposition of the metallization onto its substrate. This will provide an alternative and, perhaps, more realistic starting configuration for the simulated conductor than so far provided by MDEM.

Step 3. describes the sorts of data that we plan initially to extract from the simulation.

Step 4. is the pivotal, feedback step. There is already a wealth of experimental data that document the onset and evolution of electromigration damage, e.g., void formation and hillock growth [1]. Although such data are microscopic in scale, the smallest features are still gross on the scale of atoms. The chief utility of the simulation will be to show the connection between processes occurring at the atomic level and the experimental data. We hope that the ability of the simulation to make this connection will suggest new lines of experimental effort at Rome Lab, and ultimately provide both fundamental and practical information that can be used to reduce circuit failures due to electromigration.

1. Completion of MDEM

The main functional modules of MDEM are listed below along with a description of the services they provide and an indication of their current status.

input_configuration   (outlined; under development)

This reads and parses the config"ration input file provided by the user. The configuration file contains all the information about the atoms required for the simulation. Apart from data relevant to the entire sample (e.g., temperature), this information is organized as a set of crystallite descriptions. The data format is very general, utilizing a system of "tags" so that the configuration file can contain documentation that is ignored by the parser, and so that it is easy to allow for additional input parameters, should they become necessary, without invalidating earlier configuration files.

Parameters currently anticipated for the configuration file are the system temperature, and for each crystallite, atom species, impurity species and concentration, vacancy concentration, the shape, size, orientation and position of the crystallite, the character of each of its boundary planes (are the atoms near that plane free or pinned, and/or do they act as thermostats?), the lattice vectors defining the shape, size, orientation and position of the primitive unit cell, and the constants that define the interaction energy between pairs of atoms and between the atoms and the electron gas.

Choosing the parameters for the configuration file is the present solution to the "sample preparation" problem. This and an alternative solution, simulation of the film deposition process itself, are discussed in section 2, below.

make_crystallite   (code written)

This uses the information gleaned by input_configuration to initialize the atomic data array, crystallite by crystallite. The atomic data array keeps track of the positions and velocities of the entire set of atoms in the sample. This routine must do a substantial amount of bookkeeping and geometric analysis to assure that the user's input data is self-consistent. Its construction consumed much of the two month Summer Research Program.

4-5

modify_crystallite   (code outlined)

This is actually a sub-module of make_crystallite that reprocesses the atomic data array to create vacancies, impurities or to perform other operations that might be desired as part of the sample initialization.

monitor   (code to be written)

This is the module that calls for integration of the equations of motion and monitors the elapsed time or other cues that have been requested by the user to trigger the collection of data.

integrate   (code written)

This performs a numerical integration of Newton's equations of motion ($F_i = m_i a_i$) for each of the atoms in the sample using forces that are described under the next code module. An efficient algorithm (4th order Runge-Kutta with adaptive step-size) has been adapted for use in MDEM [2]. The integration routine takes the system from its current state (the positions and velocities of all the atoms at time t) to its state at time t + Δt under the forces assumed to control the atomic motion.

Adaptive step-size controls the size of Δt so that the error of the numerical integration is maintained within acceptable bounds; small enough that the output is free of computational artifacts and large enough that the evolution of the system proceeds at a reasonable rate. It is expected that the natural time scale for the system motion will be on the order of the period of lattice vibrations, i.e., about 100 fs, so that Δt will necessarily be somewhat smaller than this.

Rather than include the force model in the integration routine, it is embodied in a separate routine so that new force models can be substituted easily.

emt_force   (code outlined)

This provides the force exerted on each atom according to the recently developed "effective medium theory. Forces associated with thermal control of the sample and with the electron wind force presumably responsible for

electromigration will be supplied by routines similar to this, but are discussed under this module for compactness of presentation.

In order that a molecular dynamics simulation of a metal properly reproduce the elastic and thermal properties (e.g., Young's modulus, thermal expansion coefficient) characteristic of real metallic systems, it has been found necessary by several researchers [3,4,5,6,7] to describe the system energy by functions more sophisticated than pair potentials. This is because a major portion of the cohesive energy in metal systems arises from the interaction of the metal ions with the "gas" of nearly free conduction electrons rather than from the direct interaction of the ions with their neighbors. On this account, we have begun to adapt the effective medium theory of Jacobsen et al.[3], as extended by Häkkinen and Manninen [6] to MDEM. Fortunately, this recent literature provides a background of work on aluminum and copper that will be directly applicable to the electromigration problem.

Preliminary tests of the code will be made using simple pair potentials of the Lennard-Jones or Morse type.

It is important to provide the simulation with a means to add or remove energy from the system. In nature, this is accomplished by exchanges of energy between the system and its environment, but simulations cannot afford the luxury of nearly infinite heat reservoirs. Hoover [8] has shown a simple technique for temperature control that properly preserves the fluctuations characteristic of the canonical ensemble distributions of statistical mechanics. This technique was employed by Valkealahti and Manninen [7] in the work on the melting of copper clusters.

In short, the idea is to endow some of the atoms in the simulation with an additional, velocity-dependent force that behaves like a friction force except that the friction coefficient may be either positive or negative. By linking the friction coefficient to the average system kinetic energy (temperature), energy may introduced or extracted from the system as required to stabilize the system temperature. The obvious location for "thermostat" atoms is at the face of the crystallite intended to simulate the substrate since that face would in nature serve as the main heat sink for the conductor.

For numerical experiments involving thermal effects alone, this thermostat mechanism can be used to maintain thermal equilibrium. For experiments also involving the I^2R heating associated with an electron wind, the thermostats will allow the maintenance of steady-state conditions including thermal gradients. Thus, the simulation may be able to complement the work by Rome Lab's Dr. Mark Levi on the importance of thermal gradients to the electromigration process.

Finally, the passage of an electric current through the sample will be simulated by adding random momentum increments (biased appropriately in the direction of the electron drift) to the metal atoms in the sample. It is

anticipated that these momentum increments, when applied to atoms immediately upstream from vacancies, will prove to be the agent responsible for the preferential migration of metal atoms along the direction of the electron flow and of vacancies in the opposite direction.

It is difficult in laboratory experiments to discover whether this electromigration scenario is accurate. The reason is that in normal circuit operation the process is slow. In terms of the atomic picture, the significant collisions are rare, and they are accompanied by vast numbers of electron-phonon collisions that are responsible for Joule heating but are relatively ineffectual as a cause of metal migration. Experimental attempts to accelerate the migration-stimulating collisions are accompanied by heating effects that alter the conditions under which electromigration occurs in practice.

The simulation offers the possibility of separating the effects of collisions between electrons and vacancy-proximate atoms from those that merely add heat to the lattice. It seems reasonable, at least during the nucleation stages of void formation, that these two mechanisms are independent and their effects additive. If this is so, it may be possible to accelerate this atom-plus-vacancy migration mechanism without materially increasing the Joule heating.

This could be done, for example, by applying a heavy dose of electron-wind to randomly selected atoms during each time step, moving the atoms according to Newton's laws, and then scaling the speed of each atom down to prevent the average kinetic energy (temperature) from rising. If atoms next to vacancies are, indeed, the primary transport objects, this method would reveal that fact without increasing the system temperature. The result will be to compress the time scale on which effects of this electron-wind model emerge while controlling the sample temperature independently.

Whether this artificial method of temperature control will produce the same migration effects as result from prolonged current flow at lower current densities is a question that can be explored with the simulation.

randn & gausdevn   (code completed)

These are (*pseudo*)random number generators, called upon by many routines. Although generic *linear congruential* random number generators, supplied with most compilers, might be adequate for some of the simulation's requirements (e.g., assigning random initial velocities and displacements to atoms in the system), they are always suspicious, and undoubtedly insufficient for electron wind-lattice interactions. The algorithms chosen here are adaptations of the work of Press *et al* [2]. Randn produces (*pseudo*)random numbers with uniform distribution,

gausdevn yields a Gaussian distribution.

Additional routines that are not linked with MDEM include:

comb_data   (code to be written)

This combs through the atomic data array collecting information for such calculations as mean square displacements, radial distribution functions, and density profiles.

view3d   (under development)

This presents a perspective view of the sample atoms so that the user may form visual impressions of the sample. Naturally, this is where portability breaks down since graphics standards vary with platforms. This important task, however, is not as computationally intensive as the simulation itself, so it will not be necessary to adapt it to every machine. A program that allows one to visually scan through the sample is being developed for some of the most popular operating systems.

2. Preparation of test sample configurations

The first step in simulating a conducting line is to create the line. In the manufacture of microcircuits the line is prepared, first by depositing a film of metal atoms onto a substrate, and then by etching away unwanted regions of the film. During the deposition, islands of metal grow around nucleation sites randomly located on the substrate. Each island may be thought of as a randomly oriented and imperfect crystallite or grain. The islands grow until adjacent grains meet, and the resulting interfaces become grain boundaries of a polycrystalline film. What remains after the etching stage is a polycrystalline conductor whose transverse dimensions are comparable in size with the grains.

Preparation of a simulated polycrystalline conductor may be accomplished in at least two ways. The starting configuration may be tailored to test various idealized arrangements of grain boundaries, vacancy and impurity concentrations. Alternatively, it is possible to simulate the island growth process itself.

Completed components of MDEM provide for the first alternative. The user of the program can assemble a polycrystalline sample by specifying the shape, position and orientation of each grain. Furthermore, each grain may

be specified to contain atoms of a given species and be interspersed with vacancies and impurity atoms of other species. While most grains will contain "metal" atoms, some may be populated with atoms bound more tightly than those simulating metals, and be deployed as substrate and passivation structures.

Simulation of the deposition process may be of considerable value if it can provide more realistic starting configurations. We hope to be able to include this as an additional facility.

3. Collection of simulation data

The simulation will necessarily maintain a record of the current state of the system in microscopic detail; that is, the positions and velocities of all the atoms and of the time elapsed. In order to interpret these data as they evolve in time, it will be necessary to obtain periodic summaries of various macroscopic statistics. Preliminary parameters that will be calculated from the position/velocity data include the elastic moduli and thermal expansion coefficients of the modeled metal. Quantities more closely related to the electromigration process include system energy, density and temperature gradients and fluctuations, and diffusion coefficients.

4. Comparison of simulated and experimental data

The data collected from the simulation may be displayed as tables, graphs and visualizations. After ascertaining that the simulation properly reproduces the known mechanical and thermal properties of the metal, it will be of major interest to determine the effects of the orientation of grain boundaries, relative to the direction of current flow, on the transport of atoms and on the concentration of vacancies. Are these effects consistent with observed electromigration phenomena? We anticipate that graphics to present three dimensional visualizations will be crucial to these comparisons.

Final comment

The generality of this simulation scheme is noteworthy. In addition to the experiments already indicated, other applications are easily imagined. Studies of thermal and mechanical stress effects could be made in conjunction with electromigration or separately. Looking even farther afield, nothing inherently restricts the modeled system to the solid state. Studies of interfaces between (or among) any of the phases could easily be accommodated.

A project of this magnitude could not have been completed in the twelve weeks of the Summer Research Program. Consequently, specific results are not reportable. We have, instead, presented an outline of the project, and indicated the progress made toward completion. These accomplishments have led to a the negotiation of a contract to complete the simulation at Rome Lab between September, 1993 and July 1994. The results of preliminary studies are to be reported at the 1993 Fall Meeting of the Materials Research Society [9].

References

[1]   See, for example, *Phenomenological Observations on Electromigration*,
      R.W. Thomas and D.W. Calabrese,
      21st Annual Proceedings IRPS, p. 1, (1983).
      and the numerous micrographs obtained in association with this work.

[2]   *Numerical Recipes*,
      W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling,
      Cambridge University Press.

[3]   *Interatomic Interactions in the Effective Medium Theory*,
      K.W. Jacobsen, J.K. Nørskov and M.J. Puska,
      Phys. Rev. B35, 7423 (1987).

[4]   *Monte Carlo Calculation of the Thermal Expansion Coefficient of Al*,
      P. Stoltze, K.W. Jacobsen and J.K. Nørskov,
      Phys. Rev. B36, 5035 (1987).

[5]   *Disordering and Melting of Aluminum Surfaces*,
      P. Stoltze, J.K. Nørskov and U. Landman,
      Phys. Rev. Lett. 61 440 (1988).

[6]   *The Effective-medium Theory Beyond the Nearest-neighbor Interaction*,
      H. Häkkinen and M. Manninen,
      J. Phys.: Condens. Matter 1 9765 (1989).

[7]   *Melting of Copper Clusters*,
      S. Valkealahti and M. Manninen,
      Computational Materials Science, 1, 123 (1993).

[8]   *Canonical Dynamics: Equilibrium Phase-space Distributions*,
      William G. Hoover,
      Phys. Rev. A31 1695 (1985).

[9]   *A Molecular Dynamics Simulation of Electromigration*,
      H.F. Helbig, T. Bartelt, L.H. Walsh and J.V. Beasock
      Submitted to MRS Fall Meeting, Boston MA, 29 Nov. - 3 Dec., 1993.

AN EFFICIENT METHOD AND CODE FOR DETERMINING TE AND TM
PROPAGATION CONSTANTS OF INFINITE ARRAYS OF PRINTED
ANTENNAS ON DIELECTRIC SHEETS

Daniel K. Lee
Student
Department of Electrical Engineering

Southern Illinois University - Carbondale
Carbondale. IL 62901

and

Jean-Pierre R. Bayard, Ph.D
Associate Professor
Department of Electrical Engineering

California State University - Sacramento
6000 J Street
Sacramento, CA 95819-6019

AN EFFICIENT METHOD AND CODE FOR DETERMINING TE AND TM
PROPAGATION CONSTANTS OF MODES OF INFINITE ARRAYS OF
PRINTED ANTENNAS ON DIELECTRIC SHEETS

Daniel K. Lee
Student
Department of Electrical Engineering
Southern Illinois University - Carbondale

Jean-Pierre R. Bayard
Associate Professor
Department of Electrical Engineering
California State University - Sacramento

Abstract

This report presents an efficient method for finding the roots, TM and TE mode propagation constants for a periodic array of semi-infinite dielectric slabs backed by a ground plane. An adjustable step-searching method combined with a Newton-Raphson procedure gives very reliable and accurate results even for large dielectric thickness or permittivites. CPU time required for finding the roots is reduced considerably in comparison with the work in [1]. Results are presented for various thicknesses and permittivities of the substrate, as well as scan angles and frequencies, and are validated graphically.

AN EFFICIENT METHOD AND CODE FOR DETERMINING TE AND TM PROPAGATION CONSTANTS OF INFINITE ARRAYS OF PRINTED ANTENNAS ON DIELECTRIC SHEETS

Daniel K. Lee   and   Jean-Pierre R. Bayard

Introduction

The full-wave solution using the method of moments described in [1] for infinite arrays of printed dipoles on finite height dielectric sheets is to be considered. In particular, the search for the TM and TE propagation constants is to be improved for the periodic geometry shown in figure 1. In [1], the values for the TM and TE mode propagation constants are found where the determinants of the matrices in (19) and (20) vanish. The mode propagation constants are calculated using the Newton-Raphson method, with an initial guess from a quasi-TEM approximation. For a particular n(the Floquet index in y), the initial guess for the qth root for both TM and TE modes is found as equation (39) in [1]. Although, from the numerical result shown in [1], the initial guess using a quasi-TEM approximation worked well for relative dielectric constant less than five and for small values of t. for high relative dielectric constants or large values of t. the method did not necessarily converged to a root. Furthermore, the Newton-Raphson method is not guaranteed to find all the characteristic mode propagation constants even with the careful steps followed in [1].

Hence, the research will mainly focus on finding an efficient way of locating all the possible roots for both the TM and TE modes. This is of

out most importance to the numerical solution for these arrays, as missing a root or introducing a nonexistent root may signify, if such a mode is excited. that the dielectric-free space electromagnetic boundary conditions are not satisfied. Also. reducing the CPU time for locating these roots will be another important aspect of this research. In order to find an efficient method. the function will be evaluated analytically and its behavior will be investigated graphically and numerically. Using a variable step across the region in the complex plane where the roots should be located. initial guesses fairly close to the exact roots are provided to the Newton-Raphson procedure for accuracy refinement. The CPU time gain is achieved by designing a step search whose value is a function of both frequency and the magnitudes of the TM and TE characteristic equations. Once all the roots can be successfully located in relatively short amount of time. then the method can be used to study the characteristics of antennas such as dipoles printed on the protruding dielectric slab.

Method Formulation

In order to find the characteristic modes, for the array in figure 1, the fields in the dielectric and free-space regions are expressed using the TM_x and TE_x mode sets. As in [1], matching the electric and magnetic field components tangential to the dielectric-free space interface of the array unit cell results in 8 homogeneous equations, 4 of them in terms of the TM mode coefficients, and the others in terms of the TE coefficients. For the TM matrix,

$$
\begin{bmatrix}
-\dfrac{jk_{r1}}{\epsilon_r} e^{-jk_{r1}t} & \dfrac{jk_{r1}}{\epsilon_r} e^{jk_{r1}t} & jk_{01}e^{-jk_{01}t} & -jk_{01}e^{jk_{01}t} \\[2em]
e^{-jk_{r1}t} & e^{jk_{r1}t} & -e^{-jk_{01}t} & -e^{jk_{01}t} \\[2em]
-\dfrac{jk_{r1}}{\epsilon_r} & \dfrac{jk_{r1}}{\epsilon_r} & jk_{01}e^{-j(k_{01}-u_0)a} & -jk_{01}e^{j(k_{01}+u_0)a} \\[2em]
1 & 1 & e^{-j(k_{01}-u_0)a} & e^{-j(k_{01}+u_0)a}
\end{bmatrix}
$$

(1)

and for the TE matrix,

$$
\begin{bmatrix}
e^{-jk_{r2}t} & e^{jk_{r2}t} & -e^{-jk_{02}t} & e^{-jk_{02}t} \\[2em]
-jk_{r2}e^{-jk_{r2}t} & jk_{r2}e^{jk_{r2}} & jk_{02}e^{-jk_{02}t} & -jk_{02}e^{jk_{02}t} \\[2em]
-jk_{r2} & jk_{r2} & jk_{02}e^{-j(k_{02}-u_0)a} & -jk_{02}e^{j(k_{02}+u_0)a} \\[2em]
1 & 1 & -e^{-j(k_{02}-u_0)a} & e^{j(k_{02}+u_0)a}
\end{bmatrix}
$$

(2)

where

$$k_{r1} = \sqrt{\epsilon_r k_0^2 - V_n^2 - k_{z1}^2} \ , \tag{3}$$

$$k_{r2} = \sqrt{\epsilon_r k_0^2 - V_n^2 - k_{z2}^2} \ , \tag{4}$$

$$k_{01} = \sqrt{k_0^2 - V_n^2 - k_{z1}^2} \ , \tag{5}$$

$$k_{02} = \sqrt{k_0^2 - V_n^2 - k_{z2}^2} \ . \tag{6}$$

$$U_0 = k_0 \sin \theta_0 \cos \phi_0 \ , \tag{7}$$

$$V_n = k_0 \sin \theta_0 \sin \phi_0 + \frac{2n\pi}{b} \ . \tag{8}$$

$$k_0 = \omega \sqrt{\mu_0 \epsilon_0} \ , \tag{9}$$

θ_0 and ϕ_0 are the scan angle of the antenna beam, and k_{z1} and k_{z2} are the TM and TE mode propagation constants to be found. The solutions for these unknowns are the values, real or imaginary but not both, where the determinants of matrices (1) and (2) go to zero, permitting a non-trivial solutions to the homogeneous matrix equations.

Although the method used for root searching in [1] was reliable and accurate, it worked well for cases under certain limitations such as low dielectric constants($\epsilon_r < 5$) and very thin dielectric substrate($t < a/20$). Furthermore, numerical difficulties arose for solving the determinant when the exponents were large. In order to obtain a method to find a reliable and accurate root searching method that works for diverse cases while avoiding such numerical difficulties, the determinants of the TE and TM matrices were solved analytically using the co-factor method. Solving for the determinant of the two matrices analytically and using mathematical manipulations and rearrangements resulted in the following simple expressions:

5-6

The determinant of the TM matrix is,

$$TM(k_{z1}) = -\left\{ k_{01} - \frac{k_{r1}}{\epsilon_r} \right\}^2 \cos((t-a)k_{01} + tk_{r1})$$

$$+ \left\{ k_{01} + \frac{k_{r1}}{\epsilon_r} \right\}^2 \cos((t-a)k_{01} - tk_{r1}) - 4\frac{k_{r1}k_{01}}{\epsilon_r} \cos(u_0 a) = 0$$

$$(10)$$

Since the derivative of the function is also required in the Newton-Raphson method, it is found as:

$$TM'(k_{z1}) = -\left\{ k_{01} - \frac{k_{r1}}{\epsilon_r} \right\}k_{z1}\left\{ 2\left\{ \frac{1}{\epsilon_r k_{r1}} - \frac{1}{k_{01}} \right\} \cos(x_2) + \left\{ k_{01} - \frac{k_{r1}}{\epsilon_r} \right\} \right.$$

$$\left. \left\{ \frac{(t-a)}{k_{01}} + \frac{t}{k_{r1}} \right\} \sin(x_2) \right\} - \left\{ k_{01} + \frac{k_{r1}}{\epsilon_r} \right\} k_{z1} \left\{ 2\left\{ \frac{1}{k_{01}} + \frac{1}{\epsilon_r k_{r1}} \right\} \right.$$

$$\cos(x_1) + \left\{ k_{01} + \frac{k_{r1}}{\epsilon_r} \right\} \left\{ \frac{t}{k_{r1}} - \frac{(t-a)}{k_{01}} \right\} \sin(x_1) \right\} + 4\frac{k_{z1}}{\epsilon_r} \left\{ \frac{k_{01}}{k_{r1}} + \frac{k_{r1}}{k_{01}} \right\}$$

$$\cos(U_0 a) \qquad (11)$$

where

$$x_1 = (t-a) k_{01} - t k_{r1} \qquad (12)$$
$$x_2 = (t-a) k_{01} - t k_{r1} \qquad (13)$$

The determinant for the TE case is,

$$TE(k_{z2}) = -(k_{02}+k_{r2})^2\cos((t-a)k_{02} - tk_{r2})+(k_{r2} - k_{02})^2\cos((t-a)k_{02} + tk_{r2})$$

$$+ 4 k_{r2}k_{02} \cos(U_0 a) = 0 \qquad (14)$$

with its derivative given by,

$$TE'(k_{z2}) = (k_{r2}+k_{02})k_{z2}\left\{ 2\left\{ \frac{1}{k_{02}} + \frac{1}{k_{r2}} \right\}\cos(y_1) + (k_{02}+k_{r2})\left\{ \frac{t}{k_{r2}} - \frac{(t-a)}{k_{02}} \right\} \right.$$

$$\sin(y_1) + (k_{r2} - k_{02}) k_{z2} \left\{ 2\left\{ \frac{1}{k_{02}} - \frac{1}{k_{r2}} \right\} \cos(y_2) + (k_{02} - k_{r2}) \right.$$

$$\left\{ \frac{t}{k_{r2}} + \frac{(t-a)}{k_{02}} \right\} \sin(y_2) - 4 \cos(U_0 a) k_{z2} \left\{ \frac{k_{02}}{k_{r2}} + \frac{k_{r2}}{k_{02}} \right\} \qquad (15)$$

where

$$y_1 = ((t-a) k_{02} - t k_{r2}) \qquad (16)$$
$$y_2 = ((t-a) k_{02} + t k_{r2}) \qquad (17)$$

In both TM and TE cases, since the structure is loss-free, the root values are either real(propagating modes) or imaginary(evanescent modes). It can also be shown that $TM(k_{z1})$ and $TE(k_{z2})$ are either real or imaginary for particular values of k_{z1} and k_{z2}. After careful numerical study and graphical observation of the functional behavior, an algorithm which consists of four major steps was developed to find the roots: 1) The range within which all the roots reside for given parameters(f, a, b, t, ϵ_r, number of modes, number of roots) is calculated. 2) Within the calculated range, the step size is adjusted in terms of function's magnitude and frequency. 3) Upon detection of a sign change (or an abrupt shift between real and imaginary in the function value) between two adjacent steps, an initial guess located halfway between the two step is fed into a Newton-Raphson method to accurately find the root. 4) All the roots found from the Newton-Raphson procedure are then verified by considering the functions' values in the neighborhood of the roots. The range where the roots are to be searched can be limited by the following two equations:

$$k_z^a = \sqrt{\epsilon_r k_0^2 - V_n^2 - U_1^2} \tag{18}$$

$$k_z^b = \sqrt{k_0^2 - V_n^2 - U_1^2} \tag{19}$$

where, $U_1 = k_0 \sin \theta \cos \phi + \dfrac{1 \ 2 \ \pi}{a} \tag{20}$

The equations (18) and (19) are limiting cases where the unit cell is either completely filled with the dielectric material or is completely empty(free-space). Since the propagation constants are complex, the regions delimited by (18) and (19) are on either the real axis, the imaginary axis, or both real and imaginary axis. Thus, the root searching

region is divided into the above three criteria.

By observing the function graphically, it was found out that an increase in frequency results in an increase of magnitude of the function, as well as a wider across which the roots exist. Thus, if constant step size is to be used as in [1], the CPU time increases proportional to the increase in the operating frequency. In order to reduce the CPU time specifically at high frequencies, an adjustable step size in terms of frequency and the magnitude of the function was developed. First the frequency is separated into three categories: a frequency less than 30 MHz(not a realistic value for these types of antennas), a frequency between 30 MHz and 1 GHz, and a frequency above 1 GHz. The step size is then made proportional to frequency. Additionally, the step size is adjusted according to the magnitude of the function. A finer step size is assigned when the magnitude of the function approaches zero indicating the onset of a root. And as the magnitude of the function increases, a bigger step size is used to skip over uninterested region more rapidly(refer to table 1). By using larger step size in the root searching, the less amount of function evaluation are required thus reducing the CPU time.

Once the sign change has been detected between the two adjacent steps or an abrupt shift between real and imaginary in the function value (root does not necessarily exist), the halfway value between the two steps is fed into a Newton-Raphson method as an initial guess to find a more accurate value of the root. If a Newton-Raphson method converges rapidly (in less than 50 iterations) to within the assigned tolerance value (typically 10^{-4}) then there is a root; if it does not converge then there is no true root within those steps. The roots are verified by insuring

that the functions do go through a sign change at these locations. This procedure is continued until the range is searched completely. Lastly, for a particular n. the roots found are ordered. the real ones first then the ones with the lowest imaginary values.

Analysis of Results

The results of various cases using the new method are shown in this section. The comparison of the results obtained using the method in [1] and the new method will be discussed. specifically the accuracy of the roots found and the CPU time required for finding these roots. First, the discussion of the functional behavior is presented.

The magnitude of the functions (determinant of TM and TE) with respect to the propagation constant for some examples of cases are shown in fig.2 to fig.7. All of the figures from 2 to 7 show a very smoothly varying behavior. This behavior was expected since the analytically solved determinants were the combination of well behaved functions, cosine and sine. Although, the functions may vary smoothly, they have very steep slopes, causing very careful considerations in assigning step sizes to prevent from missing any roots. Figure 4 illustrates the steepness of the slope very well; a change in $Im \{k_z\}$ from 4070 to 4180 resulted in a magnitude drop of 1.7×10^7. Thus, because of such sharp variation, very fine steps were required(see table 1) in order to avoid missing any roots especially as the function approaches to zero.

To find the roots accurately, it is important to be able to distinguish the roots that are located close to one another. It was found out that very often roots are located extremely close to one another as seen in figures from 5 to 7 (TE mode). Each dip represented by an arrow and a number indicate the existence of a root and the number of roots found respectively. Note that there are two roots in most of the dips for the TE mode. This implies that the TE rooots are closely spaced. How closely some of the roots may located can be visualized through figure 5. In

figure 5, the fifth and sixth roots are located at 69.771 and 69.775 respectively. Thus, the required step size to locate both roots successfully is at least 2×10^{-3}, otherwise one of the two or even both roots can be missed. Nevertheless, looking at tables 2 and 3, the new method gives much more accurate results as compared to [1]. Roots found using the new method are accurate where as the old method misses a number of roots for both TM and TE modes.

Probably the most important aspect of this research can be seen from the same tables which is the reduction in CPU time. Both tables show a significant reduction in CPU time for finding the roots. It is more significant as the operating frequency increased. The CPU time was reduced much more for the TM case than the TE case. This is due to the fact that the closely located roots for TE mode forced finer step sizes to be used.

Finally, both tables show results that work well for a broad range of parameters. Frequencies tested from 300 MHz to 12 GHz did not show any particular problem. Also the thickness of the substrate tested from 0.01λ to 0.3λ did not cause any error and likewise for the relative permittivity tested up to 12.

Conclusion

The root searching using the new method proved to be much more robust and faster than [1]. Some of the limitations and problems mentioned in [1] were eliminated by using the new method. The new method seemed to be more reliable in accuracy and practical in time. However, the new method was tested extensively for only E-plane, H-plane , and D-plane. Therefore, in order for the new method to be fully reliable and concrete, other scan angles should be tested as well.

TABLE 1. STEP SIZE CHANGE WITH RESPECT TO MAGNITUDE RANGE AND OPERATING FREQUENCY

FREQUENCY	MAGNITUDE	STEP SIZE $\cdot$
f > 1GHz	mag. > f*1E5/3E8	f*0.05/3E8
	f*1E4 < mag. $\leq$ f*1E5/3E8	f*0.01/3E8
	f*1E3 < mag. $\leq$ f*1E4/3E8	f*0.005/3E8
	f*1E2 < mag. $\leq$ f*1E3/3E8	f*0.002/3E8
	mag. $\leq$ f*1E2/3E8	f*0.001/3E8
30MHz < f $\leq$ 1GHz	mag. > f*1E5/3E8	f*0.2/3E8
	f*1E4 < mag. $\leq$ f*1E5/3E8	f*0.1/3E8
	f*1E3 < mag. $\leq$ f*1E4/3E8	f*0.05/3E8
	f*1E2 < mag. $\leq$ f*1E3/3E8	f*0.02/3E8
	mag. $\leq$ f*1E2/3E8	f*0.01/3E8
$\cdot$f $\leq$ 30MHz	ALL	0.0001

* For TE mode step size, f = f * 0.1

TABLE 2. TM ROOT SEARCHING RESULT AND CPU TIME COMPARISON BETWEEN OLD AND NEW METHOD *

f	a = b(m)	t(m)	Er	n	CPU (OLD) sec	CPU (NEW) sec	F, M (OLD) **	F, M (NEW) **
300 MHz	0.5	0.01	2.2	1			F	F
				2	43.2	7.3	F	F
				3			F	F
		0.1	4.9	1			M	F
				2	1:06.4	10.7	M	F
				3			F	F
		0.3	12.0	1			M	F
				2	1:07.9	11.9	F	F
				3			F	F
2GHz	.075	.0015	9.0	1			M	F
				2	7:34.3	35.4	F	F
				3			M	F
		.015	4.9	1			M	F
				2	7:03.7	29.5	M	F
				3			F	F
		.045	12.0	1			M	F
				2	7:21.9	31.7	F	F
				3			F	F
5GHz	.03	.0006	4.9	1			M	F
				2	20:03.	22.1	M	F
				3			M	F
		.006	2.2	1			M	F
				2	13:57.	17.7	M	F
				3			M	F
		.018	9.0	1			M	F
				2	22:37.	23.8	F	F
				3			F	F
12GHz	.0125	.00025	9.0	1			F	F
				2	47:39.	18.2	M	F
				3			M	F
		.0025	4.9	1			F	F
				2	40:33.	15.5	M	F
				3			M	F
		.0075	12.0	1			M	F
				2	45:13.	18.5	M	F
				3			F	F

* data used: theta - 0.1, phi - 89.9 (degrees) , n = 6, #of roots = 13
** F - all roots found are correct   M - all roots found are not correct
(either missing or erroneous)

TABLE 3. TE ROOT SEARCHING RESULT AND CPU TIME COMPARISON BETWEEN OLD AND NEW METHOD *

f	a = b(m)	t(m)	Er	n	CPU (OLD) sec	CPU (NEW) sec	F, M (OLD) **	F, M (NEW) **
300 MHz	0.5	0.01	2.2	4			F	F
				5	41.4	39.4	F	F
				6			F	F
		0.1	4.9	4			F	F
				5	43.6	39.9	F	F
				6			F	F
		0.3	9.0	4			F	F
				5	59.1	40.2	M	F
				6			F	F
2GHz	.075	.0015	4.9	4			F	F
				5	6:23.8	3:13.5	F	F
				6			F	F
		.015	9.0	4			F	F
				5	5:13.3	3:13.7	F	F
				6			F	F
		.045	12.0	4			M	F
				5	7:15.3	3:16.2	M	F
				6			M	F
5GHz	.03	.0006	2.2	4			F	F
				5	11:43.	2:19.5	F	F
				6			F	F
		.006	4.9	4			F	F
				5	11:46.	2:19.3	F	F
				6			F	F
		.018	9.0	4			F	F
				5	1≦:35.	2:20.2	M	F
				6			F	F
12GHz	.0125	.00025	4.9	4			F	F
				5	40:41.	1:14.2	F	F
				6			F	F
		.0025	9.0	4			F	F
				5	30:43.	1:36.4	F	F
				6			F	F
		.0075	12.0	4			M	F
				5	42:29.	1:38.2	M	F
				6			M	F

* data used: theta - 0.1, phi - 89.9 (degrees) , n = 6, #of roots = 13
** F - all roots found are correct   M - all roots found are not correct
(either missing or erroneous)
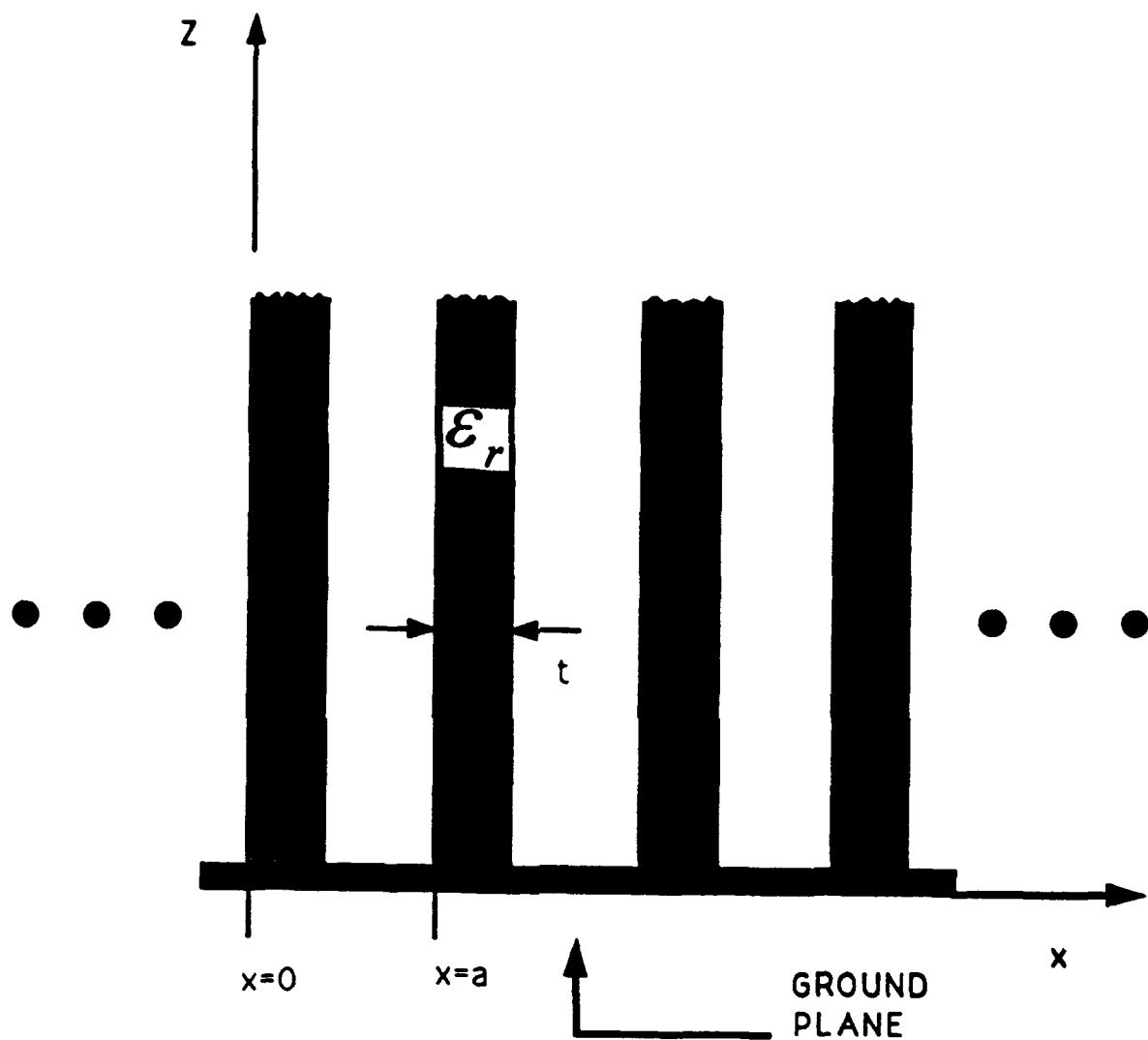
Figure 1. The geometry of a periodic array of
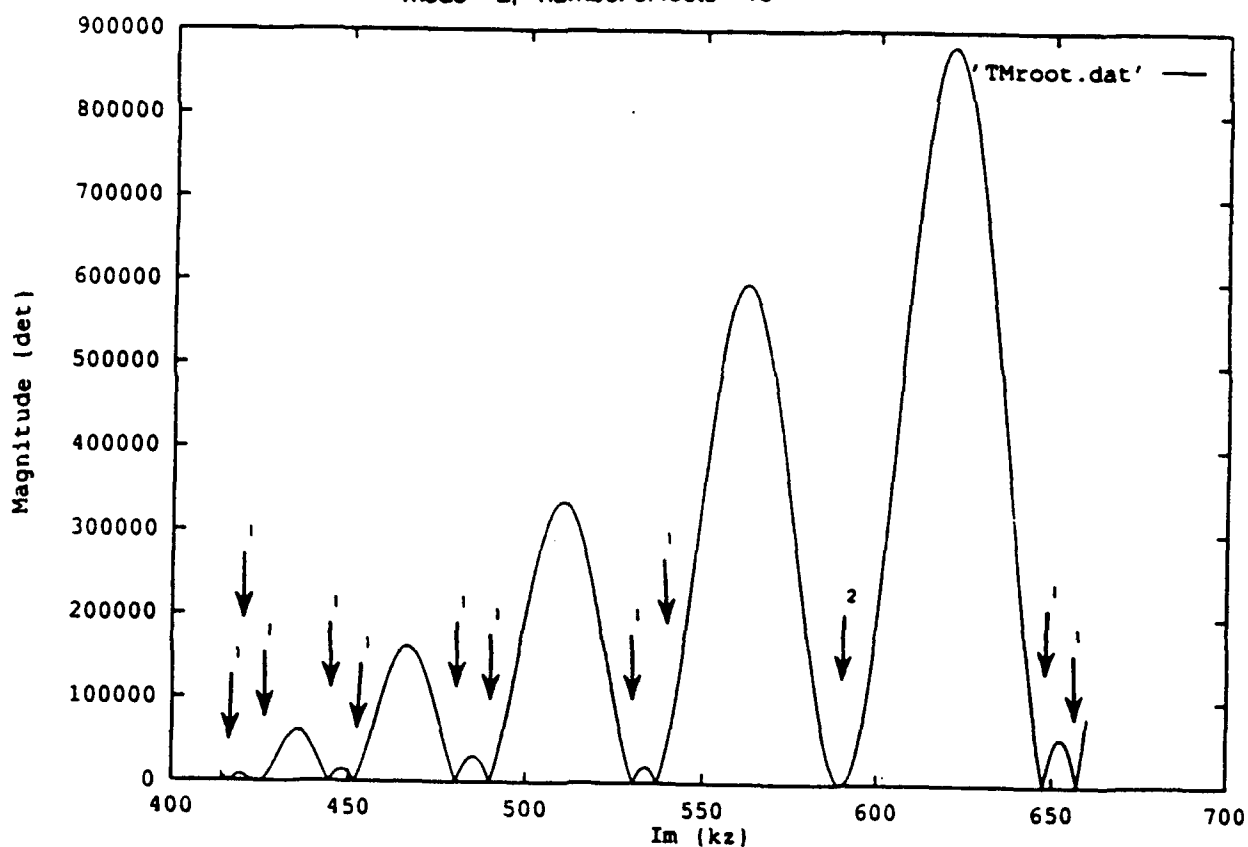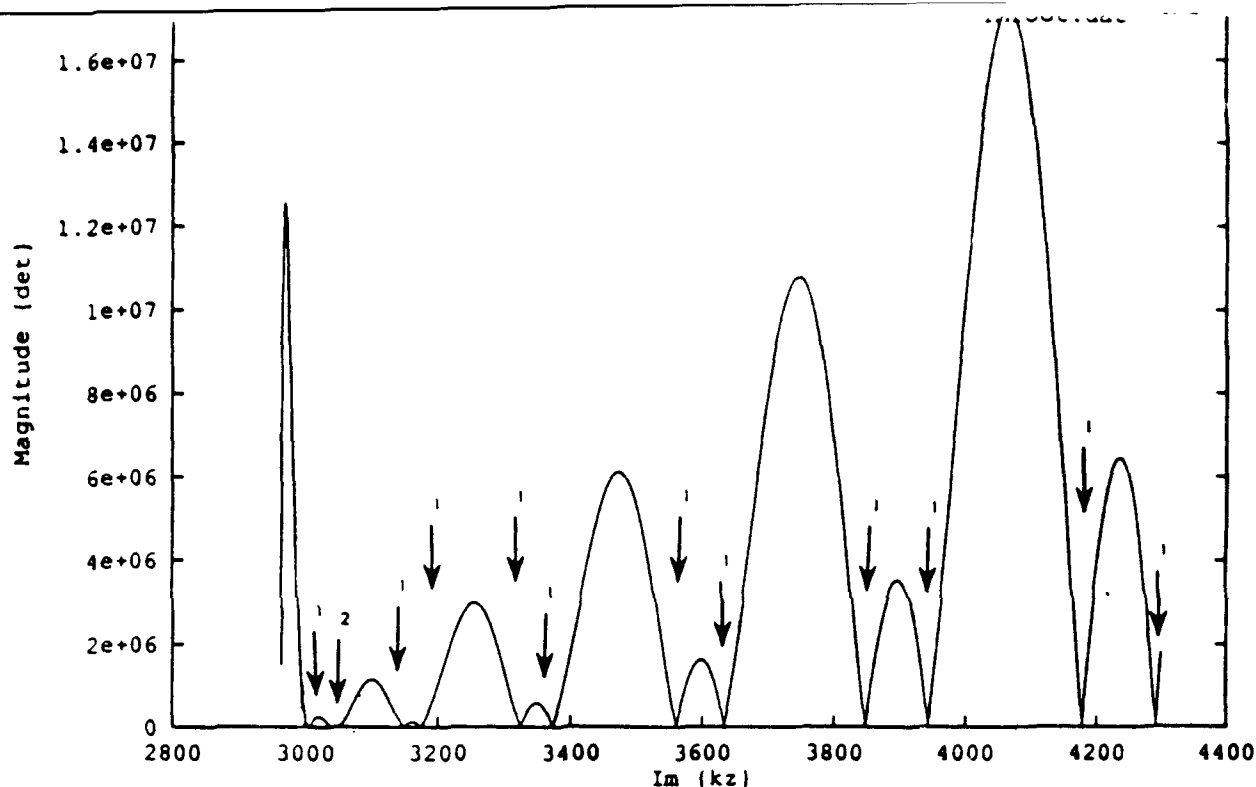dielectric walls backed by a perfect
conductor.
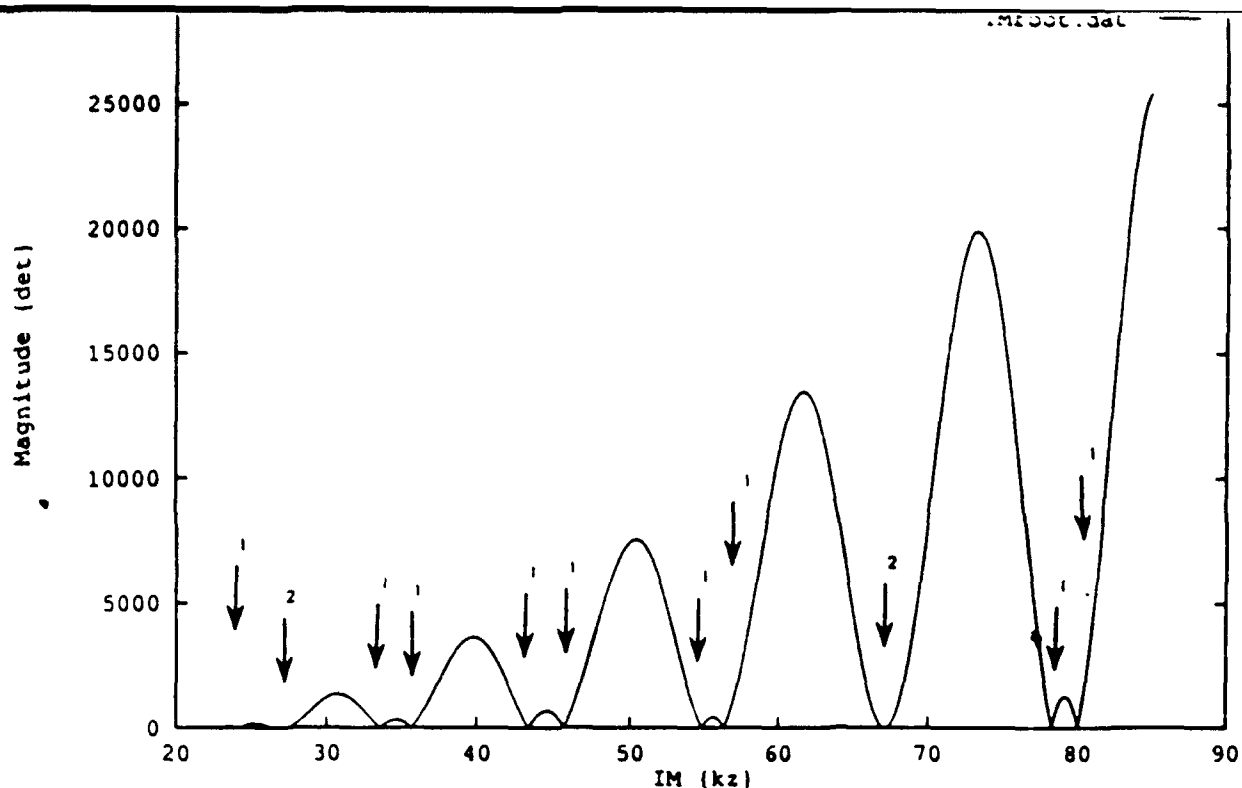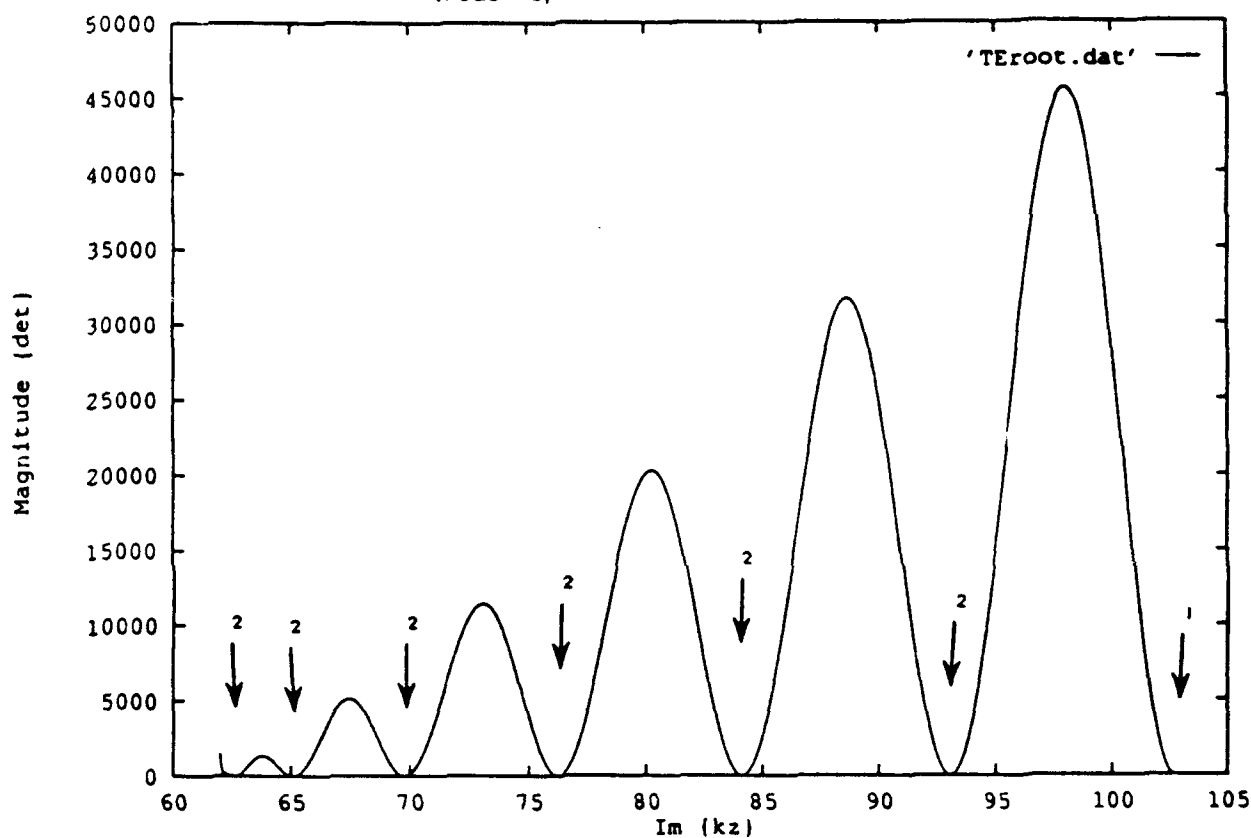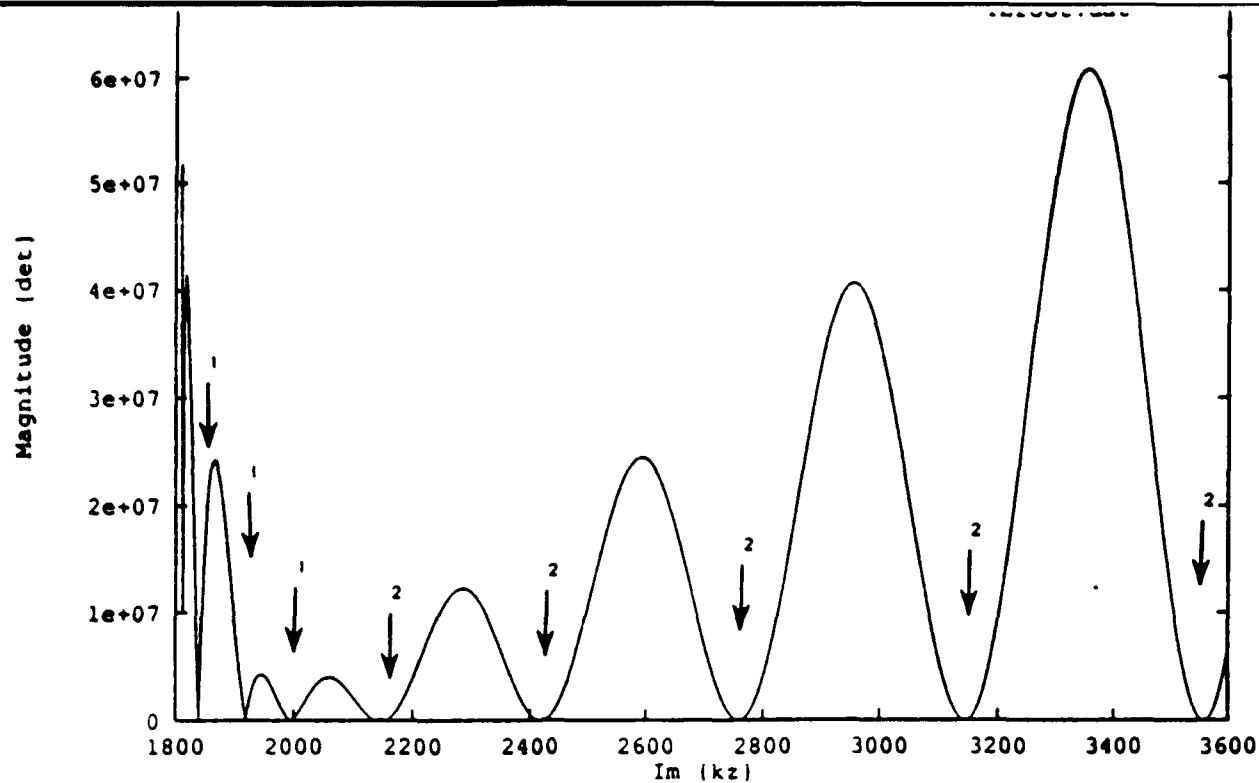
Fig. 2   Magnitude of the determinant of TM mode versus
imaginary part of the propagation constant.
Frequency = 300MHz, t = 0.3m, a = b = 0.5m
Theta = 0.1, phi = 89.9, relative permittivity = 2.2
mode = 2, number of roots = 13

Fig. 3   Magnitude of the determinant of TM mode versus
imaginary part of the propagation constant.
Frequency = 2 GHz, t = 0.045m, a = b = 0.075m
Theta = 0.1, phi = 89.9, relative permittivity = 2.2
mode = 5, number of roots = 13

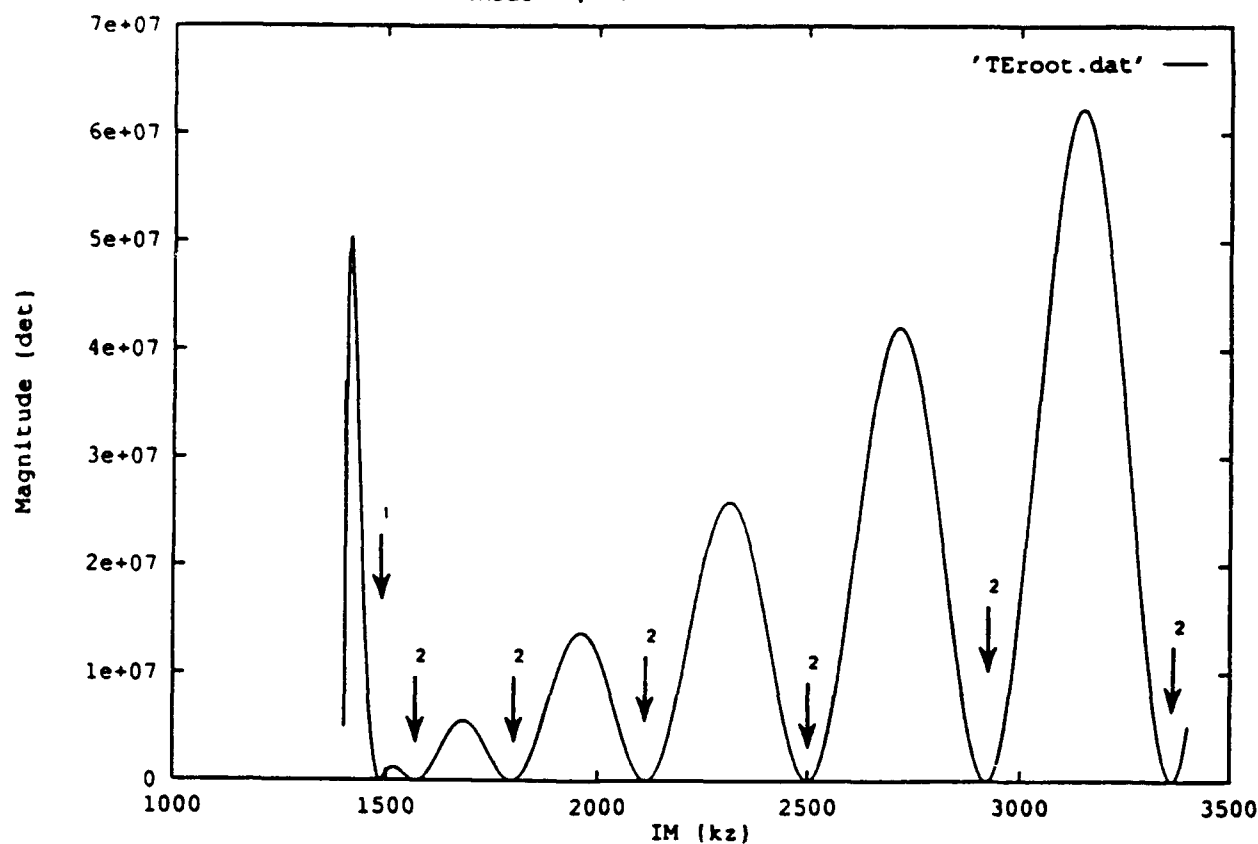Fig. 4 Magnitude of the determinant of TM mode versus
imaginary part of the propagation constant.
Frequency = 12 GHz, t = 0.00025m, a = b = 0.0125m
Theta = 0.1, phi = 89.9, relative permittivity = 2.2
mode = 6, number of roots = 13

Fig. 5 Magnitude of the determinant of TE mode versus
imaginary part of the propagation constant.
Frequency = 300MHz, t = 0.3m, a = b = 0.5m
Theta = 89.9, phi = 0.1, relative permittivity = 2.2
mode = 5, number of roots = 13

5-19

Fig. 6   Magnitude of the determinant of TE mode versus
imaginary part of the propagation constant.
Frequency = 12 GHz,   t = 0.0075m,   a = b = 0.0125m
Theta = 0.1,   phi = 89.9,   relative permittivity = 12.0
mode = 4,   number of roots = 13

Fig. 7   Magnitude of the determinant of TE mode versus
imaginary part of the propagation constant.
Frequency = 12 GHz,   t = 0.00025m,   a = b = 0.0125m
Theta = 0.1,   phi = 89.9,   relative permittivity = 4.9
mode = 3,   number of roots = 13

[1] J-P. R. Bayard. M.E. Cooley. and D.H. Schaubert, "Analysis of Infinite Arrays of Printed Dipoles on Dielectric Sheet Perpendicular to a Ground Plane." IEEE Tran. Antennas Propagat., vol. 39, no. 12. pp 1722-1732, Dec. 1991.

$Cr:Al_2O_3$ THIN FILMS AS

OPTICAL WAVEGUIDES AND AMPLIFIERS

Bethanie J. H. Stadler

Ph.D. Candidate

Department of Materials Science and Engineering

Massachusetts Institute of Technology

77 Massachusetts Avenue

Cambridge. MA 02139

$Cr:Al_2O_3$ THIN FILMS AS OPTICAL WAVEGUIDES AND AMPLIFIERS

Bethanie J. H. Stadler
Ph.D. Candidate
Department of Materials Science and Engineering
Massachusetts Institute of Technology

ABSTRACT

The purpose of this project was to evaluate the potential for developing thin film waveguide materials from nontraditional glasses. In the initial phase, completed this summer, Al_2O_3 was investigated. The dependence of the optical properties, refractive index and loss, were examined as a function of the processing conditions. The nonequilibrium nature of reactive rf magnetron sputtering was found to allow deposition of amorphous Al_2O_3 films within a wide processing window. Both x-ray and electron diffraction verified that the films were amorphous in the as-deposited state for substrate temperatures up to 500°C and for a wide range of oxygen flow rates. These films had excellent optical and mechanical properties, with an average refractive index of 1.65 at 500nm, optical losses of 1-4 dB/cm throughout the visible, transmission windows that extend well into the infrared, and they are hard, chemically-stable, wear-resistant, and adherent. Both refractive index and loss were found to be dependent on processing conditions. The refractive index was found to increase with substrate temperature, but to be independent of oxygen flow rate. The loss, however, was sensitive to oxygen flow rate.

As a final check on the amorphous structure, Cr-doped films were prepared by co-depostion. No fluorescence was detected in the amorphous films. Annealing the films at 1000°C yielded crystalline γ-Al_2O_3 films that demonstrated weak fluorescence. Annealing at 1200°C produced α-Al_2O_3 films that exhibited the strong characteristic fluorescence of ruby. The results of this study demonstrated the feasibility of producing thin film waveguides with good optical properties from oxides which are not traditional glass formers, and suggests that materials which serve as good rare-earth hosts (Y_2O_3, YAG) warrant further investigation.

$Cr:Al_2O_3$ THIN FILMS AS

OPTICAL WAVEGUIDES AND AMPLIFIERS

Bethanie J. H. Stadler

INTRODUCTION

As optical communications continue to be developed, there is an increasing need to switch and process optical signals directly, that is without converting them into electrical signals. Current optical devices include fiber and planar devices such as Er-fiber amplifiers and $LiNbO_3$ electro-optical switches. For the most part, fiber devices are limited to traditional amorphous systems, although some work is done on single crystal fibers despite difficulties in fabrication (Klocek,1986; Artjushenko, 1990). Planar devices are also limited to single crystals and glasses because diffusion and ion-exchange are used to define waveguide structures.

In this study, novel amorphous and polycrystalline ceramic planar devices were investigated. This project involved fundamental research on several levels, the first of which was the investigation of a new class of materials. Through sputtering, materials that are typically crystalline can be produced in an amorphous form. Also, some non-equilibrium polymorphs , such as γ-Al_2O_3, may easily be fabricated. In this way, new amorphous and polycrystalline oxides, like Y_2O_3 which is transparent well into the infrared, can be studied that have potential to surpass traditional oxides in optical performance. Secondly, this project yielded a fundamental understanding of the effects of processing on the microstructure and resulting optical properties of thin films. Although the initial application of interest was planar waveguides, this understanding will be valuable in the development of a variety of thin film applications, such as amplifiers and switches.

Sputtering will allow the deposition of films onto a variety of substrates, such as an LED (light emitting diode) for direct laser pumping. Or, waveguides can eventually be stacked by depositing alternating waveguide/SiO_2 layers. Interference between waveguides can be minimized

since a higher index difference between waveguide and cladding $(\Delta n=0.12)$ can be achieved than that which is available through current techniques $(\Delta n=0.02-0.05)$.

To date, most of the optical research on oxide thin films has focused on coatings, where the direction of light propagation is perpendicular to the plane of the film. (Pawlewicz et. al. 1980, Rainer et.al. 1985) Relatively high losses have been tolerated because the path length is short, and the relationship between optical losses and microstructure has not been adequately explored. This research will help to fill a substantial gap in the current understanding of optical films through a systematic study of this relationship.

The Al_2O_3 system was chosen because it has excellent intrinsic optical properties: losses comparable to SiO_2 $(\sim 0.1$ dB/km) and a high index of refraction (~ 1.765). Because it is important in current optical materials (e.g., Ti:sapphire and ruby lasers), there is a wide data base available on Al_2O_3 and $Cr:Al_2O_3$. Al_2O_3 has excellent mechanical properties; it is hard, chemically stable, and wear-resistant. Cr was chosen as a dopant for the Al_2O_3 films because it is similar in size and behavior to Al. Since Cr is a transition metal, however, it is not a good active ion for an amorphous system. Therefore, Cr was mostly expected to yield information about the Al_2O_3 host, such as crystallinity and coordination data. This work will be useful in future films that will be doped with more promising rare-earth active ions, such as $Er:Y_2O_3$.

The devices were made by rf (radio frequency) magnetron sputtering, because it is an excellent technique for the fabrication of single-or multi-component insulating thin films that have high densities, uniform thicknesses, and good adhesion to the substrate. Microstructures ranged from amorphous to polycrystalline, and the optical properties of interest were index of refraction, dispersion, attenuation and fluorescence.

FABRICATION AND CHARACTERIZATION

The films were made by reactive rf magnetron sputtering with 2-inch, metallic targets (99.999% pure Al, and 99.995% pure Cr). The substrates (1" x 1" x 1mm plates of fused SiO_2 with a scratch/dig finish of 80/50) were located 12 cm from the targets and were rotated to maintain homogeneous films. The deposition conditions were: a base pressure of $2x10^{-7}$ Torr, deposition pressure of 4.0 mTorr, rf power of 300W ($96W/cm^2$). The total gas flow rate was fixed at 40sccm, while the ratio of O_2/Ar flow rates were varied from 1/99 to 10/90, the Cr content was varied between 0 and 1%, and the films were annealed from 100-1200°C. Films were ~1μm thick.

The film microstructure was characterized by x-ray diffraction, both θ-2θ using a Rigaku 300 and glancing angle using a Rigaku 200. Transmission Electron Diffraction (TED) in an Akashi 002B Transmission Electron Microscope (TEM) was also used to look at microstructure. Chromium content was measured using a Jeol 840 Scanning Electron Microscope (SEM) together with a Tracor Northern Energy Dispersive Spectroscopy (EDS) package. However, since the lowest amounts of Cr that could be detected was ~1%, optical absorption was used to estimate the Cr content in films with < 1% Cr.

Index of refraction was measured by two techniques: the envelope method (Manifacier et.al., 1976) and m-lines (Ulrich, 1970). In the envelope method, light was passed normal to the film, and the transmission through the film was detected as a function of wavelength in a Perkin-Elmer Lambda 9 spectrometer. Since the optical thickness of the films was of the same order of magnitude as the wavelength of the light, several maxima and minima occurred due to the interference of the reflected light from both surfaces of the film. The index of the film was then calculated from the dependence of the extrema on wavelength. This method also gave a measure of dispersion since the index was calculated at several wavelengths. The m-lines method used a totally reflecting prism (LaSF-N9 flint glass) that was separated from the film by a narrow gap of reduced refractive index (air). Distributed coupling was accomplished through the evanescent fields to the modes of the thin-film waveguide (Fig. 1).
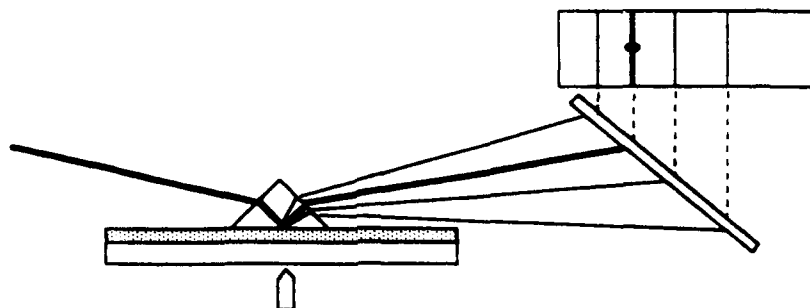
Fig. 1- The m-lines technique

Several configurations for measuring attenuation were tested, including detection of the streak of scattered light at the surface of the film, end-coupling into and out of the film, and prism-coupling into and out of the film (both with right-angle and symmetric prisms). Prism-coupling light into and end-coupling the light out of the film proved to be the most suitable geometry. The set-up is similar to that for m-lines, only the end of the sample was cleaved and a 20X objective lens was used to gather the output and focus it onto a Si photodiode. The prism was translated away from the edge so that the light would be guided through various lengths of the film. This allowed the cancellation of coupling losses and hence a more accurate measurement of actual waveguide losses. HeNe and Ar lasers were used to detect the losses at several visible wavelengths.

Fluorescence in the $Cr:Al_2O_3$ films was at first detected by illuminating a film with the 514.5 nm line of an Ar laser and viewing it through Ar safety glasses to see if any red could be observed. The films that appeared to fluoresce were later setup such that the 514.5 nm light was passed through a fiber that was butted up against the film. Any fluorescence was gathered back through the fiber and focused onto the slit of a monochrometer with a LN2-cooled Ge detector at the output.

RESULTS AND DISCUSSION

From previous work in our laboratory, it was known that yield x-ray-amorphous-Al_2O_3 can be fabricated within a wide processing window that extends from room temperature to 500°C substrate temperature, and from 4 to 10% O_2 flow rate. Below 4% O_2, films become defective and even metallic at very low O_2. Above 10% O_2, the Al target oxidizes and the plasma is extinguished. These amorphous Al_2O_3 films were shown to be quite stable as they survived anneals up to 800°C without any crystallization. However, films annealed at 1000°C crystallized into γ-Al_2O_3, and a 1200°C anneal yielded the equilibrium phase, α-Al_2O_3. The x-ray diffraction patterns of these films are shown in Fig. 2. To further confirm that the films are truly amorphous, they were analyzed by TEM and TED. The films were found to be amorphous as-deposited. but they were crystallized by the electron beam of the TEM. Micrographs and TED patterns are shown for as-deposited films and electron-beam-annealed films in Fig. 3.

The optical properties of the Al_2O_3 films also indicated that they were amorphous. The index of the films averaged 1.65 in the visible and was observed to increase slightly with substrate temperature, but remain constant with % O_2. This index is roughly 94% of the index of crystalline Al_2O_3, as shown in Fig. 4. Although a reduced index sometimes indicates porosity, this index reduction is the same as that between fused and crystalline SiO_2, and is therefore attributed to the amorphous nature of the films.

The Al_2O_3 films had wide transmission windows that were measured from 200-3200nm (Fig. 5), but losses were found to increase with increased oxygen content of the sputtering gas (Fig. 6). From 465.8 to 632.8 nm, the films deposited with 5% O_2 had relatively constant losses of 1-4 dB/cm. The films deposited with 10% O_2 had losses of 15-20 dB/cm at 632.8nm, 16-25 dB/cm at 514.5nm and > 50 dB/cm at shorter wavelengths.

Films with 0-1% Cr were fabricated by reactive co-sputtering of Cr and Al targets. Cr content was analyzed by EDS, but the lower limit for this technique was 1% (Fig. 7). Optical absorption was used to determine the Cr content of the other films. The addition of small amounts

of Cr did not change the index of refraction of the Al_2O_3 films, but at the highest concentrations, the index was raised from 1.65 to 1.67 at 500nm (Fig. 8). The waveguide losses that of the films at 632.8nm are summarized in Fig. 9. Again, there appears to be losses with stoichiometric roots, and also losses due to increased Cr content, the latter of which can be seen again in Fig. 10.

None of the as-deposited films tested for fluorescence showed any response. The films did exhibit some absorption in the UV which has been attributed to Cr^{6+} in the literature. Therefore, the films were annealed at various temperatures to try to change the valence state of the Cr and to see how it affected the fluorescence. The Cr-O phase diagram indicates that the Cr should have been in the Cr^{3+} state at any temperature above 600°C (Fig. 11). However, no fluorescence was observed in any of the films until the annealing temperatures were such that the films crystallized. Films with the defect spinel structure (γ-Al_2O_3) were observed to fluoresce red, but the fluorescence was such that it was not substantially above the noise level of the Ge detector (Fig.12). Films with the corundum structure (α-Al_2O_3) exhibited a sharp fluorescence peak at 694 nm with a shoulder at 693 nm which is the identical to the fluorescence of ruby.

The fluorescence in ruby is due to the $^2E \to {}^4A_{2g}$ transition of Cr^{3+} in an octahedral site. The fact that this fluorescence is present even in small amounts in the γ-Al_2O_3 films agrees with two expectations: the Cr is in the Cr^{3+} state since the films were annealed above 600°C (actually, to 1000°C), and that octahedral sites are present in this defect spinel structure. The fact that the fluorescence is so weak could indicate that some of the Cr is in the tetrahedral sites of the γ-Al_2O_3 structure, which would be reasonable since Cr_2O_3 can exist in an isostructural γ-Cr_2O_3 phase (Watari et.al. 1981, Mitchell et.al., 1983). The lack of fluorescence in the amorphous films is yet another indication that they are truly amorphous, and not just extremely fine grained. Some reports indicated that amorphous Al_2O_3 is comprised of mixed octahedral and tetrahedral sites (Popova, 1979. El-Mashri et.al., 1982). It is possible that this is true for the present films, with each octahedral site being distorted such that the electronic structures are not similar enough for net fluorescence at a particular wavelength. Higher absorption losses would be an indication of Cr

occupying tetrahedral sites, since the lack of a center of symmetry allows the electron transitions ($^4T_1 \to {}^4T_2$, or $^4T_1$, or $^4A_2$) to be stronger than the ($^4A_{2g} \to {}^4T_{2g}$, or $^4T_{1g}$) transitions in the octahedral site which are not allowed. However, the films are too thin to see any absorption, so future measurements will be required. It should also be mentioned that there is a dispute as to whether Cr^{3+} would occupy an octahedral site at all because the tetrahedral ligand field splitting only provides two lower energy sites for three Cr^{3+} valence electrons. So, Cr^{3+} may be much more likely to occupy a neighboring octahedral site or to cause Jahn-Teller effects that would interfere with the theory that is based on regular splitting.

CONCLUSION

In this project, it was determined that ther is a great potential for developing thin film waveguide materials from nontraditional glasses. In particular, amorphous Al_2O_3 was found to have excellent optical and mechanical properties that will be necessary for developing successful processing of and applications for waveguiding devices. Sputtered amorphous Al_2O_3 films have an average refractive index of 1.65 and optical losses of 1-4 dB/cm throughout the visible, along with transmission windows that extend well into the infrared. They are also hard, chemically-stable, wear-resistant, and adherent. Also, the optical properties of amorphous Al_2O_3 can be taylored by varying the processing conditions. The refractive index was found to increase with substrate temperature, but to be independent of oxygen flow rate. The loss, however, increased as a function of oxygen flow rate. The results of this study demonstrate the feasibility of producing thin film waveguides with good optical properties from oxides which are not traditional glass formers, and suggests that materials which serve as good rare-earth hosts (Y_2O_3, YAG) warrant further investigation.

REFERENCES

Klocek, Paul, *MRS Bulletin*, May/June , p. 41 (1986).

Artjushenko, V.G, *SPIE vol. 1228, Infrared Optics II*, p. 12 (1990).

Pawlewicz,W.T., K.K.Hays and P.M.Martin,*Thin Solid Films*, **73**, 169-75 (1980).

Rainer, F, W.H. Lowdermilk, D. Milam, C.K. Carniglia, T.T. Hart and T.L. Lichtenstein, *Applied Optics*, **24** no. 4, 496-500 (1985).

Manifacier, J.C, J. Gasiot and J.P.Fillard, *Journal of Physics E: Scientific Instruments*, **9**, (1976).

Ulrich, R. and R. Torge, *Applied Optics*, **12** no.12, 2901-2908 (1973).

Popova, I.A, *Inorganic Materials*, **14**, 1503-1505 (1978).

El-Mashri, S.M. A.J. Forty in <u>Electron Microscopy and Analysis 1981</u>, The Institute of Physics, Bristol (1981).

Watari and Cowley, *Surface Science*, 105, p.240 (1981).

Mitchell et.al, *Journal of Vacuum Science Technology A*, April/Jun, p.1006 (1983).
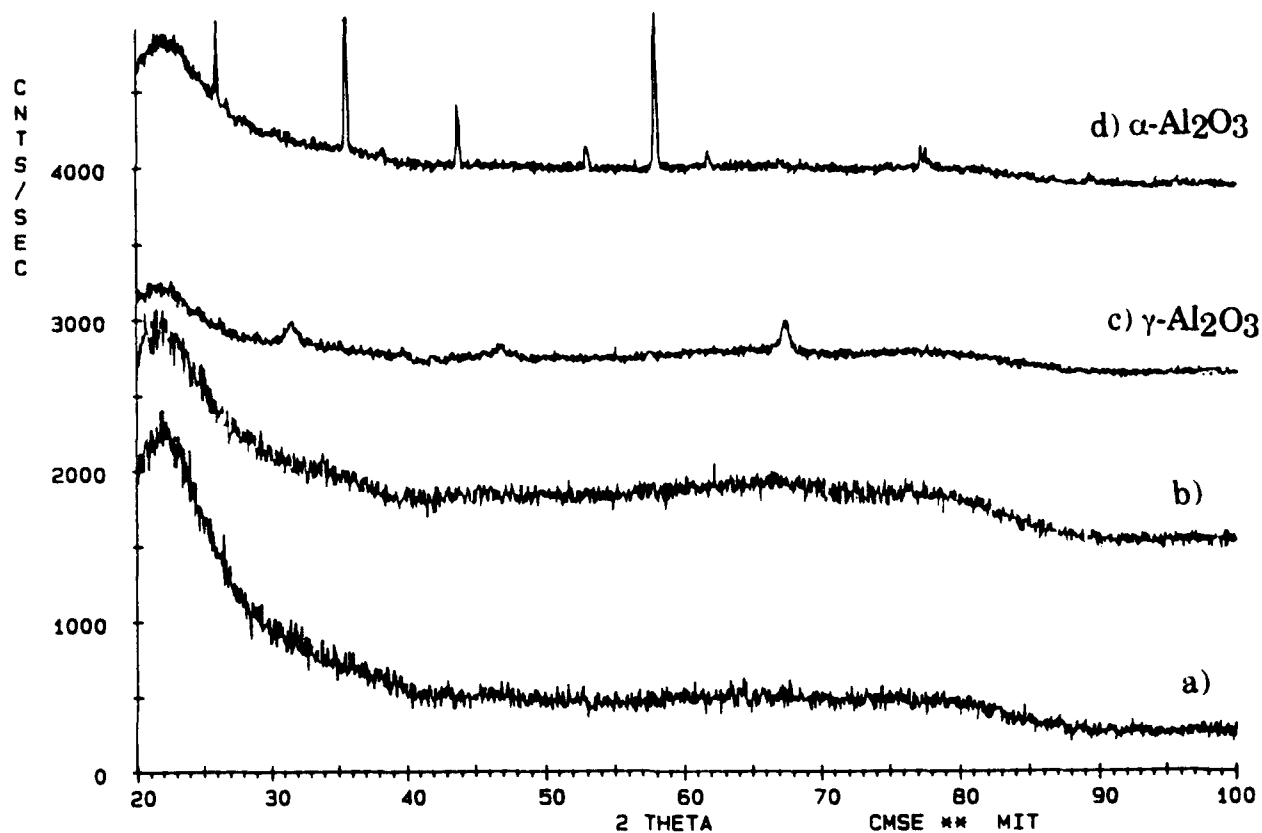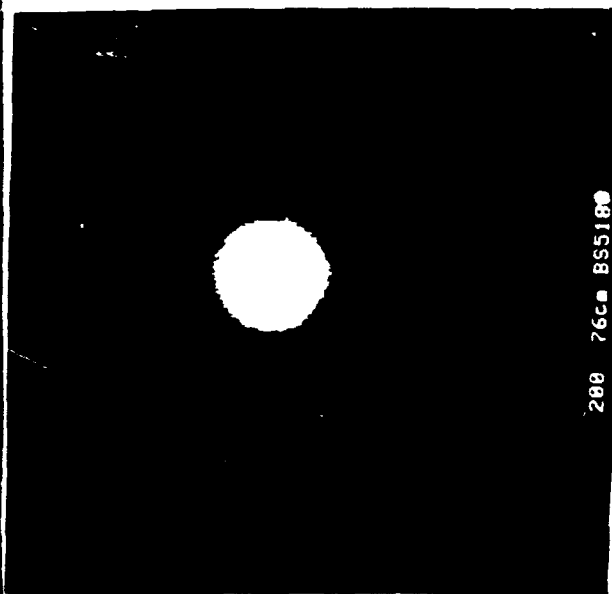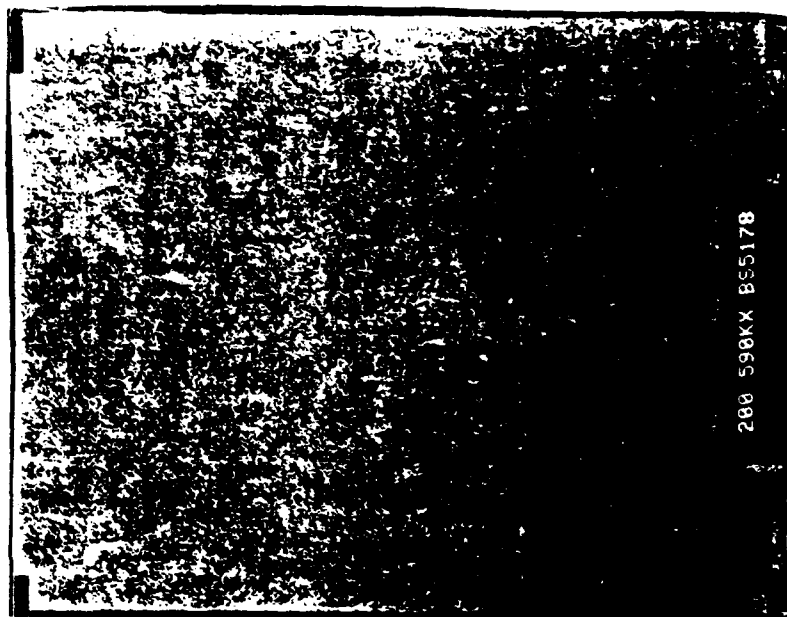
Fig. 2- X-ray diffraction patterns of
a) an as-deposited film   b) a film annealed at 800°C
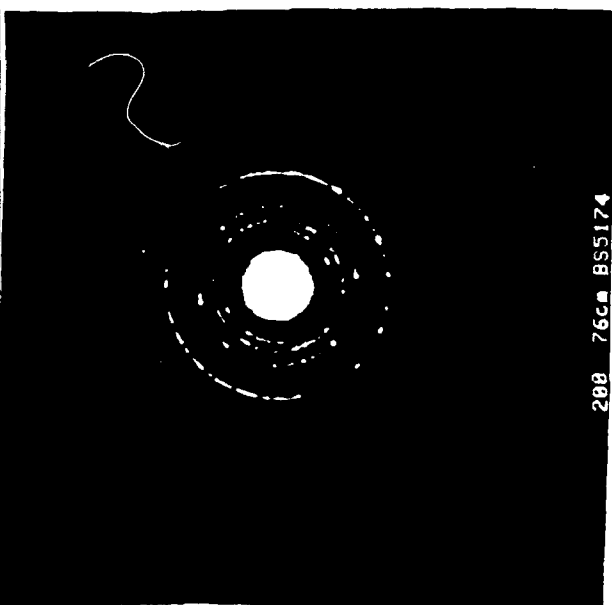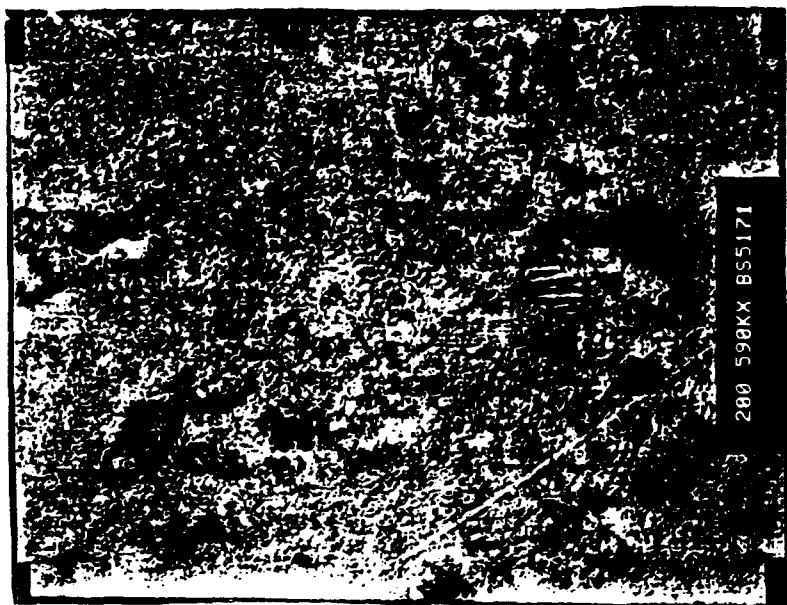c) a film annealed at 1000°C   d) a film annealed at 1200°C

6-10

a)

←——→
20 Å

b)

Fig.   Transmission electron micrographs and diffraction patterns of
as-deposited film   film annealed by the electron beam only.

6-11
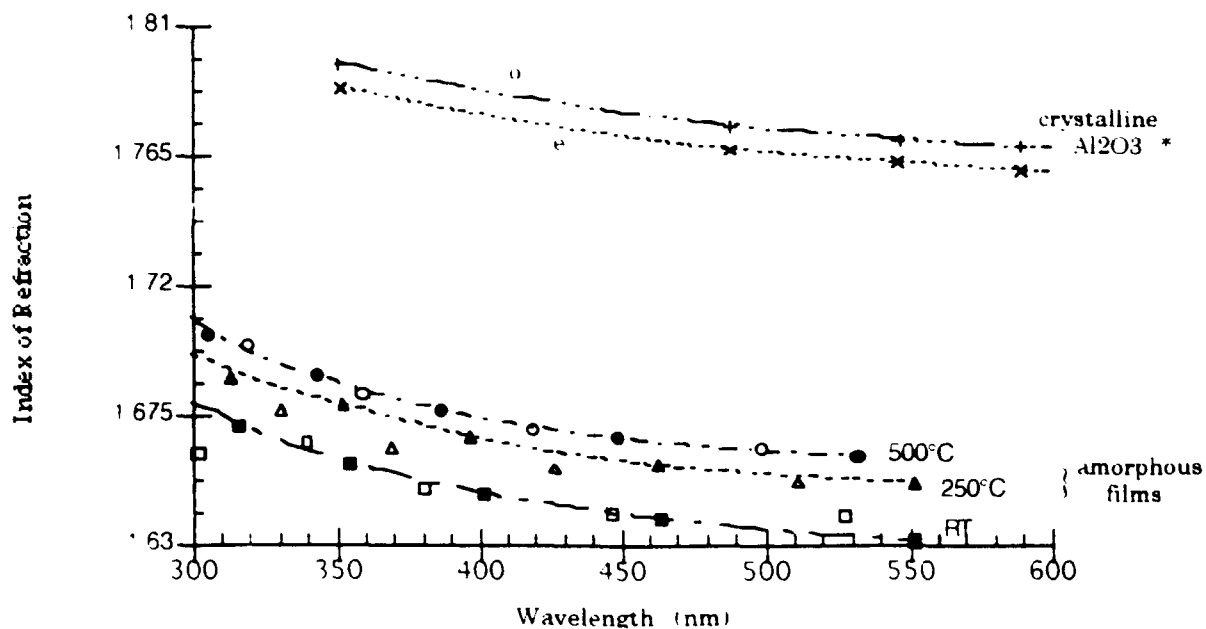
Fig 4- Amorphous Al₂O₃ has a lower index because its structure is more open than the crystalline phase.
Films deposited at similar substrate temperatures have similar shaped points
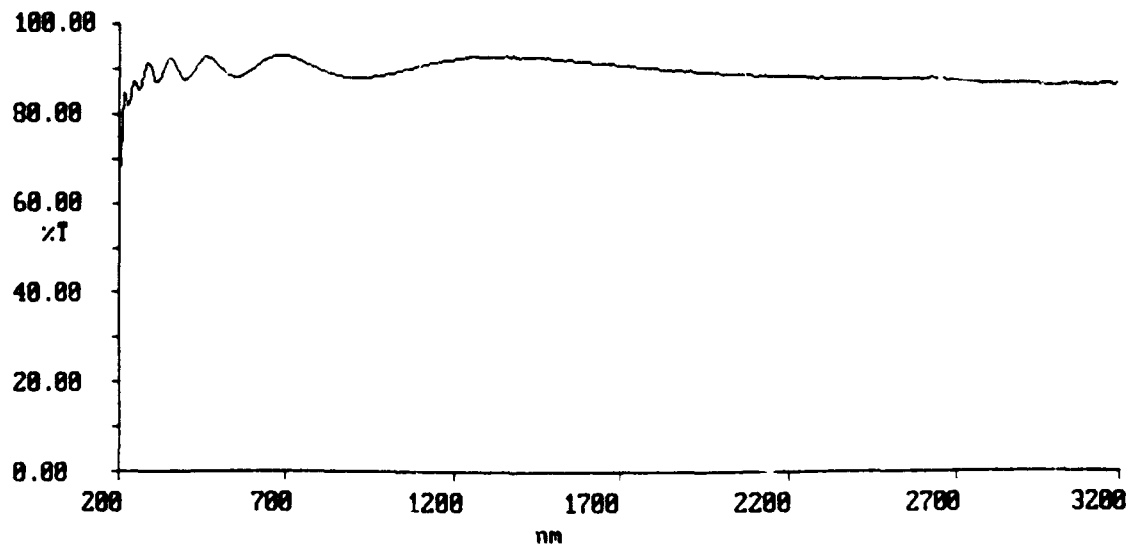Filled and hollow points belong to films deposited at 5% and 10% O₂, resp (*CRC Laser Hndbk)

Fig 5- The transmission window of amorphous Al₂O₃ films extends well into the IR.
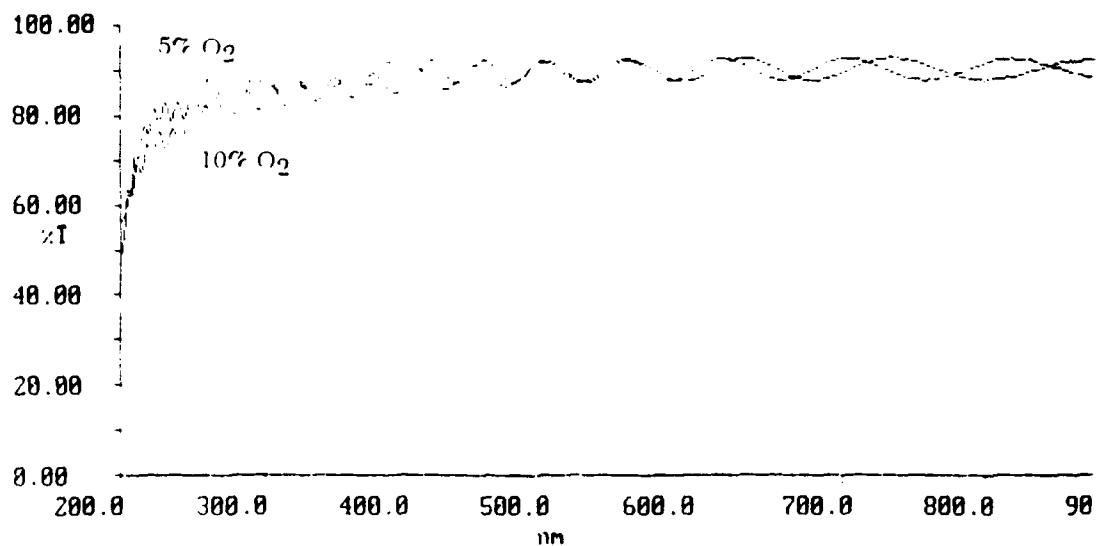Note the maxima and minima in the UV and visible that are due to interference fringes

6-12

Fig 6- Losses in amorphous Al_2O_3 increase with oxygen content of the sputtering gas.
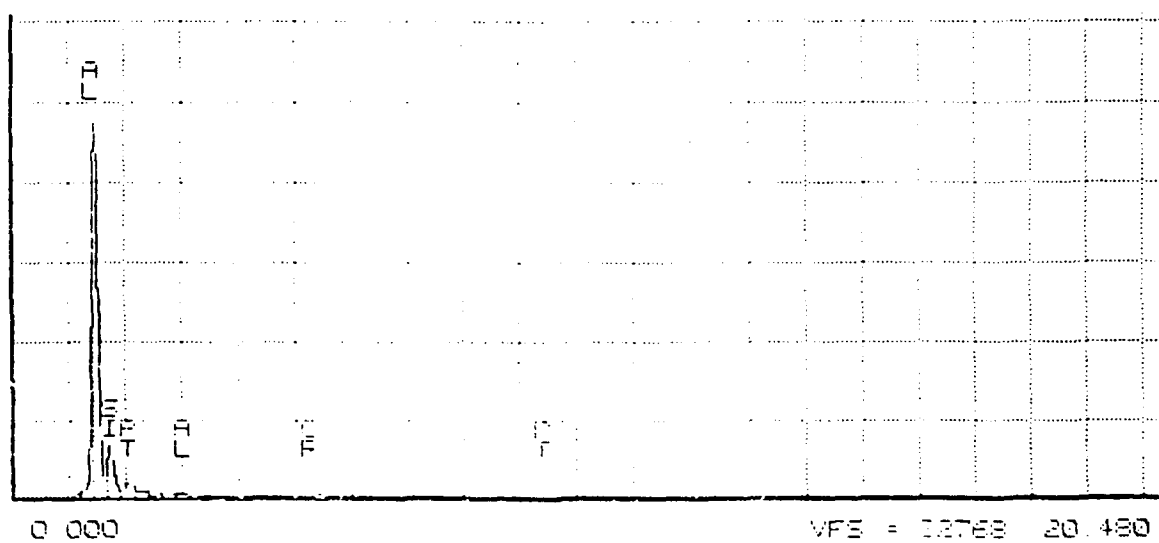

Fig. 7- 1% Cr was the lower limit of detection by EDS $I/I_0 = exp(-\epsilon cx)$ was used to determine %Cr< 1%   (I=intensity, ϵ=extinction coefficient, and c=concentration)

6-13
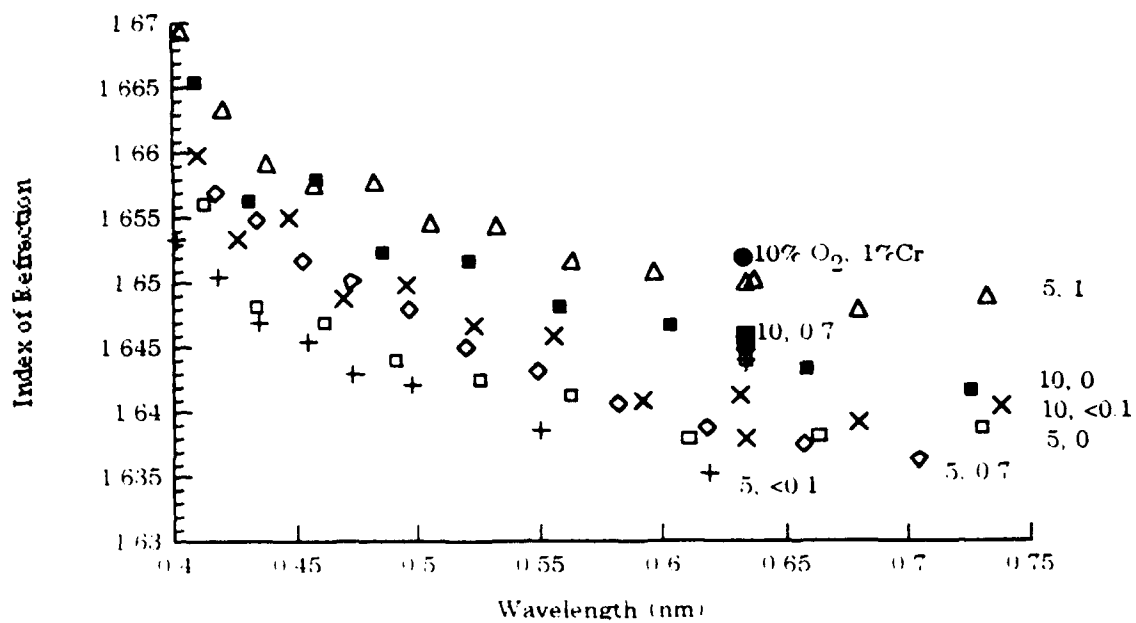
Fig 8- The index of Cr Al$_2$O$_3$ films increases with Cr ≥ 1%
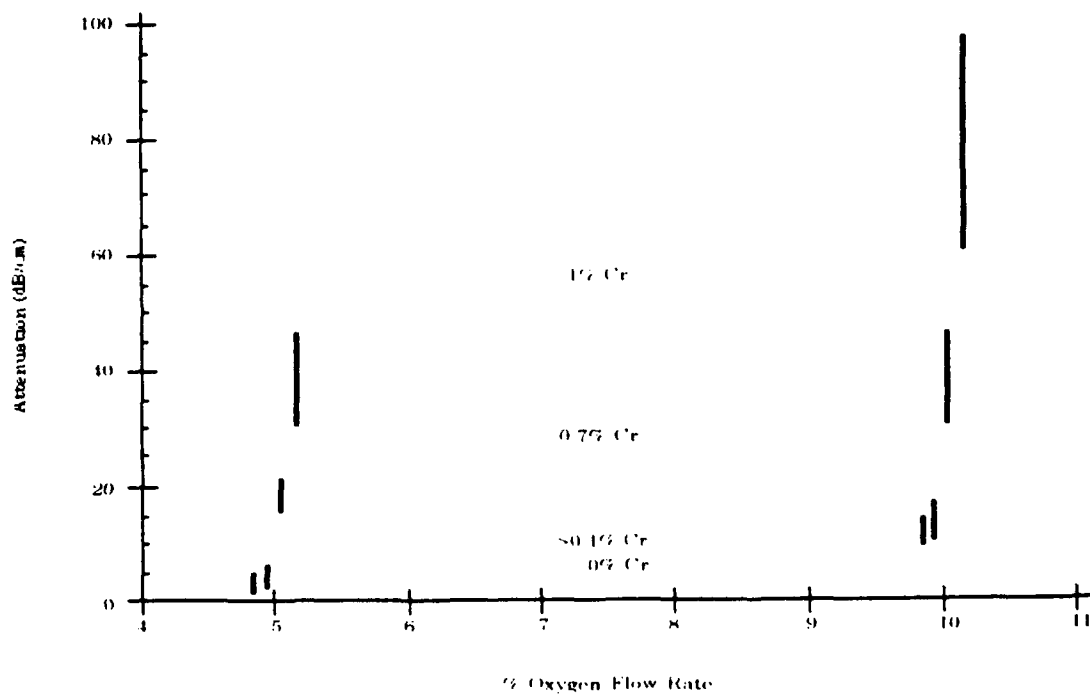

Fig 9- Waveguide losses in Cr Al$_2$O$_3$ films increase with increased Cr content.
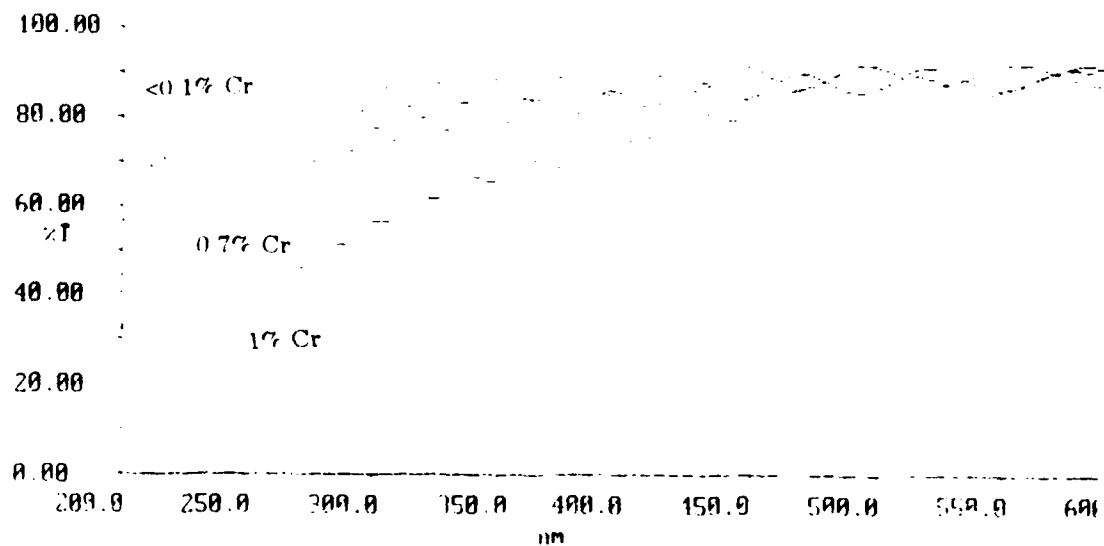
Fig. 10- Absorption increases with increased Cr content.
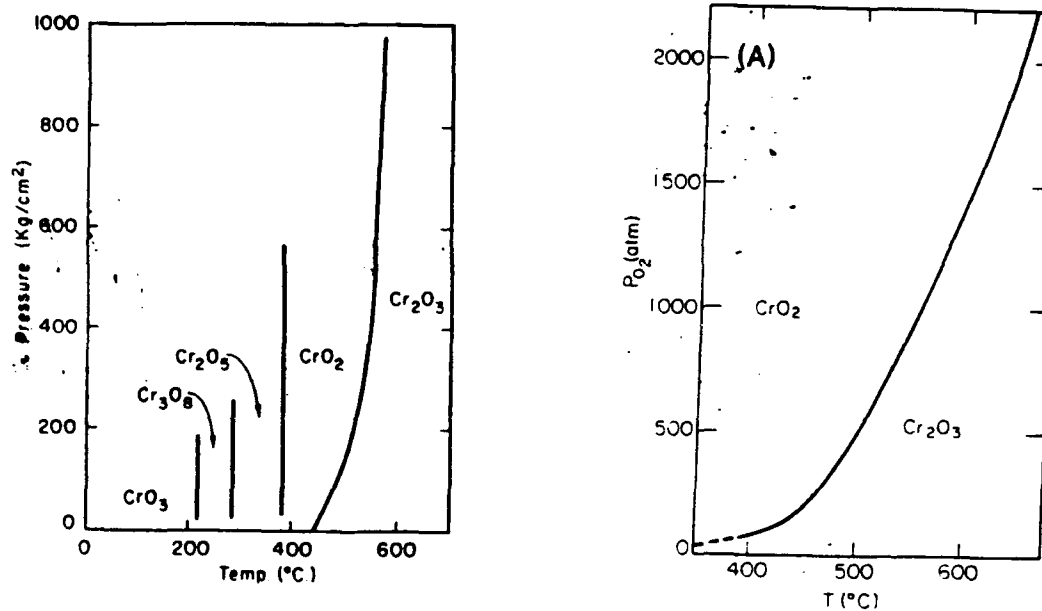
Fig. 11- The Cr-O phase diagram (Phase Diagrams for Ceramists)

6-15
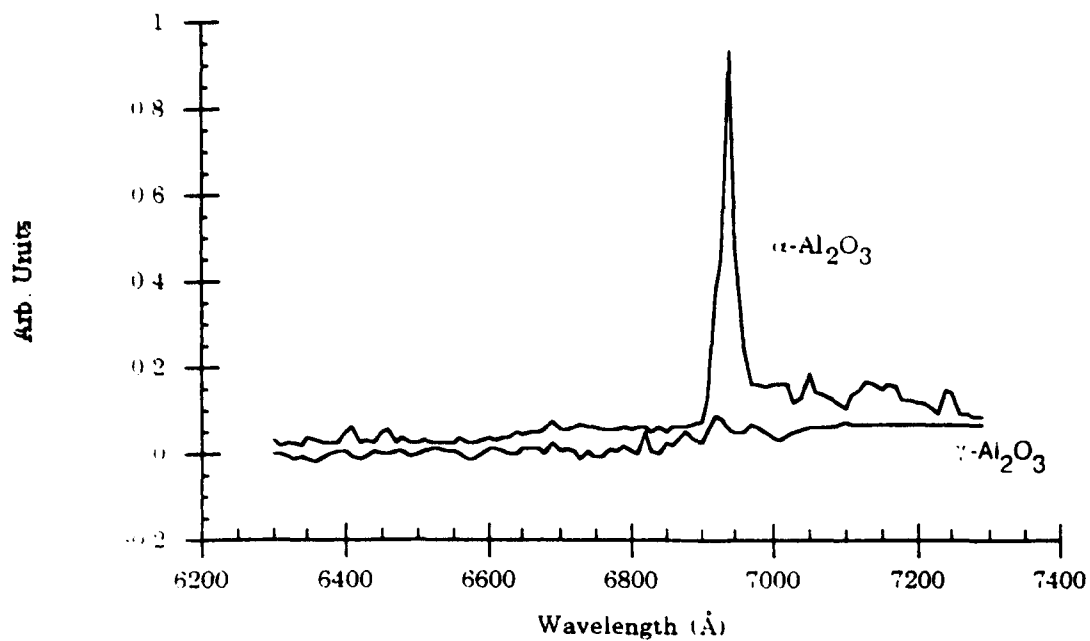
Fig. 12- Cr Al2O3 films crystallized to γ-Al2O3 fluoresce weakly.
but those crystallized to α-Al$_2$O$_3$ show the strong fluorescence of ruby

VORTEX DYNAMICS IN PATTERNED

SUPERCONDUCTING FILMS

Balam A. Willemsen

Graduate Student

Department of Physics

Northeastern University

360 Huntington Ave.

Boston, MA 02115

Final Report for

Graduate Student Research Program

Rome Laboratory

Hanscom Air Force Base

VORTEX DYNAMICS IN PATTERNED
SUPERCONDUCTING FILMS

Balam A. Willemsen
Graduate Student
Physics Department
Northeastern University

Abstract

The complex surface impedance $Z_S(\omega, H, T)$ of patterned Nb and $Y_1Ba_2Cu_3O_{7-\delta}$ thin films has been studied by measuring the resonant frequency and quality factor (Q) of suspended self-resonant patterned structures. Measurements have been carried out on various geometries for several resonances below 20 GHz, in fields up to 6 T and for temperatures down to 4.2 K. These experiments provide a sensitive probe of vortex dynamics, and can be used to investigate the vortex transition from pinning to flow, and possibly distinguish between the various theories which describe the transition region. The field dependence at fixed frequency and temperature for $Y_1Ba_2Cu_3O_{7-\delta}$ samples is analyzed in detail and is found to be consistent with a "single vortex" picture. From this data we also extract useful material parameters such as pinning forces and the flux flow viscosity. By comparison with the results on unpatterned films, only a minimal influence of patterning on the sample characteristics is observed.

VORTEX DYNAMICS IN PATTERNED
SUPERCONDUCTING FILMS

Balam A. Willemsen

Introduction

A proper understanding of material characteristics is crucial to successful applications of superconductors. For microwave applications such as filters, phase shifters, antennas, etc. it is inadequate to merely study the device characteristics alone, for example at high power or in the presence of large magnetic fields, since it is usually difficult to separate the intrinsic material properties from the device response. However the device structures that are often used are particularly useful, since they often allow us to concentrate microwave fields/currents on the sample in a way that is not often possible in measurements on unpatterned films.

In this work we studied two such patterned structures, a multiply-resonant meander line as well as a single resonant line. Rather than packaging these as one would for use as a delay line, we employed in a specially developed package which enables us to study material characteristics via measurements of the complex microwave surface impedance $Z_s(\omega,T,H)$ of $Y_1Ba_2Cu_3O_{7-\delta}$ (YBCO) as a function of frequency, temperature and applied DC magnetic field. Nb patterns have also been studied to investigate the behavior of a "conventional" Type II superconductor.

Other experiments by our group and others [Refs. 1-8] have examined the microwave surface impedance of $Y_1Ba_2Cu_3O_{7-\delta}$ thin films and crystals. Typically, these experiments usually have access to only part of the information required, choosing to focus on either the resistive or inductive parts, temperature, frequency, and DC or RF field dependencies. Our experiment has notable advantages

over those listed above in that it can *simultaneously* probe the temperature, DC field dependence of $Z_s = R_s + iX_s$ at various discrete frequencies. Also, since our technique is contactless, we need not worry about any effects due to contact resistances. Furthermore, only one film is used, so that the DC field configuration is not complicated by any demagnetization effects associated with superconducting ground planes.

Methodology

A number of films and patterns have been studied as a part of this study, we will focus on the following four samples:

Sample A: A 0.600 inch square 4500 Å thick $Y_1Ba_2Cu_3O_{7-\delta}$ film sputtered on 0.020 inch thick $LaAlO_3$ substrate (obtained from Conductus, Inc. as part of a 2 inch diameter wafer, diced into 0.600 inch squares) patterned by standard photolithography (wet-etching) into a meander line 0.0125 inch wide and 2.5 inch total length. (See schematic representation in Fig 1.). [This sample was ex

Sample B: A 0.600 inch square 4500 Å thick $Y_1Ba_2Cu_3O_{7-\delta}$ film deposited by Pulsed Laser Deposition (laser ablated) on 0.020 inch thick $LaAlO_3$ substrate (obtained from Neocera, Inc.) also patterned by standard photolithography into a meander line 0.0125 inch wide and 2.5 inch total length.

Sample C: A 0.600 inch square 1900 Å thick Nb film sputtered on 0.010 inch thick $LaAlO_3$ substrate also patterned by standard photolithography into a meander line 0.0125 inch wide and 2.5 inch total length.

Sample D: A 0.600 by 0.200 inch 1900 Å thick Nb film sputtered on 0.010 inch thick $LaAlO_3$ substrate patterned by standard photolithography into a line 0.0135 inch wide and 0.500 inch long.

These patterned film were then mounted in a copper package leaving 0.060 inch air gaps above and below the sample and a 0.045 inch "border" around the patterned portion of the sample.

The package is then loaded into the sample chamber of a Janis Super Varitemp cryostat where the temperature of the Helium flow around it can be varied from below 4.2 K to above room temperature. The sample sits at the center of a Cryomagnetics 6 T superconducting coil, so that DC magnetic fields can readily be applied perpendicular to the sample. Microwave energy is coupled in to and out of the system by means of two coaxial leads. The position of these leads relative to the sample (and hence the coupling strength) can be varied at all times by a micrometer stage mounted at the top of the cryostat. Thus, we can always ensure that we are weakly coupled to the self-resonant structure under examination. The microwave transmission amplitude S_{21} is then measured using a Hewlett Packard HP8510C Automatic Network Analyzer.

For a straight line of length l suspended in this manner one would simply expect that the resonant frequencies be equally spaced and simply given by $f_n = (nc/2l)(\epsilon_{eff})^{-1/2}$, where ϵ_{eff} is the effective dielectric constant of the medium in which the line is embedded.

For each resonance, we measure the resonant frequency (f_0) and the 3 dB bandwidth (Δf). From these we calculate the quality factor (Q), $Q = f_0/\Delta f$. We can then extract the relevant material parameters from the changes in f_0 and Q. The surface resistance of the film is extracted from the measured data by $R_s(T,H) = \omega_0 L_G/Q$. The surface reactance is similarly given by $\delta X_s(T,H) = \omega_0 L_G(\delta f_0(T,H)/f_0)$ or equivalently in terms of the penetration depth $\delta\lambda(T,H) = (L_G/\mu_0)(\delta f_0(T,H)/f_0)$, where δf_0 are the changes in resonant frequency. Here L_G is the geometrical inductance of the meander line [Ref. 9].

A key feature of our design is the variable coupling which ensures that we are

always measuring an unloaded Q. Losses in this system can arise from at least three sources. Losses due to the nonzero surface impedance of the superconductor, dielectric losses due to the substrate, and losses due to the normal metal package. We have chosen this geometry, so that the microwave fields (and hence also currents) are concentrated for the most part on the sample. Thus we can generally ignore any effects due to the package walls (as long as we stay below the cutoff frequency at which the package and substrate begins to have resonances of its own) as well as the substrate losses.

Results and Discussion

By examining the package loaded with only a blank 0.020" $LaAlO_3$ substrate, we can determine that modes which are not determined by the sample appear around 12.5 GHz. Thus we have restricted our initial investigation to only those modes below this "cutoff" frequency.

A representative trace is presented in Fig. 2 displaying some of the various accessible modes for Sample B. As can be seen from the plots of mode frequency versus mode number for the in Fig. 3, the resonances of both patterns are well represented by the above equation, and yield $\epsilon_{eff} \sim 10$ for the meander lines (Samples A, B and C) which seems reasonable, since the $LaAlO_3$ substrate with ϵ = 25 is only present on one side of the meander line. What is a little surprising is the apparent insensitivity of ϵ_{eff} to the sample thickness. The deviations from linearity occur near the frequency for which the length of an individual arm of the meander line corresponds to a half wavelength, so that the presence of the bends leads to interference which most strongly affects those modes near frequencies $f_n = (nc/2l_{arm})(\epsilon_{eff})^{-1/2}$, i.e. at multiples of 4.66 GHz. For these modes, we are no longer looking at the entire structure, and are avoided in the present work. For a possible application of this kind of interference to microwave antennas see Ref. 10.

We then proceed to examine the temperature dependence of Z_s in zero applied magnetic field. R_s appears to have a fairly conventional temperature dependence in all samples studied to date, except for a slight plateau near 60 K observed in Sample A which is similar to that observed by other groups in high quality films [Fig 4].

The results for the temperature dependence of the penetration depth [Fig. 5] measured for Sample A seem consistent with our measurements at Northeastern on unpatterned portions of the same wafer using a tunnel diode oscillator (TDO) at 4 MHz as well as measurements in a Nb cavity at 10 GHz. This would appear to indicate that patterning has not significantly altered the characteristics of the film.

Since this method is least sensitive near T_c, it is difficult to pinpoint T_c itself to any level of certainty. By extrapolation from the data measured on the two Nb samples, we found that T_c must have been somewhere around 7 K. Since T_c for Nb is usually above 9 K this was rather disturbing. In order to properly determine Tc and thus be able to improve it, we set up a 4 point DC resistance probe which could be immersed directly in the liquid helium storage dewars to rapidly determine T_c and feed this back into the sample preparation loop. By means of this tight sample loop, we have improved the T_c of our Nb from 7.2 K to 8.6 K, but are still unable to get T_cs above 9 K.

If we now keep the sample at fixed temperature and ramp the magnetic field up to a maximum value (6 T) and then back to zero again, we find that R_s and λ increase as expected with increasing applied field. Surprisingly they return to their initial values when the field is reduced back to zero, with only minimal hysteresis observable on the scale of the overall changes. As can be noted by plotting R_s^2 vs. H [Fig. 7].

This field dependence can be explained in terms of a simple viscoelastic picture

of vortex motion. Starting from the equation of motion for a pinned vortex one can show that $Z_s(\omega,T,H) = \omega[\phi_0\mu_0/(\alpha+i\omega\eta)]^{1/2}(H)^{1/2}$ (where α and η are the vortex pinning forces and viscosity respectively), leading to $R_s^2(\omega,H,T) = \gamma^2(\omega,T)H$.

From the slopes of the R_s^2 vs. H plots, we can extract the viscoelastic coefficient $\gamma(T)$ (a combination of α and η) [Fig. 8]. When 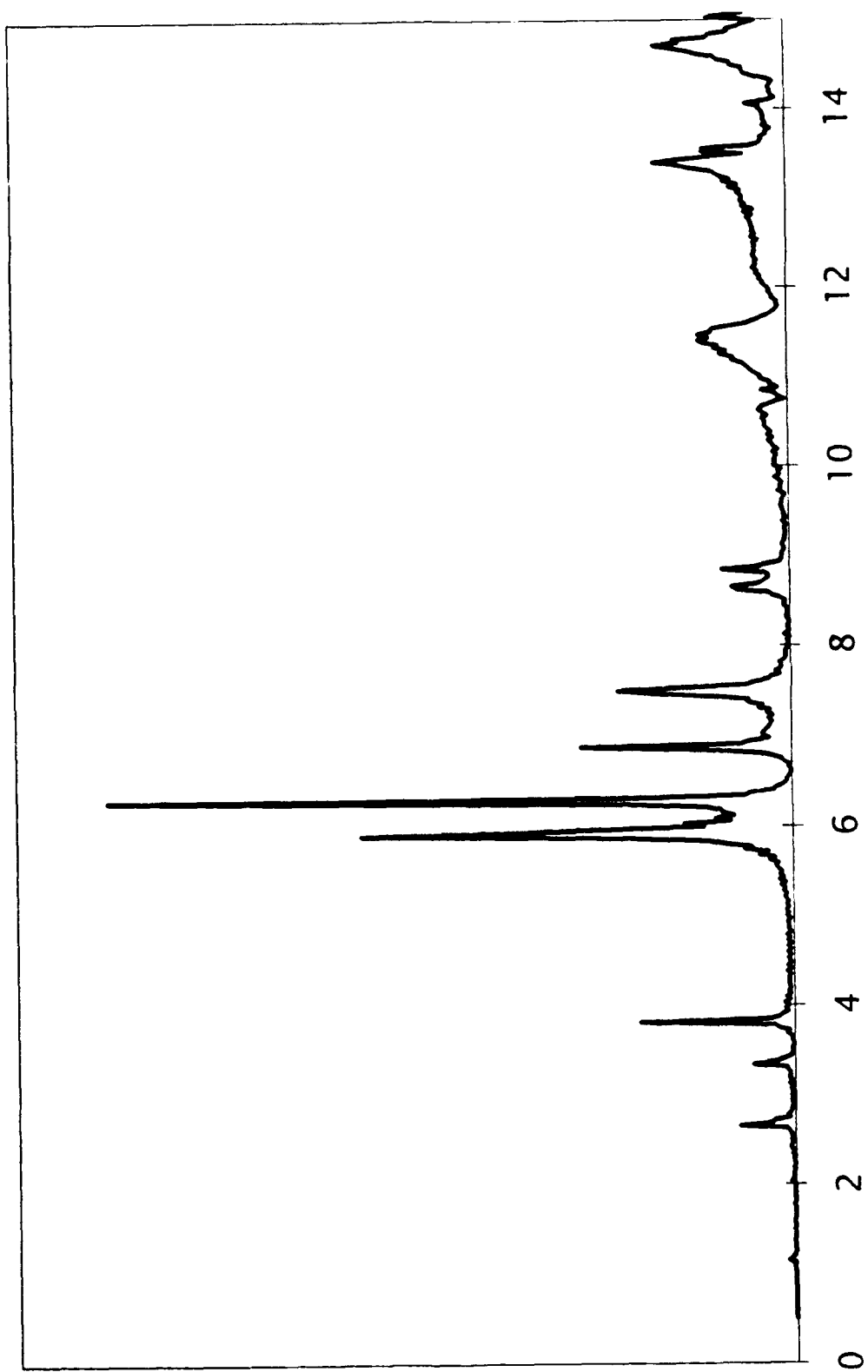compared to the results previously obtained by our group on unpatterned films [Ref. 2] and single crystals [Ref. 3], we find an excellent agreement over a broad range of temperatures (10 - 80K). An important aspect of the present work is the ability to measure the vortex parameters over a wide range of temperatures, in contrast to other methods limited to near Tc.

The frequency shifts [Fig. 9] appear to rise *linearly* with increasing field. This is in contrast to the simple viscoelastic model described above which would predict that $\lambda^2(\omega,H,T) = \phi_0\alpha(\omega,T)H$. This odd temperature dependence of the frequency shift is apparent in all 4 samples discussed here. Initially, we suspected that the frequency shift was being dominated by extraneous effects, possibly due to the dielectric. Yet, taking another look at the temperature dependence of λ, we find that the measured frequency shifts here are very well described by $1/(1-t^n)$ with $n=4$ for the Nb films, and $n=2$ for YBCO [Ref. 7]. Note that the two fluid model would predict that $\lambda(T) = \lambda(0)[1/(1-t^4)]^{1/2}$ which should hold for Nb, and $\lambda(T) = \lambda(0)[1/(1-t^2)]^{1/2}$ has often been reported for YBCO. This, coupled with the measured field dependence, would tend to indicate that $\delta\lambda^2(T,H) \propto (\delta f_0(T,H)/f_0)$ rather than $\delta\lambda(T,H) = (L_G/\mu_0)(\delta f_0(T,H)/f_0)$ as was discussed above. Lacking any outside motivation for this behavior, we must hesitate to take this leap. It is also important to note that we still cannot rule out any extrinsic effects which may be causing such odd behavior in the frequency shifts.

7-15

7-17
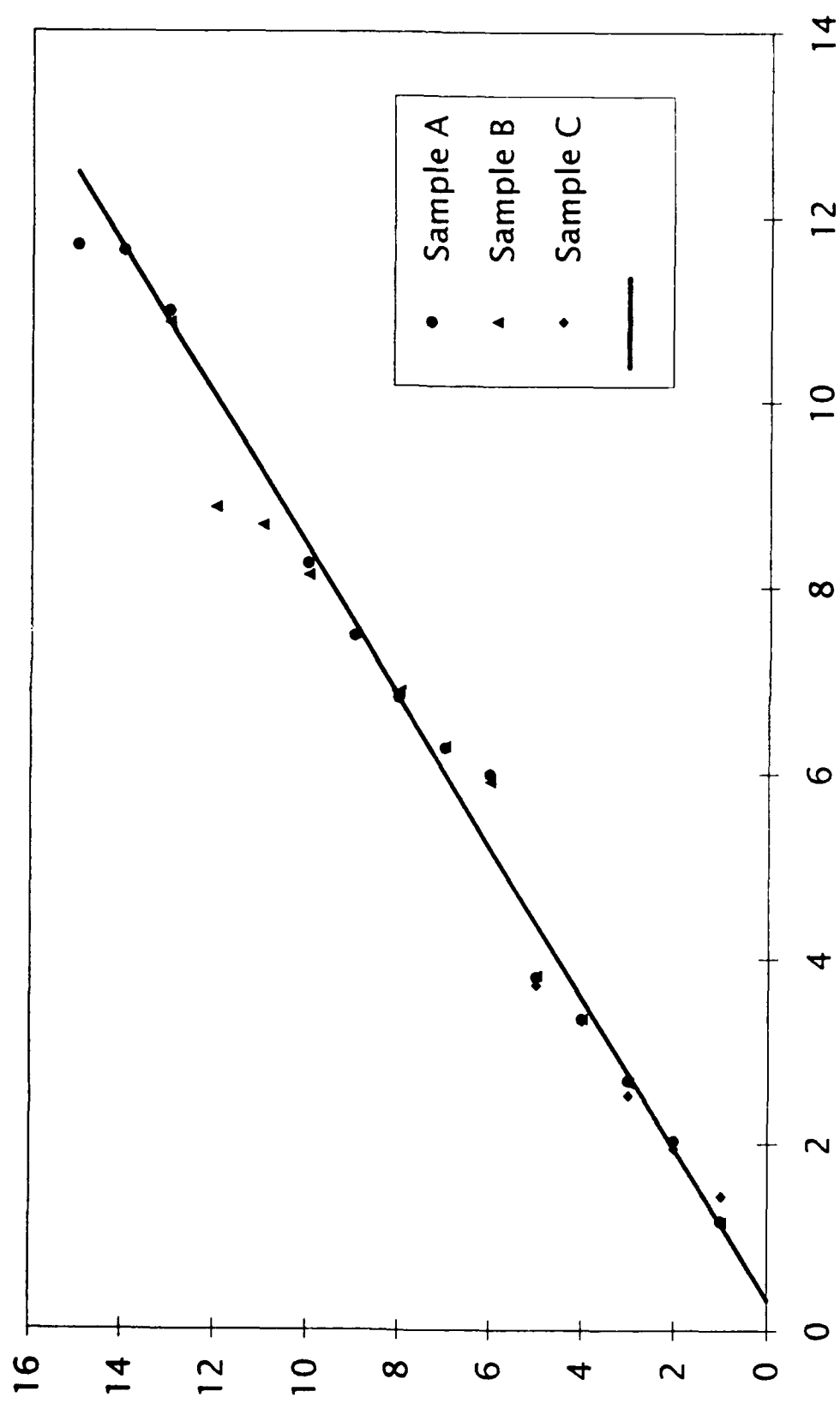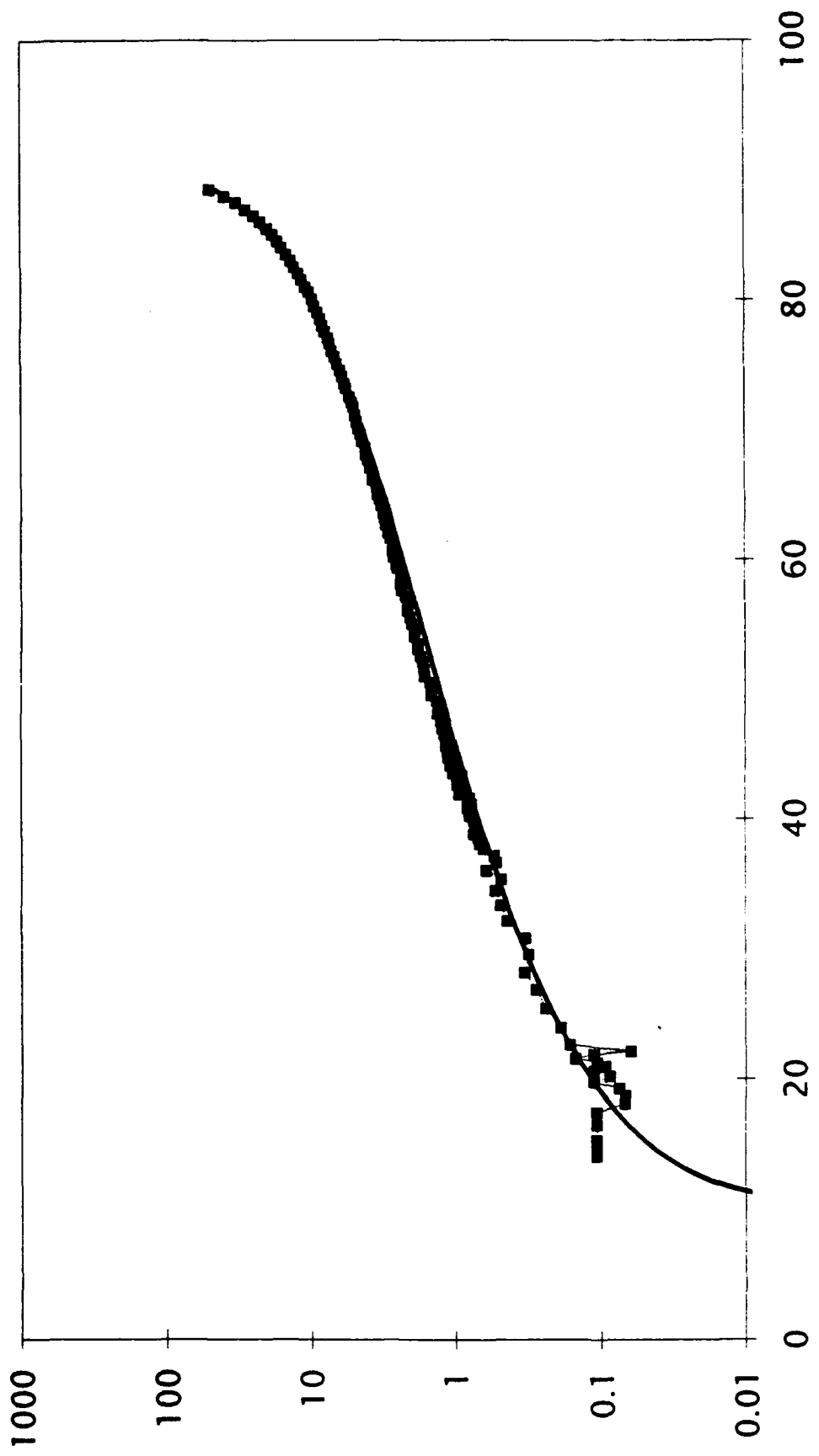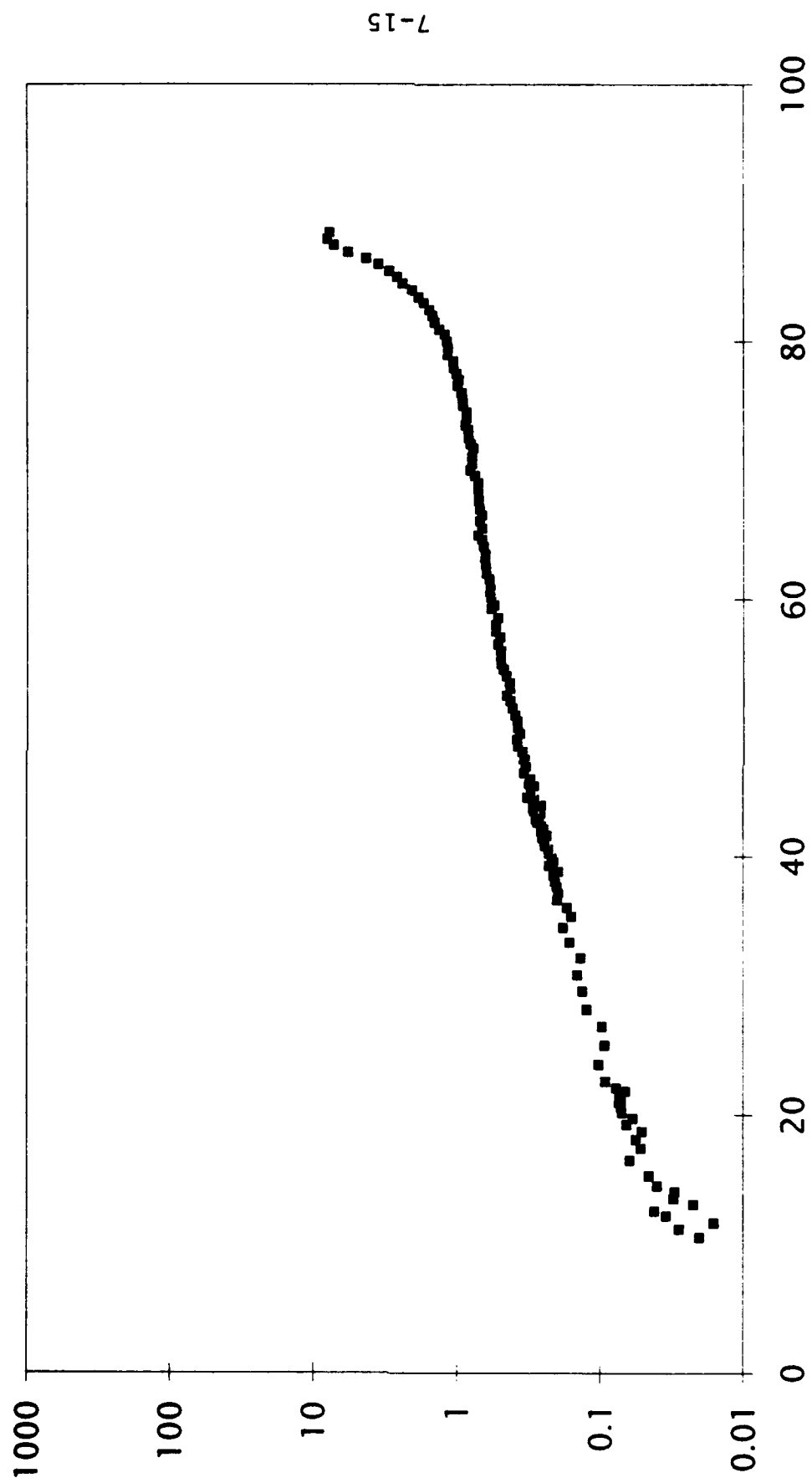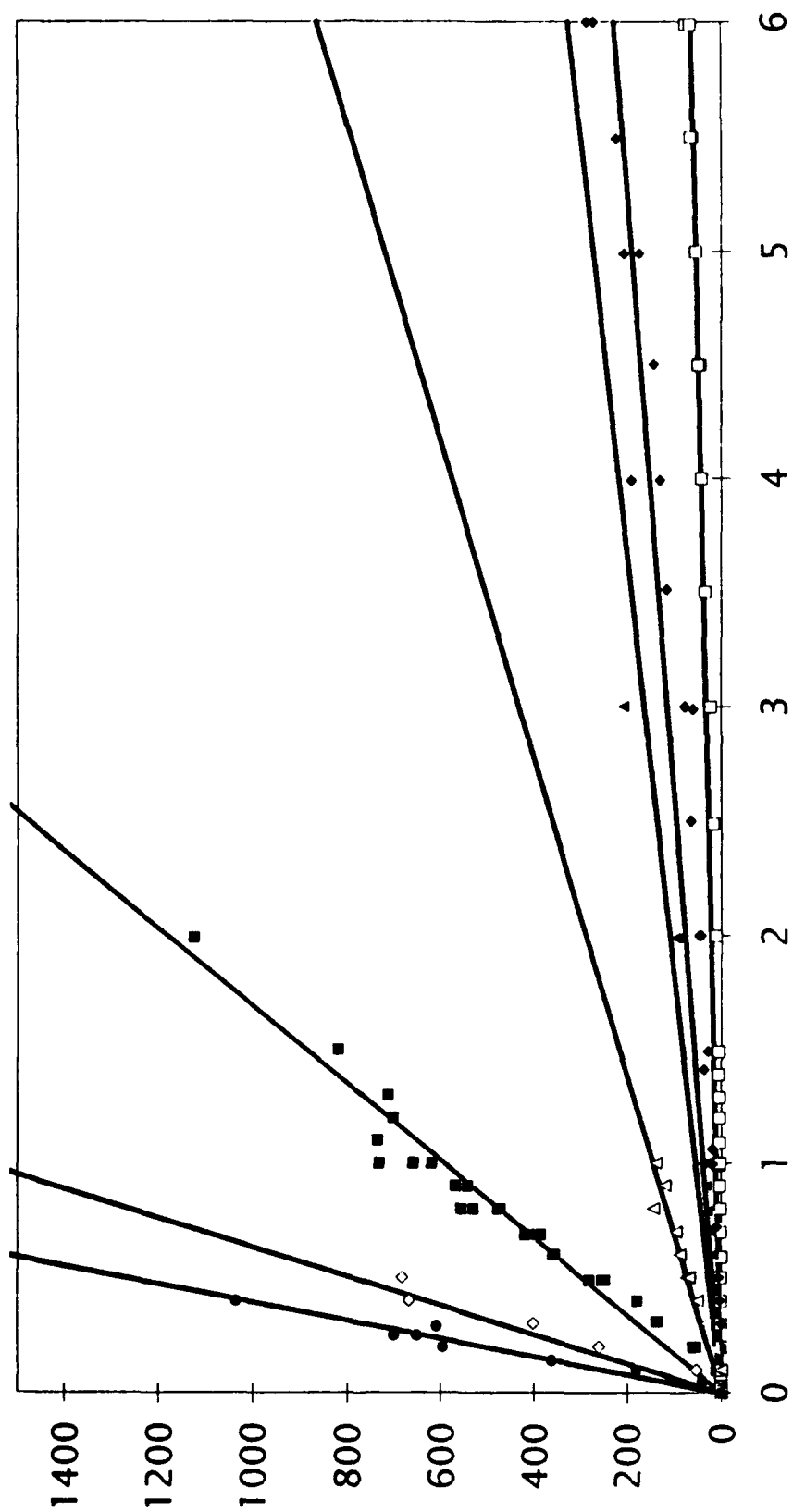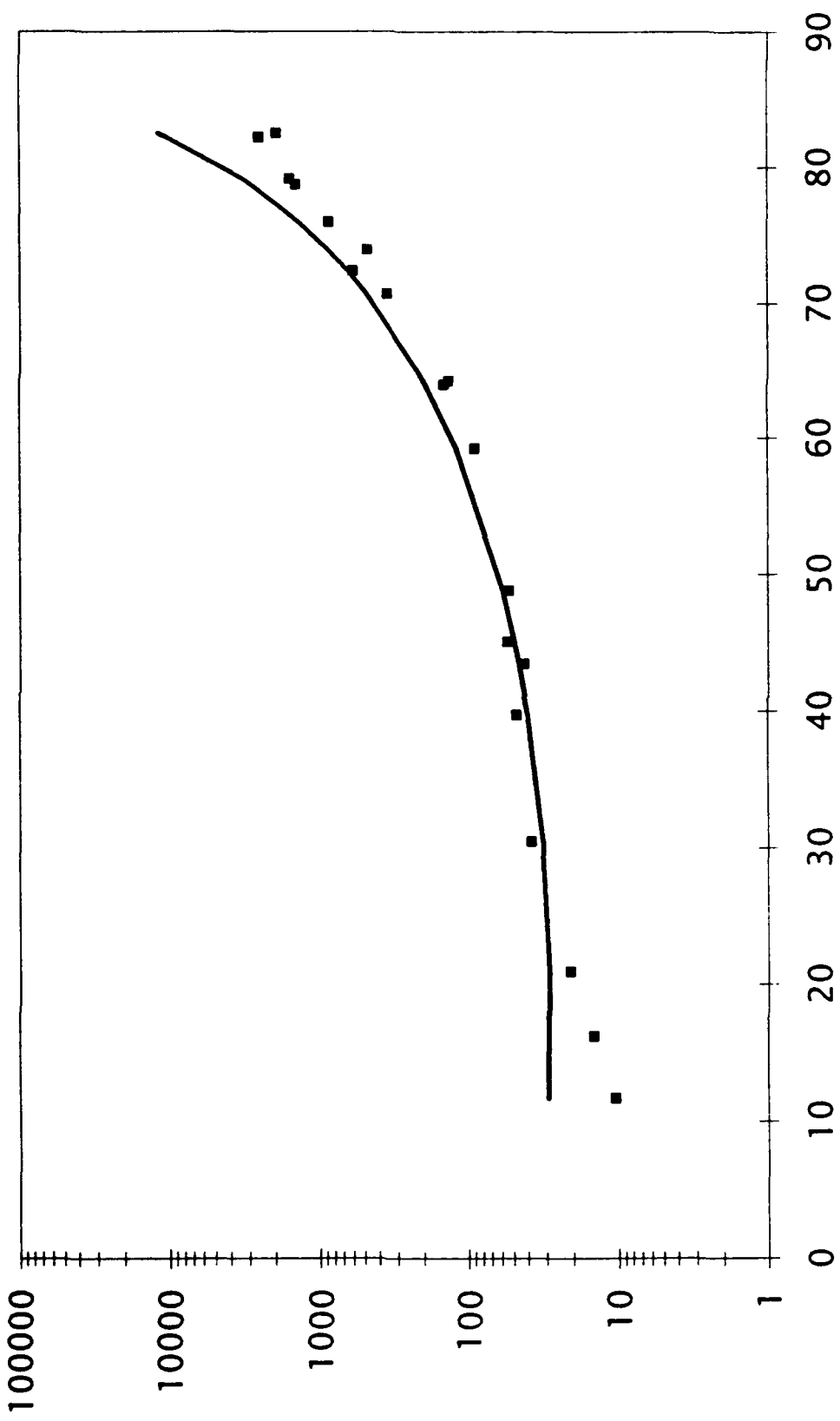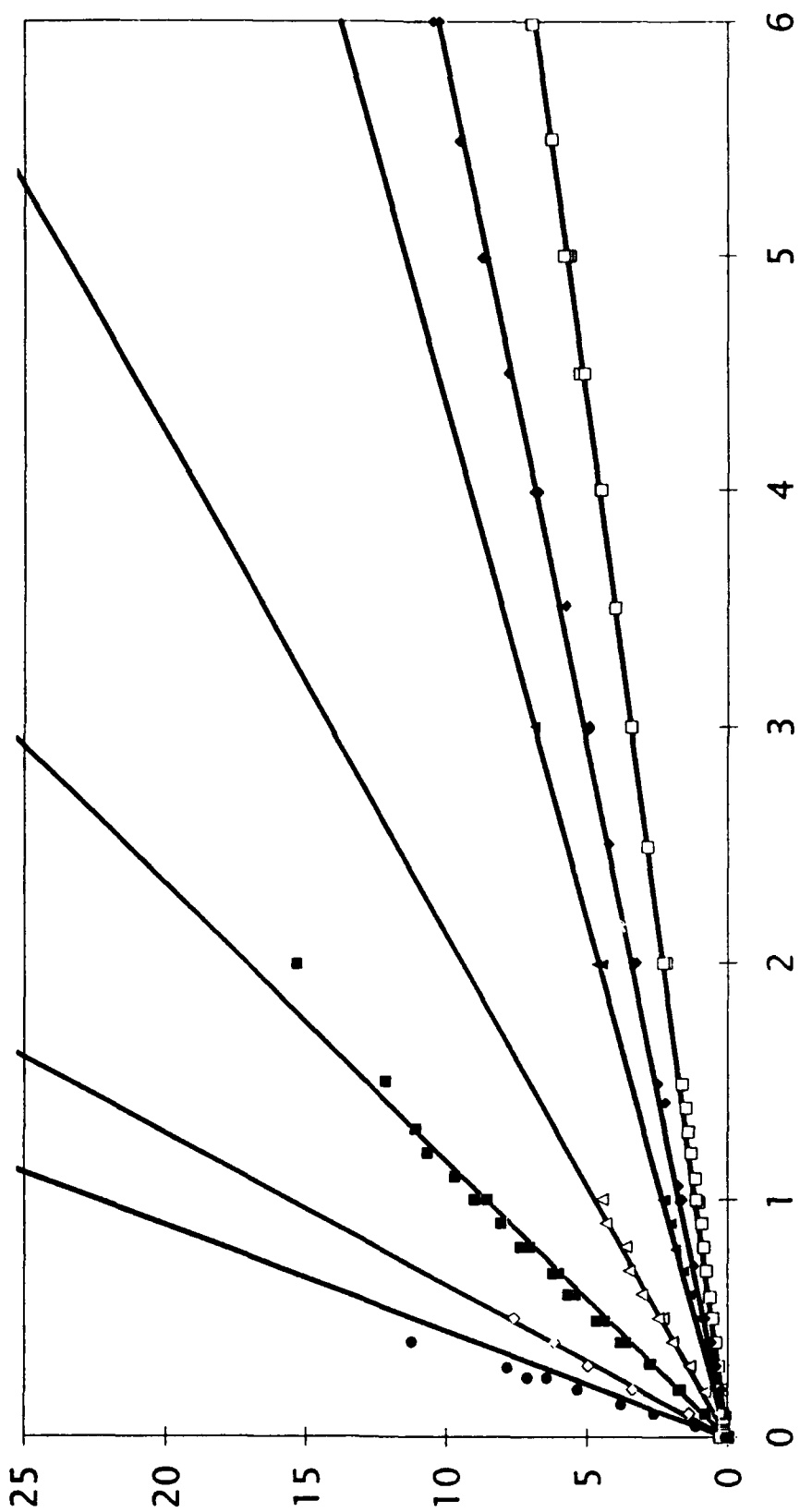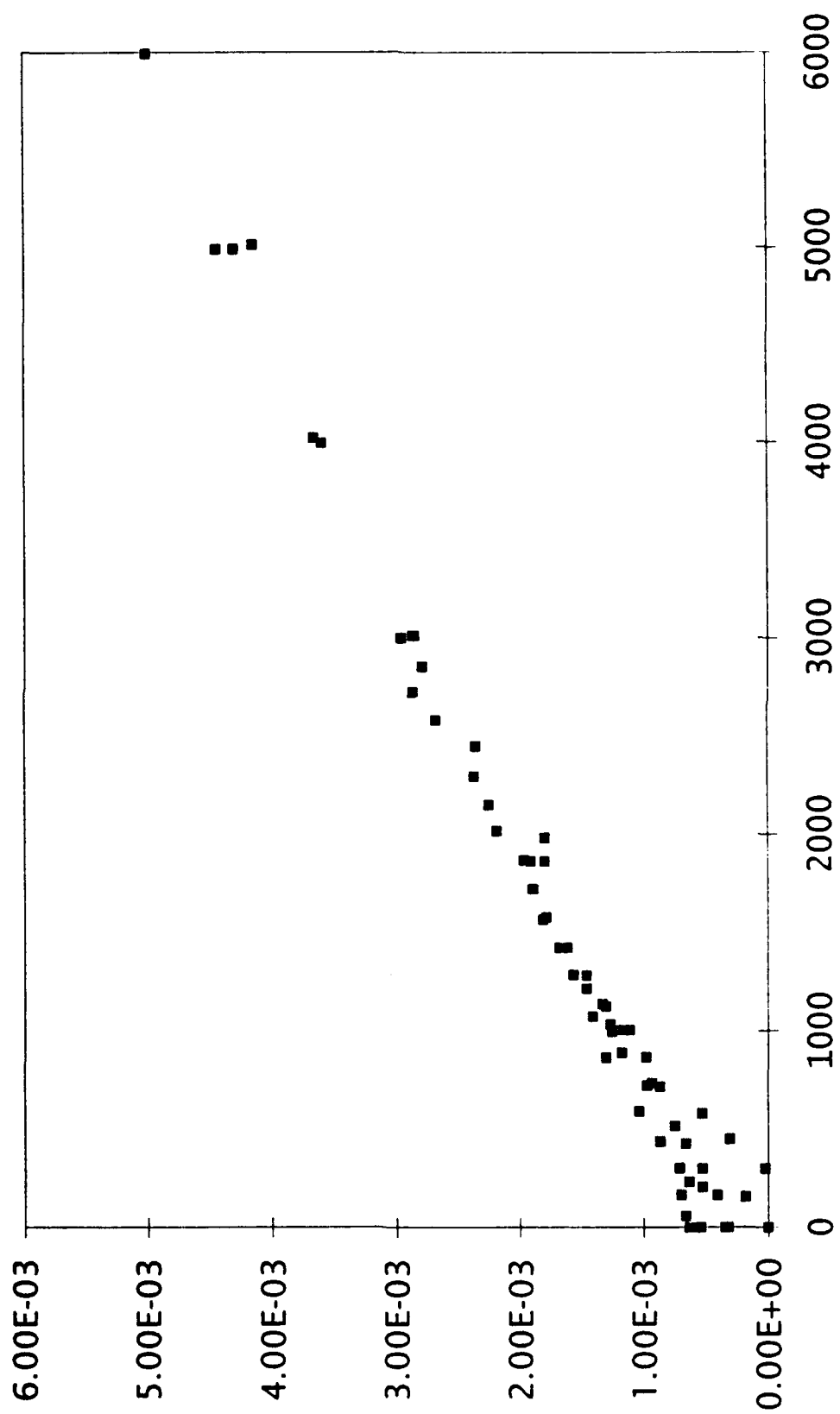
7-19

Wavelet Transforms, Ambiguity Functions, and Radar

Charity A. Carter
Graduate Student
Department of Electrical Engineering
Stevens Institute of Technology
Hoboken, NJ 07030

Wavelet Transforms, Ambiguity Functions, and Radar

Charity A. Carter
Graduate Student
Department of Electrical Engineering
Stevens Institute of Technology

Abstract

In radar systems, the response of the correlation receiver to the target return is typically described by the ambiguity function. The plot of this function is used in the extraction of information from the return signal. When the receiver is not the idealized matched filter, the function is referred to as the cross-ambiguity function. Wavelet transforms are essentially wideband cross-ambiguity functions (WBCAF) and can be used in a more efficient implementation structure of this function. Extending the concept of the wavelet transform is the chirplet transform. This transform embodies many other time-frequency (TF) and time-scale (TS) transforms and may provide greater resolution capabilities than transform methods generally employed.

Wavelet Transforms, Ambiguity Functions, and Radar

Charity A. Carter

Introduction

Many methods exist for transforming or decomposing a signal into components whose properties can then be analyzed. One method which is currently receiving much attention is that of the wavelet transform. Essentially, a continuous wavelet transform (CWT) is a time-integrated product of a signal x(t) with a set of analyzing basis functions. The form of the basis functions used is such that the one-dimensional time signal is mapped into a two-dimensional time-scale space. Wavelet signal analysis has already proven useful in acoustic and seismic data analysis and image compression. Other areas which are promising for wavelet theory applications are biomedical and biological systems [1] and radar signal return analysis [2]. The latter will be the focus of this report.

The ambiguity function or ambiguity diagram is commonly used in radar systems to describe the ability of a waveform and its processing filter to resolve two or more radar reflectors at arbitrarily different ranges and velocities. An ambiguity function is basically a correlation of an unmodified signal with a second modified signal. When the radar receiver takes the form of a matched filter, the result is the auto-ambiguity function or simply ambiguity function. If the idealized matched filter receiver is not used, the function is termed the cross-ambiguity function [3]. The output of an ambiguity function is a surface of correlation values

which can then be used to obtain information about the range and velocity of a radar reflector. Recently, wideband ambiguity function processing has been more extensively researched due to its relationship to the wavelet transform. It has been shown that the wideband cross-ambiguity function can be reformulated with wavelet transforms and such an implementation would prove to be efficient [2].

Current research into time-frequency and time-scale methods has yielded a multidimensional space containing a " time-frequency-scale volume". This space would include functions such as the short-time Fourier transform , Fourier transform, wavelet transform and wideband ambiguity function, in addition to other TF and TS signal representations. Signals in this multidimensional space can be obtained by a new transform which has been named the chirplet transform [4].

The Continuous Wavelet Transform

The Fourier transform, a commonly used signal processing transform, represents functions as weighted sums of exponentials at different frequencies. A Fourier integral is given by:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-j\omega t) \, dt \qquad \textbf{(1)}$$

The Fourier transform is used in the analysis of signals that are stationary, that is, whose properties do not change over time. Since the sine and cosine basis functions used do not change their behavior over time, they cannot represent events that might occur

abruptly and then disappear. Such changes in time are spread out over the entire frequency range and can only be recaptured when the signal is put back together in the time domain. Therefore, this transforming technique would not prove useful for many real-world signals [1,5]. Typically, the solution to this problem has been to use the short-time Fourier transform (STFT). In this case, the FT is taken over a small time segment, or window, of the signal s(t). The STFT is defined as:

$$S(\omega, \tau) = \int_{-\infty}^{\infty} s(t) w(t-\tau) \exp(-j\omega t) dt \qquad (2)$$

The window function, w(t), restricts the Fourier analysis to a small time segment of the total signal. Unlike the Fourier transform, the STFT captures information in the time and frequency domains simultaneously. By selecting the shape of the window function, it may be possible to construct a STFT that will be particularly sensitive to segments of the signal that have a similar shape. However, certain disadvantages are encountered in using this signal processing technique. The STFT is based on the use of windows of fixed duration. Since the window duration (time resolution) is fixed, the frequency resolution is also fixed, in agreement with the uncertainty principle. A transform that permits variable resolution as a function of time and frequency would be able to resolve many signal events which cannot be resolved by the STFT. Such a capability is provided by the wavelet transform (WT) [1].

The continuous wavelet transform is expressed as the time-

integrated product of a signal with a set of analyzing basis functions. These basis functions are shifted and translated versions of a signal called the mother wavelet g(t). A WT may be written as:

$$CWT(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(x) \, g^*(\frac{x-b}{a}) \, dx \qquad (3)$$

The wavelet transform results in a two-dimensional function of the variables " a" (scale) and " b" (translation) obtained from the one-dimensional function f(x). The WT can also be written as a function of frequency and time [6] but is more commonly a time-scale function. When implementing the WT as a TS function, translation represents time while scale is analogous to frequency. While the STFT uses windows of a fixed size, the resolution in the time-scale space of the WT changes since the windows are not of a fixed size. As the scale parameter of the mother wavelet varies, the effective window width does as well. However, the change in window size does not vary in an arbitrary manner but rather changes so as to maintain a constant-Q property, or constant relative bandwidth. Due to this constraint, the wavelet transform provides better translation resolution as the scale value decreases and improved scale resolution as the scale parameter increases. Most importantly, the wavelet transform representation and its properties are determined by the choice of mother wavelet. Choosing one mother wavelet may provide an advantage over another mother

wavelet depending on the form of the signal being analyzed. While many functions may be mother wavelets, all must be oscillatory and quickly decay to zero in both the positive and negative directions, therefore many bandpass signals are suitable. A more detailed formulation of the conditions placed on mother wavelets may be found in [2]. Signals satisfying these conditions are called admissible functions.

In summary, the WT has a correlator structure that indicates how well signal matches a particular scaled and translated mother wavelet. The greater the match between the two signals, the greater the wavelet coefficient CWT(a,b) [2]. Such a capability is also provided by the radar ambiguity function and investigations into the similarities of the two functions have been made [2,5].

The Radar Ambiguity Function

In radar system analysis, it is desired to obtain information about a target by inspecting the signal that has been reflected from it. This radar return signal contains range and velocity information which can be acquired from the ambiguity function.

The ambiguity function is obtained from a correlation integral which performs a similarity measurement between two signals. By taking the squared magnitude of this correlation integral, the ambiguity function is formed. A plot of this function is the ambiguity diagram. The greater the degree of match between the signals, the larger the value of the correlation integral. In radar applications, the two signals being correlated are the transmitted and received signals. By plotting the value of the

ambiguity function for different values of range and velocity, the ambiguity diagram can be constructed. Peaks in this diagram allow for identification of areas of high correlation and determination of the actual velocity and range values [3,7].

Ambiguity functions are generally expressed either in narrowband form or in wideband form. The narrowband form can be used provided that the signal's energy is concentrated about its central frequency. When implementing the narrowband ambiguity function, the Doppler shift can be approximated by a frequency shift which is constant across the signal bandwidth [8]. The narrowband cross ambiguity function for two signals r(t) and s(t) can be obtained from the expression

$$NBCAF(\omega_D, \tau) = \int_{-\infty}^{\infty} r(t) s^*(t-\tau) \exp(-j\omega_D t) dt \qquad \textbf{(4)}$$

Where τ is the time delay and ω_D is the Doppler frequency shift. It can be seen that this function performs the same operation as the STFT. Therefore, a STFT could be used where it is necessary to implement a narrowband cross-ambiguity function [2].

However, for many applications utilizing wideband signals, the narrowband ambiguity function cannot be used. In general, a signal is considered wideband if its time-bandwidth product is much greater than one. A large signal Q value, that is, a fractional bandwidth $BW/f_c > 0.1$, also indicates that a signal should be treated as wideband. Since the deviation from the central frequency

is large, the Doppler shift should not be approximated by a frequency shift. For those signals considered wideband, such an approximation will lead to an incorrect implementation structure of the ambiguity function. If the narrowband processor is invalid over part of the processing interval, then its gain will be decreased in that interval. Rather, a wideband ambiguity function which considers the actual time compression or expansion of the Doppler effect must be used [9]. Such a wideband ambiguity function may be formulated as

$$WBCAF(a,b) = \sqrt{a} \int_{-\infty}^{\infty} r(t) s^*(at-b) \, dt \qquad (5)$$

This function is similar in structure to the wavelet transform and a reformulation of the ambiguity function using wavelet transforms is possible [2].

Reformulation of the Wideband Cross-Ambiguity Function

Just as the narrowband cross-ambiguity function has been implemented with Fast Fourier Transforms (FFTs) to improve computational efficiency, the wideband cross-ambiguity function can be reformulated in the wavelet transform domain. In addition, the reformulation of the WBCAF is exactly analogous to that of the NBCAF [2].

Since wideband signals have a large time-bandwidth product, they also have high resolution. This requires any characterization describing a wideband signal or system to have high resolution as well. When estimating WBCAFs, it becomes necessary to use a very

dense evaluation set to avoid omitting important points from the estimation. Therefore, many WBCAF evaluations are essential in order to include all possible scale and translation values. This large number of evaluations can become burdensome for conventional WBCAF generation algorithms. Consequently, it would be advantageous to employ a fast, efficient algorithm capable of carrying out the calculations required by the high resolution of the system.

The WBCAF of equation five can be expressed as a WT by allowing the following substitutions : $a=1/s$, $b=\tau/s$, $f(x)=r_1(t)$ and $g(x)=r_2(t)$ and replacing the variable x by the time variable t. Therefore,

$$WBCAF(s,\tau) = \langle r_1, r_{2, \frac{1}{s}, \frac{\tau}{s}} \rangle = (W_{r_2} r_1) \left(\frac{1}{s}, \frac{\tau}{s} \right) \tag{6}$$

This allows the WBCAF to be characterized as the WT of $r_1(t)$ with respect to the mother wavelet $r_2(t)$ where the scale factor is $1/s$ and the shift is represented by τ/s. In addition, this definition requires that $r_2(t)$ is an admissible function as described previously.

An important result of wavelet transform theory is that it allows the WBCAF to be defined in the transform domain rather than in the time domain as it typically has been. By wavelet transforming the signals $r_1(t)$ and $r_2(t)$ with respect to a chosen mother wavelet, two sets of wavelet transform coefficients are generated. Obtaining the WBCAF from these coefficients involves the

use of a formula known as the resolution of identity. Essentially, this formula allows a linear mapping of a product of two inner products to a single inner product. For this application, the resolution of identity permits the two sets of wavelet transform coefficients, obtained from the wavelet transform inner product, to be mapped to the WBCAF inner product.

Several advantages exist for computing the WBCAF in the wavelet transform domain rather than in the time domain. Wavelet transforms are useful in the analysis of signals and systems that possess non-stationary characteristics. The wavelet transform coefficients provide a more stable representation of the signals and can be used to investigate signal properties in the transform domain as opposed to the time domain. The multidimensional space-time wavelet transform formulated in [2] maps two 2-dimensional functions to one 2-dimensional function. This mapping does not increase the dimensionality and complexity of processing the signals of interest. As stated previously, the WT reformulation of the WBCAF has the exact same implementation structure as the FFT representation of the NBCAF. Therefore, it may be possible to utilize the WT structure in a similar manner for wideband applications where the narrowband conditions are invalid [2]. In addition, Young [2] has developed an operator, termed the mother mapper, which maps a WT with respect to one mother wavelet to a new WT with respect to a different mother wavelet. This allows several mother wavelets to be considered and compared. Recalling that the mother wavelet determines the resolution characteristics, utilizing

the "best" choice of mother wavelet would provide increased efficiency.

The Chirplet Transform

The Fourier transform, short-time Fourier transform and wavelet transform are examples of some of the many time- frequency and time- scale representations used in signal processing. While these transforms may appear to be dissimilar, it has been shown that they can be unified in the time-frequency-scale volume by the chirplet transform [4]. For example, this multidimensional space contains the short-time Fourier transform as a slice along the time and frequency axes, and the wavelet transform as a slice along the time and scale axes. In the same manner that a wavelet may be regarded as a "piece of a wave", a chirplet may thought of as a "piece of a chirp". A chirp is actually a linear FM waveform which may be written as

$$f(t) = \cos(\omega_c t + \frac{a}{2} t^2) \qquad \text{(7)}$$

In this equation "a" is the chirp rate and ω_c is the carrier frequency. In addition to time, frequency, and scale axes, there exist two other coordinate axes within this transform space: shear-in-time (convolution with a chirp signal) and shear-in-frequency (multiplication by a chirp signal). While the basis functions, or mother wavelets, of the wavelet transform are wavelets that have been translated and scaled, the basis functions of the chirplet transform are chirplets that have undergone translations, rotations

and shears. This allows the chirplet basis functions to have greater flexibility in their form as compared to functions such as the mother wavelets. Therefore, by selecting a particularly shaped basis function, it may be possible to adapt the chirplet transform to be especially sensitive to certain features of the system under analysis and improve the resolution characteristics.

Conclusion

The wavelet transform has been presented with an emphasis placed on wideband applications especially useful for radar systems. Both the wideband cross-ambiguity function and the wavelet transform are inner products that have a correlation type structure and provide a measurement of similarity between two signals. While the WBCAF operation has usually been performed in the time domain, this implementation is not efficient for wideband systems requiring high resolution. By executing the WBCAF in the transform domain, the dimensionality required for processing does not increase. In addition, a more stable representation may be utilized for analyzing non-stationary signals.

The efficient implementation structure of the WBCAF in the WT domain may prove to be useful when employed in the receive end of a phased array radar system. Recent developments in photonics have resulted in optical processors capable of providing continuously variable time delay allowing the beampointing error of squint to be overcome [10,11]. This capability allows wideband use of systems traditionally restricted to narrowband usage due to this inaccuracy. By incorporating wavelet transform methods with these

optical processors, it may be possible to realize wideband phased array systems with greater accuracy than previously possible.

The chirplet transform, which operates in the time-frequency-scale volume, has been presented. Due to the flexibility of the chirplet transform basis functions, it may be possible to employ them with greater efficiency in applications where other time-frequency or time-scale transforms have been used. Recent investigations into possible uses for wavelet transforms may indicate areas which would also be suitable for chirplet transform implementation.

References

[1] M.R.Raghuveer, V.J.Samar, K.Swartz, "Wavelet Transforms: An Introduction to Theory and Applications", to be included in "A Practical Guide to Image Processing", Marcel-Dekker, New York

[2] R.K.Young,"Wavelet Theory and Its Applications", Kluwer Academic Publishers, 1993

[3] E.Brookner,ed.,"Radar Technology", Artech House, Inc.,1977

[4] S.Mann and S.Haykin, "Beyond Wavelets:The Chirplet Transform, a new time-frequency signal representation

[5] O.Rioul and M.Vetterli,"Wavelets and Signal Processing", IEEE Signal Processing Magazine, Vol.8,No.4,October 1991,pp 14-38

[6] F.Hlawatsch and G.F.Boudreaux-Bartels,"Linear and Quadratic Time Frequency Signal Representations",IEEE Signal Processing Magazine,April 1993

[7] M.Skolnik,"Introduction to Radar Systems", McGraw-Hill, Inc.1962,1980

[8] L.Auslander and I.Gertner, "Wide-Band Ambiguity Functions and ax+b Group"

[9] D.Swick,"A Review of Wideband Ambiguity Functions",NRL Report 6994, 1969

[10] H.Zmuda and E.N.Toughlian,"Adaptive Microwave Signal Processing:A Photonic Solution",Microwave Journal, Vol.35,No.2,February 1992,pp 58-71

[11] H.Zmuda and E.N.Toughlian,"A Variable Time Delay System for Broadband Phased Array and Other Transversal Filtering Applications',Optical Engineering,Vol.32,No.3 March 1993,pp 613-617

ARTIFICIAL NEURAL NETWORK INVESTIGATION
IN AUTO SOURCE UPDATE PROGRAM

Walter P. Sweeney Jr.   Mohamad T. Musavi
Graduate Student   Associate Professor

Department of Electrical & Computer Engineering

University of Maine
5708 Barrows Hall
Orono, ME 04469-5708
Phone : (207) 581-2243
Fax: (207) 581-2220
E-mail: wsweeney@eece.maine.edu

Final report for:
Graduate Student Research Program
Rome Laboratory

July 1993

ARTIFICIAL NEURAL NETWORK INVESTIGATION IN AUTO SOURCE UPDATE PROGRAM

Walter P. Sweeney Jr.   Mohamad T.Musavi
Graduate Student   Associate Professor
Department of Electrical & Computer Engineering
University of Maine

Abstract

An investigation of artificial neural networks (ANN) in the candidate selection of auto source update (ASU) program was conducted. Among many types of ANNs the emphasis was given to the probabilistic neural network (PNN) architecture. As compared to other types of ANNs, PNN has proved to be reliable and superior in terms of speed of operation and simplicity of adaptation process. A PNN network was designed and implemented to find the best match and its confidence estimate for a given message among several possible candidates. Different experiments were conducted on a set of messages and candidates from the Defense Mapping Agency (DMA) databases to check the performance of the PNN. The results have been successful and promising. Future enhancement of the proposed PNN and other ANN methodologies in ASU program has also been presented.

ARTIFICIAL NEURAL NETWORK INVESTIGATION IN AUTO SOURCE UPDATE PROGRAM

Walter P. Sweeney Jr. and Mohamad T. Musavi

1. INTRODUCTION

The objective of this project is to investigate application of artificial neural networks (ANN) in support of the Auto Source Update (ASU) program which is currently under development by the Grumman Data Systems Corporation, Woodbury, New York. The ASU is a prototype software capability for maintaining the accuracy and currency of vectorized geographical information using Defense Mapping Agency (DMA) data [1] and other source material as input.

1.1 Auto Source Update (ASU)

The accuracy and currency of the DMA databases prove to be very critical in accurately identifying different segments in geographical regions of interest. The accuracy issue depends primarily on the original DMA databases and the currency issue concerns rapid and dynamic changes in a geographical area due to sudden events during time of conflict. Some changes occur so rapidly that the MCG&I (Mapping, Charting, Geopositioning, and Imagery) database supplied by DMA are altered significantly in a matter of minutes or hours. Under these conditions, one can not solely rely on the original DMA database for planning purposes. An updating system is required to incorporate new changes to be used with the original database in a time efficient manner. Dependence of a large number of Air Force systems on the cartographic information, signifies the importance of an updating system.

The objective of the ASU program [2, 3] is to create such an updating environment for the vectorized information supplied by DMA. From an external input-out point of view, the ASU program can be viewed as a black box that accepts inputs from two different sources of information: (1) the update message data and (2) the original database. The message data is an update request obtained from the dynamic source materials. The original database can be any vectorized data from the DMA databases. The output of the ASU is the appropriate correction and/or update of the original database for the given message. The ASU internal organization consists of three main building blocks: (1) scenario generation, (2) candidate retrieval, and (3) candidate selection. Central to the ASU design is the ASU database which serves as a data depository for different operations and contains data from different external sources in Neutral Update Format (NUF). NUF is a format used by the Grumman Data System in the ASU project.

The scenario generation block provides a simulation environment for generation of update messages to be used in training and testing of the ASU program. It consists of a message generator and message handler. Both will work with data in Neutral Update Format (NUF). The candidate retrieval block will accept update requests as input and will produce a set of candidates for the given message. The candidates are retrieved from the ASU database. The candidate selection block, which is the focus of this study, will accept the message and its associated candidates and identify the most appropriate candidate for the given message.

1.2   Candidate Selection

The candidate selection block of the ASU program requires intelligence and adaptation process to select the best possible candidate among all given candidates. The objective of this research is to investigate an optimal artificial neural network architecture that can effectively be utilized in this task.

2.   PROJECT DESCRIPTION

The objective of this project is to apply artificial neural network (ANN) technology in candidate selection of the ASU program.

2.1   Message and Candidates

The candidate selection (CS) module should accept message data and candidates' data as input and then find the best match for the given message among all possible candidates. The message data is extracted from dynamic update sources such as evaluated intelligence products, Electronic Chart Updating Manual (ECHUM), raw imagery (e.g. LANDSAT TM, panchromatic SPOT, scanned NHAP, infrared RADAR, SAR, and etc.), native cartographic source materials (e.g., paper road maps), intelligence databases, and any other update source within the Air Force. These update messages are normally in varying United States Message Text Format (USMTF) types (e.g., RECCEXREP, IIR, etc.). The ASU program will reformat the information into the Neutral Update Format (NUF) before submitting it to the candidate selection.

A message data in NUF is normally a set of points, given in latitude (LAT) and longitude (LON) degree coordinates, from segments such as roads, runways, lakes, buildings, and etc. in a geographical area of interest. Each message consists of one of four features: (1) point, (2) line, (3) polyline, or (4) area. Each feature can have both numerical and categorical attributes associated

with it. These attributes include feature attribute coding system (FACS) identifier, surface material category (SMC), predominant height (HGT), width (WID), and etc. For example, for a forest area FACS = 5C030 (i.e. woodland), SMC = 12 (i.e. trees), and HGT = 15 meters above ground level (AGL). While the attributes of a bridge can have FACS = 1Q040 (i.e. bridge/overpass/viaduct), SMC = 2 (i.e. part-metal), HGT = 20 meters AGL, and WID = 10 meters. Feature attributes are normally given in the beginning of a NUF message for identification purposes. The appendix at the end of this report gives a print-out of a NUF message and Figure 1 presents the same message coordinate points plotted as a curve.

The candidates' data for a given message are provided by the candidate retrieval. The candidate retrieval extracts appropriate candidates from the ASU database based on the NUF message. The candidates also are in NUF and have the same structure as the message. A set of candidates for the message of Figure 1 is shown in Figure 2. In this specific example there are 11 different candidates (numbered from 0 to 10). The candidate selection has to find the candidate that best matches the given message.
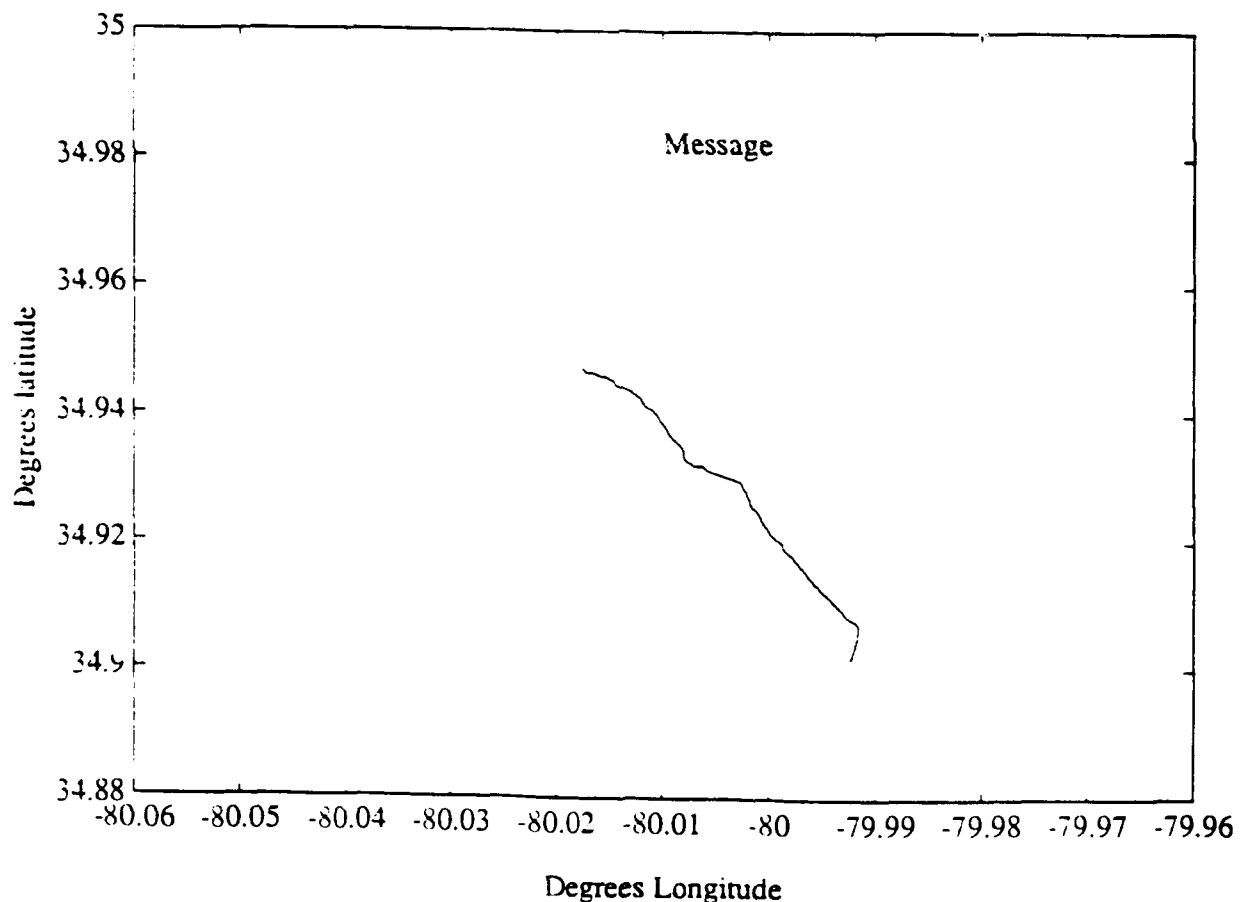
Figure 1. A NUF message plotted by connection of message points.
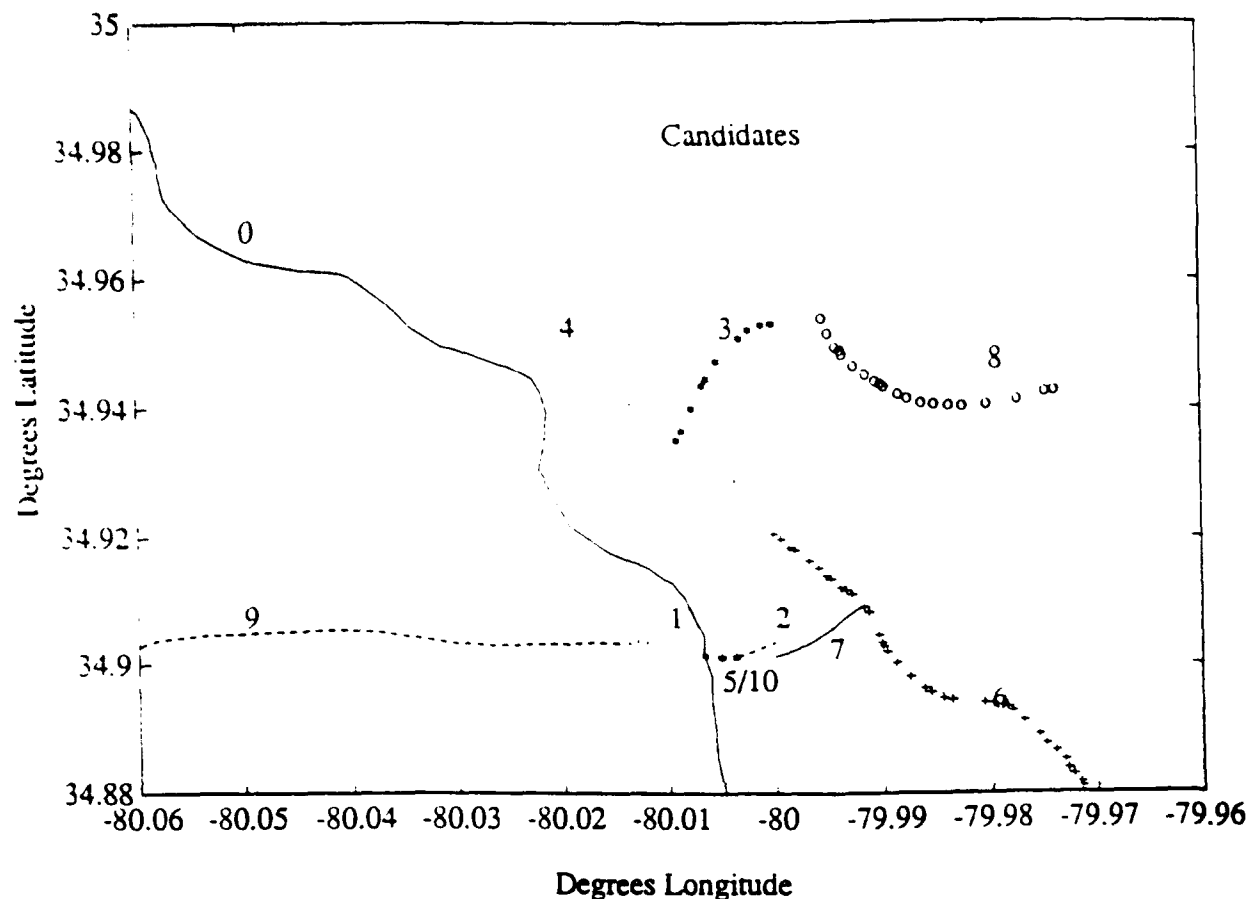
Figure 2. Eleven different candidates associated with the message of Figure 1.

2.2 Probabilistic Neural Network (PNN) Architecture

Our experience with the probabilistic neural network (PNN) [4, 5, 6, 7] has indicated certain advantages of this network over other types of artificial neural networks. At the lower level of architectural design, PNN would be sufficient to act as an independent building block. However, for higher level decision making and integration of individual blocks in the overall system design, more sophisticated architecture and learning paradigm may be needed.

The PNN [8] has been recently recognized as a viable alternative in dealing with different classification problems. PNN [9, 10] is the combination of a kernel based estimator for estimation of probability densities and the Bayes rule for classification decision. For example in a two class (A and B) problem the probability densities are found by:

$$f_A(\mathbf{x}) = \frac{1}{T} \sum_{n=1}^{T} \Phi_n \left(\|\mathbf{x} - \mathbf{c}_{an}\| \right) \tag{1a}$$

$$f_B(\mathbf{x}) = \frac{1}{T} \sum_{n=1}^{T} \Phi_n \left(\|\mathbf{x} - \mathbf{c}_{bn}\| \right) \tag{1b}$$

where T is the total number of training points, $\Phi_n (\|.\|)$ is a local kernel function, $\mathbf{x}$ is the input to the network, and $\mathbf{c}_{an}$ and $\mathbf{c}_{bn}$ are the training data from class A and B respectively. Equations (1a) and (1b) measure the closeness (not necessarily Euclidean) of input $\mathbf{x}$ to the training data and then a simple comparison of the two determines if $\mathbf{x}$ belongs to class A or B. Similar strategy can be adopted for more than two classes of data.

The training process in PNN is that of finding a set of appropriate kernels $\Phi_n (\|.\|)$ and there is no weight adaptation process. To simplify the training process, Specht makes the following three assumptions:

i) All kernels are the same. In other words,

$$\Phi_n \left(\|\mathbf{x} - \mathbf{c}_n\| \right) = \Phi \left(\|\mathbf{x} - \mathbf{c}_n\| \right). \qquad \text{for} \quad n = 1, 2, \ldots, T \tag{2a}$$

ii) The kernel is a Gaussian function given by:

$$\Phi \left(\|\mathbf{x} - \mathbf{c}_n\| \right) = (2\pi)^{-M/2} |\Sigma|^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{c}_n)^T [\Sigma]^{-1} (\mathbf{x} - \mathbf{c}_n) \right]. \tag{2b}$$

Where M is the dimension of the input space and Σ is the covariance matrix.

iii) And that the covariance matrix is diagonal and has equal eigenvalues ($\Sigma = \sigma^2 I$),

$$\Phi \left(\|\mathbf{x} - \mathbf{c}_n\| \right) = (2\pi)^{-M/2} \sigma^{-M} \exp \left[-\frac{\|\mathbf{x} - \mathbf{c}_n\|^2}{\sigma^2} \right] \tag{2c}$$

Using the above assumptions, the problem of training PNN is simply reduced to selecting the parameter σ that is referred to as the smoothing parameter. This parameter is normally found by trial and error or the nearest neighbor criterion. Musavi et al. [4] have recently presented a

technique for optimal selection of all covariance matrices. In our future studies, this technique will be utilized to enhance the accuracy of PNN network.

The significant advantages of the PNN classifier are its speed and training process. First, the training process is one-pass and there is no need for weight adaptation, hence yielding great processing speed as compared to the back propagation (BP), radial basis function (RBF), and other types of neural networks. Second, the network generalizes to the new incoming patterns without having to repeat the training process. These characteristics are ideal for real time applications [11, 12, 13]. The architecture of PNN is very simple, thus, software/hardware implementation of the network on conventional or parallel computers is easy. The processing speed of PNN is achieved at the expense of increased memory size. This is due to the fact that the training data set has to be available for classification of every incoming datum. Specht, in defense of his network, argues justifiably that memory is abundant and affordable.

2.3 PNN in Candidate Selection Module

Accurate selection of a candidate, among many candidates, to best match a given message relies on many attributes. Some of these attributes are proximity, orientation, shape, FACS ID, width, height and etc. A robust system should take advantage of all available attributes to make the best possible decision.

To apply artificial neural networks in the design of candidate selection, PNN will be used as an individual building block. One PNN will be designed for each attribute. The following four sections will discuss structures of proximity PNN (PPNN), orientation PNN (OPNN), shape PNN (SPNN), and segment PNN (SGPNN).

2.3.1 Proximity PNN (PPNN)

For the proximity attribute, a network like that of Figure 3 is constructed. In this network, the input vector is two dimensional, consisting of latitude (LAT) and longitude (LON) coordinates. The output is a scalar giving a numerical value presenting the closeness of the input point to the message under consideration. The points of the message (training data) are set to be the centers of the Gaussian functions. Therefore, the number of nodes in the network equals the number of points in the message. In case of excessive message points, it is possible to use some kind of clustering algorithm without significantly affecting the generalization ability of the network. The smoothing parameter, σ, can be either entered by the operator using trial and error, or found automatically by the nearest neighbor criteria.

The following procedure is followed to find a measure of confidence on the closeness of each candidate to a given message. A block diagram of the procedure has also been given in Figure 4.

Procedure:

Step 1.   Set up the PNN network for the given message.

Step 2.   Select a candidate for the given message.

Step 3.   Input each point of the candidate to the PNN and calculate the output.

Step 4.   Add the output to the previous output.

Step 5.   Repeat Steps 3 and 4 until all points of the selected candidate are covered. The total summation value at Step 4 is a measure of confidence on the closeness of the selected candidate to the given message.

Step 6.   Repeat Steps 2 though 5 until all candidates of the selected message are covered.

Step 7.   Perform a normalization on the total summations of Step 4 for different candidates. Comparison of these normalized values give a measure of confidence on each of the candidates.

Step 8.   Repeat Steps 1 through 7 for a different message.
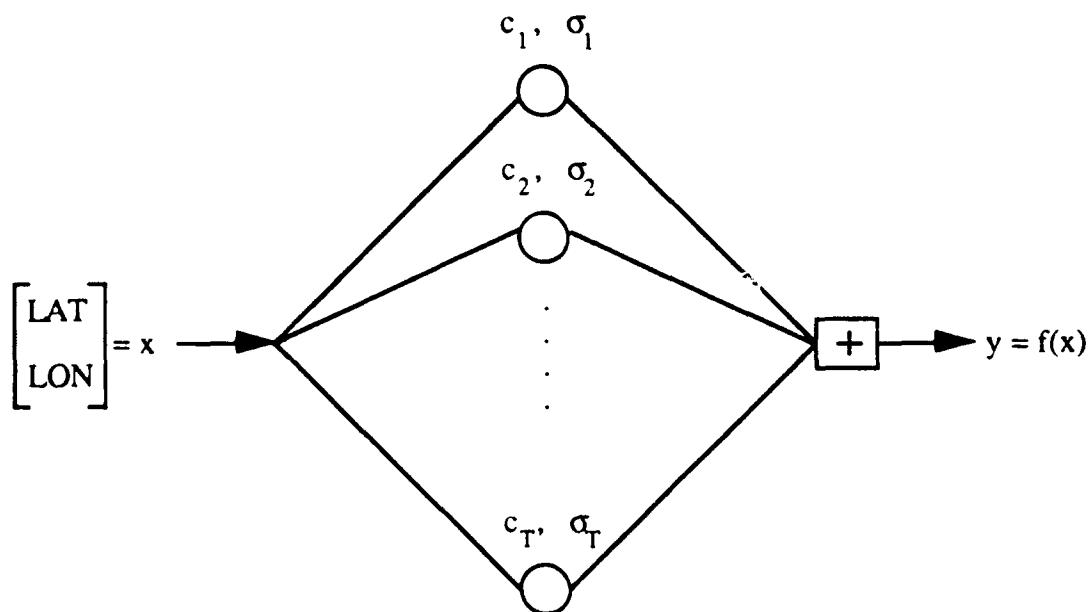
Figure 3. The PNN architecture used in the candidate selection block.
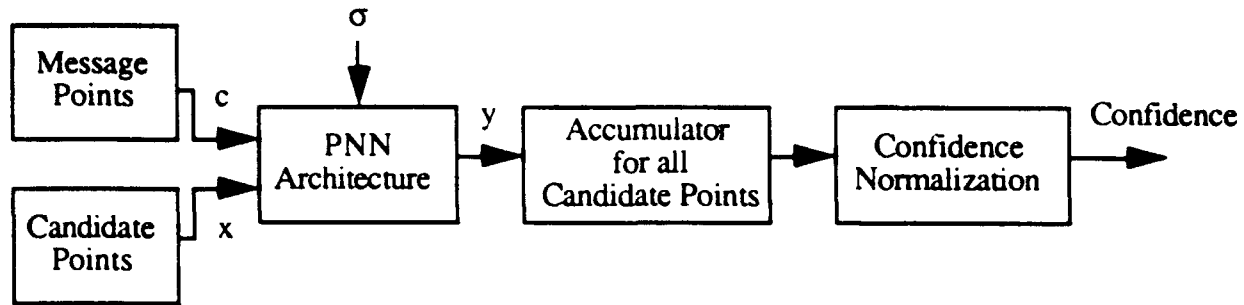
Figure 4. A block diagram of the confidence generator for each given candidate.

2.3.2 Orientation PNN (OPNN)

A PNN similar to the one given in Figure 3 was also used for the orientation attribute (OPNN). In this case, the input to the network is one dimensional, representing the orientation of a small segment of a candidate object. The centers of the Gaussian nodes are orientations of segments that construct the message object. The output of the OPNN is a measure of similarity between the orientations of the message with those of its candidate objects.

2.3.3 Shape PNN (SPNN)

A PNN was also designed to be used for shape comparison (SPNN). The input to this network is a vector whose components are the orientations of small segments. The input object is either the message or a candidate, depending on which is respresented by the least number of segments. The number of nodes used in this network is the same as the number of orientations of the input object. The centers of these nodes are the orientations of segments constructing the other candidate or message object. The output of the SPNN is a measure of similarity between the shape of the message/candidate (input) with the candidate/message object segment (of network) as it moves across its length.

2.3.4 Segment PNN (SGPNN)

A PNN like that of Figure 3 was also be used for segment comparison (SGPNN). In this case the input to the network is one dimensional representing the orientation of one segment of the candidate object. The candidate is split into many segments dependent on its size. The centers of the Gaussian nodes are the orientations of segments that construct the message object. The output

of the SGPNN is a measure of similarity between one candidate segment with the message object segments.

2.4   Candidate Selection Architecture

The overall intelligent system for candidate selection module is built by the integration of PNN blocks for different attributes. Figure 5 presents a block diagram of the overall candidate selection system design.
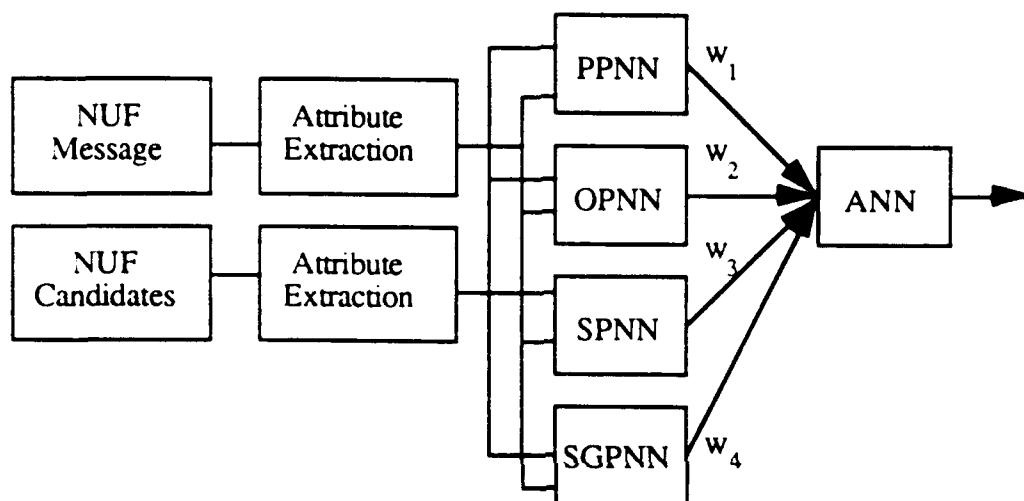
Figure 5. The overall system design for candidate selection module.

3.   TEST RESULTS

A set of messages and their associated candidates were obtained from Grumman Data System for testing the methodology presented in Section 2. The message file consisted of 19 different messages and the candidate file consisted of 76 different candidates for the 19 messages. The 19 messages and 76 candidates are shown in Figure 6, the messages are in solid and the candidates are in dotted lines. Since all the messages and candidates have been plotted on the same graph the beginning and end of a message and its corresponding candidates are not clearly identified. A message with its corresponding candidates are shown in Figures 1 and 2 respectively. Another message with its corresponding candidates are shown in Figure 7.

Table 1 presents the best results from the experiments with the given messages and candidates. This was accomplished using only the PPNN and SGPNN networks, connected by equal weights. The first column gives the message number, the second column gives the associated candidate numbers for that message, column three presents the confidence of the network in selecting each of the candidates, and the last column indicates if the network response is correct.
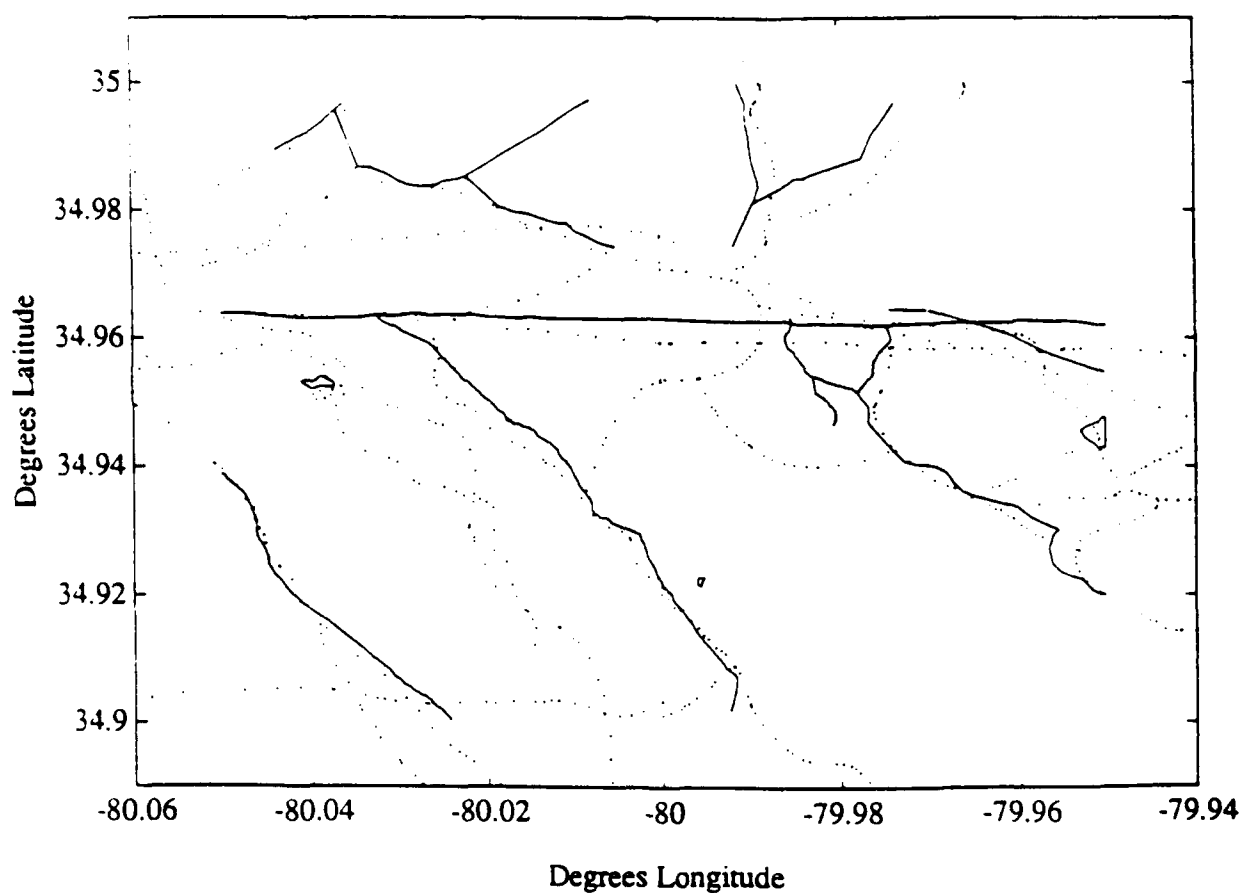
Figure 6. Nineteen messages and their corresponding candidates.
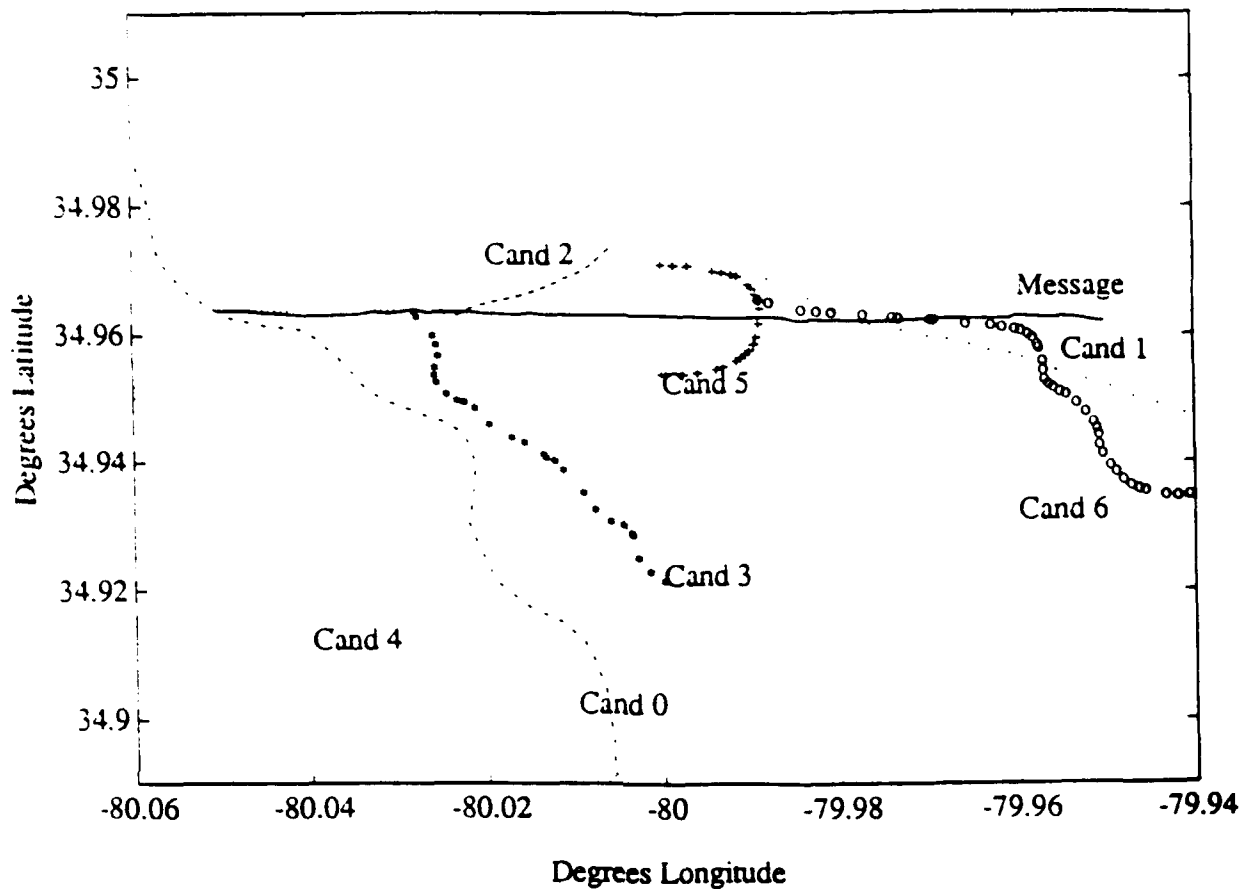
Figure 7. A messages and its corresponding candidates.

All programs for this project were written in C on a SUN SPARCstation IPC. It should be mentioned that for all the 19 messages and 76 candidates the program generated the confidence measures of Table 1 in under 1 second. Smoothing parameters of 0.001 and 5 were used in the PPNN and SGPNN networks, respectively.

Message	Candidate	Confidence Level	Correct
1	1	90%	Yes
	0	6%	
	2	4%	
2	1	78%	Yes
	0	22%	
3	2	62%	No
	0	30%	
	3	5%	
	1	4%	
4	4	45%	Yes
	6	28%	
	0	17%	
	8	3%	
	7	2%	
	3	2%	
	2	2%	
	1	0%	
	5	0%	
	10	0%	
	9	0%	
5	4	76%	Yes
	0	11%	
	5	6%	
	1	3%	
	6	3%	
	2	2%	
	3	0%	
6	1	69%	Yes
	0	30%	
	2	1%	
7	4	27%	Yes
	6	22%	
	5	18%	
	2	11%	
	3	11%	
	0	6%	
	7	4%	
	1	0%	

Table 1-a. The results of the candidate selection on the test messages.

Message	Candidate	Confidence Level	Correct
8	3	82%	Yes
	2	14%	
	0	4%	
	1	0%	
9	2	100%	Yes
	0	0%	
	1	0%	
10	0	48%	No
	2	28%	
	1	23%	
11	No Candidates	Were	Available
12	No Candidates	Were	Available
13	No Candidates	Were	Available
14	0	85%	Yes
	1	15%	
15	2	76%	Yes
	1	16%	
	0	8%	
16	1	86%	Yes
	2	14%	
	0	0%	
17	0	39%	Yes
	6	34%	
	4	19%	
	3	7%	
	2	0%	
	5	0%	
	1	0%	
18	0	100%	Yes
19	0	84%	Yes
	1	16%	

Table 1-b. The results of the candidate selection on the test messages.

IV. CONCLUSION and FUTURE WORK

5.1 Conclusion

In this study, the probabilistic neural network (PNN) architecture was successfully applied to the candidate selection part of the ASU program. At the present time, the candidate selection results are solely based on the proximity and segment orientation. However, the proposed architecture should be tested with more test cases and the other two PNNs should be included for system robustness.

5.2 Future Work

To enhance the candidate selection program further investigation needs to be conducted. The following subject areas can be considered.

- Inclusion of Feature Attributes
- Shape Description
- Hardware Implementation

5.2.1 Inclusion of Feature Attributes

An important enhancement of the proposed work reported here is the inclusion of feature attributes. Additional artificial neural network (ANN) modules can be designed to handle these attributes. Each module can be designed to work internally as an independent component and externally as an integral part of the overall system. The ANN design depends of the complexity and specifications of the feature. Radial basis function (RBF) and probabilistic neural networks (PNN) can be investigated in this phase.

5.2.2 Shape Analysis

Although inclusion of attribute information is important for more accurate identification of the candidates, the information might sometimes be unknown, incomplete, or even wrong to be of any use. The overall system design should therefore be able to handle conflicting scenarios in a self organizing manner. One of the strong means for achieving this objective is object identification through shape analysis. Therefore a shape descriptor algorithm needs to be devised and implemented. The algorithm has to be size and rotation invariant. The design of the shape

descriptor requires a supervised neural network. However, its integration in the overall candidate selection design has to be self-organized and based on an unsupervised network. The radial basis function (RBF) and probabilistic neural network (PNN) can be utilized for this purpose.

5.2.3 Hardware implementation

For real-time application of the candidate selection program a hardware implementation of the proposed ANN methodology is necessary. A commercial hardware platform that has recently been designed for implementation of PNN and RBF networks can be utilized for this investigation. The result of our investigation may lead into development of a portable intelligent hardware to be used by the Air Force.

ACKNOWLEDGMENT

REFERENCES

[1]   Digitizing the Future, Third Edition, Defense Mapping Agency, Fairfax, Va.

[2]   Draft Software Design Document for Auto Source Update: Part I. Prepared for Rome Laboratory Intelligence and Reconnaissance (IRRP) by Grumman Data Systems Corporation, 7 August 1992.

[3]   Draft System/Segment Specification for Auto Source Update. Prepared for Rome Laboratory Intelligence and Reconnaissance (IRRP) by Grumman Data Systems Corporation, 9 January 1992.

[4]   Musavi, M.T., Kalantri, K., Ahmed, W., & Chan, K.H. (1993). A minimum error neural network (MNN), *Neural Networks*, Vol. 6, 397-407.

[5]   Sweeney, W.P., & Musavi, M.T. (1993). Probabilistic Neural Networks as Chromosome Classifier. submitted to *Cytometry*.

[6]   Musavi, M.T., Chan, K.H., Hummels, D.M., Kalantri, K., & Wahid, A. (1992). A probabilistic model for evaluation of neural network classifiers. *Pattern Recognition*, Vol. 25, No. 10, 1241-1251.

[7]   Musavi, M.T., Kalantri, K., & Ahmed, W. (1992). Improving the Performance of Probabilistic Neural Networks. Proc. of Int. Joint Conf. on Neural Networks (IJCNN), Vol. I, pp. 595-600, Baltimore, MD, June 7-11.

[8]   Specht, D.F. (1967). Generation of polynomial discriminant functions for pattern recognition. *IEEE Trans. on Electronic Computer*, vol. EC-16, 308-319.

[9]   Specht, D.F. (1990a). Probabilistic neural networks. *Neural Networks*, vol. 3, 109-118.

[10]   Specht, D.F. (1990b). Probabilistic neural networks and the polynomial adaline as complementary techniques for classification. *IEEE Trans. on Neural Networks*, vol. 1, no. 1, 111-121.

[11]   Maloney, P.S. (1988). An application of probabilistic neural networks to a hull-to-emitter correlation problem. Washington: 6-th Annual Intelligence Community AI Symposium.

[12]   Maloney, P.S., & Specht, D.F. (1989). The use of probabilistic neural networks to improve solution times for hull-to-emitter correlation problems. San Diego: International Joint conference on Neural Networks, I, 289-294.

[13]   Washburne, T.P., Okamura, M.M., Specht, D.F., & Fisher, W.A. (1991). The Lockheed probabilistic neural network processor. Seattle: International Joint conference on Neural Networks, I, 513-518.

APPENDIX

The following is the coordinate information and partial attribute information of a NUF message.

```
Message #4
SCAL = 50000
HAC  = 50.000000
CONF = 90.000000
UTC  = 2
FACS = 1P030
TYPE = L

34.9637 ,-80.0499
34.9638 ,-80.0471
34.9634 ,-80.045
34.9633 , 80.0431
34.9631 ,-80.0413
34.963  ,-80.0398
34.963  ,-80.0384
34.9631 ,-80.037
34.9631 ,-80.0356
34.9632 ,-80.0342
34.9636 ,-80.0326
34.9634 ,-80.0314
34.9634 ,-80.0304
34.9636 ,-80.0294
34.9639 ,-80.0282
34.9638 ,-80.0275
34.9634 ,-80.0264
34.9636 ,-80.0254
34.9636 ,-80.0237
34.9636 ,-80.022
34.9633 ,-80.0202
34.9632 ,-80.0184
34.9631 ,-80.0163
34.9634 ,-80.015
34.9631 ,-80.0136
34.963  ,-80.0067
34.963  ,-79.9996
34.9626 ,-79.9933
34.9627 ,-79.9885
34.9627 ,-79.9861
34.9622 ,-79.9847
34.9623 ,-79.9828
34.9622 ,-79.9803
34.962  ,-79.9755
34.9624 ,-79.969
34.9624 ,-79.9645
34.9624 ,-79.9609
34.9629 ,-79.9602
34.9626 ,-79.9567
34.9627 ,-79.9548
34.9625 ,-79.9531
34.9622 ,-79.9518
34.9619 ,-79.9506
34.962  ,-79.9502
```

TEMPLATE BASIS FOR 1-D

HIGH RANGE RESOLUTION

TARGET IDENTIFICATION

Brian N. Barber

Graduate Student

Department of Electrical Engineering

Washington University

Campus Box 1127

St. Louis, MO 63130

TEMPLATE BASIS FOR 1-D
HIGH RANGE RESOLUTION
TARGET IDENTIFICATION

Brian N. Barber
Graduate Student
Department of Electrical Engineering
Washington University

Abstract

One dimensional high range resolution radar profiles for use as a template set for the identification of non cooperative targets were studied. In particular, three things were accomplished. A literature search was performed to find previous investigations in this area. The Xpatch software routine, developed by Wright Laboratory, was investigated for potential use as a tool for generating template sets. Real, high range resolution data was analyzed for use in further investigations. The results at this time are incomplete, although Xpatch shows much promise as a generator of the high range resolution template sets required. Further work needs to be done.

TEMPLATE BASIS FOR 1-D

HIGH RANGE RESOLUTION

TARGET IDENTIFICATION

Brian N. Barber

Introduction

The use of high resolution radar responses for the identification of non cooperative targets has been investigated by many people over the years. The methods currently published in the open literature have used inverse synthetic aperture radar (ISAR) to generate an image of the target, and use this image to identify the target [1],[2]. This work is part of a project being performed at Washington University in St. Louis, under contract with Rome Laboratory at Griffiss Air Force Base, New York, to develop a similar system of target tracking and identification. This work differs from previous methods in that this system will employ one dimensional profiles for target identification, rather than two dimensional ISAR images. Use of one dimensional rather than two dimensional radar responses will provide a faster target estimate update, since the target can be identified at each pulse interval, provided that the increased template set will take less time to search than the comparable ISAR image generation time. This gives the one dimensional method the potential for real time target tracking and identification.

Methodology

The investigation of the effectiveness of Xpatch as a tool to generate high resolution profiles of the target began with a review of the literature provided with the package. The examples in the literature were a good starting point for generation of profiles appropriate to the problem at hand. The literature search on this area of research indicated that one dimensional target identification was considered too sensitive to aspect angle to provide a useful target identification procedure [1],[2]. This seems to be the key difficulty in using one dimensional techniques, therefore an investigation was conducted using Xpatch's estimated profile sets.

Having chosen an arbitrary scenario of a slant range of 50 nautical miles and an elevation of 30000 feet, range profiles were generated at this elevation angle and an azimuth angle ranging from 0 to 180 degrees, stepping every degree, for a representative target. Having generated this data, correlations between the 90 degree mark (broadside) and the remaining data, as well maximum likelihood tests (assuming additive, white, Gaussian noise), were performed.

Additionally, the rate of change of the response was investigated to determine the resolution of the template set required. This was done by averaging the responses to get an "average" 90 degree response. In other words, symmetric averaging of the data set was done (89 with 91, 88 with 92, etc....). These "average" responses were compared to the 90 degree response in the same manner as previously described (correlation and maximum likelihood). See Appendix B for results.

The code that did both sets of calculations appears in Appendix A.

Results

After running the code to generate the test values for the 0-180 degree range, it was found that a significant improvement occurred in the results for one degree off true for both the straight test and the averages test.

For the straight test of the data, not doing the averaging, anything more than two degrees away on a one degree scale showed as noisy uncorrelated or unlikely results. However, upon getting to within one degree of the true aspect angle, the results stepped roughly halfway to the results achieved for the true value. Similar results appeared for the averages. Upon observing these results, it was decided that perhaps a smaller step size in the vicinity of the true value would be illuminating.

What was hoped to be found was a line of demarcation, where the correlations and/or maximum likelihoods took a large step, so some limit could be placed upon the step size used in generating a template set. To find this line, a step size of 0.05 degrees was used for the range of 88 to 92 degrees. The same tests were done as on the previous range of azimuth angles. Upon examining the results, it was found that both the correlation and maximum likelihood results moved smoothly from the off true results to the true results; no break point was found.

Conclusion

The papers by Steinberg et al. ([1],[2]) indicating that one dimensional target responses are too aspect angle dependent to be of use is perhaps too pessimistic. Given that a rapid means of comparison

between template sets and real responses is possible, perhaps using a massively parallel processor, this same aspect angle dependence could be a positive feature of this technique, providing a more precise estimate as to the spatial orientation of the target. Combining this accuracy of orientation with software that incorporates certain basic facts of flight dynamics (for example, a target cannot change orientation more than by some small amount within a single radar pulse interval) should yield an accurate tracking/identification system.

The results of Xpatch at this time are inconclusive. It seems to be an excellent tool for generating estimated radar responses, though it does have certain limitations. The current version runs at too slow of a rate to generate the responses real time, so a template set is called for. Also, the responses generated by Xpatch have only magnitude, not phase, components, so some potentially valuable information is unused. See Appendix C for a sample range profile generated by Xpatch3.

Having recognized that there is no break point between correct and incorrect orientation, rather there is a smooth transition, the choice of template set size can be based upon many other things. For example, if the tracking algorithm cannot track any better than to within a five degree orientation, it may not be necessary to provide an orientation estimate which is good to within one half of a degree, allowing a savings to be made in terms of processing time and memory required for the template sets.

At the current time, little investigation has been made of the real high range resolution tracks gathered. These tracks were not gathered new, due to equipment problems, but rather were previously

recorded tracks of commercial targets of opportunity. The information

known about these tracks (range, altitude, weather conditions,

polarization, aircraft type, etc...) is such that a facet model could be

generated for use by Xpatch. With this facet model, estimated responses

can be generated and comparisons and tests can be done using the real

track, the template sets generated, and the tracking and identification

algorithm previously mentioned.

References

[1]   B. D. Steinberg, "Target Recognition and Detection Sensitivity In
      Two-Dimensional Microwave Imaging," VFRC Final Report on Contract
      No. E-21-T04-0025 for Georgia Institute of Technology, Atlanta,
      Georgia, Valley Forge Research Center, University of Pennsylvania,
      January 1990.

[2]   B. Kang, B. D. Steinberg, S. B. Kesler, "Comparison of 1-D and 2-D
      Aircraft Images," IEEE International Conference on Systems
      Engineering, August 1989.

[3]   S. W. Lee DEMACO & Dennis J. Andersh, WL/AARA, USAF "An Overview
      of Xpatch," September 16, 1992.

APPENDIX A

```
        program main

c-----------------------------------------------------------
c-----Program to average and correlate results  ------
c-----from xpatch.  Considers different cases of------
c-----normalization and different methods of     ------
c-----comparison.                                ------
c-----Written by Brian Barber                    ------
c-----------------------------------------------------------
        character*3         x1c,x2c
        character*1         x1hc,x1dc,x1oc
        character*11        file1,file2,file3,file4
        character*3         root

        integer             x1i,x2i,mode,temp,temp1
c-----------------------------------------------------------
c-----The following portion pulls the pertainent------
c-----columns of data from the .out3 files       ------
c-----generated by xpatch.                       ------
c-----------------------------------------------------------
        open(5)
        open(6)
        do 100 x1i=0,180
          temp1=x1i/100
          write(x1hc,901)temp1
          temp=x1i-temp1*100
          temp1=temp/10
          write(x1dc,901)temp1
          temp=temp - temp1*10
          write(x1oc,901)temp
          file1='d.'//x1hc//x1dc//x1oc
          file2='ds.'//x1hc//x1dc//x1oc
          file3='dn.'//x1hc//x1dc//x1oc
          call reformat(file1,file2)
          call reformat(file1,file3)
          call normalize(file3)
100     continue
c-----------------------------------------------------------
c-----The next portion averages the source files------
c-----generated in the previous portion of code.------
c-----------------------------------------------------------
        do 110 x1i=0,90
          x2i=180-x1i
          temp=x1i/10
          write(x1dc,901)temp
          temp=x1i - temp*10
          write(x1oc,901)temp
          x1c='0'//x1dc//x1oc
          temp1=x2i/100
          write(x1hc,901)temp1
```

```
            temp=x2i-temp1*100
            temp1=temp/10
            write(x1dc,901)temp1
            temp=temp - temp1*10
            write(x1oc,901)temp
            x2c=x1hc//x1dc//x1oc
            file1='ds.'//x1c
            file2='ds.'//x2c
            file3='as.'//x1c//'.'//x2c
            call average(file1,file2,file3)
            file1='asn.'//x1c//'.'//x2c
            call copy(file3,file1)
            call normalize(file1)
            file1='dn.'//x1c
            file2='dn.'//x2c
            file3='an.'//x1c//'.'//x2c
            call average(file1,file2,file3)
            file1='ann.'//x1c//'.'//x2c
            call copy(file3,file1)
            call normalize(file1)
110     continue
c----------------------------------------------------
c-----The remaining portion of the code does the-----
c-----numerous correlations on the data sets    -----
c-----previously generated.                     -----
c----------------------------------------------------
        mode=1
        file1='ds.090'
        root='ds.'
        file3='cords'
        file4='lklds'
        call correlate(file1,root,file3,mode)
        call likelihood(file1,root,file4,mode)
        root='dn.'
        file3='cordn'
        file4='lkldn'
        call correlate(file1,root,file3,mode)
        call likelihood(file1,root,file4,mode)
        mode=2
        root='as.'
        file3='coras'
        file4='lklas'
        call correlate(file1,root,file3,mode)
        call likelihood(file1,root,file4,mode)
        root='an.'
        file3='coran'
        file4='lklan'
        call correlate(file1,root,file3,mode)
        call likelihood(file1,root,file4,mode)
        root='sn.'
        file3='corasn'
        file4='lklasn'
```

```
          call correlate(file1,root,file3,mode)
          call likelihood(file1,root,file4,mode)
          root='nn.'
          file3='corann'
          file4='lklann'
          call correlate(file1,root,file3,mode)
          call likelihood(file1,root,file4,mode)
          close(6)
900       format(I2)
901       format(I1)
          end
c----------------------------------------------------------

          subroutine reformat(infile,outfile)

c----------------------------------------------------------
c-----Routine to strip all but the range and      -----
c-----linear response information from the         ------
c-----original .out3 files created by xpatch.     ------
c----------------------------------------------------------
          character*11 infile, outfile
          real*8 info(256,6)
          integer x1,x2
          open(10,file=infile,status='old',err=800)
          do 120 x1=1,16
          read(10,55)
120       continue
          open(11,file=outfile,status='new',err=801)
          do 130 x1=1,256
          read(10,*)(info(x1,x2),x2=1,6)
          write(11,902)info(x1,1),info(x1,3),info(x1,5)
130       continue
          close(10)
          close(11)
          return
800       write(6,*)'Error in opening original .out3 = ',infile
          stop
801       write(6,*)'Error in opening reformated file ',outfile
          stop
55        format(A55)
902       format(2x,3(f15.5,2x))
          end
c----------------------------------------------------------

          subroutine normalize(filex)

c----------------------------------------------------------
c-----Routine to normalize a datafile in the      ------
c-----format range--vert(lin)--hori(lin).        ------
c----------------------------------------------------------
          character*11 filex
          real*8 info(256,3),temp1,temp2
```

```fortran
      integer x1,x2
      open(12,file=filex,status='old',err=815)
      do 150 x1=1,256
      read(12,*)(info(x1,x2),x2=1,3)
150   continue
      close(12)
      temp1=0
      temp2=0
      do 160 x1=1,256
      temp1=temp1+info(x1,2)**2
      temp2=temp2+info(x1,3)**2
160   continue
      temp1=temp1**0.5
      temp2=temp2**0.5
      open(13,file=filex,status='delete')
      open(13,file=filex,status='new',err=802)
      do 170 x1=1,256
      info(x1,2)=info(x1,2)/temp1
      info(x1,3)=info(x1,3)/temp2
      write(13,903)info(x1,1),info(x1,2),info(x1,3)
170   continue
      close(13)
      return
802   write(6,*)'Error in reopening ',filex,' when writing new values'
      stop
815   write(6,*)'Error in opening ',filex,' to find normal values'
      stop
903   format(2x,3(f15.5,2x))
      end
c------------------------------------------------------

      subroutine copy(filein,fileout)

c------------------------------------------------------
c------Routine to duplicate a datafile in the    ------
c------format --range--vert(lin)--hori(lin).     ------
c------------------------------------------------------
      character*11 filein, fileout
      real*8 info(256,3)
      integer x1,x2
      open(14,file=filein,status='old',err=803)
      open(15,file=fileout,status='new',err=804)
      do 180 x1=1,256
      read(14,*)(info(x1,x2),x2=1,3)
      write(15,905)info(x1,1),info(x1,2),info(x1,3)
180   continue
      close(14)
      close(15)
      return
803   write(6,*)'Error in opening source file ',filein
      stop
804   write(6,*)'Error in opening duplicate file ',fileout
```

```
      stop
905   format(2x,3(f15.5,2x))
      end
c─────────────────────────────────────────────

      subroutine average(in1,in2,mean)

c─────────────────────────────────────────────
c──────Routine to generate the averages of .out3 ──────
c──────files for comparison with the result of   ──────
c──────calculating the values using xpatch.       ──────
c─────────────────────────────────────────────
      character*11 in1,in2,mean
      integer x1,x2
      real*8 d1(256,3),d2(256,3),dav(256,3)
      if (in1 .EQ. in2) then
          open(16,file=in1,status='old',err=805)
          do 190 x1=1,256
          read(16,*)(d1(x1,x2),x2=1,3)
          do 200 x2=1,3
          d2(x1,x2)=d1(x1,x2)
200       continue
190       continue
          close(16)
          goto 210
      else
          open(16,file=in1,status='old',err=806)
          open(17,file=in2,status='old',err=807)
          do 220 x1=1,256
          read(16,*)(d1(x1,x2),x2=1,3)
          read(17,*)(d2(x1,x2),x2=1,3)
220       continue
          close(16)
          close(17)
      endif
210   open(18,file=mean,status='new',err=808)
      do 230 x1=1,256
      dav(x1,1)=(d1(x1,1)+d2(x1,1))/2.0
      dav(x1,2)=(d1(x1,2)+d2(x1,2))/2.0
      dav(x1,3)=(d1(x1,3)+d2(x1,3))/2.0
      write(18,906)dav(x1,1),dav(x1,2),dav(x1,3)
230   continue
      close(18)
      return
805   write(6,*)'Error in opening repeated file to avg',in1
      stop
806   write(6,*)'Error in opening file1 to avg ',in1
      stop
807   write(6,*)'Error in opening file2 to avg ',in2
      stop
808   write(6,*)'Error in opening average output file ',mean
      stop
```

```
906     format(2x,3(f15.5,2x))
        end
c--------------------------------------------------------------

        subroutine correlate(fin1,fin2root,fout,mode)

c--------------------------------------------------------------
c------Routine to do correlation between fin1 and------
c------the set of files initially labeled with    ------
c------fin2root.  Mode indicates if there are one------
c------or two extensions.                          ------
c--------------------------------------------------------------
        character*11 fin1,fin2,fout
        character*3 fin2root
        character*3 xc1,xc2
        character*1 x1hc,x1dc,x1oc
        integer mode,x1,x2,x3,temp,temp1
        real*8 din1(256,3),din2(256,3),cor1,cor2
        open(19,file=fin1,status='old',err=809)
        do 240 x1=1,256
        read(19,*)(din1(x1,x2),x2=1,3)
240     continue
        close(19)
        open(21,file=fout,status='new',err=810)
        do 250 x1=0,180
        if (mode .EQ. 1) then
            temp1=x1/100
            write(x1hc,911)temp1
            temp=x1-temp1*100
            temp1=temp/10
            write(x1dc,911)temp1
            temp=temp - temp1*10
            write(x1oc,911)temp
            xc1=x1hc//x1dc//x1oc
            fin2=fin2root//xc1
        else
            x2=180-x1
            if (x1.GT.x2) then
                goto 250
            endif
            temp=x1/10
            write(x1dc,911)temp
            temp=x1 - temp*10
            write(x1oc,911)temp
            xc1='0'//x1dc//x1oc
            temp1=x2/100
            write(x1hc,911)temp1
            temp=x2-temp1*100
            temp1=temp/10
            write(x1dc,911)temp1
            temp=temp - temp1*10
            write(x1oc,911)temp
```

```fortran
          xcl=xlhc//xldc//xloc
          fin2=fin2root//xcl
      else
          x2=180-x1
          if (x1.GT.x2) then
              goto 250
          endif
          temp=x1/10
          write(xldc,911)temp
          temp=x1 - temp*10
          write(xloc,911)temp
          xcl='0'//xldc//xloc
          temp1=x2/100
          write(xlhc,911)temp1
          temp=x2-temp1*100
          temp1=temp/10
          write(xldc,911)temp1
          temp=temp - temp1*10
          write(xloc,911)temp
          xc2=xlhc//xldc//xloc
          if (fin2root.EQ.'sn.' .OR. fin2root.EQ.'nn.') then
              fin2='a'//fin2root//xcl//'.'//xc2
          else
              fin2=fin2root//xcl//'.'//xc2
          endif
      endif
      open(30,file=fin2,status='old',err=811)
      do 260 x2=1,256
      read(30,*)(din2(x2,x3),x3=1,3)
260   continue
      close(30)
      cor1=0.0
      cor2=0.0
      do 270 x2=1,256
      cor1=cor1+din1(x2,2)*din2(x2,2)
      cor2=cor2+din1(x2,3)*din2(x2,3)
270   continue
      write(21,908)cor1,cor2
250   continue
      close(21)
      return
809   write(6,*)'Error in opening base correlation file'
      stop
810   write(6,*)'Error in opening correlation output file'
      stop
811   write(6,*)'Error in opening one of the sequence of'
      write(6,*)'correlation files.  Name = ',fin2
      stop
907   format(I2)
908   format(2x,2(f15.5,2x))
911   format(I1)
      end
```

```
c-----------------------------------------------

      subroutine likelihood(fl,f2root,fo,m)

c-----------------------------------------------
c------Routine to do the maximum likelihood      ------
c------comparison  between fl and the set of     ------
c------files initially labeled with f2root.  Mode------
c------indicates if there are one or two         ------
c------extensions.  Since only looking at the    ------
c------exponent portion of the likelihood (with  ------
c------AWGN), actually looking for the smallest   ------
c------value when looking at the results.        ------
c-----------------------------------------------
      character*11 fl,f2,fo
      character*3 f2root
      integer m,x1,x2,x3,temp,templ
      real*8 dl(256,3),d2(256,3),max1,max2
      character*3 xc1,xc2
      character*1 xlhc,xloc,xldc
      open(22,file=fl,status='old',err=812)
      do 280 x1=1,256
      read(22,*)(dl(x1,x2),x2=1,3)
280   continue
      close(22)
      open(23,file=fo,status='new',err=813)
      do 290 x1=0,180
      if (m.EQ.1) then
          templ=x1/100
          write(xlhc,912)templ
          temp=x1-templ*100
          templ=temp/10
          write(xldc,912)templ
          temp=temp - templ*10
          write(xloc,912)temp
          xc1=xlhc//xldc//xloc
          f2=f2root//xc1
      else
          x2=180-x1
          if (x1.GT.x2) then
              goto 290
          endif
          temp=x1/10
          write(xldc,912)temp
          temp=x1 - temp*10
          write(xloc,912)temp
          xc1='0'//xldc//xloc
          templ=x2/100
          write(xlhc,912)templ
          temp=x2-templ*100
          templ=temp/10
          write(xldc,912)templ
```

```
                temp=temp - temp1*10
                write(xloc,912)temp
                xc2=xlhc//xldc//xloc
                if (f2root.EQ.'sn.' .OR. f2root.EQ.'nn.') then
                    f2='a'//f2root//xc1//'.'//xc2
                else
                    f2=f2root//xc1//'.'//xc2
                endif
            endif
            open(24,file=f2,status='old',err=814)
            do 300 x2=1,256
            read(24,*)(d2(x2,x3),x3=1,3)
300         continue
            close(24)
            max1=0.0
            max2=0.0
            do 310 x2=1,256
            max1=max1+(d1(x2,2)-d2(x2,2))**2
            max2=max2+(d1(x2,3)-d2(x2,3))**2
310         continue
            write(23,910)max1,max2
290         continue
            close(23)
            return
812         write(6,*)'Error in opening base likelihood file',f1
            stop
813         write(6,*)'Error in opening likelihood output file',fo
            stop
814         write(6,*)'Error in opening one of the sequence of'
            write(6,*)'likelihood files.  Name = ',f2
            stop
909         format(I2)
910         format(2x,2(f15.5,2x))
912         format(I1)
            end
```
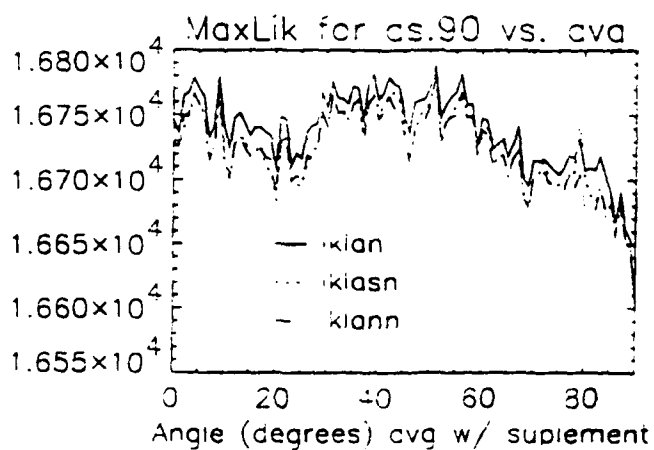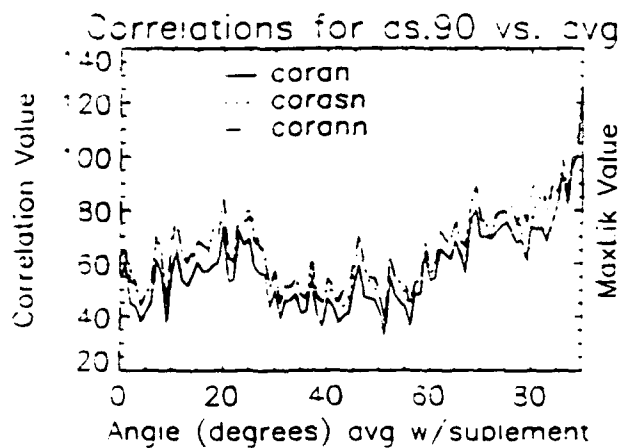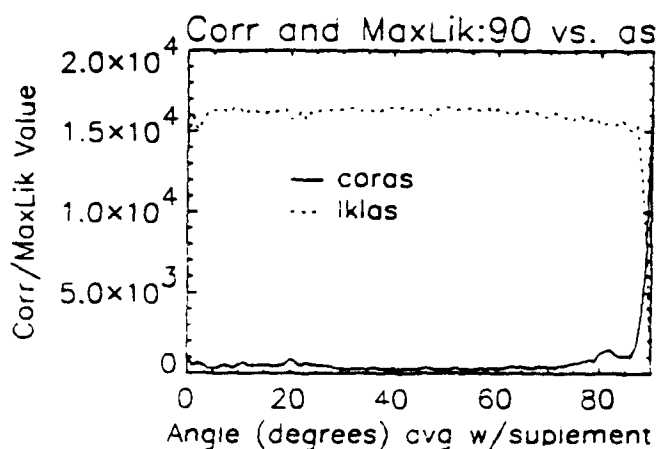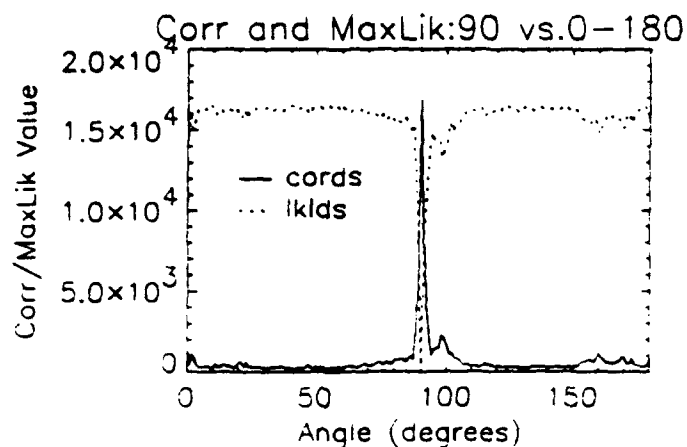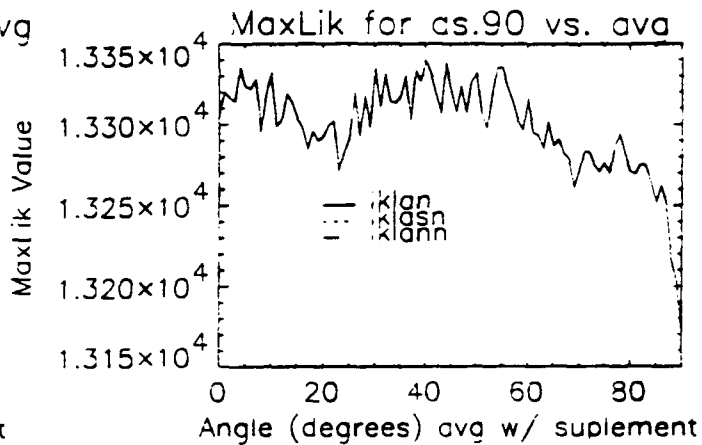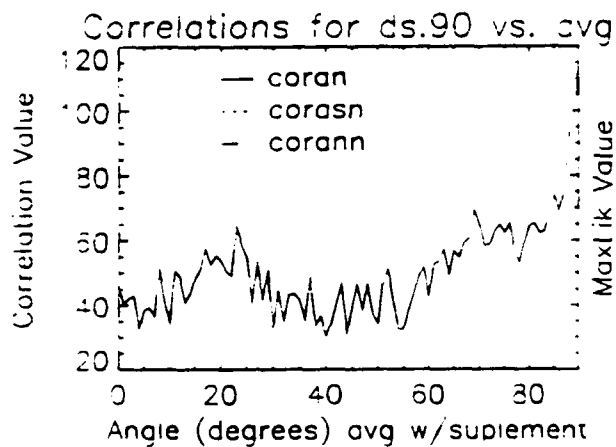---

Format of File Labels

d.XXX	.out3 files from Xpatch3
ds.XXX	range and linear E-field responses
dn.XXX	normalized responses
as.XXX.YYY	response made by averaging angle XXX results with YYY results, no normalization
asn.XXX.YYY	same as as.XXX.YYY except normalized
an.XXX.YYY	same as as.XXX.YYY except average done using normalized values
ann.XXX.YYY	same as an.XXX.YYY except renormalized
cor*	correlation of ds.(center angle), not normalized with the appropriate set *
lkl*	same as cor* except maximum likelihood, assuming addative white Gaussian Noise

APPENDIX B.1: VERTICAL POLARIZATION RESPONSES

Corr and MaxLik:90 vs.0−180

Corr and MaxLik:90 vs. as

Correlations for ds.90 vs. avg

MaxLik for as.90 vs. ava
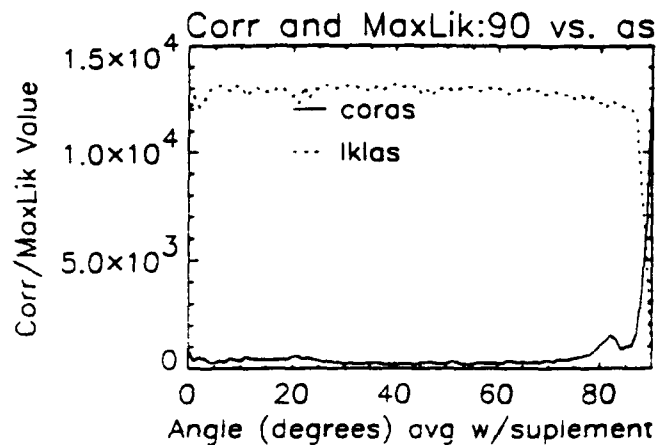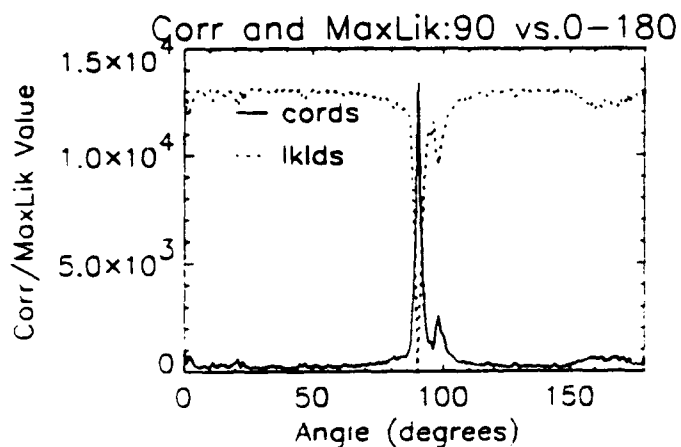
APPENDIX C: SAMPLE RESULTS FROM XPATCH3

Sharper dBsm response due to horizontal polarization, smoother response due to vertical polarization.
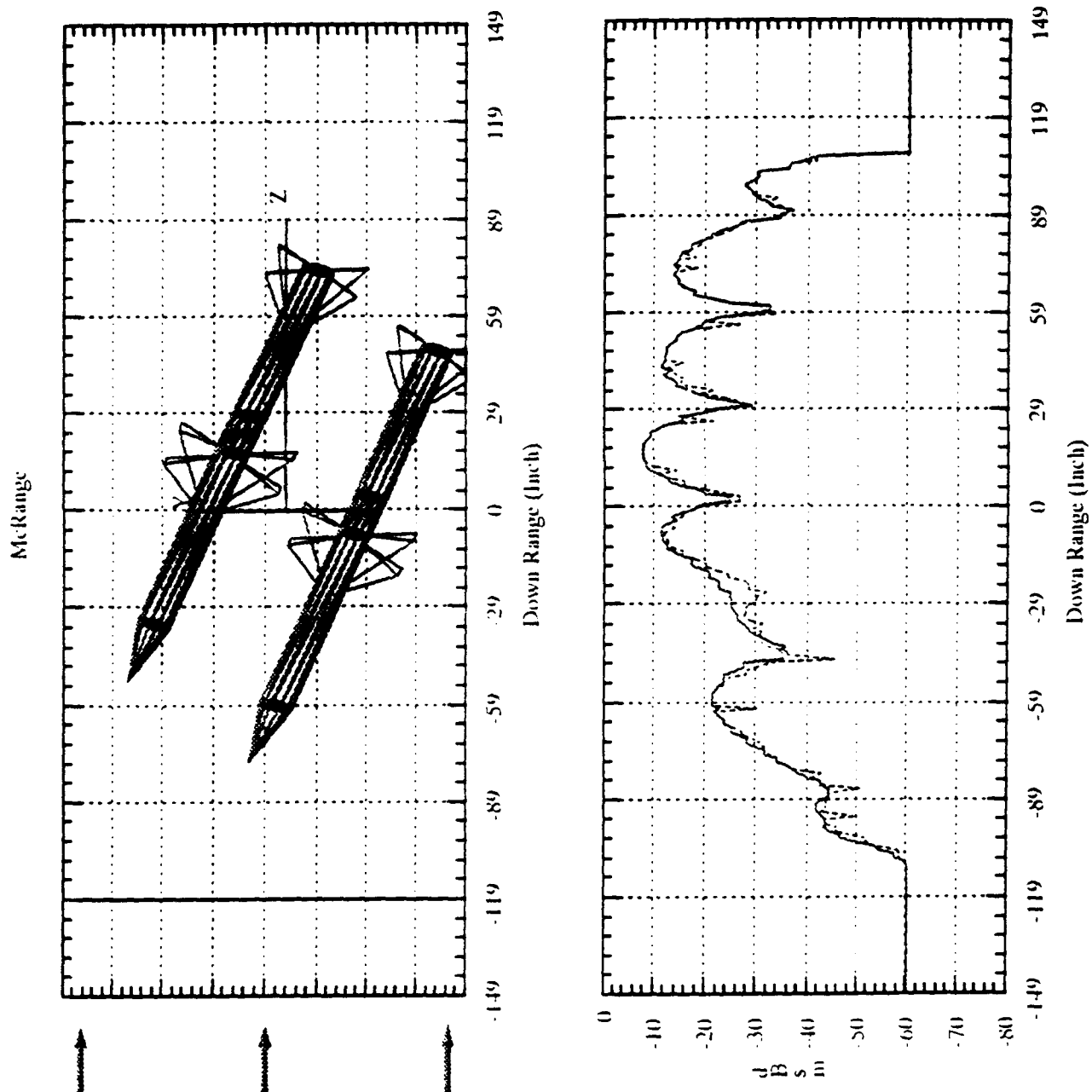
HIGH SPEED OPTICAL DETECTORS FOR MONOLITHIC MILLIMETER WAVE INTEGRATED CIRCUITS

Kerry L Litvin
School of Electrical Engineering

Cornell University
Phillips Hall
Ithaca, NY 14850

Final Report For:
Summer Research Extension Program
Rome Laboratory, Griffiss AFB

September 1993

High Speed Optical Detectors for Monolithic Millimeter Wave Integrated Circuits

Kerry Litvin
School of Electrical Engineering, Cornell University
Ithaca, New York 14853

ABSTRACT

Metal-semiconductor-metal photo diodes with interdigitated Schottky barrier fingers are being developed for applications in monolithic optical receiver circuits with the purpose of detecting millimeter wave modulation signals being transmitted via an optical carrier. The devices are planar and incorporate submicron finger spacings and a thin absorption region for speed with a buried stack of tuned Bragg reflectors for enhanced sensitivity at the carrier wavelength. These devices are being integrated with short-gate MODFET amplifiers to form the complete monolithic integrated optical receiver circuit.

High Speed Optical Detectors for Monolithic Millimeter Wave Integrated Circuits

Kerry Litvin

DEVICE DESCRIPTION

Metal-semiconductor-metal photo diodes with interdigitated Schottky barrier fingers are being fabricated in the AlGaAs material system using electron beam lithography techniques. Both linear and circular detector finger geometries have been examined (Fig. 1). Such photo detectors are capable of detection bandwidths in excess of 100 GHz [1-3]. The devices have a planar construction making them ideal candidates for monolithic integration [4,5]. The detector structures were designed with the aid of a computer program and then grown by molecular beam epitaxy. The photo diodes in this design incorporate sub-micron electrode spacings and a thin absorption layer of GaAs, these features limit the transit time of the optically generated carriers thus providing the mechanism for millimeter frequency operation (Fig. 2).

Device sensitivity to low intensity light is enhanced with the inclusion of a highly reflective (94%) stack of tuned Bragg reflectors buried below the absorption layer. The Bragg reflector stack is composed of nonabsorbing alternating quarter wavelength layers of $Al_{0.9}Ga_{0.1}As$ and $Al_{0.35}Ga_{0.65}As$.

The reflective stack causes incident photons, of the appropriate wavelength, to make a second pass through the absorption layer thus creating additional carriers for collection at the electrodes. Additional sensitivity refinements were made by including a top surface optical impedance matching layer of $Al_{0.35}Ga_{0.65}As$. This layer reduces top surface reflectance to nearly negligible values (0.5%). The interfaces between the GaAs absorption layer and the matching layer above it and the bottom of the absorption layer and the top surface of the Bragg reflector stack also form a potential energy well for the photo generated carriers. This is due to the energy band discontinuity of the material layers. The potential energy well serves to confine the photo generated carriers to the thin absorption region where they can be rapidly collected by the electrodes on the top surface. (Also prevents carrier loss by rapid surface recombination (v_s @ $3x10^6$ cm/sec)). The combined effects of the finite line width of the Bragg reflector stack, absorption layer and top surface matching layer introduces some wavelength discrimination capability into the device because there is an increased optical sensitivity at the design wavelength and a decreased sensitivity at other wavelengths away from this line (Fig. 3).

MODEL and MEASUREMENT RESULTS

The topology of a microwave equivalent circuit model for the detector diodes was determined using S-parameter measurements of the devices and of associated short circuited test structures (Figs. 4 & 5). The S-parameter

measurements on the short circuit test structures readily allow for the extraction of the finger electrode resistance and inductance. In addition to the dc resistance in the fingers there is an additional frequency dependent skin effect resistance . The skin effect resistance is proportional to the square root of the frequency. The proportionality constant can be determined by comparing the frequency dependence of the measured device input impedance with that of the model. The MSM detector capacitance is extracted from the device measurements and the excess parasitic capacitance is determined by comparing the measured capacitance to that calculated for an MSM structure using the elliptic integral method [6]. The parasitic capacitance is the result of the finite length of the interdigitated fingers and the close proximity of the contact pad of opposite polarity to the open ends of the fingers. The equivalent circuit model is used to determine the passive element frequency response and time domain impulse function (Fig. 6) .

A Monte Carlo simulation of the transport behavior of the optically generated carriers, which takes into account the material dimensions, electrode geometry, bias conditions ,and optical absorption properties, is used to produce modeling information about the time domain response of the carriers [7,8]. The FFT of the Monte Carlo results provides the carrier frequency response function (Fig. 7). This information is combined with the circuit model information (multiplied in the frequency domain or convolved

in the time domain) to give the composite frequency response and impulse function of the MSM detector (Fig. 8).

Static optical tests were done with a white light source and monochromator in order to characterize the spectral response of the diodes. The response signal was normalized to the white light spectrum (Fig. 9). A large size detector was used in the experiment for better sensitivity. Figure 9 shows that the expected peak in the response for a 800 nm wafer design was down-shifted slightly, this probably indicates that the angle of incidence was not quite normal to the sample. Photospectrometer measurements of the wafer reflectance support this conclusion.

Optical measurements using a short pulse (120 $\tilde{i}$s) mode locked laser operating at 780 nm with a beam diameter of 10 μm indicated a carrier transit time limited bandwidth in excess of 40 GHz (this was the limit of the measurement system). The detector response was observed using a HP2782 spectrum analyzer. The measurement on a detector diode with 23 μm long fingers, 0.5 μm finger width and 0.5 μm gap between the fingers fabricated on a GaAs wafer designed for 750 nm light had the following results: At 7 volts applied bias and 80 mW average incident power, it showed a flat response up to at least 40 GHz. When the average incident light power was doubled to 160 mW the response at 39 GHz was -6 dB down due to excess carrier generation. The dynamic range of the detectors was then measured using a Ti:sapphire laser operating at 770 nm with a variety of neutral density filters . The dynamic range was found to be 33 dB.

Optical measurements employing a semiconductor mode locked laser as a short pulse optical source were also performed and the detector response was observed in the time domain using a sampling oscilloscope. The full width half maximum of the output pulse was 15 ps. This value was again essentially the limit of the measurement system's capability.

DEVICE INTEGRATION

The MSM photodiodes are being integrated with short-gate MODFETs for use as monolithic integrated optical receiver circuits capable of detecting a 44 GHz modulation signal being transmitted via an optical carrier [9]. The transistor layers are grown on top of the detector layers in the MBE system. A slow etch-rate citric acid solution is used for the transistor mesa isolation. The wafer top surface reflectance is monitored with a photospectrometer until the reflectance minima shifts to the design wavelength. The MODFETs will be used as low noise amplification devices in the optical receiver circuits. The current transistor design has a source-drain spacing of 2 μm with a gate length of 0.25 μm and gate width of 100 μm. We are in the process of developing a transistor with a gamma-gate and self-aligned ohmic metalization in order to reduce the source resistance and therefore improve the noise and frequency characteristics of the devices [10,11]. In test structures we have achieved gates with 0.25 μm footprint and 0.6 μm cross-section along the top. The gate had an overhang ratio of more

then 2:1 and a source-drain spacing of 0.6 μm. With self-aligned metalization the source-gate spacing was 0.1 μm and the drain-gate spacing was 0.25 μm (Fig. 10). Coplanar waveguide interconnections are being used in the circuit design.

TEST and MEASUREMENT PROCEDURE

As alluded to earlier in the text, the microwave equivalent circuit element model must be extracted from measurements of the device s-parameters. This section briefly reviews the mechanics of this technique.

When the size of the system or device being investigated is comparable to the wavelength of the electromagnetic signal being processed it becomes convenient to measure the characteristics of the electromagnetic waves rather then measure voltages or currents directly. This is the case at microwave frequencies. The electromagnetic wave properties are described

in terms of transmission and reflection coefficients. The transmission and reflection coefficients are determined by the system or device under test, they therefore provide a means of complete characterization of the system or device being measured. The coefficients are the system s-parameters and all other network parameters (z, y, h, and T for example) can be determined from them provided the system normalizing impedance is known. In microwave systems the normalizing impedance is usually chosen to be 50

ohms. In free space systems 377 ohms may be used as a matter of convenience. In this discussion the system normalizing impedance is always taken to be 50 ohms. For a given network (device or system) with n electrical ports there are n^2 s-parameters. Thus a one-port device has only one s-parameter, the input reflection coefficient. The input reflection coefficient is determined by the device's input impedance and system normalizing impedance. A one-port does not have a transmission coefficient because the input wave and output wave both propagate through the same port and the output wave is therefore the reflected wave. A two-port device has four s-parameters two transmission coefficients (transmission from port one to port two and transmission from port two to port one) and two reflection coefficients (reflection at port one and reflection at port two).

The s-parameters are described mathematically as follows:

1. (for a one-port device)

$$S_{11} = \frac{b_1}{a_1}$$

2. (for a two-port device)

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \bullet \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

3. and

$$S_{11} = \frac{b_1}{a_1}\bigg|_{a_2=0} \qquad S_{12} = \frac{b_1}{a_2}\bigg|_{a_1=0}$$

$$S_{21} = \frac{b_2}{a_1}\bigg|_{a_2=0} \qquad\qquad S_{22} = \frac{b_2}{a_2}\bigg|_{a_1=0}$$

Where a_1 and a_2 are the input waves at ports one and two, b_1 and b_2 are the output waves at ports one and two. The s-parameters can easily be converted to other network parameters.

4. for one-port conversion from s to z or y parameters

$$Z_{11} = Z_o\left(\frac{1+S_{11}}{1-S_{11}}\right) \text{ input impedance (z-parameter)}$$

$$Y_{11} = \frac{1}{Z_o}\left(\frac{1-S_{11}}{1+S_{11}}\right) \text{ input admittance (y-parameter)}$$

5. for two-port conversion from s to z (impedance) parameters

$$Z_{11} = Z_o\left(\frac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}}\right) \qquad Z_{12} = Z_o\left(\frac{2S_{12}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}}\right)$$

$$Z_{21} = Z_o\left(\frac{2S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}}\right) \qquad Z_{22} = Z_o\left(\frac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}}\right)$$

6. for two-port conversion from s to y (admittance) parameters

$$Y_{11} = \frac{1}{Z_o}\left(\frac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}\right) \qquad Y_{12} = \frac{1}{Z_o}\left(\frac{-2S_{12}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}\right)$$

$$Y_{21} = \frac{1}{Z_o}\left(\frac{-2S_{21}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}\right) \qquad Y_{22} = \frac{1}{Z_o}\left(\frac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}\right)$$

Where Z_0 is the system normalizing impedance (50 ohms). Using equations 4, 5 and 6 along with some insight regarding the device physics it is possible to develop an equivalent circuit element model of the structure being studied.

The actual measurement of the s-parameters is carried out using a

microwave vector network analyzer. The network analyzer measures the magnitude and phase of the s-parameters by applying an input wave at each port of the device and then measuring what comes out of the other ports (transmission) and what is reflected back from the input port. The transmitted and reflected waves are divided by the input wave to produce the s-parameters. Before a measurement can take place the network analyzer must be calibrated. Calibration is accomplished by measuring several structures with known s-parameters (calibration standards) and comparing the measured results with the expected values in order to determine error coefficients which are then corrected for in the actual measurements. The network analyzer uses internal software routines to carry out the calibration corrections, these algorithms are designed to provide maximum accuracy. It is important to note that the corrections for on-wafer device probing with

coplanar waveguide probes are NOT the same as those used for coaxial measurements they require different calibration software.

A typical calibration standard set consists of the load (50 ohm standard), the short circuit, the open circuit, and the through line of known length. A different standards set might be used depending upon the calibration algorithm that is available, for example, a set of transmission lines of variable length can also be used for calibration with some methods. The calibration corrections are stored in the network analyzer's memory and applied automatically to all measurements that follow. With t' ə passing of

time the calibration may drift do to changes in conditions such as warming up of the instrument or mechanical adjustments made to the probes, cables, or connectors. A calibration should be performed at least once a day when the instrument is in use. More frequent calibrations will be needed if measurement inaccuracies are noticed with time. It is recommended that the instrument be allowed to warm up for at least two hours before a calibration and measurements are performed. It is therefore recommended that the instrument not be turned off during periods of frequent use in order to ensure maximum stability during the measurements. Finally, the network analyzer has built in bias tees which allow active devices to be biased during microwave measurements. The DC bias is brought in through two ports in the rear of the instrument. Be sure not to exceed the maximum DC power and voltage limits of the bias tees.

SUMMARY

Metal-semiconductor-metal photodetector diodes have been designed with the aid of a computer program. The material was grown by MBE and the devices were fabricated with electron beam lithography. The detectors incorporate features for high speed operation, (submicron electrode spacings, thin absorption layer, and a potential well to confine optically generated carriers to the absorption layer), with features to enhance the device sensitivity, (buried Bragg reflector stack, and a top surface optical impedance matching layer).

Device measurements indicate top surface reflectance of 0.5% with an operation bandwidth in excess of 40 GHz and a dynamic range of 33 dB. The devices are being integrated with short-gate MODFETS for applications in monolithic millimeter wave optical receiver circuits.

ACKNOWLEDGMENTS

REFERENCES

[1] B. J. Van Zeghbroeck, W. Patrick, J. Halbout, and P. Vettiger, "105-GHz Bandwidth Metal-Semiconductor-Metal Photodiode," IEEE Electron Device Lett. , vol. 9, pp. 527-529, 1988.

[2] S.Y. Chou, Y.Liu, W.Khalil, T.Y. Hsiang, and S. Alexanddrou, "Ultrafast nanoscale metal-semiconductor-metal photodetectors on bulk and low temperature grown GaAs," Appl. Phys . Lett., vol 61, no. 7,

pp. 819-821, 17 Aug. 1992.

[3]   S.Y. Chou, Y. Liu, and P.B. Fischer,"Tera-hertz GaAs metal-
semiconductor metal photodetectors wiyh 25 nm finger spacing and
finger width," Appl. Phys. Lett., vol. 61, pp. 477-479, 1992.

[4]   S. J. Wojtczuk, J. M. Ballantyne, S. Wanuga, and Y. K. Chen,
"Comparative study of easily integrable photodetectors," IEEE J.
Lightwave Technol., vol. LT-5, pp. 1365-1370, 1987.

[5]   D. L. Rogers, "Integrated optical receiver using MSM detectors," in
Proc. LEOS '90.

[6]   Y. C. Lim, and R. A. Moore, "Properties of Alternately Charged
Coplanar Parallel Strips by Conformal Mappings," IEEE Transactions
on Electron Devices, vol ED-15, pp. 173-180, March 1968.

[7]   L. Reggiani (ed), *Hot-Electron Transport in Semiconductors* ,
Springer-Verlag, 1985

[8]   J. B. D. Soole,and H. Schumacher, " Transit Time Limited Frequency
Response of InGaAs MSM-Photodetectors," IEEE Tranactions on
Electron Devices, vol ED-37, p. 2285, 1990

[9]   L. D. Nguyen, P. J. Tasker, D. C. Radulescu, and L. F. Eastman, "Design, Fabrication, and Characterization of Ultra High Speed AlGaAs/InGaAs MODFET's," IEDM Tech. Digest, pp. 176-179, 1988.

[10] L.D. Nguyen, A.S. Brown, M.A. Thompson, L.M. Jelloian, L.E. Larson and M. Matloubian, " 650 A Self-Aligned-Gate Pseudomorphic AlInAs/GaInAs High Electron Mobility Transistors," IEEE Electron Device Letters, vol 13, no. 3, pp 143-145, March 1992.

[11] R.C. Tiberio, J.M. Limber, G.J. Galvin, and E.D. Wolf, "Electron Beam Lithography and Resist Processing for the Fabrication of T-Gate Structures", SPIE Vol. 1089, Electron-Beam, X-ray and Ion Beam Technology: Submicrometer Lithographies VIII, pp. 124-131, 1989.
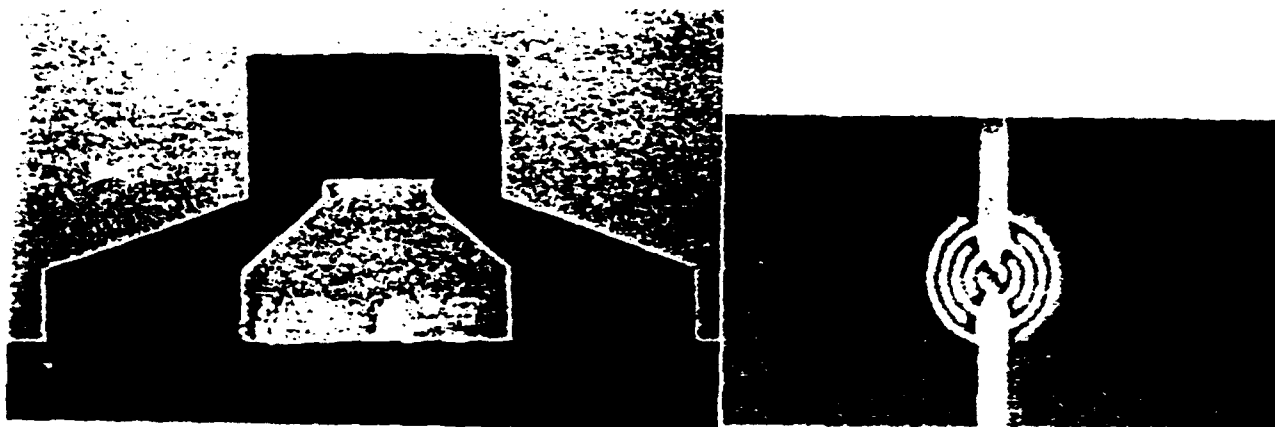
Figure 1. Two MSM Photodiode diode geometries. The linear device has 0.35μm fingers and gaps with a 20μm overlap. The circular device has 0.5μm fingers and gaps and a 10μm diameter.
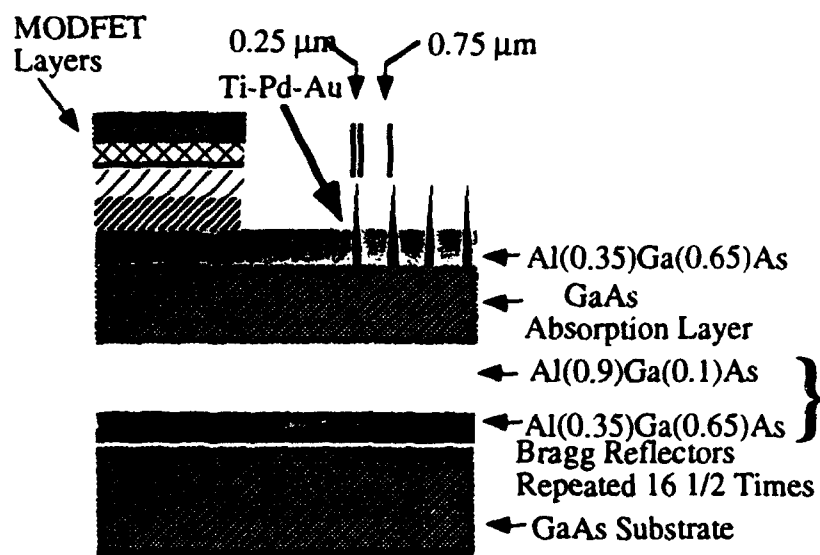
Figure 2. Cross-sectional view of the MSM photodiode material structure showing the top surface impedance matching layer, absorption layer, and Bragg reflector stack.
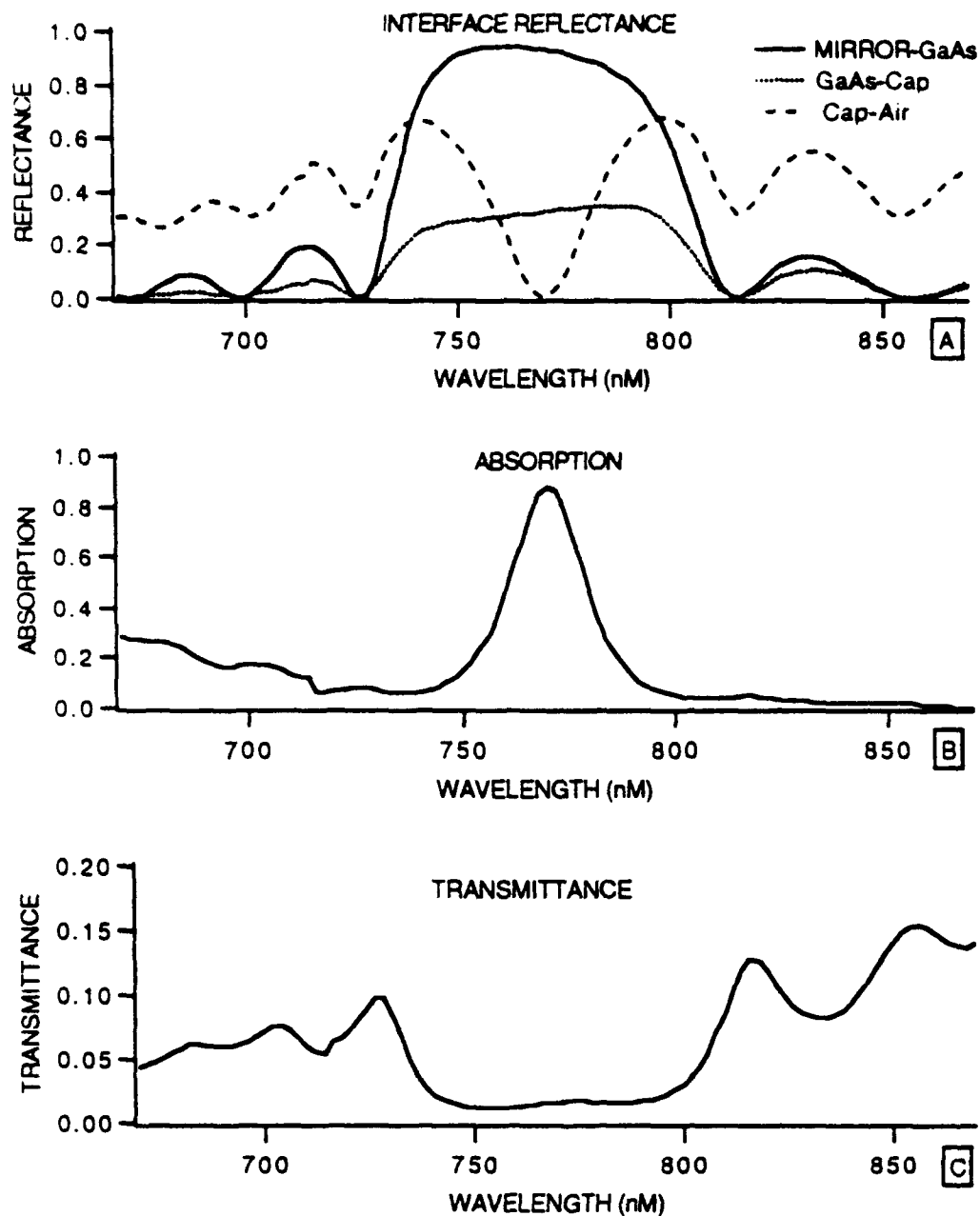
Figure 3. Calculated optical response of the detector material layers for a 770 nm light design: A). reflectance at important interfaces of the structure. B). fraction of incoming light absorbed. C). fraction of incoming light transmitted into the substrate.
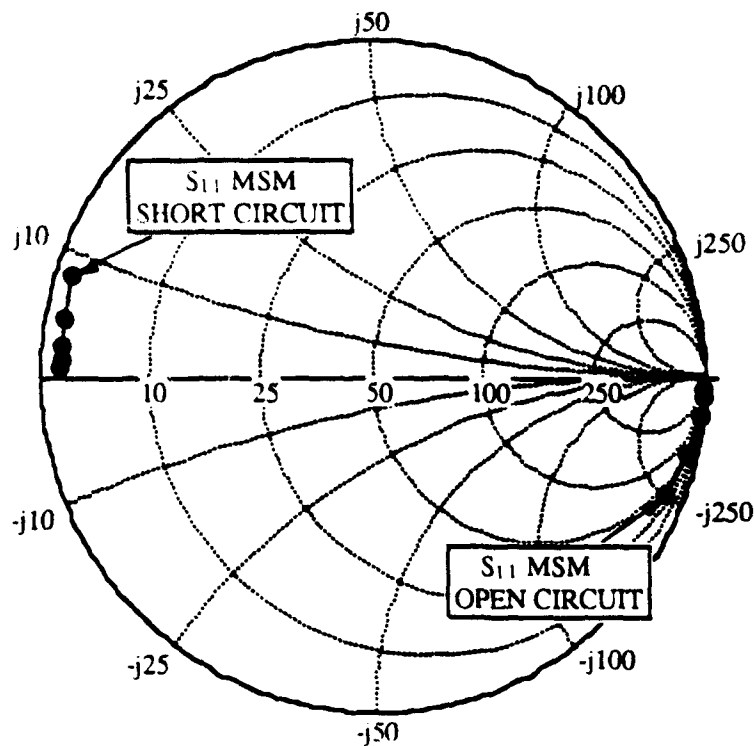
Figure 4. Measured S-parameters, (45 MHz-26.5 GHz), for a MSM detector and its associated short circuit test device. The geometry is linear with 15 fingers, 0.5 μm wide, 0.5 μm gap and 23 μm overlap

$N = N_1 + N_2$ (Number of Fingers)
$G = N - 1$ (Number of Gaps)

$$\frac{\{R_f + R_{skin} * (F/F_o)^{0.5}\}}{N_1}$$

$$\frac{R_g}{G}$$

$$\frac{L_f}{N_1}$$

$$G*C_g + N*C_p$$

$$i(h\omega, |E|^2)$$

$$\frac{L_f}{N_2}$$

$$\frac{\{R_f + R_{skin} * (F/F_o)^{0.5}\}}{N_2}$$
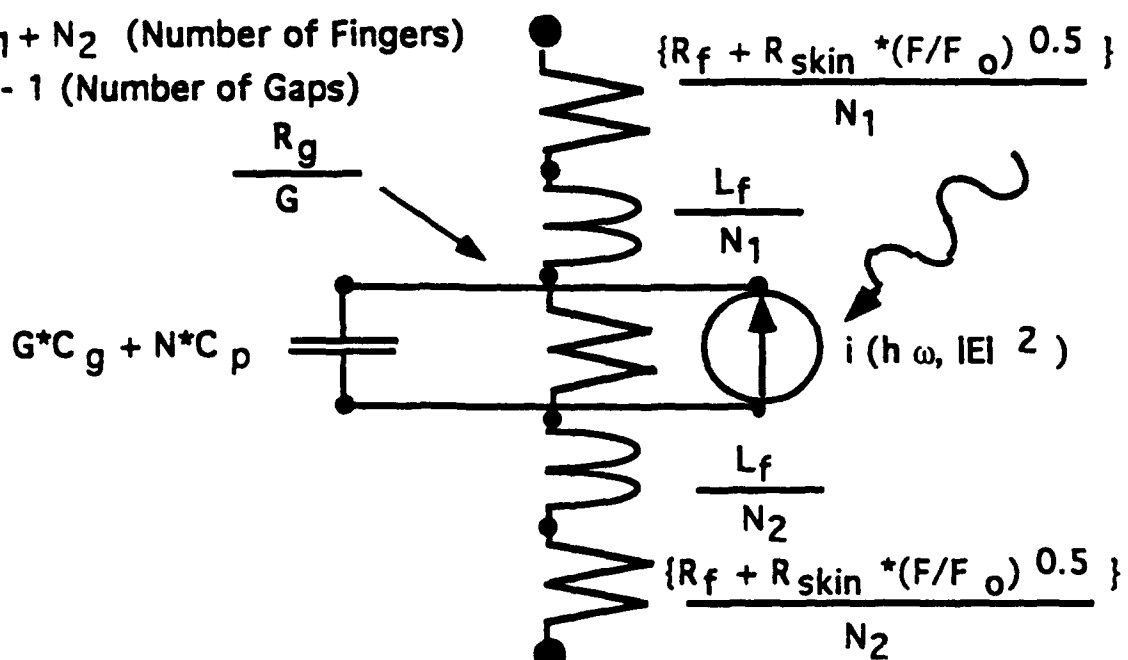
Figure 5. MSM detector diode equivalent circuit model.

TIME (ps)
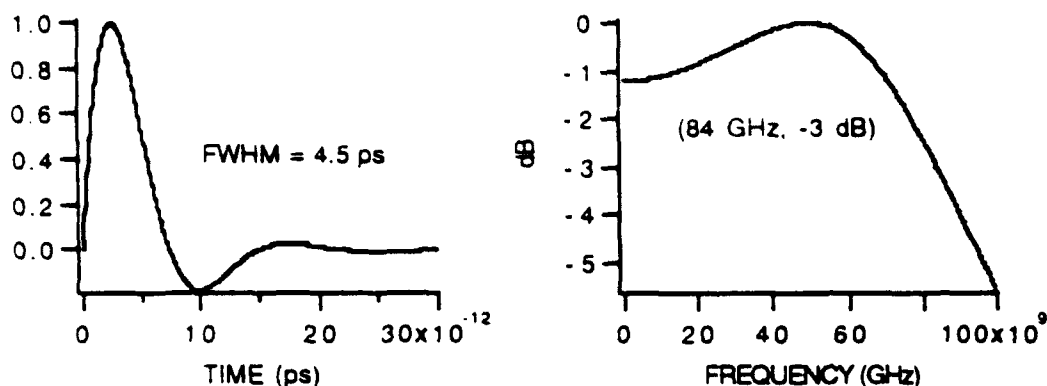
FREQUENCY (GHz)

Figure 6. Impulse function and frequency response of the MSM detector passive circuit element model (50 Ω load). For the geometry described above.

FWHM = 4.5 ps

(84 GHz, -3 dB)

TIME (psec)

FREQUENCY (GHz)
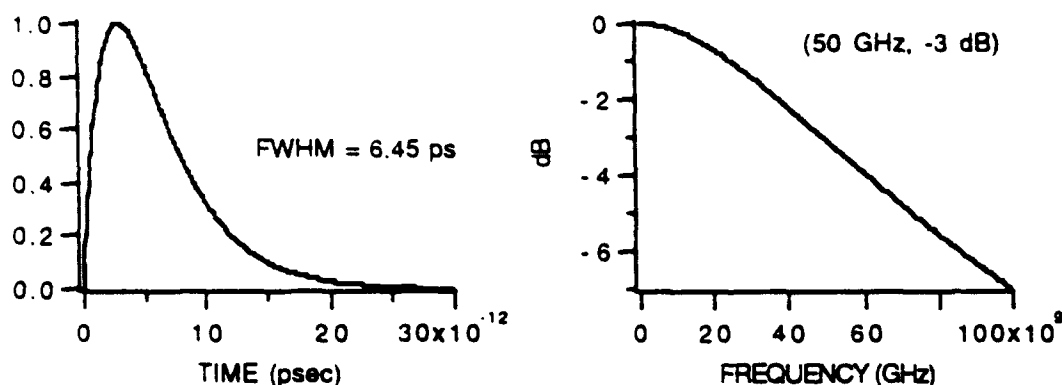
Figure 7. Time domain Monte Carlo simulation of the optically generated carriers in the MSM structure and the FFT of this function.

FWHM = 6.45 ps

(50 GHz, -3 dB)

TIME (psec)
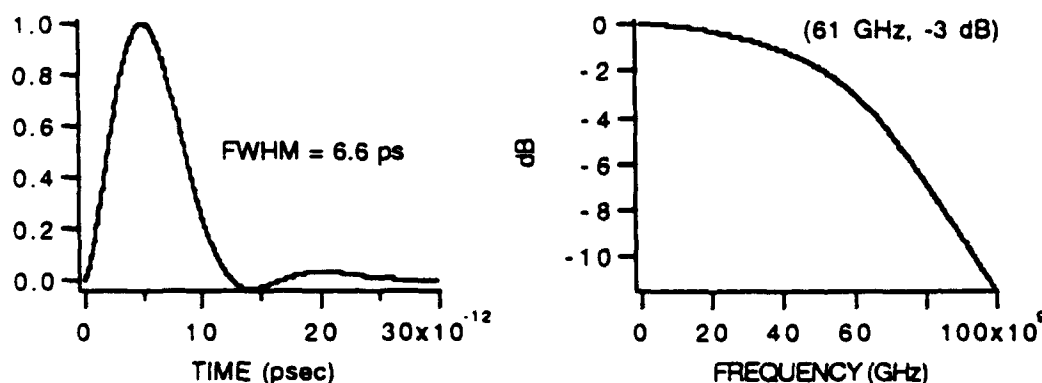
FREQUENCY (GHz)

Figure 8. Composite response of the MSM detector. The impulse response is the convolution of the time domain curves of figures 6 and 7. The frequency response is the product (addition of dB) of the frequency responses of figures 6 and 7.

FWHM = 6.6 ps

(61 GHz, -3 dB)

11-19
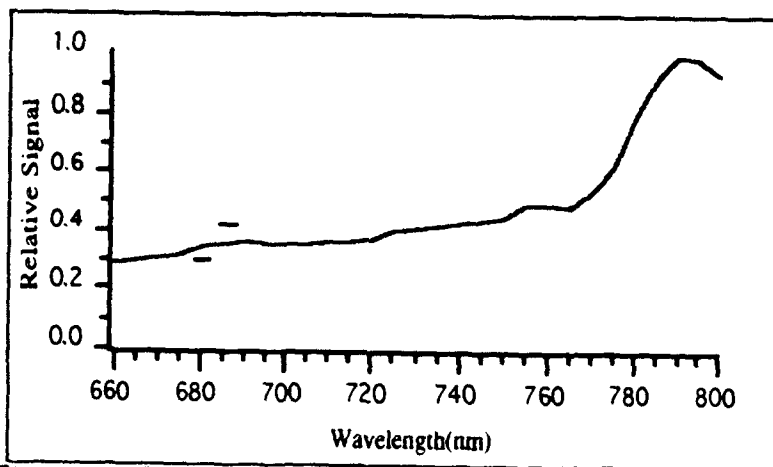
Figure 9. Relative signal from a detector vs. optical wavelength. The detector was fabricated on a wafer with the Bragg reflectors and top surface optical impedance matching optimized for 800 nM light.

Figure 10. Gamma-gate with 0.25 μm footprint and 2:1 overhang ratio. The photograph also shows the 0.6 μm drain - source spacing after the self-aligned ohmic metalization step.

QUANTUM WIRE STRUCTURES GROWN BY MOLECULAR BEAM EPITAXY

Jason D. Reed, Y.-P. Chen, William J. Schaff, and L. F. Eastman

School of Electrical Engineering

Cornell University

Ithaca, New York 14853

QUANTUM WIRE STRUCTURES GROWN BY MOLECULAR BEAM EPITAXY

Jason D. Reed, Y.-P. Chen, William J. Schaff, and L. F. Eastman
School of Electrical Engineering, Cornell University

Abstract

Arrays of $In_{.20}Ga_{.80}As$/GaAs quantum wires were fabricated using electron beam lithography, reactive ion etching and citric acid etching, wet chemical surface preparation, *in situ* annealing, desorption, migration enhanced epitaxy, and molecular beam epitaxy. The wires were examined by photoluminescence (PL) and transmission electron microscopy (TEM). TEM photographs show that the wires have smooth interfaces and excellent sidewall coverage with no visible dislocations or defects. The successful regrowth was achieved only after development of the combination of reactive ion etching and wet chemical preparation of the wafers. Arrays of 190nm wires exhibited a PL efficiency of 24% compared to the unpatterned quantum well and a 11.1meV shift of the luminescence to higher energy, which may partially be due to increased quantum confinement. The largest part of the shift is likely due to defects incorporated on the wire surfaces during the patterning process. The relatively high efficiencies make the quantum wires candidates for embedding in an optical cavity to form quantum wire lasers.

QUANTUM WIRE STRUCTURES GROWN BY MOLECULAR BEAM EPITAXY

Jason D. Reed, Y.-P. Chen, William J. Schaff, and L. F. Eastman

Introduction

Interest in high quantum confinement semiconductor lasers has been driven by predictions of ultrafast modulation bandwidths that are as much as three times higher than the fastest p-doped multi-quantum well lasers.[1,2,3] This improvement is shown graphically in figure 1, reproduced from reference 2, calculated for 50Å confinement dimensions. The graph shows that as the quantum confinement is increased from quantum well (QW) to quantum wire (QWR) to quantum box (QB), the relaxation oscillation frequency, and thus the modulation bandwidth, increases by a factor of about three. The lower dashed line on the figure shows the current performance record for QW lasers, at about 30 GHz. For the same photon density, a QWR laser is predicted to exhibit a 60 GHz frequency. The upper dashed line is at 100 GHz, which the figure indicates can be achieved at comparable photon density by a quantum box laser.

Other predictions, such as reduced temperature sensitivity,[4] extremely narrow linewidths and nearly threshold-less laser operation[5,6,7] have contributed to this interest in QWRs and QBs. Figure 2 shows the densities of states for bulk, quantum well, wire and box material and should help to explain the reasons behind these potential improvements in performance. It can be seen in the figure that as the degree of quantum confinement is increased, the density of states becomes narrower and higher around the state energies. This "sharpening" of the density of states leads to higher differential gain.[1] Higher differential gain leads to increased modulation speed and reduced threshold current, since carriers are bound more tightly around the transition energy, where they participate more efficiently in the lasing

process. Also, the emission linewidths become narrower and the temperature sensitivity of threshold current is reduced due to less thermal spreading of carriers.

Recently, there have been improvements in nanolithography and epitaxial regrowth techniques that bring high quantum wire and box performance within reach, specifically for quantum wire (QWR) lasers.[8,9] Methods such as selective regrowth by molecular beam epitaxy (MBE), organometallic chemical vapor deposition (OMCVD), or organometallic vapor phase epitaxy (OMVPE) on pre-patterned substrates,[10,11,12,13] lithographically-defined wet or dry etching followed by liquid phase epitaxy (LPE), MBE or OMVPE regrowth,[14,15] ion implantation,[16] lateral top barrier modulation,[17] strain-induced lateral confinement,[18] and several in situ techniques[19,20] have been used to fabricate quantum wire structures of varying quality. Quantum wire lasers have only been made by the selective regrowth technique.

There are two major problems that remain to be fully addressed in the fabrication of quantum wire structures. First, patterned lateral dimensions must be reduced to the 10-50nm regime while maintaining the high wire density necessary for sufficient gain. The geometry of the most successful fabrication technique, which involves selective regrowth on substrates patterned along certain preferential crystalline planes, may preclude achievement of high wire densities. Since the substrate patterning before wire growth is 1µm or more, while the wire sizes are about 1/10 µm, the maximum horizontal density is approximately 10%, and only if wide wires and wide laser structures are made[10]. This low density will not yield a high enough gain to allow short-cavity devices to be used, which are necessary for high speed operation. Stacking the wires vertically may alleviate the geometry problem somewhat, but it remains to be demonstrated that fluctuations in the sizes of the stacked wires can be brought under control. Furthermore, the QWR lasers made in this way have the wires oriented along the laser cavity, which is an arrangement that suffers a reduction by a factor of two in maximum gain due to the improper orientation of the cavity electric field relative to the wire axis.[21] The gain will be double if the photon cavity of the laser is oriented perpendicularly to the long axis of the

quantum wires, for reasons outlined in reference 21. In the case of the pre-patterned QWR lasers, the long axis of the wires is along the photon cavity, and the gain is half of the possible maximum.

Second, damage and fluctuations in wire size induced during the patterning must be eliminated, or reduced.[22,23] The required tolerance depends on the dimensions of the quantum wire, with tolerances becoming tighter as the dimensions shrink. Roughly, however, QWR size must be controlled to better than approximately 10% to maintain the benefits gained by the "sharper" density of states[24]. For example, for a quantum wire with 150Å dimensions a fabrication tolerance of 15Å (i.e. 10%) will yield a sharpening of the density of states, whereas with a tolerance of 30Å (i.e. 20%) the density of states is smoother and would not yield an improvement in differential gain or threshold current.[24] This requirement has posed and will continue to pose a major challenge for nanolithography to create arrays of thousands of 100-200Å structures that do not deviate from each other by more than 10-20Å.

In the short term (next 4 years), in situ growth techniques, such as growth on pre-patterned substrates, will dominate, since they can achieve damage free growth of small numbers (1-10) of 150Å to 500Å wires on certain AlGaAs crystalline planes. In the long term (4-8 years), direct nanolithographic patterning will dominate because of the inherent quantum mechanical advantage of having the correct orientation between wires and cavity, and because of the much greater flexibility of the technique. For example, direct nanolithography is the most likely candidate to create quantum box structures, where the most dramatic improvements in the density of states will be achieved. Processes that depend on certain crystalline plane orientations to give the wires their shape cannot be extended to make quantum boxes, because growth can be confined only vertically and along the bottom of the patterned groove. Quantum boxes require confinement in three dimensions, so it is unlikely that preferential growth techniques can be used to achieve them.

The direct approach also allows quantum wires and boxes of specific aspect ratio to be made, thereby tailoring the subband shapes and separations to create new optical materials. In an integrated opto-electronic device it may be advantageous to make high density arrays of wires and boxes in many different orientations on a wafer. This would create specifically tailored optical environments for non-linear interactions (holographic memories, phase conjugation or four-beam coupling), delta-beta waveguide couplers, optical amplifiers, or on-chip laser/LED emitters and detectors. Each of these will make different demands on material properties that can be met with quantum wire and box arrays.

The major challenge for direct nanolithographic patterning is to refine the techniques for creating lateral confinement. Compared with MBE growth, where in the vertical direction control over layers can be maintained to better than a half monolayer, i.e. approximately 2.5Å, nanolithography is still relatively unrefined. The best e-beam lithography systems today can achieve 20Å dimensions if all conditions are optimal. X-ray lithography, which may still be a decade away, holds the potential of 8Å dimensions. This is still a factor of four away from what is achieved regularly by MBE in the vertical dimension.

After lithography, etching must be done with a process that involves the transfer of kinetic energy from gas particles (reactive or not) to the exposed areas of the wafer. Atoms of gas are accelerated across a potential difference, after which they bombard the wafer. If the accelerated particles are not reactive themselves, then reactive gases such as chlorine may be introduced to etch the surface. This technique (called RIE, IBAE, CAIBE, and others, depending on details) yields an anisotropic etch--crucial for many of the new opto-electronic devices and materials such as short cavity lasers, integrated lasers and electronics, ring lasers, waveguides, and quantum wires and boxes. The major drawback of the etch is the same feature that makes it useful at all, namely the transfer of kinetic energy, which causes damage to the crystal structure of the GaAs wafer itself, and also causes many types of reactions with surface atoms and atoms of gas in the etch chamber. Contaminants from previous etches and atoms

adsorbed onto the surface of the wafer can also yield unpredictable results after etching. Both of these effects leave damage in the crystal and residues on the crystal surface that can be impossible to remove once the wafer has been reloaded into the MBE machine. *In situ* etching may be an answer to the problems of contaminants, but it is very expensive to build and run a joint MBE/etcher.

The approach adopted in this work involves lithography and etching in the National Nanofabrication Facility (NNF), followed by annealing and desorption in the ultra high vacuum environment of the MBE chamber in the Eastman Lab. The etching has been optimized to minimize the amount of incorporated damage, and the annealing and desorption are designed to remove contamination.

To conclude this introduction into the issues of QWR fabrication, brief mention should be made of strained quantum wires. The Eastman group has made great improvements in quantum well lasers with strained $In_{.20}Ga_{.80}As$ quantum wells.[25,26,27] Strain lifts the degeneracy of the light and heavy hole bands, allowing for a more balanced ratio between electron and hole effective masses, and thus densities of states in the material, thereby increasing the differential gain of injected carriers. There has been a theoretical prediction that the introduction of strain into quantum wires, which are strained in two dimensions, will lead to an improvement in performance compared to the unstrained case.[28] The practical challenges in fabricating strained quantum wires, specifically the quality of the regrown strained interfaces, are unknown.

Methodology:

During this summer research tour, a new two-growth quantum wire fabrication process was developed. Samples were prepared on a (100)-oriented semi-insulating GaAs substrate by MBE. As shown in figure 3, two $In_{.2}Ga_{.8}As$ quantum wells of 50Å and 100Å thickness were grown, lattice-mismatched to the surrounding 1000Å GaAs cladding layers (the term

"mismatched" is used to emphasize the point that the well layers are pseudomorphic due to the strain). The top layer is 800Å of GaAs. Following the lithography and RIE, the samples were dipped into a citric acid wet etch that selectively removed the exposed $In_.2Ga_.8As$. The benefits of using a wet etch as the last step before regrowth are that damage or contaminants left on the InGaAs material by the RIE process are removed and the wet etch itself introduces little damage or contaminants. In addition, the low-damage RIE process is selective to $In_.2Ga_.8As$, because of the low vapor pressure of indium chlorides. By stopping on the surface of the upper well, the RIE process cannot contaminate or damage the sidewalls of the quantum wires, which is crucial for maintaining high PL efficiencies in the regrown QWRs. The drawback of wet etching is the possibility of non-uniform etching, which would introduce variations in QWR size. TEM results indicate that the wires created by this process exhibit a 7% variation in size for 100nm lateral confinement, which is acceptable. As the lateral confinement dimension shrinks to approximately 20nm, which is the next processing goal, it remains to be seen if the 7% variation can be maintained. If not, then the process may require modification.

Lithography was performed with a 20kV Cambridge EBMF 10.5/CS e-beam machine, which was used to make $1mm^2$ arrays of lines and spaces as small as 100nm in 240nm of PMMA. After development, the wafers were etched in a reactive ion etcher (RIE) for 45s with 12 SCCM BCl_3 and 4 SCCM Cl_2 at a pressure of 3 mTorr. The incident power was $110mW/cm^2$ at a 85V DC self bias. The bias for this etch is as low as possible in order to reduce crystal damage while still maintaining a sufficient component of physical sputtering, which is necessary to remove the gallium and arsenic chlorides formed during the etch. As can be seen in figure 4, the bottom edges of the wires are slightly rounded. This slight non-anisotropy is due to the low bias voltage, and is the price that must be paid for holding the etch damage to a minimum. As dimensions shrink, it may be necessary to increase the plasma bias voltage to increase anisotropy, possibly requiring an improvement in damage reduction after etching.

The wafers were held for 60s in the citric acid etch solution, composed of 11 parts 1 molar citric acid, 1 part Hydrogen Peroxide (30%), and 44 parts DI water at room temperature, which removed 100Å isotropically. All traces of the resist were removed and the native oxide layer was stripped with sulfuric acid. The wafers were reloaded into the MBE machine for desorption, migration enhanced epitaxy (MEE) and MBE. The substrate temperature was raised to 640°C for 60s to remove the surface oxide. The substrate was then held *in situ* at 655°C for 120s under an As pressure of 6 μTorr, which desorbed approximately 50Å of the InGaAs, as measured from TEM micrographs taken before and after. Figure 5 illustrates the layer structure after regrowth, which consisted of 550Å of GaAs grown at 500°C at a rate of 0.2 μm/hr by MEE, with two thin AlAs layers introduced as tracers. A GaAs cap layer 2540Å thick was grown at 580°C by MBE at a rate of 1.0 μm/hr. This layer contained four AlAs tracers.

Results

After regrowth, samples were characterized by cross-sectional TEM using a novel preparation technique developed at Cornell by the authors during the summer research tour and described in reference 29. As seen in Figure 6, the quantum wires have excellent sidewall coverage, and the interface between the surface before regrowth and the regrown layer is invisible. There are no dislocations or planar defects in the structure. The dimensions of the smallest wires are 120nm, 190nm, and 410nm, with wire density fill factors of 32, 36, and 30% respectively. Finally, from the AlAs tracer layers (thin white lines in the micrograph), which are spaced 500Å apart, it can be seen that the growth front has planarized after the first 1500Å of regrowth.

Photoluminescence (PL) measurements were performed at 514nm wavelength with a 100μm beam diameter on the wire arrays to determine the peak intensity and energy of the luminescence for each wire width. As shown in figure 7, the intensity of the wire regions for

sample C declines to 24% as the lateral dimension is reduced from 100μm to 190nm. Samples A and B did not receive the wet etch, and show an intensity reduction by a factor of approximately 200 at the smallest dimension, due to RIE-induced damage on the sidewalls that gives rise to non-radiative recombination centers. The high PL peak for sample C indicates good surface characteristics of the regrown QWRs.

Peak PL energy as a function of wire size was also measured. As shown in figure 8, sample C shows a blueshift in peak luminescence energy of 11.1 meV, from 1.2985eV in the QW to 1.3096eV for the 190nm wires, as measured at 77K. The wider lines on the sample do not exhibit a shift. The FWHM of the PL peak at 77K was 17.3meV for the 190nm wires and 7.2meV for the QW.

Although the 190nm wire dimension is too large for the blueshift to be attributed solely to the additional degree of quantum confinement, which is calculated to be about 0.3meV, some component of the shift is likely due to quantum confinement. Some of the shift is possibly due to migration of the indium from the patterned wire regions into the surrounding GaAs during the 120s 655°C desorption step, although one would expect to see some shift in the wider dimension peaks in figure 8. The largest component of the shift is most likely due to surface defects that create an electric field that perturbs the energy levels in the wire. It should be possible to refine the regrowth process to increase the damage reduction, while maintaining the high radiative efficiency gained by the *in situ* annealing step. The damage masks the additional component of quantum confinement that would be shown by a shift in the transition energy. It is important to note that for QWR laser performance it is not a requirement to have a specific blue shift. On the other hand, it is important for QWR laser performance that the quantum wires have high radiative efficiency, which we have successfully demonstrated during this summer tour.

Conclusions:

We have achieved arrays of $In_{.2}Ga_{.8}As/GaAs$ quantum wires fabricated by electron beam lithography, dry and wet etching, chemical surface preparation, desorption, migration enhanced epitaxy, and molecular beam epitaxy. The wires were examined by PL and TEM. TEM photographs show that the wires have smooth interfaces and excellent sidewall coverage with no visible dislocations or defects. The successful regrowth was achieved only after development of a combination of RIE, citric-acid wet etching, and sulfuric acid preparation of the wafers. Arrays of 190nm wires exhibited a PL efficiency of 24% compared to the unpatterned quantum well and a 11.1meV shift of the luminescence to higher energy, which is due largely to defects on the etched surfaces of the smallest wire structures. The high efficiency of these wire structures makes them excellent candidates for embedding in an optical cavity to make quantum wire lasers.

Acknowledgments:

List of References

[1] Y. Arakawa, K. Vahala, and A. Yariv, Appl. Phys. Lett. **45**, 950 (1984).

[2] T. Takahashi and Y. Arakawa, IEEE J. Quantum Electron. **27**, 1824 (1991).

[3] S. Weisser, J. D. Ralston, E. C. Larkins, I. Esquivias, P. J. Tasker, J. Fleissner, and J. Rosenzweig (submitted to Electron. Lett.).

[4] Y. Arakawa and H. Sakaki, Appl. Phys. Lett. **40**, 939 (1982).

[5] H. Sakaki, Jpn. J. Appl. Phys. **19**, L735 (1980).

[6] A. Yariv, Appl. Phys. Lett. **53**, 1033 (1988).

[7] M. Asada, Y. Miyamoto, Y. Suematsu, IEEE J. Quantum Electron. **22**, 1915 (1986).

[8] S. Tiwari, G. D. Pettit, K. R. Milkove, R. J. Davis, J. M. Woodall, and F. Legoues, IEDM 1992.

[9] E. Kapon, S. Simhony, R. Bhat, and D. M. Hwang, Appl. Phys. Lett. **55**, 2715 (1989).

[10] M. Walther, E. Kapon, J. Christen, D. M. Hwang, and R. Bhat, Appl. Phys. Lett. **60**, 521 (1992).

[11] M. Walther, E. Kapon, C. Caneau, D. M. Hwang, and L. M. Schiavone, Appl. Phys. Lett. **62**, 2170 (1993).

[12] S. Tsukamoto, Y. Nagamune, M. Nishioka, and Y. Arakawa, Appl. Phys. Lett. **62**, 49 (1993).

[13] K. Komori, A. Hamano, S. Arai, Y. Miyamoto, and Y. Suematsu, Jpn. J. Appl. Phys. **31**, L535 (1992).

[14] W. Hornischer, P. Grambow, T. Demel, E. Bauser, D. Heitmann, K. von Klitzing, and K. Ploog, Appl. Phys. Lett. **60**, 2998 (1992)

[15] B. I. Miller, A. Shahar, U. Koren, and P. J. Corvini, Appl. Phys. Lett. **54**, 188 (1989).

[16] F. E. Prins, G. Lehr, M. Burkard, H. Schwiezer, M. H. Pilkuhn, and G. W. Smith, Appl. Phys. Lett. **62**, 1365 (1993).

[17] Ch. Greus, A. Forchel, J. Straka, K. Pieger, and M. Emmerling, Appl. Phys. Lett. **61**, 1199 (1992).

[18] I-H. Tan, R. Mirin, V. Jayaraman, S. Shi. E. Hu, and J. Bowers, Appl. Phys. Lett. **61**, 300 (1992).

[19] H. Kawanishi, Y. Sugimoto, T. Ishikawa, and H. Hidaka, Appl. Phys. Lett. **60**, 365 (1992).

[20] E. M. Stellini, K. Y. Cheng, P. J. Pearah, A. C. Chen, A. M. Moy, and K. C. Hsieh, Appl. Phys. Lett. **62**, 458 (1993).

[21] M. Asada, Y. Miyamoto, and Y. Suematsu, "Theoretical Gain of Quantum-Well Wire Lasers," Jpn. J. Appl. Phys. **24**, L95-L97 (1985).

[22] H. Zarem, K. Vahala, A. Yariv, IEEE J. Quantum Electron. **25**, 705 (1989).

[23] M. Notomi, M. Nakao, and T. Tamamura, Appl. Phys. Lett. **62**, 2350 (1993).

[24]see figure six in H. Zarem, K. Vahala, A. Yariv, IEEE J. Quantum Electron. **25**, 705 (1989).

[25] S. Offsey, W. Schaff, L. Lester, L. Eastman, and S. McKernan, IEEE J. Quant. Electron. **27**, 1455 (1991).

[26] S. Offsey, W. Schaff, P. Tasker, and L. Eastman, IEEE Phot. Tech. Lett. **2**, 9 (1990).

[27] S. Offsey, L. Lester, W. Schaff, and L. Eastman, Appl. Phys. Lett. **58**, 2336 (1991).

[28]S. Ueno, Y. Miyake, and M. Asada, Jpn. J. Appl. Phys. **31**, 286 (1992).

[29]Y.-P. Chen, J. D. Reed, S. O'Keefe, W. J. Schaff, and L. F. Eastman, Microscopy Res. and Tech. **26**, 157 (1993).
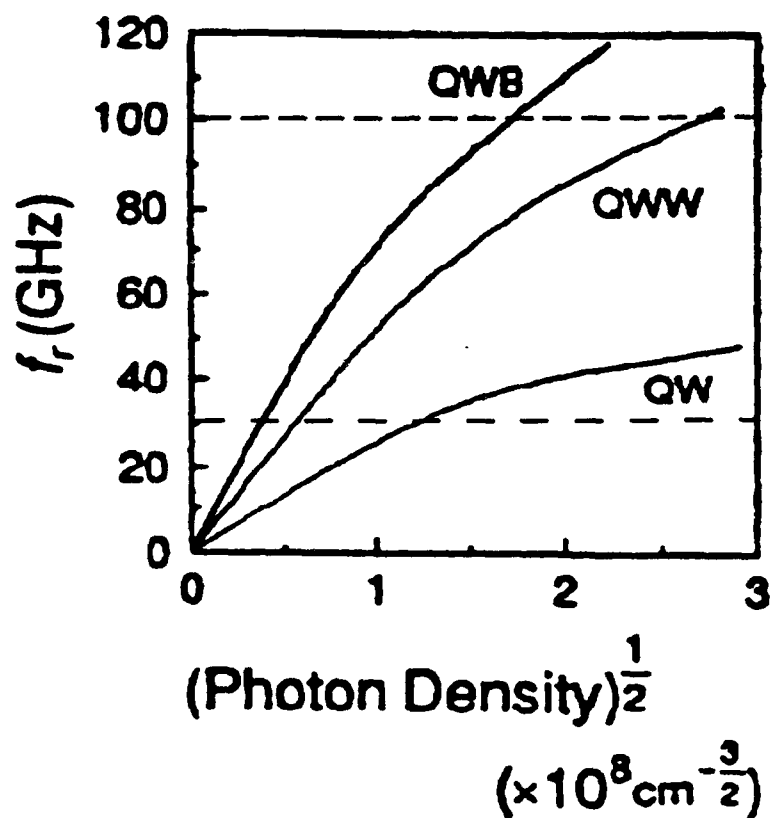
Figure 1: Relaxation oscillation frequency as a function of the square root of photon density, for different quantum structures: QW (quantum well), QWW (quantum wire, or QWR), and QWB (quantum box, or QB). The confinement dimensions are taken to be 50Å.

12-14
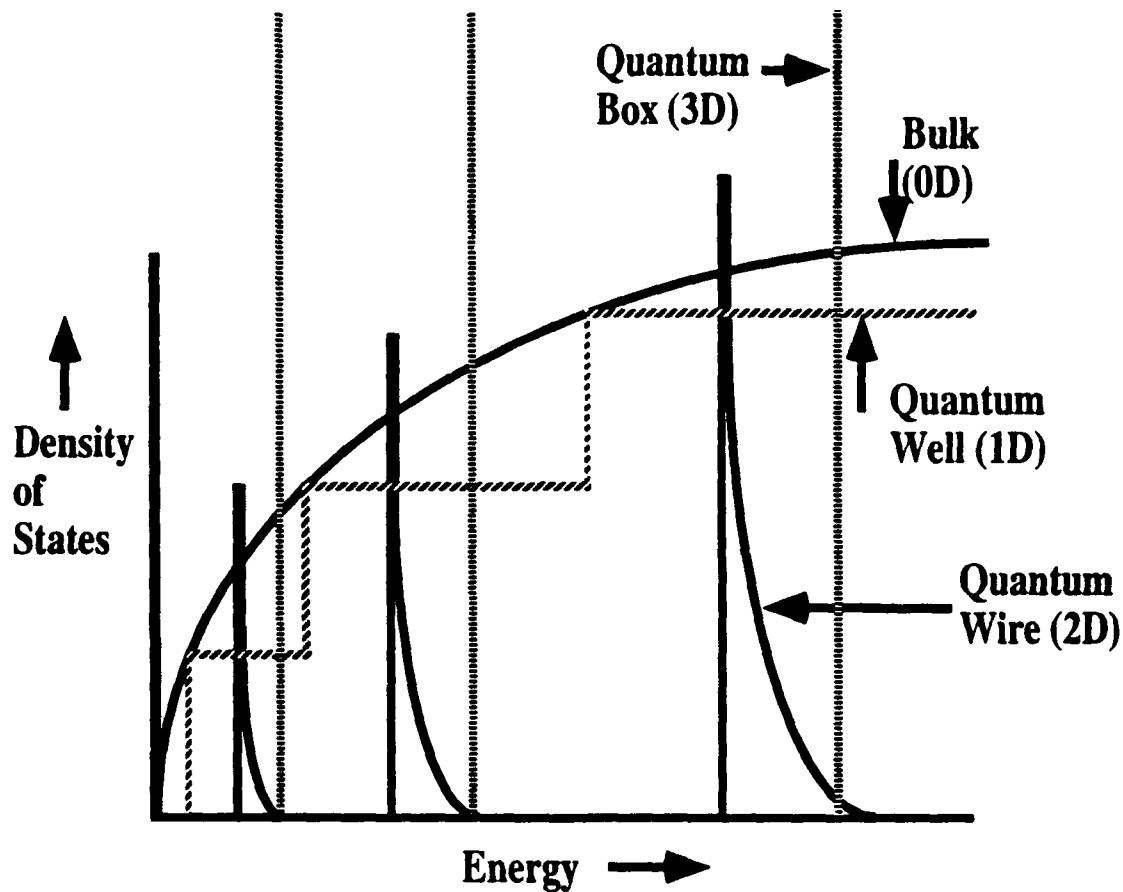
Figure 2. Densities of States for bulk, quantum well, quantum wire, and quantum box materials. As the degree of quantum confinement increases, the density of states becomes "sharper" around the transition energies, as described in the text.

Resist

800Å GaAs

100Å InGaAs

1000Å GaAs

50Å InGaAs

1000Å GaAs

500Å SL

Substrate
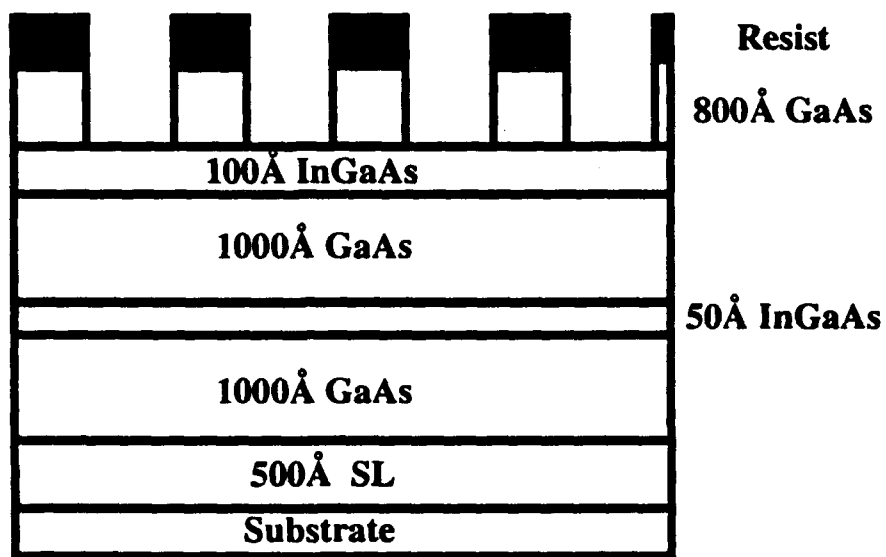
Figure 3. Schematic of sample after lithography and RIE etching (for clarity, figure is not to scale). The upper 100Å quantum well will be selectively etched by the citric acid step to form the quantum wires, which will then be regrown.

110nm

GaAs (MBE)

GaAs (MEE)

	GaAs
	100Å InGaAs
	1000Å GaAs
	50Å InGaAs
	1000Å GaAs
	500Å SL
	Substrate
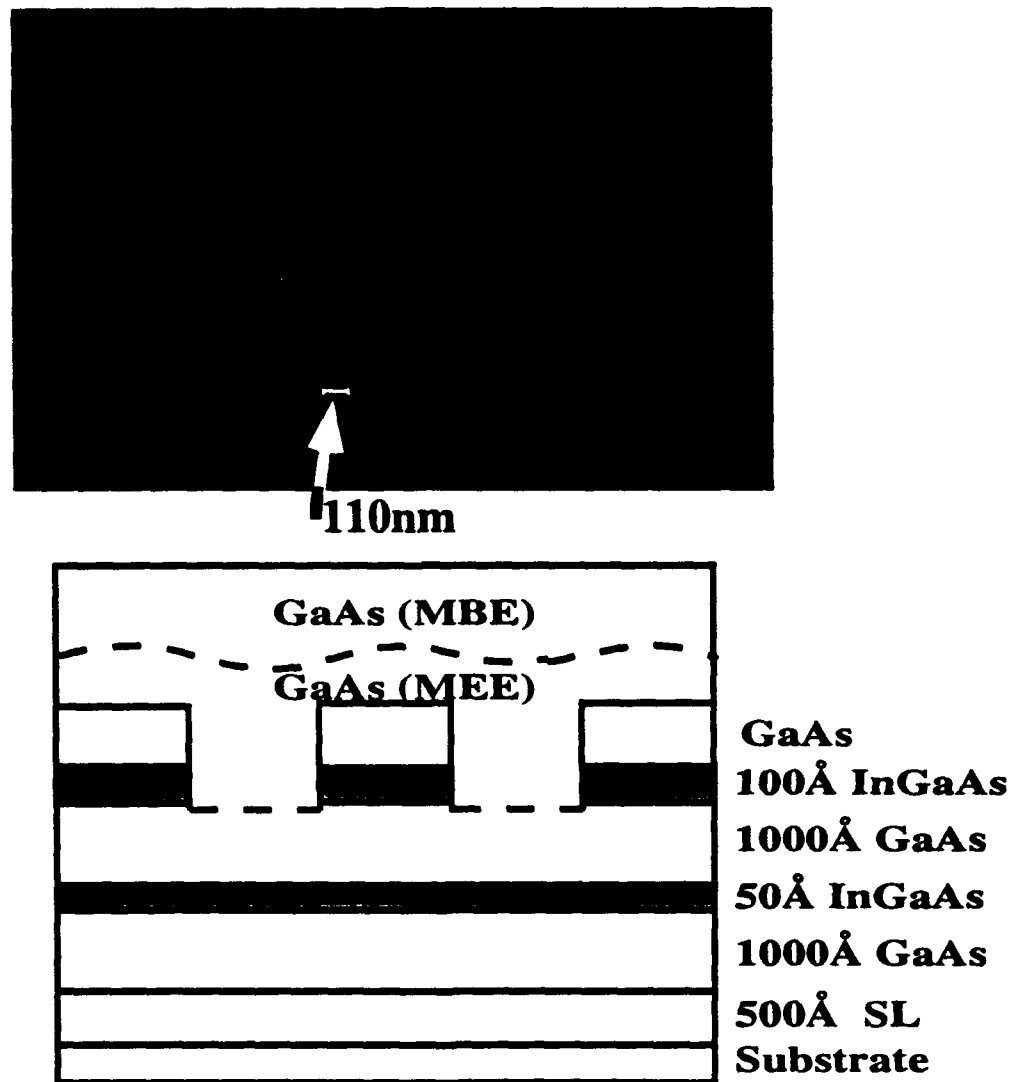
Figure 4 (above). Quantum wire sample after patterning and etching.

Figure 5 (below). Schematic of layer structure after regrowth (not to scale). The dark horizontal rectangles are the quantum wires that have been formed from the original quantum well.

12-17
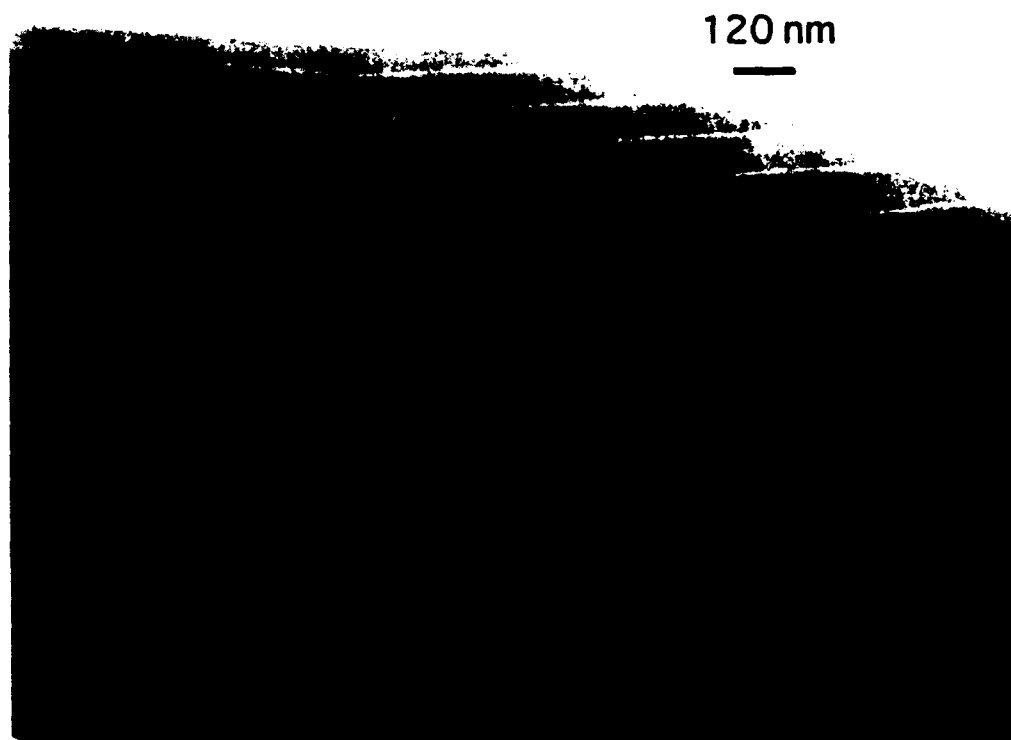
Figure 6. Cross-sectional Transmission Electron Micrograph of quantum wire structures. The dark horizontal bars are the $In_{.20}Ga_{.80}As$ wires. The surrounding lighter material is GaAs. The thin white lines are AlAs tracer layers, grown every 500Å. The remarkable point about this picture is what cannot be seen--there are no large defects or dislocations. Regrowth was smooth, and the wafer surface planarized after 1.5kÅ of regrowth.
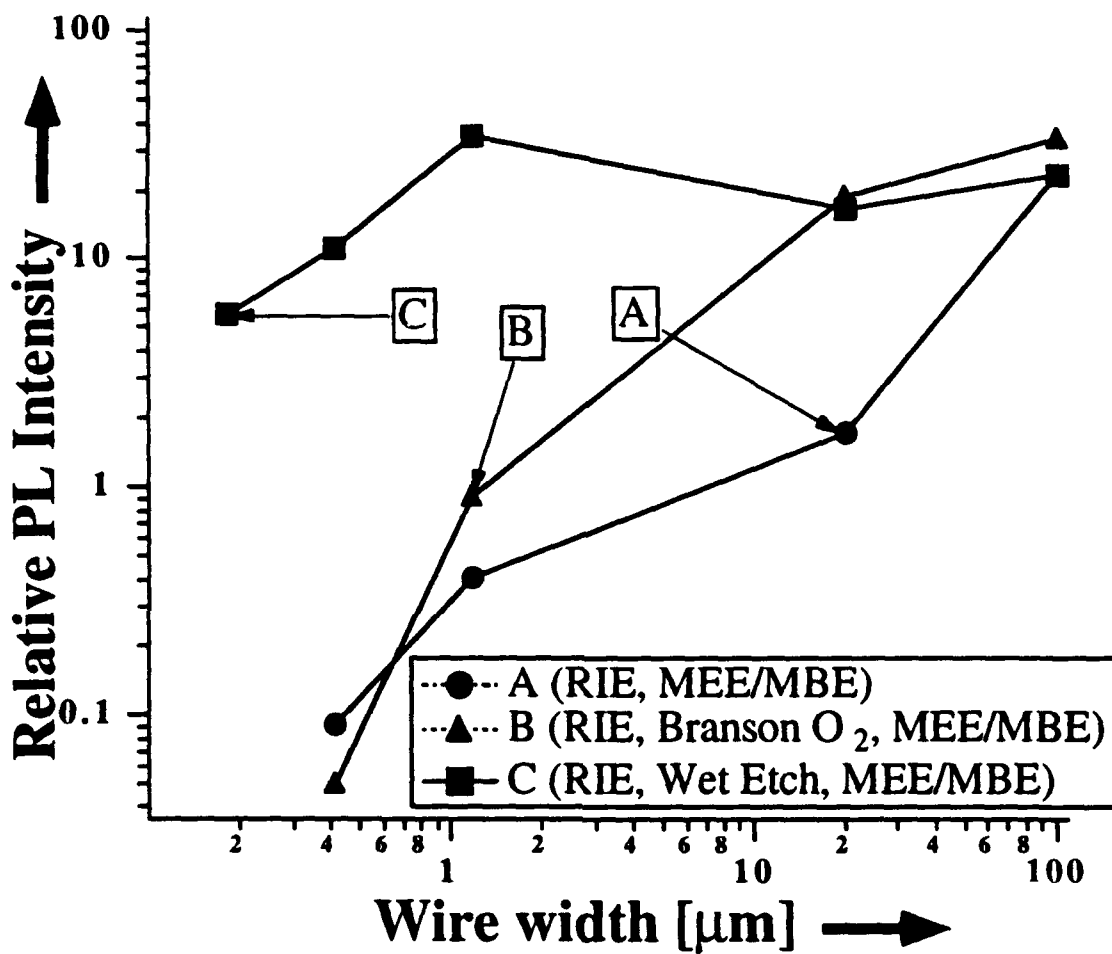
Figure 7. Photoluminescence as a function of wire width, for three samples. Sample C was given the full process as described in the text. Samples A and B did not receive the citric acid wet etch, and sample B received an oxygen plasma ash. Sample C maintains relatively good PL efficiencies down to the smallest dimensions (190nm in this case). The other samples show drastic reductions in PL, due to the presence of large numbers of defects on the wire surfaces.
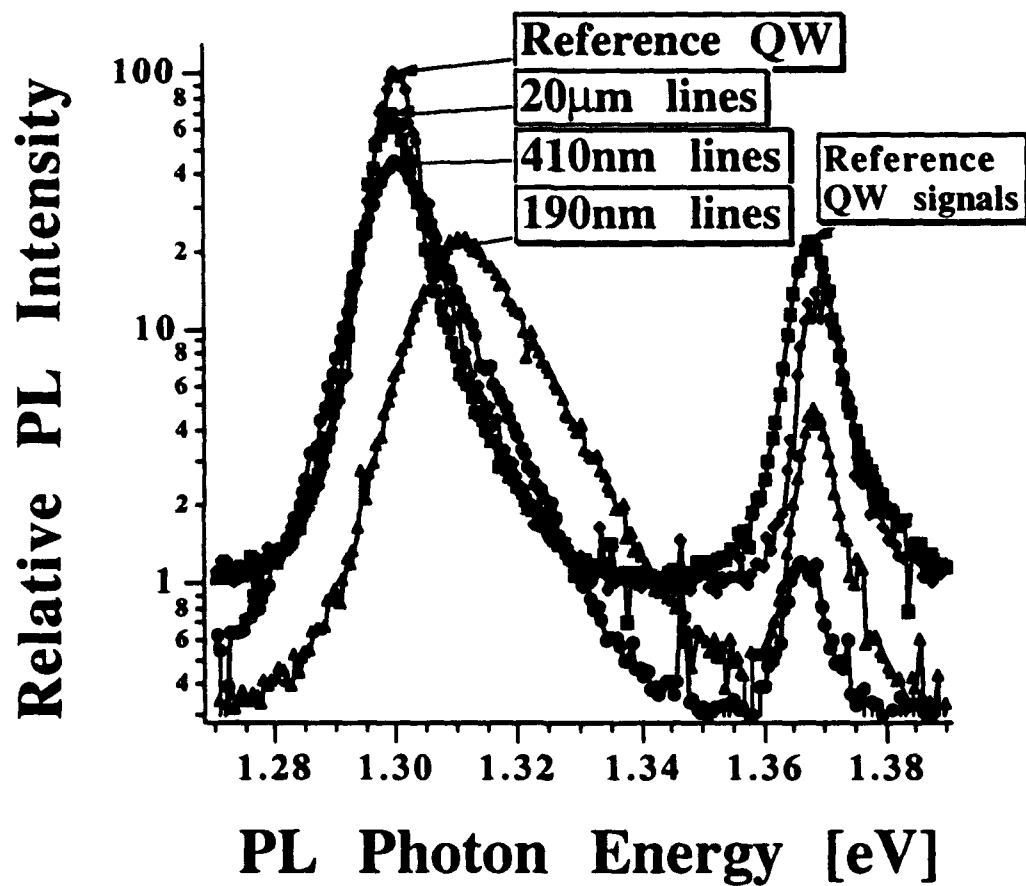
Figure 8. Peak photoluminescence energy as a function of wire size for sample C (see figure 7). As the confinement dimension becomes smaller, the peak emission shifts to the blue by 11.1meV. Mechanisms for this shift are discussed in the text. PL data was taken at 77K.

OPTICAL INTERCONNECTS

Robert Spencer

Department of Electrical Engineering

Cornell University

Phillips Hall

Ithaca, NY 14850

OPTICAL INTERCONNECTS

Robert Spencer

Department of Electrical Engineering

Cornell University

ABSTRACT

Research endeavors this summer in conjunction with the AFOSR summer research program followed two courses - chemically assisted ion beam etching (CAIBE) of compressively strained layer quantum well lasers and the design and fabrication of polymer waveguides for in- plane and plane-to-plane optical signal distribution. The mission of the group to which I was assigned at Photonics Laboratory of Rome Laboratory is the design of optical interconnect schemes for advanced computer architectures. As signal processing speed, power dissipation requirements, and packaging size requirements become more severe for new-generation Air Force electronic systems new concepts in signal distribution must be evaluated. Optical signal distribution networks for wafer scale integration and multi-chip module packing schemes are being examined by the photonics lab. In current VLSI systems, the RC parasitics associated with electrical interconnects are limiting the overall signal processing speed and are therefore the main bottleneck in a variety of novel computing architectures. Optical interconnects are limited mainly by material and waveguide dispersion and the bandwidth of the sources that are used for the optical signal. Current state of the optical detectors have bandwidths that are several times larger than the bandwidths of optical sources and are therefore not currently limiting processing speed, although the efficiency of optical detectors is still a substantial consideration in achieving low power dissipation optical interconnects. I therefore examined two aspects of optical interconnect networks - compressively strained quantum well lasers for optical sources and polymer waveguides. Compressively strained quantum well lasers have demonstrated the highest directly modulated bandwidth to date (30 GHz) of any semiconductor laser while polymer waveguides offer processing compatibility with both GaAs and Si, good chemical, mechanical and thermal stability, and excellent optical properties. A better understanding of both sources and waveguides is necessary to completely realize fully optical interconnection schemes that will permit the design and fabrication of advanced computer architectures that are required by the Air Force.

OPTICAL INTERCONNECTS

Robert Spencer

STRAINED LAYER QUANTUM WELL LASERS - THEORY

The introduction of strain within the quantum well of a semiconductor laser offers yet another design variable in the engineering of higher-performance optical sources. Research has been carried out to gain a broader understanding of the physics and fabrication of strained quantum well lasers that will further enhance our current understanding of these devices. Several ambiguities still exist in how to best optimize device structures to take advantage of the inherent performance of compressively strained quantum well lasers - only when these uncertainties are resolved can the flexibility offered by these devices be fully utilized in high speed optical sources.

To further enhance the performance of SLMQW (Strained Layer Multiple Quantum Well) lasers is it first necessary to understand how each aspect of the device affects performance. Charles Henry of AT&T Bell Labs introduced the concept of utilizing the band discontinuity of the AlGaAs/GaAs heterojunction (an unstrained system) to spatially confined electrons in a potential well in the early 1970's - only about a decade after the first successful demonstration of bulk diode lasers. Since then there have been many theoretical predictions that quantum well lasers would not only demonstrate lower threshold currents but would also have superior modulation characteristics. It was not until the early 1990's however, that the physics of quantum well lasers was understood well enough to design and fabricate devices that demonstrated superior performance to their bulk laser counterparts. Understanding the physical limitations of quantum well lasers is essential before pursuing reduced dimension devices such as quantum wire and quantum box lasers. Overcoming the difficulties associated with introducing one quantized dimension will certainly facilitate the pursuit of reduced dimensionality devices that have the potential for high-speed performance superior to even SLMQW lasers.

A semiconductor laser can be thought of as a gain medium within a resonant cavity - the purpose of the resonant cavity being to provide the necessary positive feedback. All relevant performance characteristics of the laser are related to the inherent material gain of the semiconductor and therefore an understanding of material gain is first necessary. Gain, as derived from Fermi's golden rule and modified for transitions between bands of states (as opposed to two discrete states), is as follows:

$\hbar\omega$=transition energy

e=electronic charge

n_g=material group refractive index

n=refractive index

ε_0=permitivity of free space

m_0=free electron rest mass

ρ_{red}=reduced density of states

$f_{c,v}$= electron,hole fermi functions

This function has assumed strict k-selection throughout (which therefore excludes lasers with doped active regions), and instantaneous intraband scattering. The most intuitively obvious part of this formula is the inversion factor (fc + fv - 1). This is simply a statement that the rate of stimulated emission must exceed the rate of stimulated absorption for there to be gain, i.e. there must be more electrons in the conduction band than there are in the valance band. This state is known as inversion and is a necessary precursor to lasing. Another way to state this is that the separation of the quasi-Fermi levels must be at least equal to the semiconductor bandgap in order to achieve inversion (also called the transparency condition). This can be seen as follows. The smallest transition energy between the conduction band and the valance band is equal to the bandgap. The largest transition energy is equal to the separation of the quasi-Fermi levels because at this energy fc=0.5, fv=0.5 and the inversion factor is then zero. Thus, the smallest separation in the quasi-Fermi levels must be the bandgap energy. This sets the condition for the injected current density that is necessary for transparency. Because we are only looking for the carrier concentration that separates the quasi-Fermi levels by the bandgap energy there is little difference between the transparency carrier densities of quantum well and bulk laser diodes. However, because the active region of the quantum well laser is substantially smaller, the number of injected carriers required to invert a quantum well is smaller than the number carriers necessary to invert a bulk active region. As a result, threshold currents of quantum well lasers are substantially smaller than those of bulk diodes, but this reduction has nothing to do with the quantized nature of the carriers. One additional aspect of threshold current calculations is important for understanding the advantages of SLMQW lasers as optical sources - the effect of the different masses of holes and electrons. In both the bulk and quantum well case, the density of electronic states is proportional to the effective mass of the carrier in that band. The larger the effective carrier mass, the larger the density of states and the larger the number of carriers needed for a given movement of the quasi-Fermi level away from its equilibrium position near the center of the bandgap. In all cases of unstrained systems, the hole effective mass is substantially larger than the electron effective mass. Therefore, assuming quasi-neutrality within the active region (n=p), a given density of injected carriers will move the electron quasi-Fermi level farther than the hole quasi-Fermi level. As a result, to achieve transparency (i.e. separation of the quasi-Fermi levels by the bandgap energy), the electron quasi-Fermi level will be well within the conduction band while the hole quasi-Fermi level will be positioned above the valance band edge. Moving the electron quasi-Fermi level into the conduction band requires that a larger number of carriers be injected and thus the

Compressively strained quantum wells have been utilized in both low-threshold devices and high-speed lasers. The reduction of the effective mass has several advantages for low threshold current devices. First and foremost, compressive strain further separates the LH1 and HH1 subbands in the quantum well - thus reducing the amount of mixing between these two states. This reduces the in-plane (parallel to the QW) mass of the holes in the quantum well and reduces the band-mixing induced non-parabolicity in the HH1 state. Since we are most concerned with the C1-HH1 transitions in the quantum well, the effective hole mass that essentially determines the transparency condition (i.e HH1) is reduced and the threshold carrier density is reduced. This lower effective mass of the HH1 state is preserved for larger values of hole energies due to the enhanced parabolicity of the band structure which is a direct result of the increased separation of the LH1 and HH1 states. The effect of the strain on the lasing wavelength is substantial as well, as the bandgap of InGaAs strained quantum wells can be tuned to values around 1μm. These wavelengths were previously inaccessible with any III-V lattice matched system and are necessary for the pumping of Erbium doped silica fiber amplifiers (the optical pump wavelength is must be 0.98 μm for these amplifiers). Of most concern to our project, however, are the substantial benefits of compressive strain on the high-speed performance of quantum well lasers.

High speed optical interconnects will require high-speed optical sources and thus this project has examined several aspects of strained quantum well lasers and their suitability as hi-speed optical sources. Our primary work has been on the optimization of SLMQW devices for extremely large bandwidths. Of course it may seem as if creating 30 GHz lasers is excessive for the estimated 1GHz data transmission rates of future systems, but low-power, low noise parameters are enhanced with larger bandwidth devices. For example, frequency chirping which is related to the linewidth enhancement factor is reduced for lasers with larger differential gains. Since increasing the differential gain is a primary method for enhancing the high-speed characteristics of lasers, enhancement of the modulation characteristics through structures with larger differential gains also decreases the linewidth enhancement factor, reduces the frequency chirping, and improves the noise characteristics of the devices throughout the frequency spectrum. In addition, improvement of the ultimate modulation bandwidth of theses devices means that bandwidths of 1GHz can be reached at much lower DC biases - which is important for power dissipation considerations. Therefore, extending the modulation bandwidth into the 30 GHz range will enhance the noise and power dissipation characteristics for the much lower data transmission rates that are currently under consideration.

Conventional wisdom states that the sharper density of states of a quantum well laser will provide superior modulation characteristics due to the larger differential gain of the quantum well laser. This enhanced differential gain, however, is realized only when all of the carriers are directly injected into the lowest subband (electron and hole) of the quantum well, i.e. there is a negligible population of carriers in upper subbands and barrier states.

13-5

Zhao, Chen, and Yariv [**] demonstrated that the predicted differential gain is substantially lowered when either upper subbands or barrier states are occupied. This, in part, explained why single, unstrained quantum well lasers were showing inferior modulation characteristics to their bulk laser diode counterparts. Steady state carrier distributions in the quantum wells can be influenced in several ways, the most common of which are:

(a) number of quantum wells

(b) quantum well thickness and height

Most high speed devices make use of multiple quantum well structures to achieve larger differential gains. Each quantum well has approximately the same material gain and therefore the same modal gain can be achieved with fewer injected carriers in any one well (this assumes that the injected carriers are evenly distributed in all wells). Lower injection levels lead to a decreased probability that barrier states and upper subbands will be occupied and therefore the differential gain of the laser will be larger. Quantum well design can also be altered to increase the differential gain of the device by using larger barriers. Larger barriers enhance the confinement of carriers in a finite potential well and increase the separation of the lowest subband and upper subbands. This once again increases the percentage of injected carriers that reside in the lowest subband and increases the differential gain. Only by using MQW structures can the inherent advantages of a two-dimensional density of states be fully utilized. Further enhancements in differential gain can be achieved by using strained quantum wells or p-doped quantum wells.

In the case of only one subband being occupied in both the conduction and valance band, differential gain can be written as:

$$\frac{\partial g}{\partial n} = \xi \left\{ \frac{\rho_v}{\rho_c + \rho_v}(1-f_c) + \frac{\rho_c}{\rho_c + \rho_v}(1-f_v) \right\} \qquad (2)$$

where ξ is a grouping of the physical constants that includes the momentum matrix element and $\rho_{c/v}$ is two dimensional density of states in the conduction/valance band. The two dimensional density of states is linearly proportional to the effective mass. In the case where the active region is unstrained, the first density of states ratio ($\rho_v/(\rho_c + \rho_v)$) is about 0.85 while the second ratio ($\rho_c/(\rho_c + \rho_v)$) is about 0.15. The asymmetric density of states also means that at threshold (which is relatively near transparency for most devices) the electron quasi-Fermi level is located in the conduction band while the hole quasi-Fermi level is located well outside of the valance band. Therefore, $f_c = 1$ while f_v is much less than one. It would be much more desirable for $f_v=1$ and f_c to be much less than one, which is precisely what is achieved with p-doping of the devices. Introducing a large number of holes into the active region forces the hole quasi-Fermi level into the valance band and therefore the electron quasi-Fermi level does not need to move into the conduction band in order for the separation in quasi-Fermi

levels to be equal to the bandgap of the active region. In this state transparency (and approximately threshold) is reached with $f_v=1$ and f_c much less than one and the differential gain is improved by approximately by a factor of five. In the case where effective masses of holes and electrons are equal, doping does not improve the differential gain at all. However, the reduction of the hole mass does significantly improve the differential gain as can be seen by equation two which is just $g'=\zeta f_c$ when the hole mass and electron mass are identical. With the effective masses equal, transparency occurs with electron and hole quasi-Fermi levels at their respective subband edges. For transitions at these energies each quasi-Fermi function is defined to be equal to 0.5 and therefore the differential gain is enhanced over the unstrained case by a factor of about 3. Of course, equation 2 is an oversimplification and increases in differential gain due to the enhanced parabolicity of the bands and the strain splitting of the various subbands is not taken into account, but it does give a rough idea of the improvement of differential gain by utilizing p-doping or strained layer material.

The maximum achievable bandwidth for a semiconductor lasers is given by:

$$f_{3dB} = \frac{2\pi\sqrt{2}}{K}$$

$$K = \frac{(2\pi)^2}{v_g}\left\{ \frac{\varepsilon}{\partial g/\partial n} + \frac{1}{\alpha_i + \alpha_m} \right\} \qquad (3)$$

where ε is the gain compression factor and is defined by

$$G = \frac{G_o}{1-\varepsilon S}$$

where G_o is the material gain that is independent of photon density

and S is the photon density

The gain compression factor is introduced phenomenologically and describes the dependence of gain on photon density. As can be seen from equation 3, the bandwidth has an intrinsic component (i.e. the photon lifetime) and an additional component that consists of a ratio of the gain compression factor to the differential gain. The intrinsic component is changed via the design of the cavity itself, e.g. by designing shorter cavities. The two terms in brackets are often of comparable magnitude, but the term involving gain compression is usually on the order of two to three times greater than the term involving the cavity design. Of course, the cavity design term can be substantially reduced by the use of high reflectivity coatings on the facets, but at the expense of external efficiency which of significant importance in the design of high speed sources for optical interconnects. Thus, the material

term that consists of the differential gain and the gain compression factor must be optimized to achieve larger bandwidths and superior low-noise and low power dissipation performance at low frequencies. Differential gain enhancement in strained MQW structures has already been discussed as well as minimizing the occupation of upper subbands and barrier states. There is also the possibility of improving differential gain in longer-wavelength lasers (1.3 µm to 1.5 µm) by using tensile strain lasers that operate primarily in the TM mode as opposed to compressively strained lasers which are strongly TE. Tensile strained quantum well lasers utilize transitions between C1 and LH1. As it turns out, transitions to LH states have stronger dipole matrix elements associated with TM radiation and therefore differential gain can be enhanced via this mechanism - even though the effective mass of holes in tensile strained quantum wells is larger than the hole mass in compressively strained quantum wells. In addition to optimizing the differential gain, the gain compression factor must be reduced for high speed operation. The physical mechanisms underlying the gain compression factor are not at all clear, and there are several theories that describe gain compression. Spectral hole burning and carrier heating are the two most common means of describing gain compression. Recent theoretical work suggests that carrier heating is substantially more significant than spectral hole burning, but the underlying cause of the carrier heating is not well understood. One model proposes that carrier heating is due in large part to the hot-phonon effect. This theory relates the decay of excess energy from carrier to LO phonons (which are stationary) to acoustical phonons. The LO phonon decay time is the primary bottleneck in removing excess energy from carriers and thus carriers remain in a non-equilibrium state - i.e. the carriers are 'heated'. Two of the more popular ways to describe gain compression are spatial hole burning and carrier transport induced damping. Both theories state that it takes a finite amount of time to reach the appropriate quantum well bound state. The spatial hole theory calls this delay time the 'carrier capture time' - the time that takes the quantum well to capture a carrier in the continuum above the quantum well and capture it in the QW. Carrier transport theories provide an intuitive reason for the delay between carrier injection at the ohmic contact and the capture of carriers at the quantum well - it simply takes a finite amount of time for the carrier to drift across the cladding layers and barrier layers to reach the quantum well. There are a host of theories that can be used and it is necessary to design structures that can differentiate between them. One surprising result is that p-doping the active region can reduce the gain compression factor. As was stated before , p-doping compressively strained material does not improve the differential gain significantly, but it does decrease the gain compression. Even the authors who reported this result were not sure why this happened, and there is obviously much work that still needs to be done in improving the physical models of gain compression.

EXPERIMENTAL WORK ON QUANTUM WELL LASERS

The work that was carried out this summer was the CAIBE fabrication of both p-doped and undoped compressively strained multiple quantum well lasers using material that was grown by Sean O'Keefe of Cornell University. The CAIBE processing of mirror facets is necessary to fabricate shorter (<200 μm) devices that are the optimum cavity length to use in MQW structures. In addition, shorter devices have shorter photon lifetimes which should enhance the overall frequency performance. CAIBE etching was chosen over several other dry etching methods such as RIE (Reactive Ion Etching), RIBE (Reactive Ion Beam Etching), and Ion Milling. RIE consists of chemically reactive ions that are given directionality by a DC bias while RIBE consists of chemically reactive ions that are given directionality by a ion gun. Ion Milling uses the physical sputtering of material via non-reactive ions supplied from an ion-gun. CAIBE uses a chemically inert ion species supplied from an ion gun while the reactive gas is supplied from nearby gas injection jets. In this matter, CAIBE separates the physical and chemical components of the etching process. The Cornell CAIBE etching system is an Technics Plasma Gmbh R.I.B. Etch 160 in which the gas injection jets have been added as a custom modification. The work that is being carrier out this summer complimented work that was previously done (Spring 1993) in which the high speed characteristics of p-doped and undoped lasers were compared. Those devices showed that p-doped lasers were able to achieve a given bandwidth at a lower DC bias. CAIBE etched facet mirrors are necessary for further characterization and also are necessary if these devices are to monolithically integrated with the appropriate electronic drive circuitry. Thus, a substantial amount of work was performed to characterize etching properties of the Cornell CAIBE as well as different types of etch masks that might be suitable. A robust process was sought in which the mesa would be self aligned to the p-ohmic contact. One way that this might be done is to use a p-ohmic contact that is also suitable as an etch mask. The normal ohmic contact to heavily p-doped GaAs is 500 A Ti followed by 1000 A of Pd followed by 2500 Å of Au. Gold, however, is not suitable as an etch mask due to the high sputtering rate of gold. Therefore, Cr was deposited on top of the gold to serve as an etch mask. Chromium has a low sputtering rate and does not react with chlorine, therefore giving a low CAIBE etch rate and making it suitable as an etch mask. A Cambridge electron beam lithography system was used to do the lithography in a bi-layer resist that consisted of 6000 Å copolymer (which consists of PMMA-PMAA) and 2400 Å of PMMA. The lower layer of the bi-layer resist develops out faster in the developing solution of MIBK-IPA and thus gives an undercut profile that is necessary for the liftoff processing of evaporated metals. The standard p-ohmic metalization was evaporated (4000 A total) and then 500 Å of Cr was evaporated on top of that, giving a total metal thickness of 4500 A. The following CAIBE operating conditions were used for all etches:

accelerating voltage - 500 V

current density - 0.2 mA/cm2

Cl2 injection rate - 20 ml/min

Ar injection rate - 3.5 ml/min

base pressure - 2e-5 Pa

ion beam current - 70 mA

Two types of etches were tried - one with the sample heated and the other etch with no sample heating. For samples in which no external stage heating was used the sample nonetheless did undergo ion beam heating which raised the temperature of the sample about 20°C in 6 minutes. The other etch used a heated stage in which the sample was heated to 110°C in order to enhance the etch rate of both GaAs and AlGaAs. The thermal desorption rate of AsCl2, formed by the injection of Cl2 near the sample stage, is increased substantially above temperatures of 100°C which enhance the etch rate of the laser material as well as dehydrating the sample during the etch. This dehydration of the sample's surface is important in the etching of AlGaAs. The results of both etches are shown in figure 1. As can be seen in figure 1a, the thermally assisted CAIBE etch displayed quite a bit of preferential redeposition of metal along the sides of the etched mesa and etched at a rate of 2300 A/min. In addition, it appears as if the metal mask has lifted up somewhat in several parts of the mesa. This type of scumming is substantially reduced in the unheated CAIBE etch in figure 1b and etched at a rate of only 1150 A/min. The exact source of the deposition at the sides of the mesas is unclear, but it is related to the metal masking scheme that was used since CAIBE etches that utilize resist etch mask do not display this type of pattern. Some scumming of metal masks is to be expected but the increased and preferential scumming that characterizes high temperature etches is somewhat of a surprise. The low temperature etch is relatively clean and the metal mask does provide good etch selectivity. Higher temperature etches are usually used to enhance the etch rate of GaAs/AlGaAs relative to the mask (e.g. resist masks) to create vertical sidewalls. A good rule of thumb is that the slope of the sidewall is equal to Tan-1(etch rate of substrate/etch rate of mask). The various etch rates of the materials used in this study are:

Cr - 60 Å/min

SAL601 etch mask (to be described later) - 200Å/min

GaAs with no heating - 1150 Å/min

GaAs @ 110°C - 2300 A/min

The other masking system considered was the SAL601 etch mask. SAL601 is a negative e-beam resist and can therefore be patterned as a mask directly, thus enhancing edge smoothness which is critical to achieving a high-quality facet. Due to the higher etch rate of the SAL601 mask it is necessary to use thermally assisted CAIBE etching at 110°C in order to obtain more nearly vertical sidewalls. The result of this etch are shown in figure 2. As can be seen, the etch is extremely clean and the resulting sidewalls are nearly vertical. Thus, either the low

temperature Cr-mask etch or the high temperature SAL601-mask etch would most likely be suitable for the fabrication of short-cavity CAIBE etched facet lasers. The Cr-mask does have the advantage that it is self aligned and there is no need for an additional lithography step to fabricate the p-contacts. In addition, the possibility of overlaying SAL601 on top of P-metal without Cr as a covering layer (i.e. the very easily sputtered Au was the top layer) was investigated. The resist thickness and overlay of the resist is crucial, as it is imperative not to expose the easily sputtered Au to the ion beam. This process is somewhat more complicated and the additional complexity of the processing does not provide superior results. Therefore, when laser fabrication resumes in the Fall, either the self-aligned, unheated Cr-mask process or the high-temperature SAL601-mask process will be utilized to fabricate devices with cavity lengths as short as 100 μm. Etch characterization for laser sources was completed in late June. Beginning in late June work on polymer waveguides was initiated to further explore the required components necessary for optical interconnect schemes.

POLYMER WAVEGUIDES

Advances in VLSI technology have greatly enhanced on chip processing speed, but the interconnect technology that serves to connect modules on chips and move information from chip to chip has not advanced at the same pace. The problems associated with RC charging times, power dissipation, and surface area consumption have placed severe constraints on current state -of-the-art VLSI systems. Optical interconnects, on the other hand, offer a high-bandwidth, high fan-out, interconnect scheme that is not limited by rf interference, or capacitive loading. Obviously, the electrical interconnect technology is fairly well developed and has the advantage that the associated processing techniques in forming the connections are fairy robust and sophisticated. There are a myriad of choices in the design of optical interconnect schemes, the design of the sources, waveguides, and detectors. In addition, plane-to-plane connections can be accomplished by either free-space interconnections or guided wave interconnections, each with its own advantages and disadvantages. Free-space interconnections avoid the problems associated with the mating and unmating of fiber connections and also allow signal beams to cross without interference. Work on plane-to-plane interconnect schemes is complimentary to the development of in-plane optical guides, and development of both is necessary to fully make use of all the advantages that optical interconnections offer for wafer scale integration and wafer scale hybrid integration. Thus, both multi-layer plane-to-plane interconnect schemes and in-plane optical distribution waveguides were studied.

The waveguiding material system that was studied for this project was the polyimide system. Due to the hybrid wafer scale systems that are being considered it was necessary that a waveguide could be fabricated on almost any substrate. Polyimides can be deposited on almost any substrate and are compatible with a variety of GaAs and Si processing techniques. PMMA and other polymers have optical properties that are as good as polyimides

themselves and could also be considered for waveguiding or cladding materials, but they have rather poor thermal stability. Polyimides are stable up to 350°C and have excellent mechanical properties as well. Thus it was proposed to use polyimide as the guiding material and either PMMA cladding or metal cladding. The processing of polyimide itself was examined in order fabricate the guides themselves irrespective of the cladding type chosen. In general, PMMA cladding offers ease of processing and is particularly simple if multimode guides are being considered, while metal-clad guides might be a more suitable choice for single mode guides that necessitate sharp bends. The issue of single mode and multi-mode guides might be a superfluous one, since the modal dispersion limit for moderately multi-mode guides is not really a limitation for the relatively low data transmission rates and short distances being considered. Modal and material dispersion considerations are usually more applicable to communication systems.

Optimization of the waveguides for low loss is a major consideration for low power optical interconnection networks. In general, it is desirable to have losses smaller that 1 dB/cm. To achieve such low losses polyimides have very specific heat treatment cycles. A study of the effect of the curing cycle on the optical properties of polyimide carried out by Franke and Crow of IBM's Watson Research Center show that the optical losses increase substantially for the "normal" curing cycle that is suggested by most polyimide manufactures (when the polyimide is used as an interlayer dielectric). According to Franke and Crow, the as-spun polyimides should be dried for 1 hour at 60°C plus 30 minutes at 90°C to remove moisture trapped by the solvent system used in polyimides, n-methyl-2-pyrrolidone (NMP). NMP can easily pick up moisture from the air so the drying is necessary to remove all water molecules that could produce scattering centers if the polyimide were immediately exposed to the high temperature curing cycle. Losses can be kept below 1 dB/cm if the solvent bake off temperature following the low-temp drying cycle is kept below 190°C. If the high-temp bake exceeds the boiling temperature of NMP (202 °C) than the losses can be increased up to 4 dB/cm which would be unacceptable for waveguiding material. It is always necessary to avoid the boiling of any liquids still present with the polyimide as this causes the production of voids and pinholes that serve as scattering centers. Thus, due to the low temperature bakes that are necessary to partially cure low-loss polyimides, these materials, at present, are limited to processing and operating temperatures of less than 190°C

There are several choices in the type of polyimide chosen and the means of forming a waveguide in the polyimide. The initial design of a single-mode waveguide, done on Mathematica using TE, asymmetric slab boundary conditions and the effective index method, appears below:

Figure 3- Single Mode Polyimide Ridge Waveguide
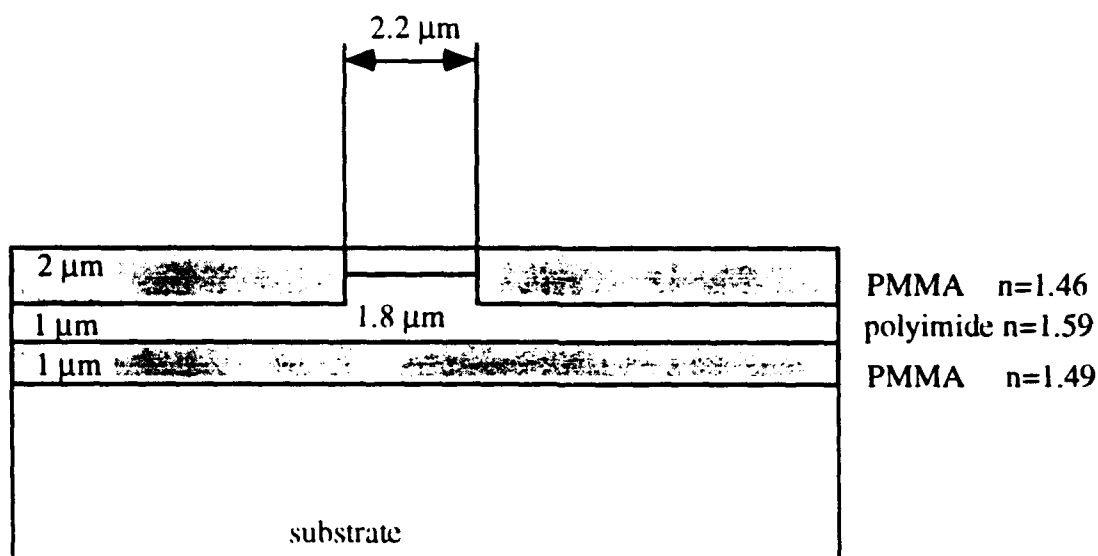
This design assumed Dupont PI-2566 as the guiding material, PMMA baked at 180°C as the lower cladding, and PMMA baked at 90°C as the upper cladding. The differences in the refractive indexes of the PMMA are due to the different bake temperatures. This design was done to get a feel for the types of dimensions that would be necessary for single mode guides and served as a starting point for the work that was to be done. As was stated previously, single mode guiding is most likely not a primary consideration for these waveguides and therefore waveguides were designed with lateral widths ranging from 10 μm to 100μm. Material systems for waveguiding structures were then examined.

A number of polyimides have been previously characterized and their optical properties evaluated. The following results for commercially available polyimides were obtained by Hewak and Jerominek of the National Optics Institute (Canada):

Tradename	Manufacturer	Transmission (@633nm)	Cutoff(nm)
Probimide 293	Ciba-Geigy	98.4%	420
Probimide 348	Ciba-Geigy	97.0%	513
Probimide 412	Ciba-Geigy	97.8%	350
Pyrallin 2566	Dupont	99.0%	370
SPI-135	G.E.	99.4%	375
Thermid 5010	National Starch	97.1%	435

Thermid 5510	National Starch	98.6%	402
T-Polyimide	Brewer Science	97.7%	365
Selectilux HTR-3 Merck 96.0%			550

The NNF carries the Ciba-Geigy polyimides, so all initial processing work was carried out using these. For low-loss structures, Dupont PI-2566 is most likely the polyimide of choice. In addition to these standard polyimides, which are often used as interlayer dielectrics in multi-level metallizations for VLSI chips, another choice of polyimide was found, the Pyralin photo-definable polyimides. These types of polyimides are sensitive to g-line radiation and can therefore be directly patterned without the necessity of first using a making material of some sort. This greatly simplifies the processing sequence involved and can be adapted to a variety of in-plane interconnection schemes. Photo-imagible polyimides give yet another processing choice that can be used for the processing of waveguides.

There are a number of waveguide structures that can be fabricated, the choice of which depends on the specific application. An important fact about the polyimides studied here is that they are attacked by most solvents, alkaline developers, and ketones (such as MIBK). Thus, the processing of resists on top of polyimide is tricky at best. In fact, Dupont suggests doing the photoresist development and etching of the polyimide at the same time. In addition, using a resist etch mask for a RIE etch of the polyimide does not enhance the anisotropy of the etch, as the resist does not provide any selectivity in a O_2 RIE etch. This is shown in figure 4 where a SAL 601 negative e-beam resist mask was used to RIE etch Probimide-287 (Ciba-Geigy). Therefore, if sidewall slope is not an issue, a wet etch during the development of the masking photoresist would most likely be the most appropriate choice. However, if sidewall slope is critical for the reduction of bending losses for the fabrication of 45 degree in-plane mirrors, then a non-erodable etch mask must be used. With these criteria in mind, two separate processing sequences were designed and are shown in figure 5. Obviously, the Ti-etch mask process is substantially more complicated than the process using the photo-imagable polyimide. This is especially true due to the necessity of using a soft-baked polyimide that is susceptible to standard developers and solvents. An e-beam bilayer resist structure (6000 Å P(MAA-MMA), 2400 Å PMMA) was used to obtain an appropriate resist profile for completing the Ti-liftoff. The development of the resist was done in AZ2131 thinner which is 85% (by weight) cellusolve acetate and 15% n-butyl acetate. It was found that a 15 second development of the bilayer was sufficient to clear the resist out of small areas that received a 100 μC/cm2 dose. This is much shorter than the standard develop time of 60 seconds in a 1:1 solution of MIBK:IPA. A 60s O_2 plasma descum was done and then 2000 Å of Ti was evaporated on the surface. At this point, an appropriate choice of liftoff solutions is required. Acetone was attempted and proved to be entirely

unsatisfactory, as shown in figure 6. As of the time of this writing the author is attempting to find a more appropriate liftoff solution, most likely some combination of n-butyl acetate:IPA or a commercial Ciba-Geigy product. Another possibility is to use a selective RIE of Ti evaporated over the entire surface of the polyimide. In this case, the Ti is evaporated immediately after the low temperature curing of the polyimide. A negative e-beam resist (SAL 601) is then used to pattern an erodable mask that will serve as an etch mask for the Ti. The Ti is selectively etched by using a CF4 RIE that will not attack the polyimide. Then the polyimide is then etched in a O2 RIE using the Ti as a non-erodable etch mask. In this manner, developers or solvents never come in contact with the soft-baked polyimide. Of course, the photo-imagable polyimide process simplifies the fabrication considerably and may be entirely adequate for the guiding purposes envisioned by the Rome Photonics Center.

In addition, ion milling experiments were carried out to determine if Focussed Ion Beam (FIB) processing of polyimide was feasible. The fabrication of 45 degree mirrors to reflect vertical optical signals into horizontal waveguides was examined on the Cornell FIB machine. These experiments showed that the sputter rate of even softbaked polyimide is rather small and that large doses are required to adequately ion mill waveguides. Cornell's FIB has demonstrated the ability to fabricate 45 degree mirrors in Si, but the large doses required for the ion milling of polyimide require excessively long and expensive exposures. As the FIB is currently not set up for the introduction of reactive gasses into the chamber, the fabrication of 45 degree mirrors will be quite difficult. Oxygen CAIBE processing with a tiltable stage offers a much quicker and less expensive processing possibility for these types of structures.

CONCLUSIONS

Optical interconnection of VLSI modules will become increasingly necessary as processing speeds are continually improved. This work provides a cursory examination of the relevant details for fabricating high-speed optical sources and in-plane polyimide waveguides. Fabrication of lasers with bandwidths in the 20 GHz range has already been accomplished and work this summer was aimed at pushing the maximum attainable bandwidth into the 30 GHz range. CAIBE etching of p-doped, compressively strained, multiple quantum well lasers is being examined for this purpose. Not only will these devices be capable of ultra-fast operation, but they will have enhanced noise, dispersion, and power dissipation characteristics at lower frequencies. Waveguide fabrication techniques were studied and a variety of issues concerning the processing of low-loss polyimide guides were discussed. Several processes were examined and the suitability of each for a particular application was mentioned. The work that was carried out in conjunction with Rome Laboratory will be continued by Rome research personal in the Fall.

Figure 1 - Self Aligned Etch Mask using Cr/Au/Pd/Ti
6 minute etch
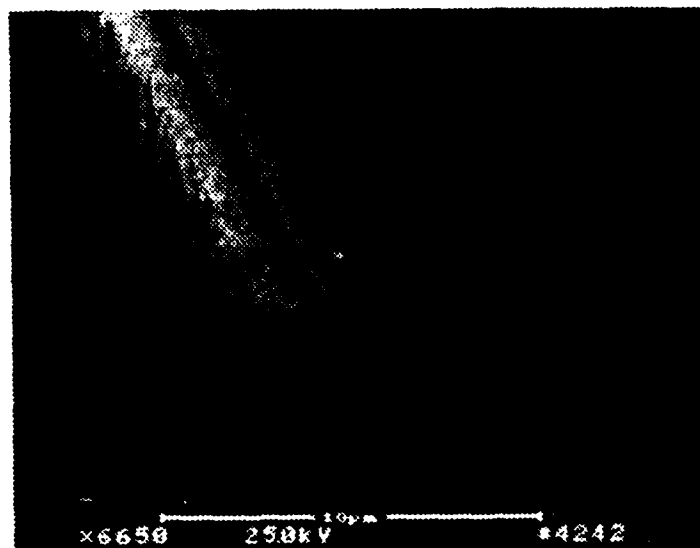
Fig. 1a - stage heated to 110°C
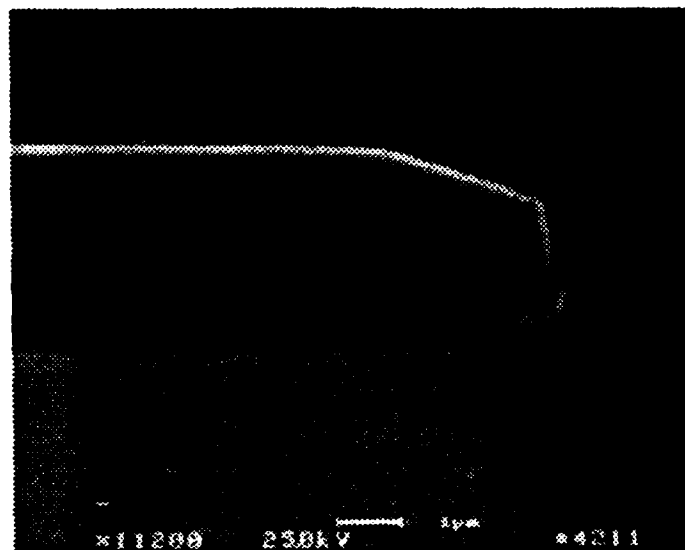
Fig. 1b - unheated stage

Figure 2 - CAIBE with SAL601 etch mask

9 minute etch

13-16
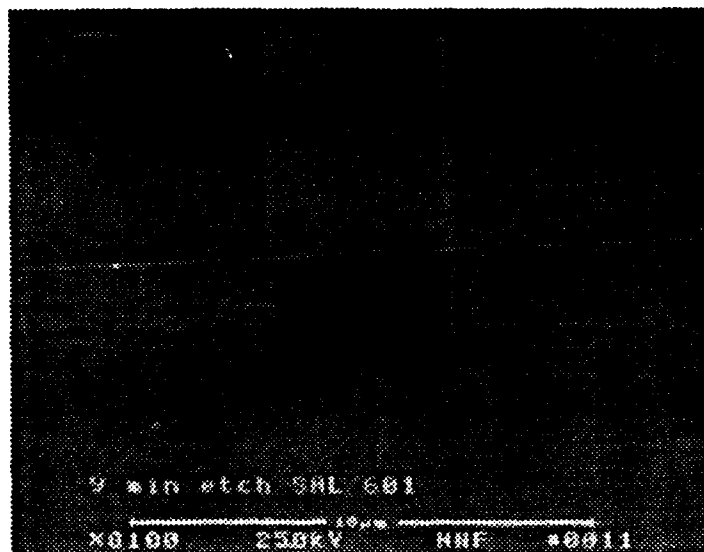
Figure 5 - Waveguide Fabrication Process Sequence

Single-Mode Waveguide Process   Multi-Mode Waveguide Process
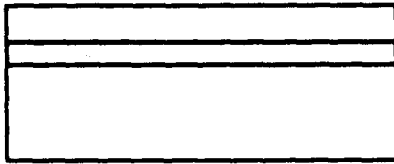
Spin Cladding and Polyimide
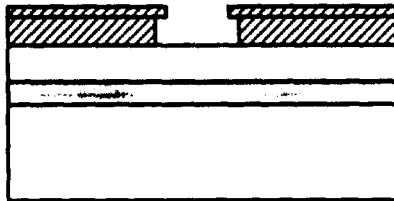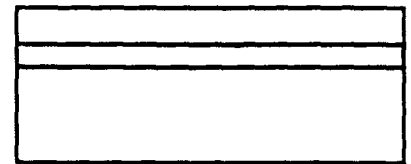
Apply Bi-Layer Photoresist
Expose and Develop Resist
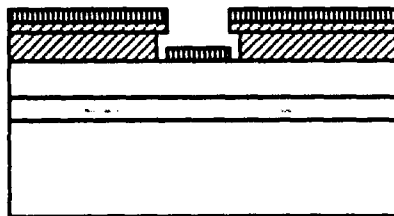
Evaporate Ti
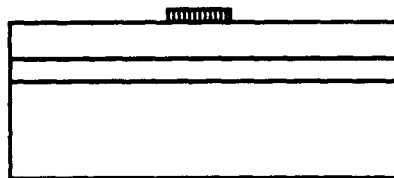
Lift Off   Expose and Develop
                            Photoimagable
                            Polyimide
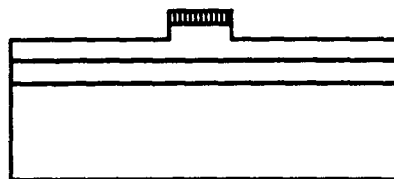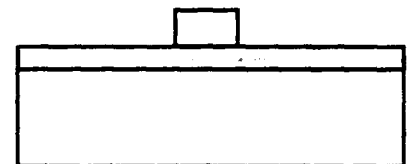
O2 RIE Ridge Waveguide
(remove Ti)
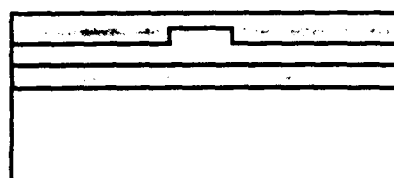
Spin Upper Cladding

13-17
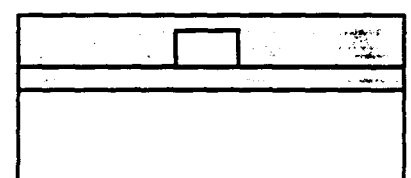
figure 4 - RIE of polyimide using SAL601 etch mask

RIE etch
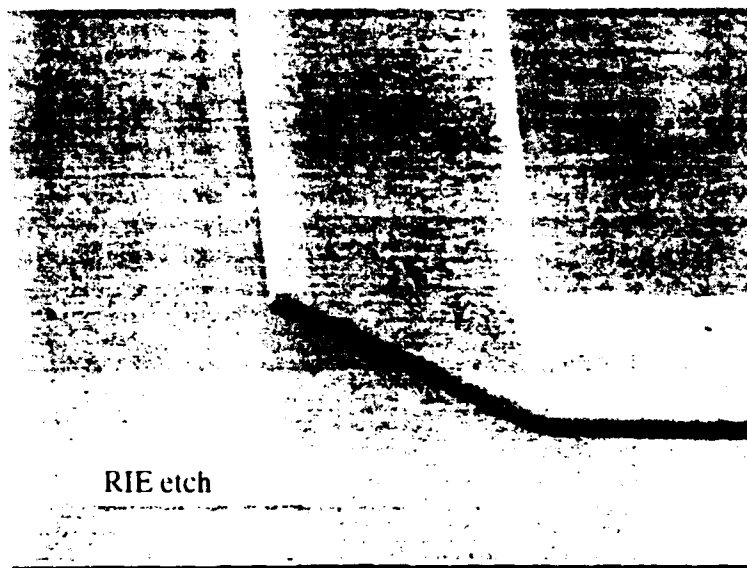
Figure 6 - Acetone Liftoff of Ti etch mask

13-18

REFERENCES

Lester, Luke - High Frequency Multiple Quantum Well Strained-Layer Lasers, Ph.D. Thesis, May 1992

Zory, Peter S (edited by). - Quantum Well Laser, ch.1-5, 1993

Zhao,B., Chen,T.R., and Yariv,A. - "The Extra Differential Gain Enhancement in Multiple-Quantum-Well Lasers", IEEE Photonics Technology Letters, Feb. 1992 pp. 124-126

Zhao,B., Chen,T.R., and Yariv,A. - "Effect of state filling on the modulation response and the threchold current of quantum well lasers", Appl. Phys. Lett., April 20 1992, pp.1930-1932

Uomi, Kazuhisa - "Modulation-Doped Multi-Quantum Well (MD-MQW) Lasers. I. Theory", Japanese Journal of Applied Physics, Jan. 1990, pp. 81-87

Ralston, J.D., Weisser, S., Esquivias, I., Larkins, E.C., Rosenzweig, J., Tasker, P.J., Fleissner, J. - "Control of Differential Gain, Nonlinear Gain, and Damping Factor for High-Speed Applications of GaAs-Based MQW Lasers", Fraunhofer-Institut für Angewandte Festkörperphysik, Freiburg, Germany - unpublished paper

Rooks, M.J., Roussell, H.V., Johnson, L.M., - "Polyimide optical waveguides fabricated with electron beam lithography", Applied Optics, Sept. 20, 1990, pp. 3880-3882

Selvaraj, R., Lin, H.T., McDonald, J.F., - "Integrated Optical Waveguides in Polyimide for Wafer Scale Integration", IEEE Journal of Lightwave Technology, June 1988, pp. 1034-1044

Hewak, D.W., Jerominek H., - "Channel optical waveguides in polyimides for optical interconnection by laser direct writing and contact printing", SPIE Vol. 1213 Photopolymer Device Physics, Chemistry, and Applications - 1990 pp. 86-99

Franke H., Crow, J.D., "Optical Waveguiding in Polyimide", SPIE Vol. 651 Integrated Optical Circuit Engineering III, 1986, pp. 102-107

HARDWARE IMPLEMENTATION OF ACTIVE JITTER COMPENSATION TECHNIQUE FOR OPTICAL COMMUNICATIONS SYSTEM

Mark A. Tascillo
PhD Student
Department of Electrical Engineering

Binghamton University
Binghamton NY 13902

HARDWARE IMPLEMENTATION OF ACTIVE JITTER COMPENSATION
TECHNIQUE FOR OPTICAL COMMUNICATIONS SYSTEM

Mark A. Tascillo
Ph.D Student
Department of Electrical Engineering
Binghamton University

ABSTRACT

An active jitter compensation scheme applicable to free space
intersatellite laser communication is developed. The theory, design and
implementation of the prototype are discussed. A performance evaluation of
the existing unit is presented. The theory and design of a second generation
self-tuned prototype is discussed including the concepts of on-line acoustic
path modeling, intelligent band-pass filtering, modular system construction
and optimization techniques. Finally, a pair of alternate technological
approaches applicable to the jitter cancellation concept are presented.

HARDWARE IMPLEMENTATION OF ACTIVE JITTER COMPENSATION TECHNIQUE FOR OPTICAL COMMUNICATIONS SYSTEM

Mark A. Tascillo

INTRODUCTION

Free-space laser communication holds considerable promise for high bit-rate data transfer. At present, several key technologies applicable to this communication strategy are under development. One such developmental technology is that of vibration control on-board the space platform. While the small vibrations caused by gyros, motors and thrusters on-board the spacecraft are of little concern to common radio and microwave communication systems, these small movements can throw a laser beam far off target as very small angular displacements of the source laser correspond to significant beam position error when amplified by the thousands of miles separating stations. This project involves the development of a prototype active jitter compensation technique applied to a laboratory based laser communication system.

SATELLITE JITTER EVALUATION

The ambient vibration on board the OLYMPUS communications satellite launched by the European Space Agency (ESA) was monitored using three orthogonally positioned microaccelerometers [1]. A small portion of this data was evaluated resulting in Figures 1-6. Figures 1-3 show a small segment of the signal from each of the three channels of vibration data. Of importance is the fact that while much of the vibration spectra lies below 200 Hz, there is a significant level of higher frequency vibration present on the orbiting platform (Figures 4-6). One may infer from this data that an existing vibration compensation scheme should have an effective bandwidth extending to at least 1 KHz to assure adequate performance.

EXISTING TECHNOLOGY

The fine steering mirror (FSM) is used to track out the small errors in beam position due in part to ambient vibration. This is currently performed

14-3

with the aid of a closed loop control using the optical feedback from the quad cell (an optical position sensitive device) [2]. This approach has been shown to adequately handle most disturbances below 500 Hz. However, quality of tracking degrades rapidly as the frequency of the disturbance should increase. It is believed that the effective control bandwidth may be increased by supplementing the existing approach with a feedforward controller [3].

ACTIVE VIBRATION CONTROL THEORY

The underlying vibration control theory applied to this project originates from the concept of "antiphase" sound and vibration cancellation. This technique involves the continuous monitoring of ambient noise near the source, and re-injecting the processed noise information back into the affected region. The noise is processed such that when re-injected it is exactly 180 degrees out of phase with respect to the source but of identical magnitude. In theory this approach completely eliminates the disturbance in the affected region [3].

The above concept has previously been applied with considerable success to the reduction of noise volume in confined (easily controlled) locations such as the motor noise[4], or in the creation of noise-cancellation headphones[5]. This approach has also been applied to the creation of active vibration isolation mounts for sensitive equipment [6].

All of the above examples rely upon the use of dedicated actuators for noise cancellation. It would be beneficial to develop a vibration isolation technique using existing actuators if possible. In the case of the PAT (position-acquisition-tracking) system, a high-bandwidth fine steering mirror (FSM) is available. In current application the FSM serves two purposes: [2].

1) Properly position the incoming laser beam upon the quad detector center after the coarse steering assembly has brought the beam to the vicinity of the quad detector center.

2) Track out external disturbances regardless of origin.

Therefore, the addition of on-line feedforward jitter-rejection allows the FSM to perform its current function, only with the advantage of having prior knowledge of all local vibration disturbances. In traditional

14-4

implementation, the FSM controller acts only in response to the detected deviation of beam position from the quad cell center. This results in any required correction being performed after the vibration has affected the communication session. The feedforward control is intended to correct for the effects of vibration before system performance is degraded. By design, once the incoming laser is properly positioned upon the quad cell by the control system, the outgoing laser will be properly positioned upon the receiving optics of the remote satellite as the beams travel along precisely the same path [2].

APPLICATION OF ACTIVE VIBRATION CONTROL THEORY TO PAT SYSTEM

The most basic theory of active vibration control would simply involve the task of regulating the gain and phase of the monitored signal such that it directly cancels the disturbance. This concept would be perfectly acceptable were the frequency response of the acoustic propagation path linear over the vibration source bandwidth (as with an ideal amplifier). In reality, the transfer function of the acoustic propagation path for typical mechanical structures would be of relatively low magnitude over most frequencies, but punctuated by a number of isolated spikes as shown in the magnitude plot of Figure 7. This requires that the vibration control algorithm incorporate a transfer function model of the acoustic path as viewed between the acoustic sensor and beam position on the quad cell.

The acoustic propagation path may be modeled as a bank of bandpass filters with each passband at a center frequency corresponding to one of the spikes in the magnitude plot. Added to the bandpass filter would be a specific gain and phase delay between sensor and each output. A mode amplifier scheme applicable to the propagation model of Figure 7 may be found in Figure 8 with a pair of bandpass filters, each channel being mapped to both azimuth and elevation FSM command inputs by unique phase and gain adjustment. In this case the center frequency of one filter would be tuned to 310 Hz and the other to 500 Hz.

The operational overview of the basic compensation scheme appears in Figure 9. Using this approach, the vibration source excites both the structural dynamics of the platform and the acoustic sensor directly. The filtering action of the structural dynamics is mirrored by the mode amplifier

14-5

of Figure 8. The mode amplifier drives the FSM to directly cancel the vibration effect on the optical path, minimizing the position error as viewed at the output of the quad cell.

An acoustic propagation model is valid for only the conditions of excitation under which it has been created. In other words, the acoustic model characteristics are dependant upon mechanical propagation paths between the disturbance location and acoustic sensor and the path between disturbance and optics. Given that the characteristics of this path may change as a function of thermal and gravitational effects, an acoustic model developed on Earth may not necessarily be relevant in orbit. Therefore, on line computation of the acoustic transfer function is highly desirable. this on-line model would be obtained by logging both acoustic sensor and quad cell data to the computer, computing the power spectrum of each and from the respective power spectra extrapolating the Bode magnitude and phase information, thus completely characterizing the acoustic transmission path in the frequency domain. The proposed system would be implemented almost completely in software with the exception of the sensors and digitally controlled analog signal processor. A more sophisticated compensator overview taking into account model variations is seen in Figure 10.

LABORATORY APPARATUS

The PAT system (built by Ball Aerospace) consists of three optical communications terminals with associated control and interface hardware. The primary station has full beamsteering ability (fine and coarse steering mirrors) and the two remote stations (separated from the primary station by approximately 40 feet) are restricted to laser communications send/receive functions [2].

The vibration effects were induced at the primary station by means of an acoustic actuator assembled from standard audio components. In this case the output of a function generator is amplified by an 80 watt amplifier which drives between 1 and 4 6X9 audio speakers. Such an approach is feasible due to the high vibration sensitivity of the optical components and has the benefit of providing the entire vibration bandwidth (10 Hz through 1 KHz) without disruption of the existing PAT system setup.

The active compensation approach discussed above was implemented using

14-6

commonly available analog hardware components and interfaced with the primary station controller, the development of which is discussed below.

BASIC ANALOG HARDWARE IMPLEMENTATION

The electronic acoustic model for jitter compensation takes the form of a specialized analog signal processor. This circuit contains the following core elements:

1) An adjustable band-pass filter with resonant frequency variable from 10 Hz to 1 KHz, 3dB bandwidth adjustable from 4 Hz to 1 KHz and adjustable gain. These units are in effect amplifiers of individual vibration modes, and are used to separate these active modes from the many other inactive modes that may be picked up by the sensors and constitute noise. In a practical implementation, several or effectively N mode amplifiers would be installed for M different sensors.

2) An allpass filter stage causes a phase delay adjustable from 0 to 180 degrees. A series pair of allpass filters allows variation of phase delay from 0 to 360 degrees. A 360 degree allpass filter module is installed between the mode amplifier and each output.

3) An adjustable DC offset applied to the output. The DC offset voltage is used to center the laser upon the quad detector, or in effect serves the purpose of the final fine tracking acquisition stage.

4) A summing junction at each output (azimuth and elevation) forming the composite of the outputs of each of the mode amplifiers. The respective gain applied to each input of each summing junction may be controlled by adjustment of the input resistances, eliminating the need for additional gain elements.

Figures 11-12 show the mode amplifier and allpass filter.

MANUAL TUNED SIGNAL PROCESSOR CONFIGURATION

The prototype vibration control module consists of a single input dual output analog mode amplifier as described above. This unit maps the output from one sensor into both azimuth and elevation compensation components. This module is manually tuned and serves the primary purpose of demonstrating and refining the developed concepts.

14-7

EVALUATION OF MANUAL TUNED SIGNAL PROCESSOR

The manually tuned mode amplifier was evaluated against the existing closed loop controller for compensation of various frequency forcing functions. During each test the output of the quad cell was digitally recorded on the PC using a data acquisition board. Results are shown for both 200 and 550 Hz forcing functions, both of which represent major resonant frequencies of the optical path.

Figures 13 and 14 compare the performance of the active feedforward vibration control with that of the existing feedback controller. In each figure, the first plot represents the system baseline noise level with no vibration or compensation. The second plot (upper right) shows the effect of uncompensated jitter on the system. The lower left plot demonstrates the effect of feedforward compensation on the output. Finally, the lower right plot of each figure shows the performance of the existing feedback control.

When driven at 200 Hz it is shown that the jitter rejection provided by the feedforward system is roughly comparable to that of the existing closed loop system (Figure 13). However, at higher frequency it is seen that the compensation level of the feedforward technique is greater (Figure 14). One will note that the jitter rejection "floor" is at roughly -10 dB. This is the pre-existing noise level as shown by the first plot of each figure entitled "noise baseline". This noise value represents the sum total of all ambient vibration, air current and electrical effects present on the system before specific external excitation and/or compensation is applied.

ELECTRONIC CONTROLLED SIGNAL PROCESSOR

Given the above-mentioned ability to compute the acoustic transmission model on-line, the next step in the jitter cancellation implementation is the addition of electronic control to the mode amplifier stages. This task entails the replacement of the potentiometers with a voltage controlled resistive element. This may be accomplished through the use of JFET transistor networks. The source-drain resistance of the JFET transistor varies with proportion to the gate voltage. Electronic controlled versions of each mode amp component are shown in figures 15-18.

Given the ability to tune the signal processor electronically, one may now connect each of the electronic controls to the digital computer in which

14-8

the acoustic transmission model is computed on-line.

ENHANCEMENT THROUGH THE USE OF OPTIMIZATION ALGORITHMS

While the computed acoustic transmission model offers a close approximation to the real acoustic path it is (as are all models) imperfect in operation, and suboptimal in performance. This is due in part to a lack of a coherent feedback to the model design algorithm some indication of the model's performance. An ideal feedback mechanism would take the computed model parameters as an initial guess and then improve upon them such that a cost function is minimized. The well known SIMPLEX algorithm performs the above task. It initializes the model parameters to those supplied by the modeling algorithm and manipulates them such that a cost function served by the quad cell position error shall be minimized.

ALTERNATE JITTER CANCELLATION APPROACH UTILIZING AN INDIRECT METHOD

The above developed jitter cancellation approach relies primarily upon the use of sensors to provide the spectral information for cancellation of noise. This approach is quite viable in that the noise must be present and detectable before there should be a need for cancellation. Furthermore, the sensors are necessary for the modeling of the acoustic transmission path which facilitates system tuning. However, should vibration spectra affect the laser position without exciting the acoustic sensors, there would be insufficient input to the mode amplifier jitter cancellation circuitry. In this case, the frequency spectra of interest may be synthesized by a bank of function generators supplementing the mode amplifiers. The phase delay, gain, signal mixing and dc offset networks would remain in place and perform exactly as with the mode amplifier configuration.

This approach has the advantage of performing completely independent of the sensor information however it has the drawback of slower performance. The computer must now determine changes in vibration magnitude by observing the quad cell error exclusively. Limiting the tuning algorithm to only quad cell input leads to difficulty in determining the exact nature of variations in vibration magnitude given that the quad cell is affected by both the real and manufactured jitter waveforms.

DIGITAL IMPLEMENTATION OF JITTER CANCELLATION APPROACH

One may observe that all of the analog hardware discussed above can be implemented in software on a digital computer. In such a system, one would not be locked into a fixed number of mode amplifiers connected in a fixed input-output structure. However, the computer imposes a number of limits of its own. The first issue is that of speed. The maximum number of input-output connections processed by the computer would be limited by its maximum speed. Furthermore, the speed of each computation loop is affected by the number of computations performed with each iteration. In implementing a fully digital system, the handling of phase delays for the compensation signal would be made complex by variations in the number of signals simultaneously processed as a function of the number of currently active vibration sources.

CONCLUSION

The concept of active vibration control through the use of feedforward compensation techniques and existing actuators has been demonstrated in the laboratory environment. It is found that the feedforward compensation technique demonstrates performance superior to a traditional fine tracking control implementation at frequencies above 500 Hz.

Practical implementation requires an intelligent computer controlled system capable of monitoring and identifying frequency components of interest. The theory and design background for such a controller has been introduced.

REFERENCES

1) Witting, M. L. Van Holtz, D.E.L. Tunbridge and H.C. Vermeulen: "In-Orbit Measurements of Microaccelerations of ESA's Communication Satellite OLYMPUS". SPIE Vol 1218 Free-Space Laser Communication Technologies II, 1990 pages 205-214.

2) Skormin, V.A. M.A. Tascillo, C.R. Herman and D.J. Nicholson: "Mathematical Modeling and Simulation Analysis of a Pointing, Acquisition and Tracking System for Laser-Based Intersatellite communication". Optical Engineering, Nov. 1993.

3) Skormin, V.A., M.A. Tascillo and D.J. Nicholson: "A Jitter Rejection Technique in a Satellite-based Laser Communication System". Optical Engineering, Nov. 1993.

4) Lundgarten, Harris: "The Cancellation of Repetitive Noise and Vibration by Active Methods". Sound and Vibration, Oct. 1987.

5) Bartels, Volker: "Headset with Active Noise-Reduction System for Mobile Applications". Journal of the Audio Engineering Society Vol. 40, No. 4 April 1992, pages 277-281.

6) Sommerfeldt, Scott D. and J. Tichy: "Adaptive control of two-stage Vibration isolation Mount". Journal of the Acoustic Society of America: Vol 88, No. 2 August 1990. pages 938-944.
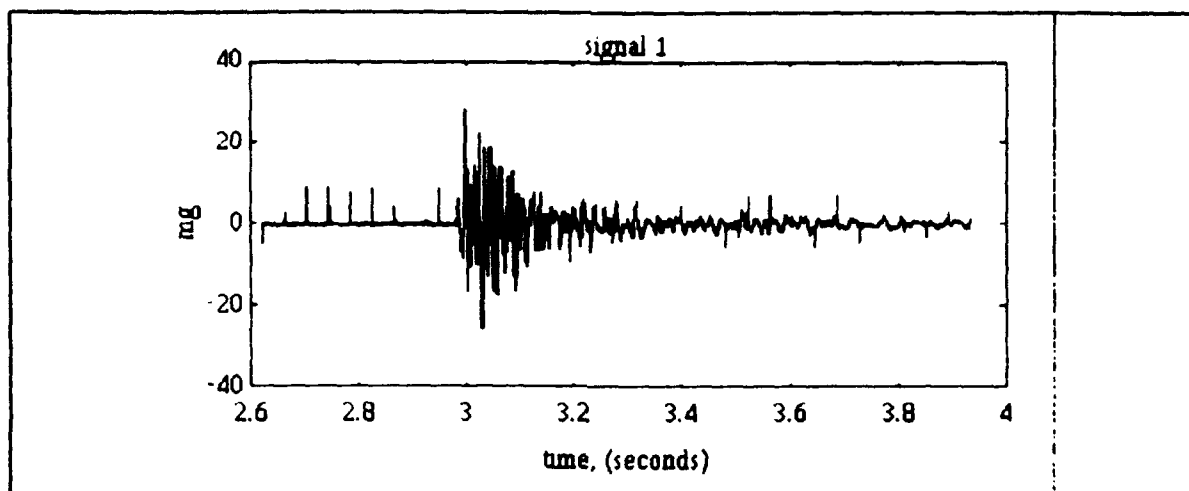
Figure 1

Figure 2

Figure 3

14-12
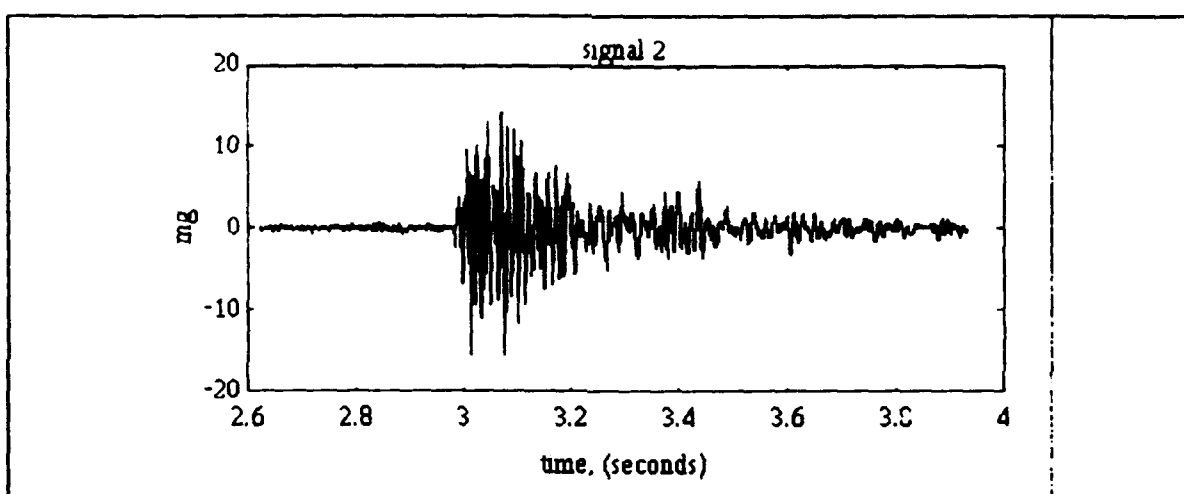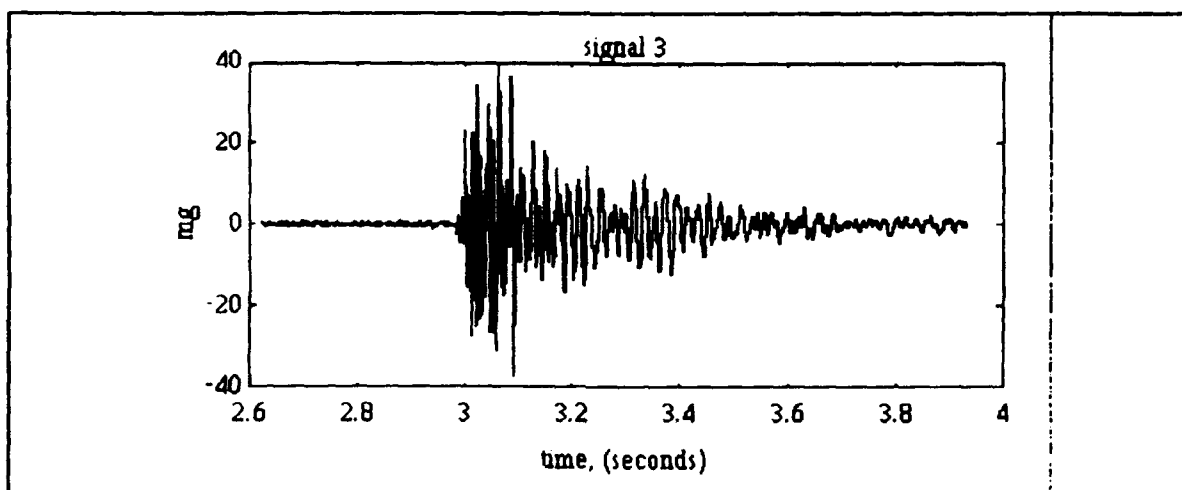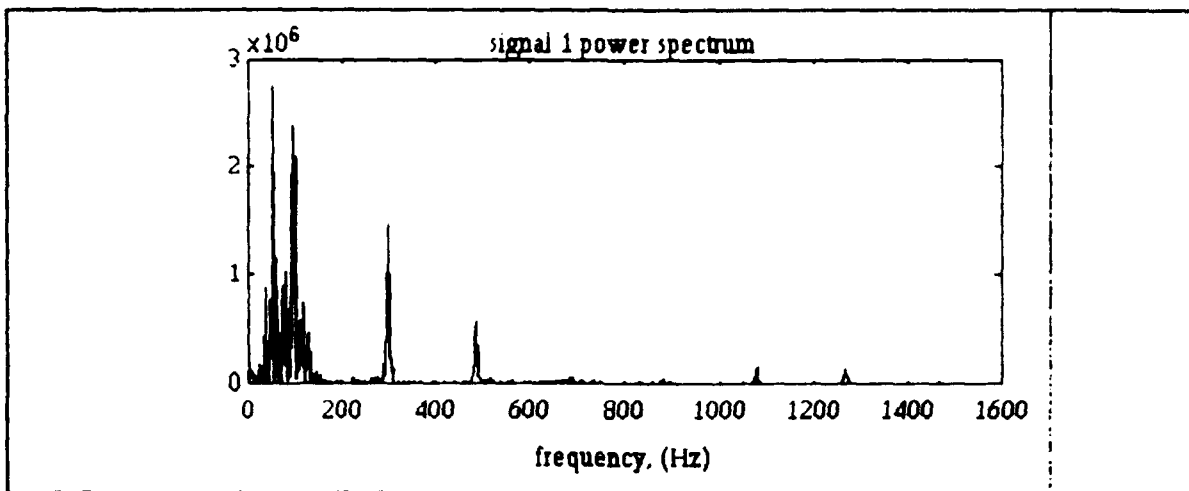
Figure 4

Figure 5

Figure 6

14-13

Figure 7

1 CHANNEL - 2 FREQUENCY   MODE AMPLIFIER

Figure 8

14-14

COMPENSATOR OPERATIONAL OVERVIEW

Figure 9

COMPENSATOR OPERATIONAL OVERVIEW

Figure 10

14-15

BANDPASS FILTER

Figure 12

STANDARD PHASE ELEMENT

Figure 11

14-16
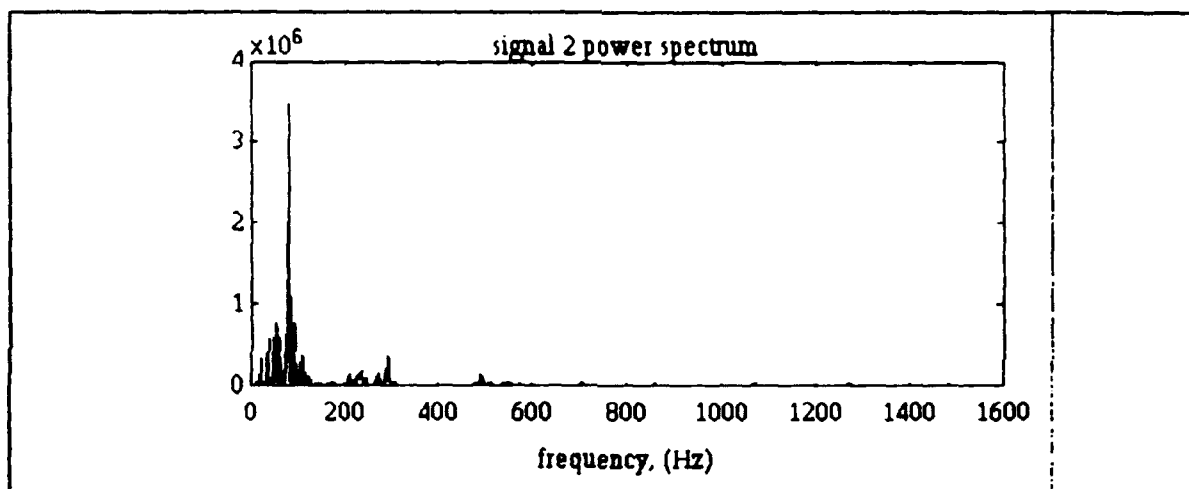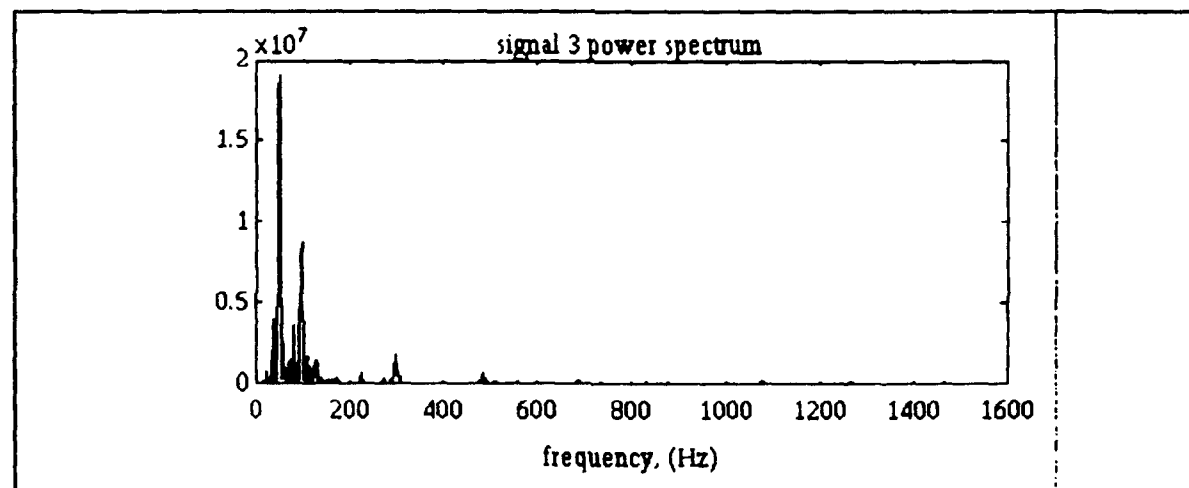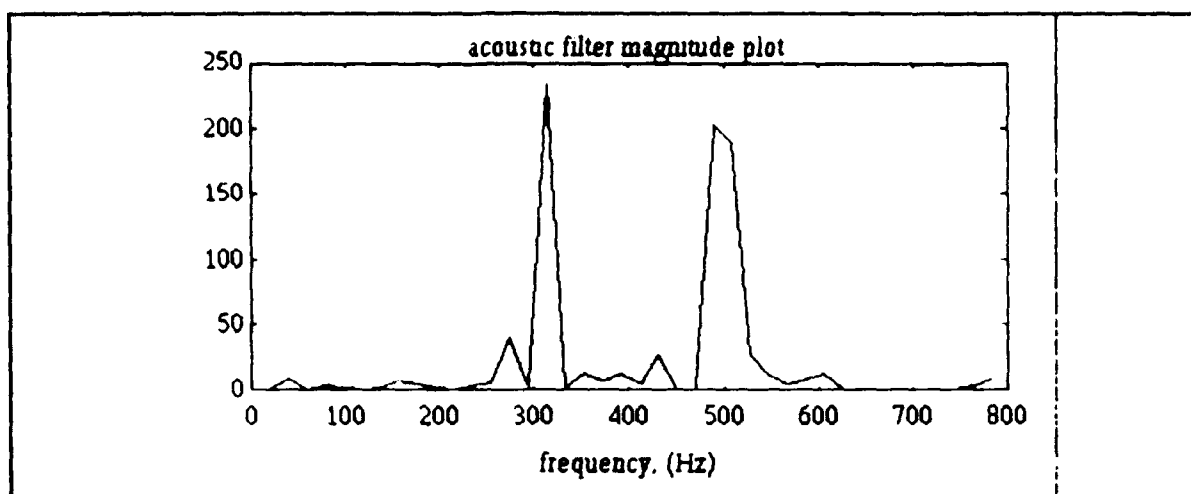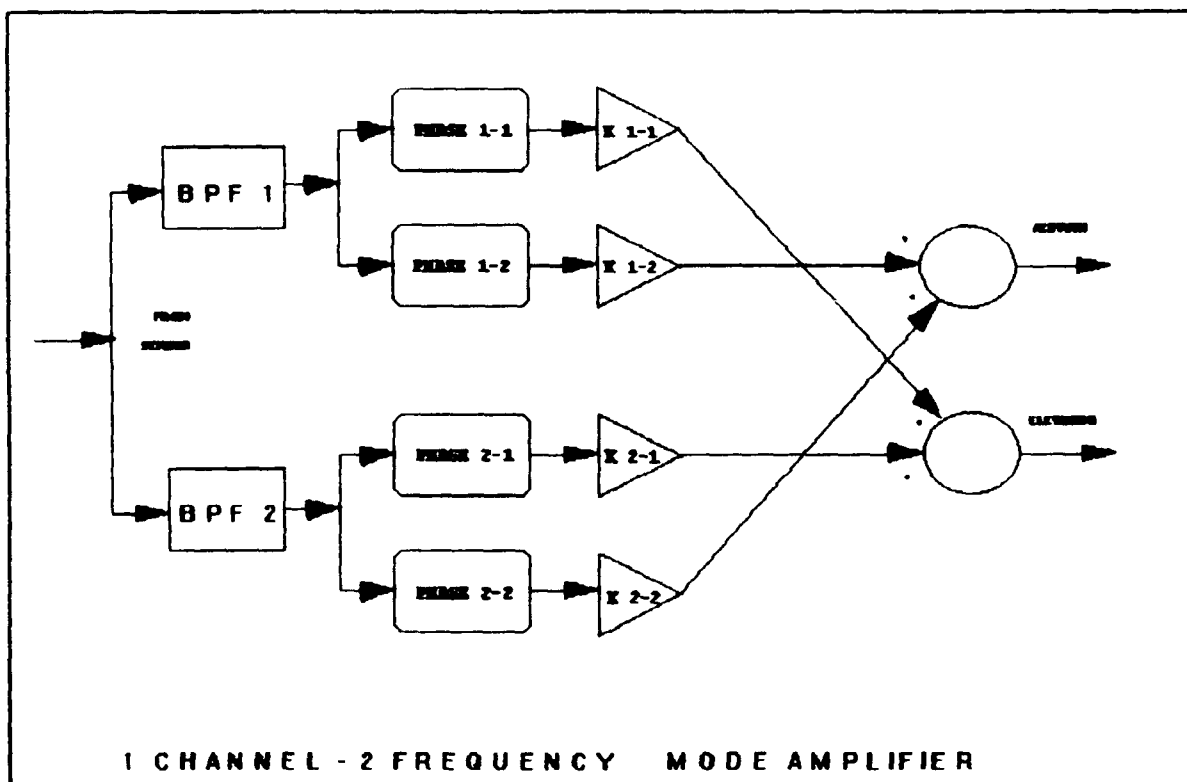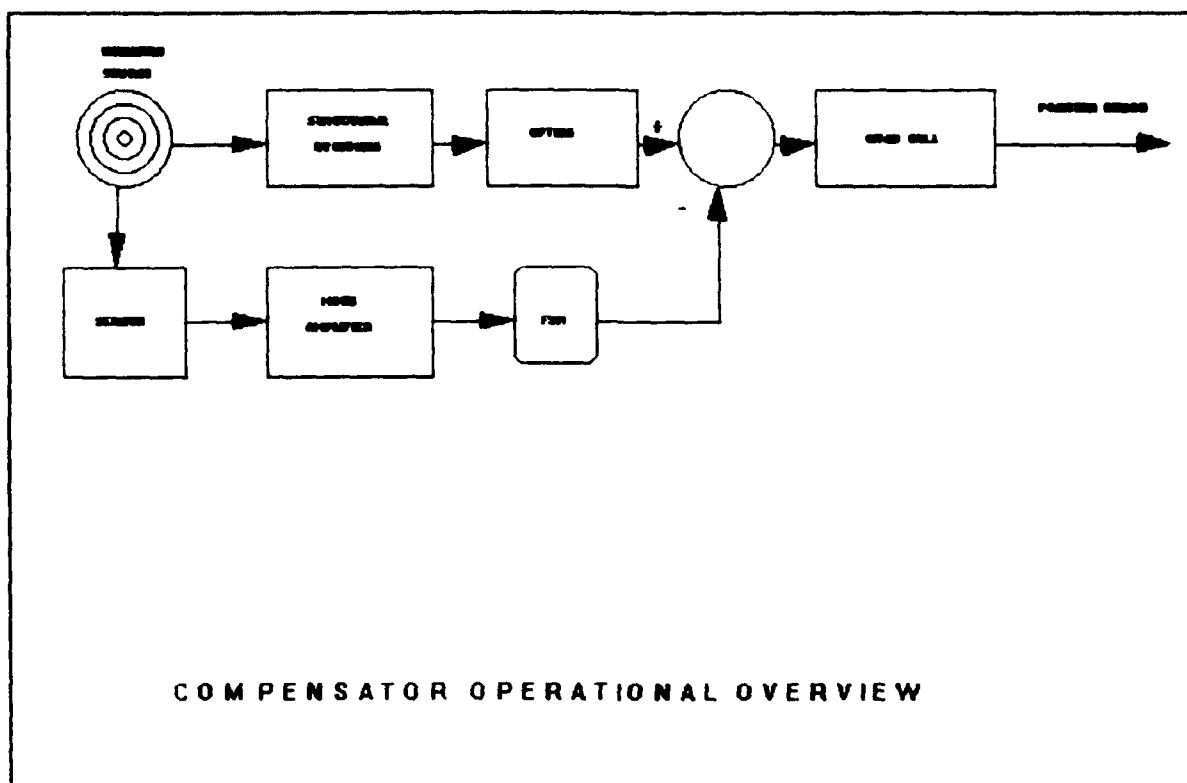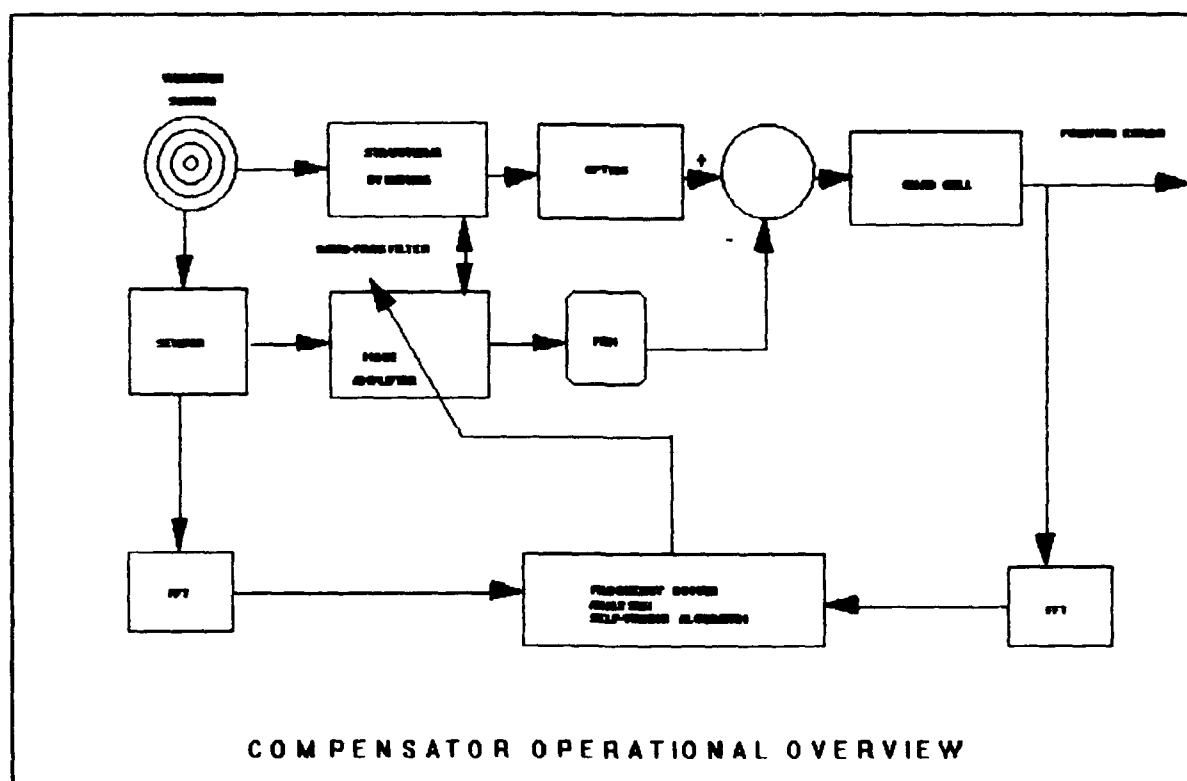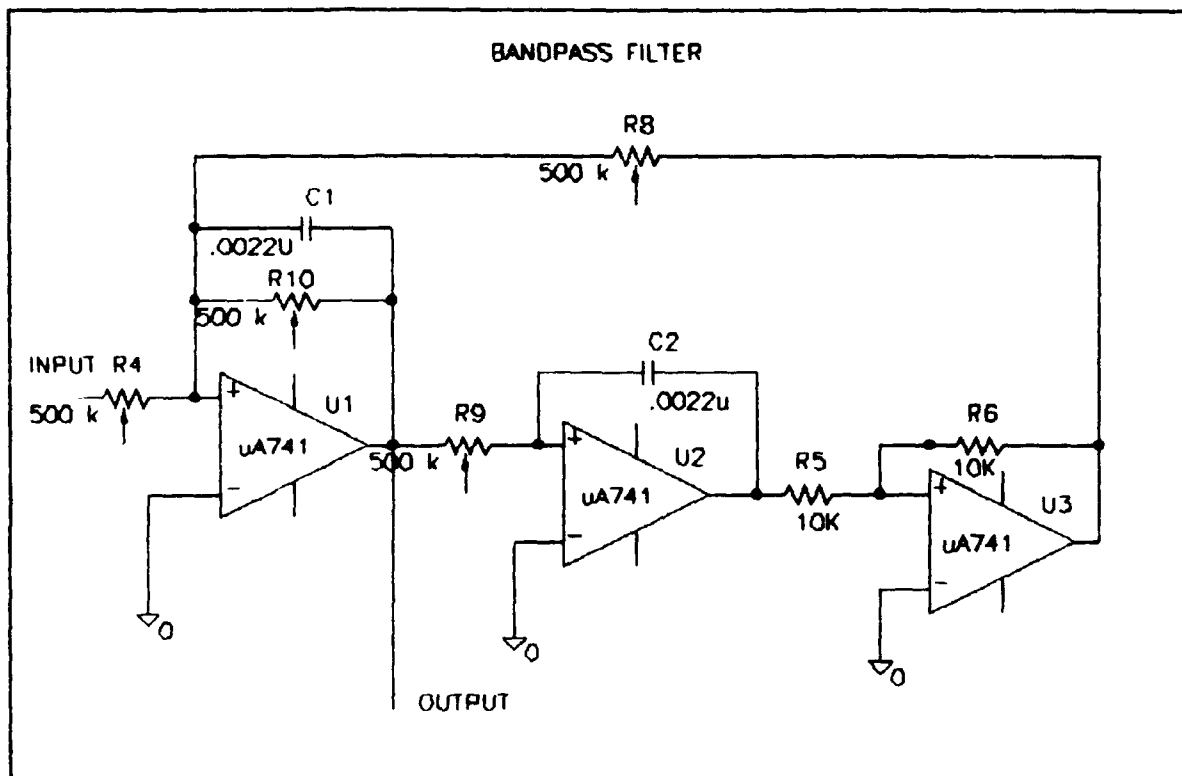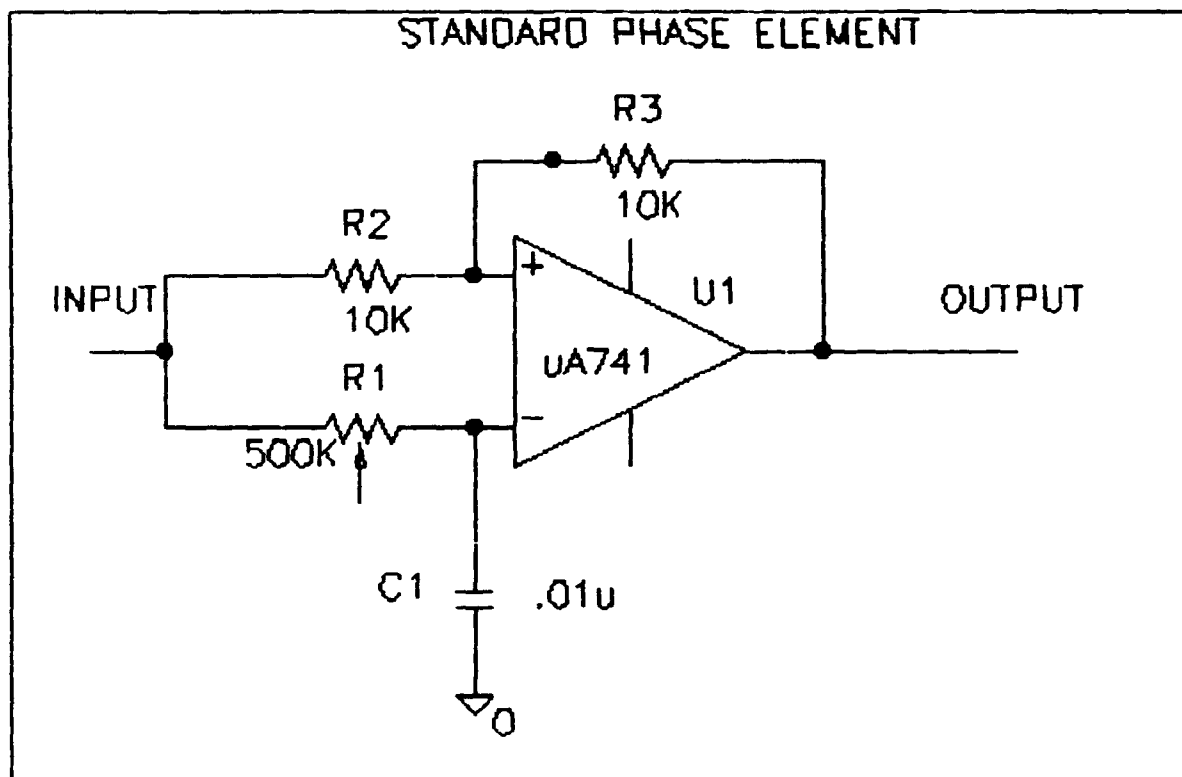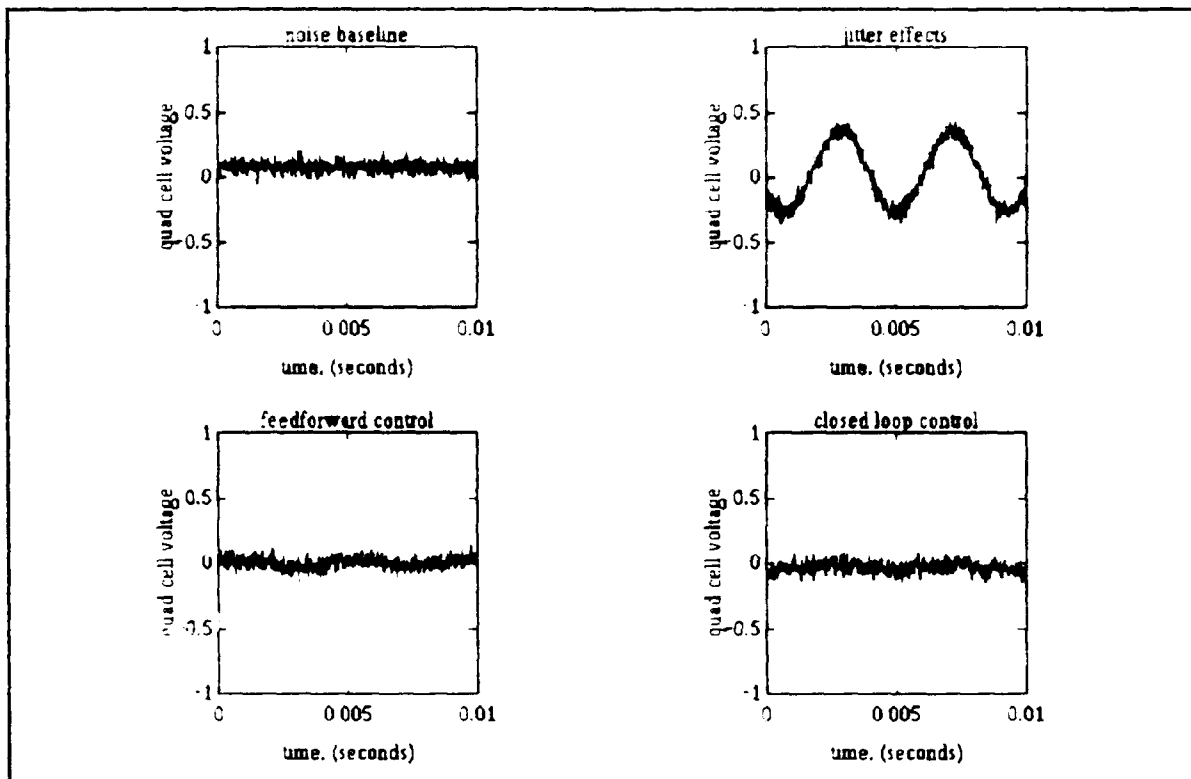
Figure 14, 200 Hz jitter source
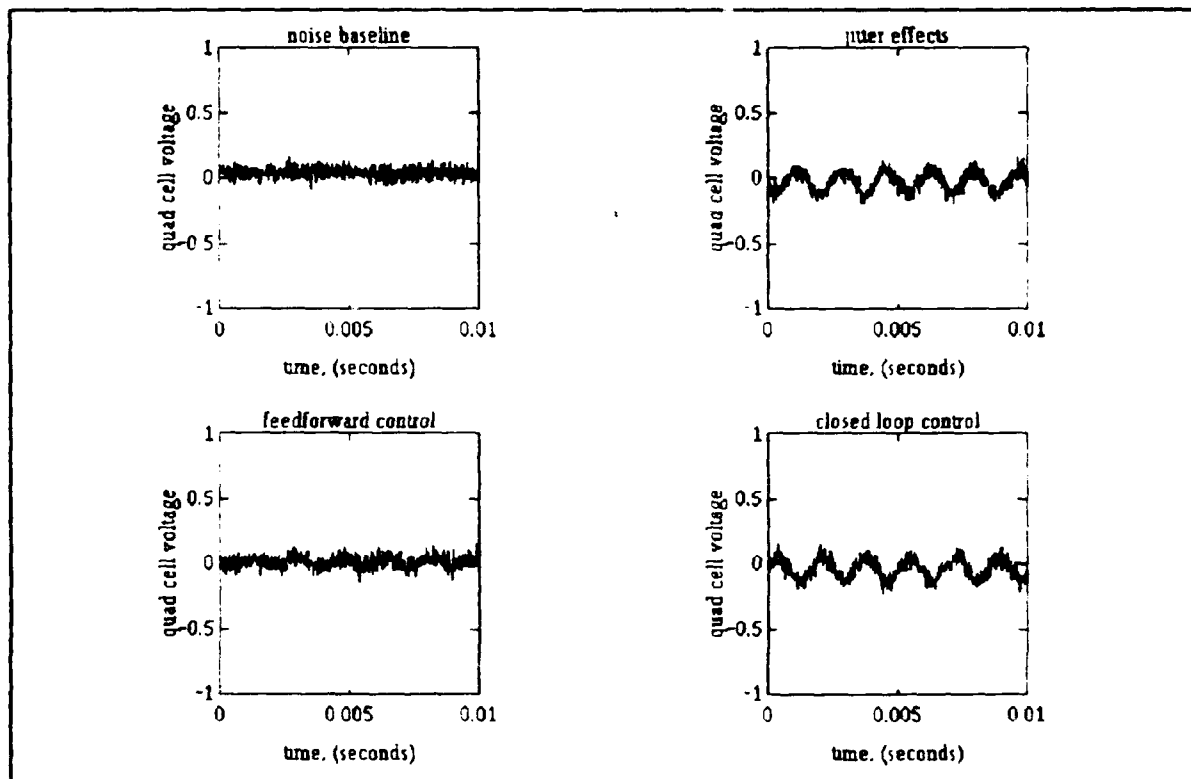
Figure 13, 550 Hz jitter source

14-17
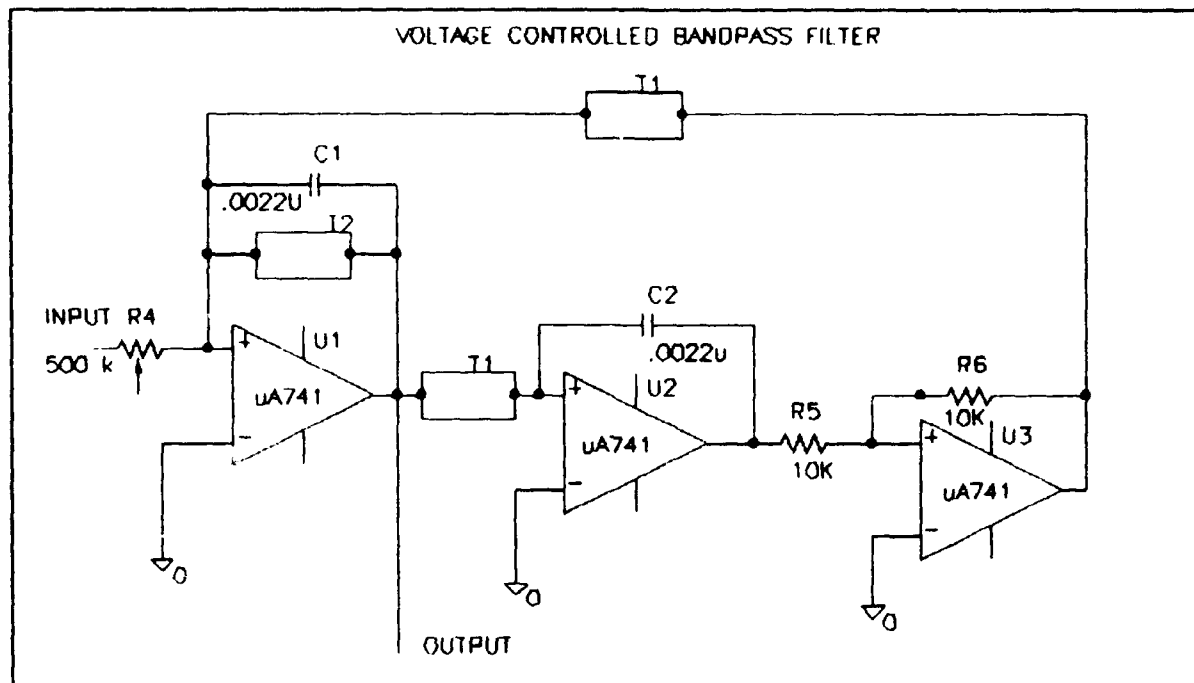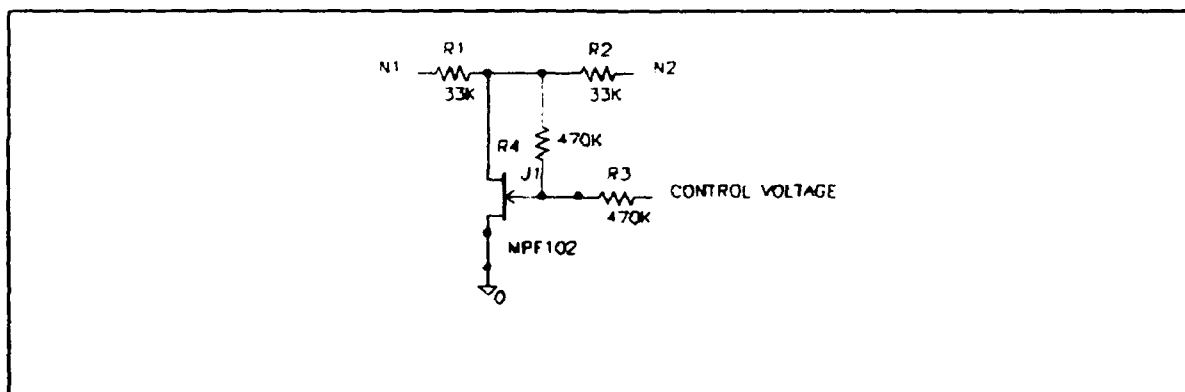
Figure 17

Figure 16, resistive element T1
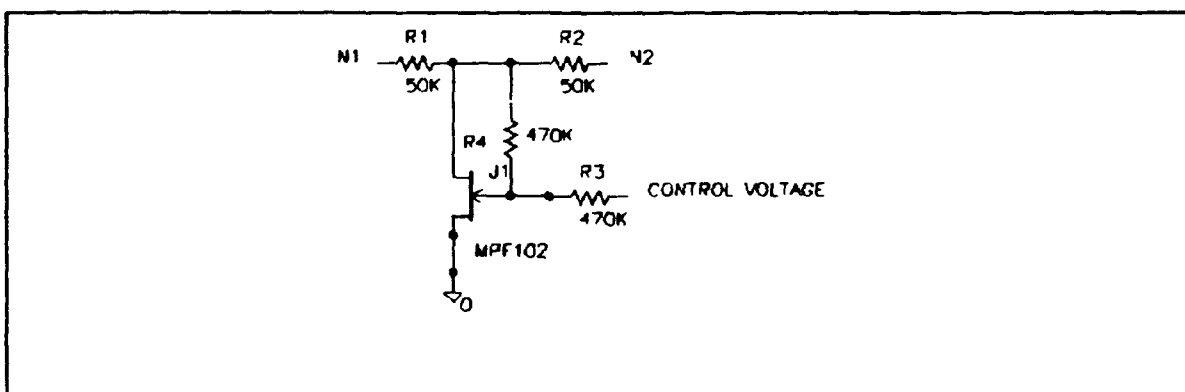
Figure 15, resistive element T2

14-18

VOLTAGE CONTROLLED PHASE ELEMENT
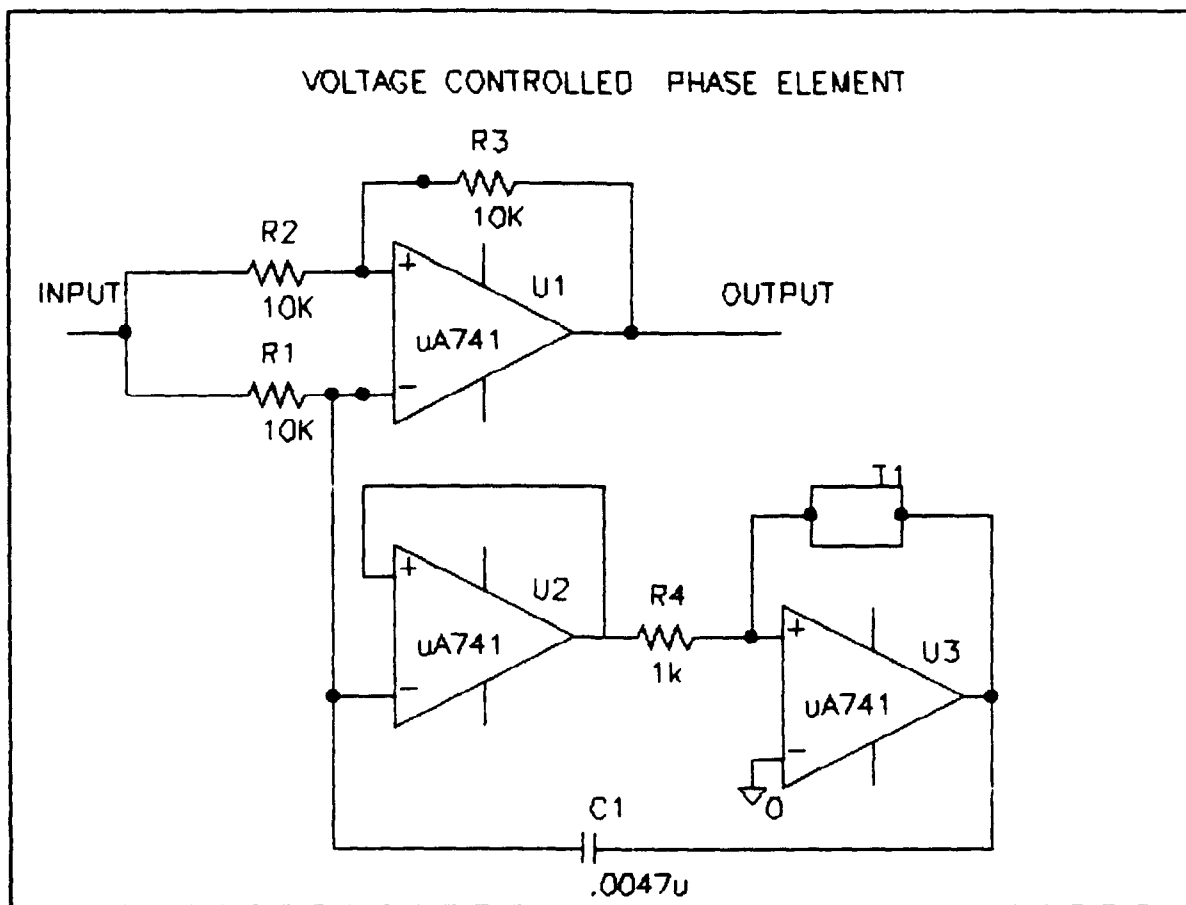
Figure 18

PROTOTYPING A MODELING AND SIMULATION
INFORMATION NETWORK

Mary Dobransky

Ph.D. Student

Department of Systems Science

Binghamton University

Binghamton, NY 13902-6000

PROTOTYPING A MODELING AND SIMULATION INFORMATION NETWORK

Mary Dobransky

Ph.D. Student

Department of Systems Science

Binghamton University

Abstract

The growing amount of information made available through advances in computer technology has created a need for systems that streamline information and make it easy to access. This paper discusses the design of such a system for computer modeling and simulation. Called a modeling and simulation information network, the system brings together information, tools, and technology within a simple interface. The design of the information network, which can be extended to other fields as well, is discussed in terms of a prototyping environment.

Introduction

The stone age has passed, the iron age has gone, the industrial age has ended. We are living in the age of information, and the computer rules. Money is managed by computer, a teenager working in a grocery store uses one, and the Federal Express delivery person brings one to the door to sign for the package on. There is no question that computers have changed the course of human history. Armstrong and Aldrin couldn't have gone to the moon without them. Computers have changed the way houses are built, cars are manufactured, and wars are fought. They have also released a wealth of information into the universe, the magnitude of which is almost unimaginable.

Unfortunately, this wealth of information does not always make our lives easier. Too much information can be just as bad as not enough. One of the biggest problems in an information-driven society is that it is not always easy to access available resources and services. Often we don't know where to look, and even if we do there is simply too much to keep track of. Is there any academician without at least one unread book, paper, and journal?

Systems that streamline information and make it easier to access are both welcome and necessary. Where would we be without telephone books? The idea of such a system can extend to a variety of areas. One interesting effort currently underway is turning the town of Blacksburg, Virginia into an "electronic village" (Mowery). Blacksburg is in the process of connecting homes and businesses to an electronic mail service, university information services, and a world-wide computing network called the Internet.

This paper gives an overview of a system which streamlines information about computer modeling and simulation and makes it easily accessible. Called a Modeling and Simulation Information Network, the system will serve as a resource for modeling and simulation education, research, development, and utilization. It will assist the user in all aspects of modeling, from design and development to test and verification. The system was designed through a summer research project at Rome Laboratory's Modeling and Simulation Office. One of the goals of this office is to advance the effective use of modeling and simulation within the Department of Defense. Many factors can contribute to this goal. One is to make the people who create models and simulations aware of available information, tools, and technology in this area. The Modeling and Simulation Information Network will help to promote this awareness.

Modeling and Simulation Information Network

The creation of the Modeling and Simulation Information Network is a team effort. The Modeling and Simulation Office has formed a consortium of experts from government, industry and academia. By combining expertise and applications found in government research labs, universities, and companies, the office has created a rich environment which serves as a base for the information network. Pulling together modeling and simulation expertise and technical developments in a flexible, readily available, easily accessible environment, will enable users to make more informed decisions about modeling and simulation. It will also facilitate advances in research and development, reductions in development time, and decreased costs.

The first step in designing the information system is to define the needs of the user. Imagine what the potential modeler might want to know. Some obvious questions come to mind, such as:

1.   What are the basic steps in developing a model?

2.   Are there any similar models to the one I need?

3.   Are there any papers or books in this area?

4.   What kind of software and hardware do I need?

Questions like these are simply expansions of the very basic question "What do I do?", and they are the foundation for the modeling and simulation information network.

In addition to helping the user with the modeling process in general, other considerations must be incorporated into the network design. The user should find the network simple to use, and the environment should be friendly and helpful. The system should promote a sharing of ideas and technology, bringing new concepts to the user in a timely manner. Additionally, the network should be designed to promote model interoperability and reusability. All of these factors have been integrated into the Modeling and Simulation Information Network.

The actual components of the network are based on the above-mentioned considerations and questions. The Modeling and Simulation Information Network is composed of several components, including:

1.   Descriptive information about existing models and simulations.

2.   A modeling and simulation reference bibliography.

3.   Descriptions and available reviews of modeling and simulation hardware and software.

4. Information about modeling and simulation. such as bulletin boards and conferences.

5. Model demos and prototypes. including shareware models.

6. Information on modeling and simulation education. including a catalog of university curricula and workshops.

Of course. this list can be changed as the system evolves.

After determining the initial components of the information network. the system was designed. One of the easiest ways to design an application that links together several components is to use a prototyping tool. In this case. I used SuperCard™ for the Macintosh. SuperCard™ is a tool that allows the user to create custom programs. from simple organizational applications to complex multimedia environments. Using a tool like SuperCard™ can speed up the production of the prototype. It can serve as a development tool and a research aid since the structure of the system is created using the tool. SuperCard™ was visual. quick. and easy to use. Using windows. menus. scripts. and buttons. I created a simple prototype of the modeling and simulation information network. The prototype is discussed below. as well as details of the information network.

Prototyping the Information Network

The prototype provides a visualization of the information network components. It consists of a series of menus which the user browses through in order to obtain information. The main menu of the information network prototype is shown below. Each component is represented as a button on the screen. and the user clicks on a button to get more information. Each of the buttons will be discussed.

Click the mouse on any topic for more information:

Bibliography	Modeling & Simulation Bibliography
Decision Aid	How to get started in Modeling & Simulation
Demos	Modeling & Simulation Demos
Education	Cirricula & Workshops
Learning Lab	Rome Laboratory Modeling & Simulation Laboratory
M & S Info	Bulletin Boards, Conferences
Model Catalog	On-line model descriptions
Videos	Modeling & Simulation Videotapes

Bibliography

When the user clicks on the bibliography button, the modeling and simulation bibliography is loaded. The user sees a screen that looks like the following:

The user can browse through the books and papers, perform a search, or print a reference. The initial modeling and simulation bibliography has been set up using FileMaker Pro™ for the Macintosh. The FileMaker Pro™ application is simply loaded with the bibliography file. When finished, the user returns to the SuperCard™ application.

Decision Aid

The Decision Aid is the most complex aspect of the information network. It consists of three components, a Network Link, Modeling and Simulation Topics, and a Search feature. The Network Link is a tool that takes the user through the information network with a minimum of user input. It is designed for the user who wants the system to make a "best guess" at what information is needed. In the Network Link, the user is asked a series of questions. Based on the answers given, the system decides where to go next. For example, the user can check the type of system being modeled and any preference of mathematical method:

What type of system are you modeling?

RETURN

- ☐ PHYSICAL
- ☐ BIOLOGICAL
- ☐ CHEMICAL
- ☐ ELECTROMAGNETIC
- ☐ COMPUTER
- ☐ OTHER

Do you have a preference of mathematical method?

- ☐ STATISTICAL
- ☐ CALCULUS-BASED
- ☐ ALGEBRAIC
- ☐ CONNECTIONIST
- ☐ FUZZY MATHEMATICS
- ☐ OTHER

Depending on the user's answers, more information and questions will appear, until the user quits. If the user check "Electromagnetic" and "Statistical", for instance, model descriptions and information about related software packages will be shown.

In the future, the Network Link will consist of a knowledge-based system which leads the user through the information network. A knowledge-based system captures expertise and makes it available to many users. In the Modeling and Simulation Information Network, a knowledge-based system can be used to help the user in all aspects of modeling and simulation. It would be especially helpful since the system will have to accommodate many levels of user, from the inexperienced modeler to the expert. The Modeling and Simulation Office will continue to develop the knowledge-based system.

The Decision Aid is also a good place to access general information about the steps involved in creating a model or simulation. If the user selects "Modeling and Simulation Topics", a screen appears showing a variety of topics:

Fundamentals of Modeling

CLICK ON ANY TOPIC FOR MORE INFORMATION

Definitions

Classification of Simulation Models

Model Structure

Criteria for a Good Model

Stages

Problem Formulation & Model Definition

Model Formulation

Model Validation

The user simply clicks on a button for more information. The user chooses Model Formulation, for example, and additional screens pop up:

STEPS IN DEVELOPING A MODEL

1. Identify the client
   - Identify other's who will be affected by the study
   - Define those to be affected positively, negatively

2. Briefly state the purpose of the model
   - How are the model and results to be used
   - What are the client's expectations

3. Determine the measures of merit to be used in evaluating
   - The performance of the model (model validity)
   - The performance of the system being modeled

4. Describe the means for validating the model
   - Model/reality comparison in original environment
   - Model/reality comparison in modified environment

5. Create scenarios to be tested with validated model

6. Develop the design plan
   - Estimate required resources including time
   - State critical design assumptions
   - Define specific validation procedure
   - Develop an experimental design to test scenarios

7. Develop the model
   - Data collection and inference
   - Visual, verbal, mathematical, computer components

8. Validate the model

9. Evaluate the scenarios

10. State findings, recommendations and conclusions
    - State assumptions on which conclusions are based
    - Provide certainty, precision and validity of results
    - Answer 'so what?' and 'where do we go from here?'

The third component of the Decision Aid is the search facility. This feature allows the user to search all or part of the network for information about a specific topic. The user types in a topic and is asked what sources should be searched:

Select one or more search sources

RETURN

☐ All Sources

☐ Bibliography

☐ Model Catalog

☐ Software/Hardware

☐ Videos

☐ Demos

Depending on the boxes checked, the components of the information network are searched and information on the user's search topic is shown.

Demos

When the user clicks on the "Demos" button, a menu called Demos appears at the top of the screen. This menu contains a list of available demos. Demos might include software packages and actual models. When a user selects a demo, information about it is given. In the future, demos will be loaded and run directly through the information network. If desired, the user may make a copy of the demo. When finished, the user returns to the Super Card application.

Education

An important component of the modeling and simulation information network is educational information. Some users will want to know what modeling and simulation classes, conferences, and workshops are available. A catalog of existing university modeling and simulation curricula will be accessible through the network, as well as information about upcoming conferences and workshops. For example, the user selects Binghamton University and chooses to look at courses in the Department of Systems Science:

CLICK ON ANY BUTTON FOR MORE INFORMATION

Binghamton University
Lehigh University
Syracuse University
SUNY TECH Utica/Rome
University of Florida
University of Massachusetts

The course listing for the Systems Science Department appears:

BINGHAMTON UNIVERSITY

SS500 Fundamental Structures
SS510 Systems Problem Solving
SS520 Modeling & Simulation
SS530 Probability & Statistics
SS540 Simple Models of Complex Systems
SS550 Heuristic Problem Solving
SS560 Fuzzy Sets & Information Theory
SS570 Biological Systems Theory
SS580 Systems Optimization
SS590 Neural Nets & Genetic Algorithms
SS600 Numerical Methods

Then the user could access specific course descriptions, if desired:

SS500 FUNDAMENTAL STRUCTURES

(RETURN)

PROFESSOR: Howard Pattee
PHONE: (607) 777-6510

**TOPICS: Topics include graph theory, set theory,
finite state machines, boolean lattices, cellular
automata, turing machines, fundamentals of computing,
artificial intelligence theory.**

Learning Lab

When the user clicks on the "Learning Lab" button, two menus appear at the top of the screen, called "Software" and "Hardware". These menus contain lists of software and hardware which may be useful in modeling and simulation. The user can get information such as a description and types of models the software or hardware is suited for. As an example, the user selects TurboCASE™ from the software menu:

TurboCASE is an Apple Macintosh tool for system modeling and software design. It supports popular structured and real-time analysis methods, including:

1. Yourdon/DeMarco Structured Analysis
2. Gane/Sarson Structured Analysis
3. McMenamin & Palmer Essential Systems Analysis
4. Chen Entity Relationship Modeling (Data Modeling)
5. Hatley/Pirbhai Real-Time Modeling
6. Ward/Mellor Real-Time Modeling
7. ESML Real-Time Modeling
8. Shlaer/Mellor Object-Oriented Analysis
9. Object-Oriented Design

RETURN

Also through the Learning Lab button, the user can obtain additional information about the Rome Laboratory Modeling and Simulation Office. The office has an actual laboratory which users may visit to try out various software and hardware, or even to work on a project.

M & S Info

The "M & S Info" button connects the user to a bulletin board resource:

```
CLICK ON ANY BULLETIN BOARD FOR MORE INFORMATION

    ┌─────────────────────────────────────────────┐
    │         Ada Joint Program Office            │
    └─────────────────────────────────────────────┘
    ┌─────────────────────────────────────────────┐
    │        AgileNet (Lehigh University)         │
    └─────────────────────────────────────────────┘
    ┌─────────────────────────────────────────────┐
    │                  CHARM                      │
    └─────────────────────────────────────────────┘
    ┌─────────────────────────────────────────────┐
    │     Defense Modeling and Simulation Office  │
    └─────────────────────────────────────────────┘
    ┌─────────────────────────────────────────────┐
    │      DoD Gateway Information System         │
    └─────────────────────────────────────────────┘
    ┌─────────────────────────────────────────────┐
    │   Interoperability Decision Support System  │
    └─────────────────────────────────────────────┘
    ┌─────────────────────────────────────────────┐
    │         Rome Laboratory Models              │
    └─────────────────────────────────────────────┘
    ┌─────────────────────────────────────────────┐
    │ University of Florida M & S Node (User Groups) │
    └─────────────────────────────────────────────┘
```

Currently, the user can obtain information about modeling and simulation bulletin boards, such as
what they contain and how to access them. If the user chooses AgileNet, for example, information
about the services available in this network is shown:

What is AgileNet?
AgileNet is a repository of information and value   added services created to
 support industries in their movement to agility. AgileNet was developed by
the Agile Manufacturing Enterprise Forum (AMEF) and Lehigh University.

How to Access AgileNet:
You can access AgileNet in two ways, via Internet or via dialup using a modem.

 Access via Internet:
   1. Enter the following command: telnet agilenet.amef.lehigh.edu
   2. Set your terminal emulator to vt100
   3. If you do not have access to the Internet, the major network services
     (Compuserve, Delphi, etc.) can now provide access.

 Access via Modem:
   1. Dial the Lehigh University network at (215) 947-4413
   2. Once the connection is established, enter the following command:
        telnet agilenet.amef.lehigh.edu

If you would like more information on AgileNet, click on one of the topics below:

User's Guide	AMEF M&S Focus Group
Point of Contact	**Agile Manufacturing**

From here the user can get more information, such as an AgileNet User's Guide. At this time, the
user receives information about how to access the various bulletin boards and services. In the
future, however, the user will be able to connect directly to services without leaving the
information network. In addition to the bulletin board resource, users can find information about
upcoming conferences within the "M & S Info" button.

Model Catalog

        In trying to understand the user's needs, one of the questions posed earlier was: "Are
there any similar models to the one I need?". When the user clicks on the "Model Catalog" button,
he or she has access to model descriptions. These descriptions come complete with purpose,

software and hardware used, lessons learned, development experiences, data requirements, applications, and points of contact. Initially, the models listed include those developed at Rome Laboratory. This will be expanded to include other models in the future.

The user has two methods of viewing model descriptions: either on the Internet or through a database. The Internet is a world-wide computer network. Modeling and simulation descriptions are available as ASCII files through Internet File Transfer Protocol (FTP). The existing mechanism for "anonymous" access through the Internet can be used to provide access to the model catalog from any computer system with an Internet connection. Internet mail addresses for responsible points of contact are provided for users desiring additional information.

The second method of accessing model descriptions is through a model database. Through a summer research project, the Modeling and Simulation Office has created a database of models developed at Rome Laboratory. The descriptive information in the database is similar to the information available on the Internet, with additional benefits. One benefit is the ability to use search queries, such as by keyword or subject. User's can search the model catalog for models that are similar to their area, for example. Another benefit is the interoperability of the information with the other database components, such as the bibliography and model demos. A user can search the model catalog, the bibliography, and the learning lab for information on a specific topic. Since all of the components are connected within the information network, the user does not have to do separate searches.

The database of Rome Laboratory models is in an object-oriented format. There are several reasons for choosing this type of database. Multimedia, such as video and graphic images, can be easily incorporated into an object-oriented environment. This makes it possible to incorporate videos and demos of the models into the database. Object-oriented technology provides a natural mechanism for the integration of network communication, multimedia support functions, and object-specific processing techniques, such as displaying, storing, and processing images, voice, and other data. Creating an object-oriented database of models has broadened the potential for data storage through the use of features like inheritability and reuse of software.

When the user clicks on the "Model Catalog" button, a screen appears with a choice of the Internet or the Database. If the user chooses the Internet, information about accessing the models on the Internet appears. In the future, the user will be connected directly to the Internet, if desired. When the user chooses the Database, the object-oriented database is loaded. The user can

15-17

perform queries and copy or print model descriptions. When finished, the user returns to the SuperCard™ application.

Videos

When the user clicks the mouse on the "Videos" button, a menu called Videos appears at the top of the screen. This menu lists available videos that are related to modeling and simulation. These might include educational videos as well as videos of actual models. Through the use of multimedia, the videos can be seen right on the user's monitor. Software packages such as Adobe Premier™, for example, make this possible.

Advantages of Prototyping Packages

Although I spent some time creating components of the network, the development of the Modeling and Simulation Information Network prototype was very quick. There are several advantages to designing via a prototype. System components can be easily linked together through one interface. It is easy to see what the system looks like, and user feedback can be incorporated during design. Also, it is not necessary to have complete information in order to create the prototype. The Modeling and Simulation Information Network prototype does not have information behind every button and window. This is not a disadvantage since ideas inevitably evolve during design. It is so easy to make changes that the prototyping environment encourages the designer to try out many different ideas.

Prototyping a system also makes it easy to see where new components can be added. The Modeling and Simulation Office is working with a team of experts from government, industry, and academia. These experts will be contributing databases in their own area of expertise to the information network. The databases will be linked with ISDN, which is the standard for future communication of all types of complex data. One other important advantage is that the prototype can become the finished product.

15-18

Summary

Information networks provide a means for channeling information and making it readily available. These networks integrate the expertise, experience, tools, and technology needed to create new applications effectively and efficiently. Advances in technology have made it simple to incorporate multimedia in an easily-to-use system. Fiber optic broadband widths already make it possible to view multimedia presentations anywhere, and they will continue to have increasing impact. In fact, one report states that "By 2006, the economic health of a country will be strongly determined by the size and effectiveness of its broadband communications network" (Nagel & Dove). Technology such as fiber optic broadband widths makes advanced information networks possible, and the birth of the information age makes them necessary. These networks can channel the information and technology needed to solve a particular problem in a timely and cost-effective manner. Already being called "information highways", information networks are clearly an important part of the future.

15-19

Bibliography

Nagel. Roger and Rick Dove (1991). "21st Century Manufacturing Enterprise Strategy Infrastructure". Volume 2. p. 41. Iacocca Institute. Lehigh University.

Mowery. Beverly P. (1993). "Electronic Village Concept Offers Diverse Information Services", in IEEE's *Signal* Magazine (July. 1993, pp. 43 - 45).

Rome Laboratory Modeling and Simulation Prototype Database

Leland W. Hovey

Master's Degree Candidate

Computer Science Department

SUNY Institute of Technology at Utica/Rome

P O Box 3050

Utica, NY 13504 - 3050

Rome Laboratory Modeling and Simulation
Prototype Database

Leland W. Hovey
Master's Degree Candidate
Computer Science Department
SUNY Institute of Technology at Utica/Rome

Abstract

The goal of the design and implementation of the Rome Laboratory Modeling and Simulation Prototype Database is to provide the Rome Laboratory Modeling and Simulation Department with a convenient means to access descriptions for all Rome Laboratory Models and Simulations. The scope of the project is a study of the issues involved in the object-oriented paradigm as they apply to the construction of the Prototype Database. The object-oriented issues explored are complex objects, behavioral data, meta knowledge, class hierarchies, aggregation, generalization, and inheritance. The performance of the resultant object-oriented Prototype Database is compared with a program written in the C language. Suggestions are made for enhancing the performance of the Prototype Database.

Rome Laboratory Modeling and Simulation

Database Prototype

Leland W. Hovey

The Modeling and Simulation Department at Rome Laboratory began in May 1993. The consensus from our first departmental meetings was, in order to eliminate redundancy and to expedite utilization, information about existing Rome Laboratory Models and Simulations should be made available to the Rome Laboratory community. At that time information about existing Rome Laboratory Models and Simulations already existed in the form of a 200 page booklet called the Rome Laboratory Models and Simulations. However, we recognized that a person researching Modeling and Simulation (M & S) would not readily search through a booklet that has no table of contents or index for Models or Simulations relevant to his or her work. Since the booklet contains information w: wish to make accessible to the Rome Laboratory community we decided to store the information as a database. Not only does such a database provide access to Rome Laboratory models and simulations but it could also be extended to include Models and Simulations from the other Air Force Laboratories[1].

After additional meetings we established the following vision. The Rome Laboratory M & S should promote community awareness of innovations in the field of M & S by providing an "access point" to a "network" that contains these innovations. The "network" is a consortium of academic, industrial and government institutions that each have an area of expertise in the field of M & S. An important component of the "network" is a database resident at Rome Laboratory, the Rome Laboratory M & S Database. The Database should contain: bibliographies about M & S literature; information on M & S software and M & S tools; video, sound and animations for demonstrations of M & S software and tools; descriptive information about all Air Force Models and Simulations; capability of resident or remote Model and Simulation execution. The "access point" is a user-friendly interface that assists the user in identification of his or her needs. If it should be determined the user needs access to the Rome Laboratory Database, the interface would provide a connection. Similarly, if it should be determined the user needs access to one of the other sites in the consortium, the interface would provide that connection as well[2].

The purpose of my summer research was to develop a Prototype for the Rome Laboratory M & S Database. The following methodology was used in the development of the Prototype Database:

Development Methodology

1. Database AnalysisPhase

2. Assemble Source Material Phase

3. Table Design ProcessPhase

4. Resultant M & S Database Prototype

Figure 1
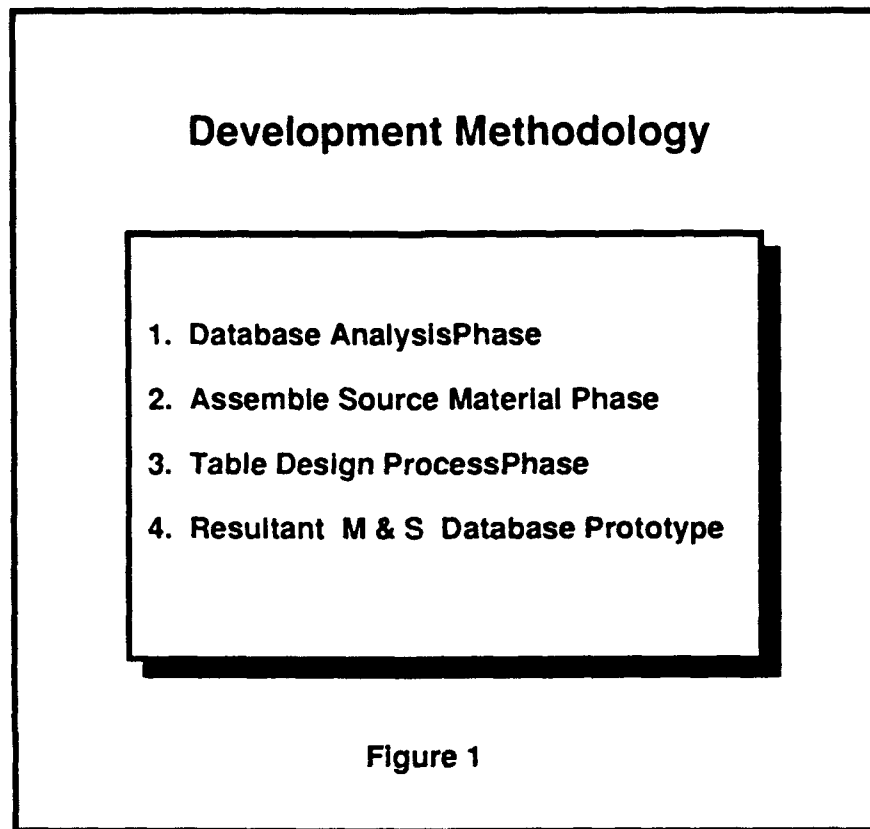
The Database Analysis Phase consisted of making a list of goals that I believed I could accomplish during the development of the Prototype Database:

Goals

1. Investigate Object-Oriented Programming using Digitalk's Smalltalk/V

2. Implement the Prototype Database using Smalltalk/V

3. Implement a Menu- DrivenUser Interface to the Prototype Database

4. Measure Performance of Prototype Database

Figure 2

The Assembling Source Material Phase consisted of reformatting the file of Rome Laboratory Models and Simulations. The problem with the file was not all lines were the same length. Reformatting would make the file easier to read. Secondarily, writing a reformatting program would give me experience with the basic technique of working with Smalltalk/V[3].

All Models and Simulations are made up of attribute category names followed by a corresponding description. Figure 3 is a list of attribute categories common to all Rome Laboratory Models and Simulations:

Attributes Names Common to all RL Models

1. Title

2. Purpose

3. Type

4. Functional Area of Application

5. Description

6. Hardware

7. Software

8. Security

9. Database

10. Network

11. Verification

12. Costs

13. Developer

14. POC

Figure 3

Using the original Model and Simulations file, my first step was to remove all carriage returns except where the a carriage return preceded an attribute category name. What this did was decompose all Models and Simulations so that each attribute category name was followed by a one line description. From this point on I refer to an attribute category name together with the corresponding description as an attribute description.

Decomposition of Models by Attribute Catagory Name

Model #X

Title:	**Description**
Purpose:	**Description**
Type:	**Description**
	• • •
POC:	**Description**
Developer:	**Description**

Figure 4

Attribute Descriptions

• **Attribute Name**

• **Corresponding Description**

TITLE: Intrasystem Electromagnetic Compatibility Analysis
Program (IEMCAP)

PURPOSE: The primary function is a characterization of
system-level ...

Figure 5

The second step in reformatting the file of Rome Laboratory Models and Simulations was to introduce carriage returns into lines longer than 72 characters. However, before I started working on this step I learned Professor Scott Spetka (also working in the Modeling and Simulation Department) intended to produce his own reformat of the original file of Models and Simulations. After learning this I decided not to do the second step and decided to move ahead with the fact that all attribute descriptions in the Prototype Database would be one long line. This should not be a problem since it is possible to scroll Smalltalk/V display windows so that lines longer than 80 characters can be viewed.

Before discussing the Table Design Phase, new database application requirements will be discussed to provide motivation for the utilization of object-oriented systems[4].

New Database Application Requirements

• **Complex objects**

• **Behaviorial data**

• **Meta knowledge**

Figure 6

It is often the case that a model (using the term model in the general sense) is not solely composed of atomic objects. A atomic object is an object that has no constituent parts. A nonatomic object has parts or refinements. Consider modeling a car. A car can be represented with the following: chassis, frame, power system, mechanical systems, and electrical systems. Because this model of a car is made up of parts it is referred to as a nonatomic or complex object.

Now consider the automatic transmission of a car. At any given time a car's transmission has one state. The actual state varies among neutral, reverse, first, second, and third. The term behavior refers to the idea that one object (in this case a car's transmission) can have more than one state. An object's behavior is defined by the set of all an object's states[5]

Meta knowledge refers to the idea that rules contribute to an object's behavior. In the case of automatic transmission one rule could be "if in first gear and rpms > 1500, upshift to second gear". Another rule could be "if in second gear and rpms < 1500, downshift to first gear"[6].

None of the three applications requirements; complex objects, behavioral data, and meta knowledge are easily represented in traditional database paradigms. However object-oriented programs and object-oriented database management systems can represent these requirements. As mentioned previously two of the goal of my summer research were "investigate object-oriented programming" and

"implement a Prototype Database for Rome Laboratory Models and Simulations".
While writing the model reformatter had the secondary purpose of enabling my
preliminary investigation into basic techniques of Smalltalk/V object-oriented
programming, writing the Prototype Database had a secondary purpose of enabling
a more comprehensive study of object-oriented programming. During the
development of the Prototype Database I examined how complex objects, behavioral
data and meta knowledge can and should contribute to this system and systems in
general. It turns out that in the Prototype Database, meta knowledge was not needed.

According to the object-oriented paradigm an object or abstract data type is the
encapsulation of code and data into a single unit. Typically, an object is formed from
a set of variables, a set of messages, and methods (bodies of code to implement each
message).

Recall that each Rome Laboratory Model or Simulation description could be
decomposed into what I refer to as an attribute description (refer to figure 5). The
first part of the Table Design Process was application of the object-oriented
paradigm to attribute descriptions[7]:

```
┌─────────────────────────────────────────────────────────┐
│                  OO Paradigm Application                │
│   ┌─────────────────────────────────────────────────┐   │
│   │        Attribute Description -- an Object        │   │
│   │   ┌─────────────────────────────────────────┐   │   │
│   │   │            Instance Variables            │   │   │
│   │   │                                          │   │   │
│   │   │  • Attribute Name -- a String of         │   │   │
│   │   │    Characters                            │   │   │
│   │   │  • Corresponding Description -- a Stream  │   │   │
│   │   │    of Characters                         │   │   │
│   │   └─────────────────────────────────────────┘   │   │
│   │   ┌─────────────────────────────────────────┐   │   │
│   │   │               Messages                   │   │   │
│   │   │                                          │   │   │
│   │   │  • putAttributeName -- stores the name   │   │   │
│   │   │    of the attribute                      │   │   │
│   │   │  • putDescription -- stores the          │   │   │
│   │   │    description                           │   │   │
│   │   └─────────────────────────────────────────┘   │   │
│   └─────────────────────────────────────────────────┘   │
│                       Figure 7                          │
└─────────────────────────────────────────────────────────┘
```
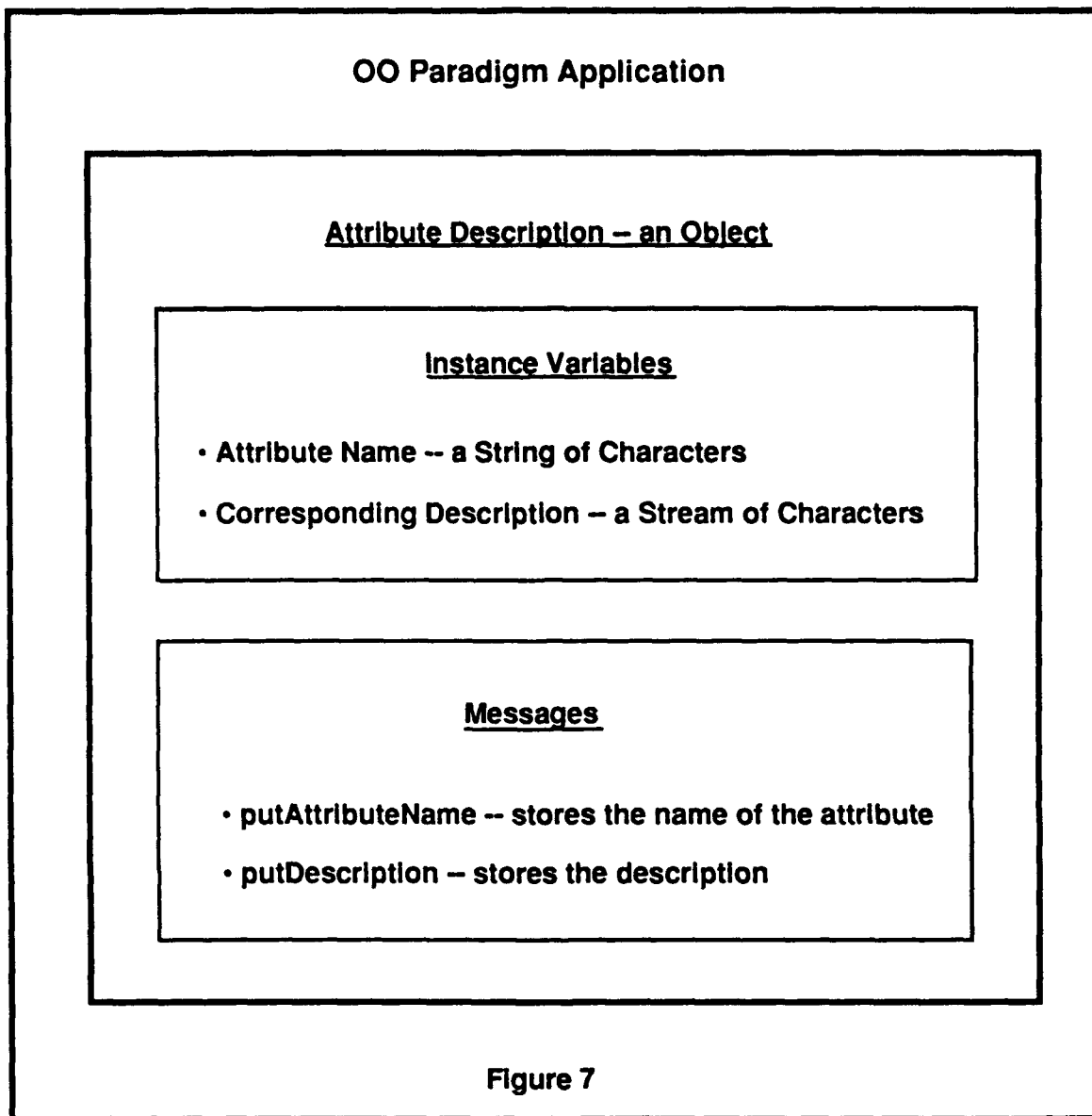
The object representation in this figure can be thought of as a generic template that can represent any one of all the possible attribute descriptions. This figure provides an example of an atomic object. The object is not composed of constituent parts or refinements, therefore is not complex.

Figure 7 lists two variables, attribute_name and description. Association of values with an these variables give the Attribute_Description object a state. An object's behavior data is the entire set of values that can be associated with an object's variables to give the entire set of all object states. Behavioral data is a central characteristic of the object Attribute_Description. In general, behavioral data is a central characteristic of all objects. Note that there is a wide variety of behavioral

data for the Attribute_Description object (e.g.: 14 unique attribute descriptions for each of the 99 Rome Laboratory models).

Each time an Attribute_Description object is created, initially it is a generic template where the variables attribute_name and description are not assigned values. Objects must have a way of assigning values to their variables. In the case of the Attribute_Description object, the message putAttributeName assigns to the variable attribute_name the string of characters that name the attribute. Likewise, the message putDescription assigns to the variable description the stream of characters that is the description of the attribute. It is in this manner that the Attribute_Description object takes on behavioral data.

Before continuing the discussion of the Table Design, Phase I will introduce and discuss three additional object-oriented concepts: aggregation, generalization, and inheritance. It was previously stated that complex objects are composed of parts or refinements. Complex objects composed of parts are referred to as aggregates and complex objects composed of refinements are referred to as generalizations[8].

A car is an example of an aggregate. The major systems composing a car are: body, chassis, power system, electrical systems, and mechanical systems. Each of these systems form a "has a" relationship with the car object. In other words every car "has a" body, every car "has a" chassis, etc. The advantage of using complex objects such as aggregates stems from the capability of constituent parts to inherit variables from the whole. For example, a car object could have variables such as make, model, and year. A particular part of a car, a power system object, has to have variables for a make, a model and a year in addition to variables for all information relevant to power systems. Inheritance enables variable sharing among objects. Variable sharing takes place in a top down manner through a class hierarchy.

A Car and its Constituent Parts

```
                          ┌─────────┐
                          │   Car   │
                          └─────────┘
          ┌──────────┬────────┼────────┬──────────────┐
    ┌────────┐  ┌─────────┐ ┌────────┐ ┌──────────┐ ┌───────────┐
    │  Body  │  │ Chassis │ │ Power  │ │Electrical│ │ Mechanical│
    │        │  │         │ │ System │ │ Systems  │ │ Systems   │
    └────────┘  └─────────┘ └────────┘ └──────────┘ └───────────┘
```
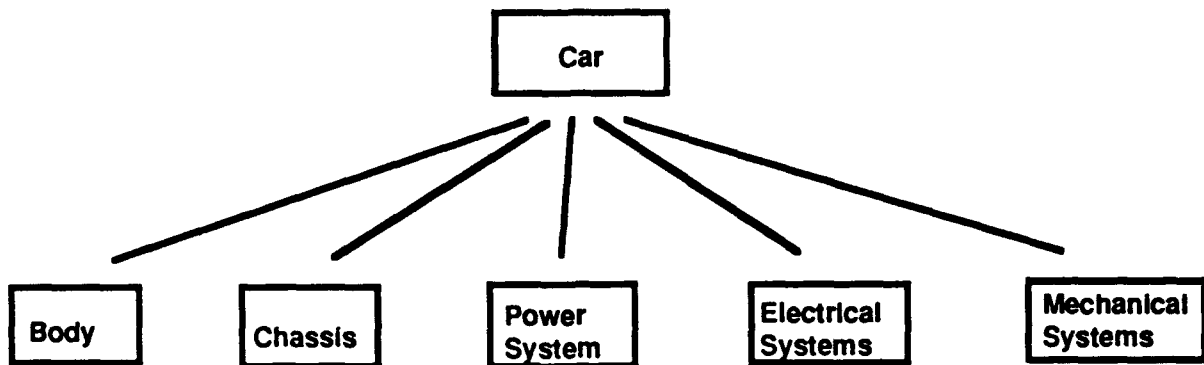
Figure 8

Class hierarchies are an important means of expressing complex objects. Object-oriented languages give every object a class, and classes are arranged into class hierarchies to define all aggregates and generalizations. The Smalltalk/V language provides an extensive predefined class hierarchy. Utilization of this class hierarchy is encouraged so the user does not have to define commonly used classes. Figure 8 illustrates the class hierarchy for a car and its systems. In this example, the power system object as well as all other constituent part objects inherit the variables make, model and year from the car object. Inheritance provides constituent parts with variables but it does not provide behavioral data.

A power system object can be created, and this is referred to as creating an object instance. An instance of the power system object inherits variables from the car object. It also possesses its own variables. Nevertheless, values must be assigned to all variables including those variables inherited form the car object. Just as object variable sharing is achieved from the class hierarchy through inheritance, object message sharing is achieved from the class hierarchy through inheritance.

Another object-oriented concept is generalization. Generalization is an abstraction that describes a collection of categorizations. For example categorizations could be Dodge, Ford, Cheverolet, and Oldsmobile, and a generalization could be cars. Inheritance works in the same manner for generalization as it does for aggregation.

During the Table Design Process of the Prototype Database development generalization and inheritance were used. Recall an object called Attribute_Description was created. Strictly speaking, an object is the most general class so an Attribute_Description is referred to as a class. Also recall that any particular Rome Laboratory Model or Simulation consists of a collection of 14 actual attribute descriptions. In traditional programming a way to represent any particular Model or Simulation is to create a "type" that is an array of Attribute_Description. Furthermore, a solution for the representation of all Models or Simulations is to create a "type" that is an array of the individual Models or Simulations. In the Smalltalk/V language, the predefined class hierarchy contains the class Indexed_Collection. An Indexed_Collection is similar to an array.
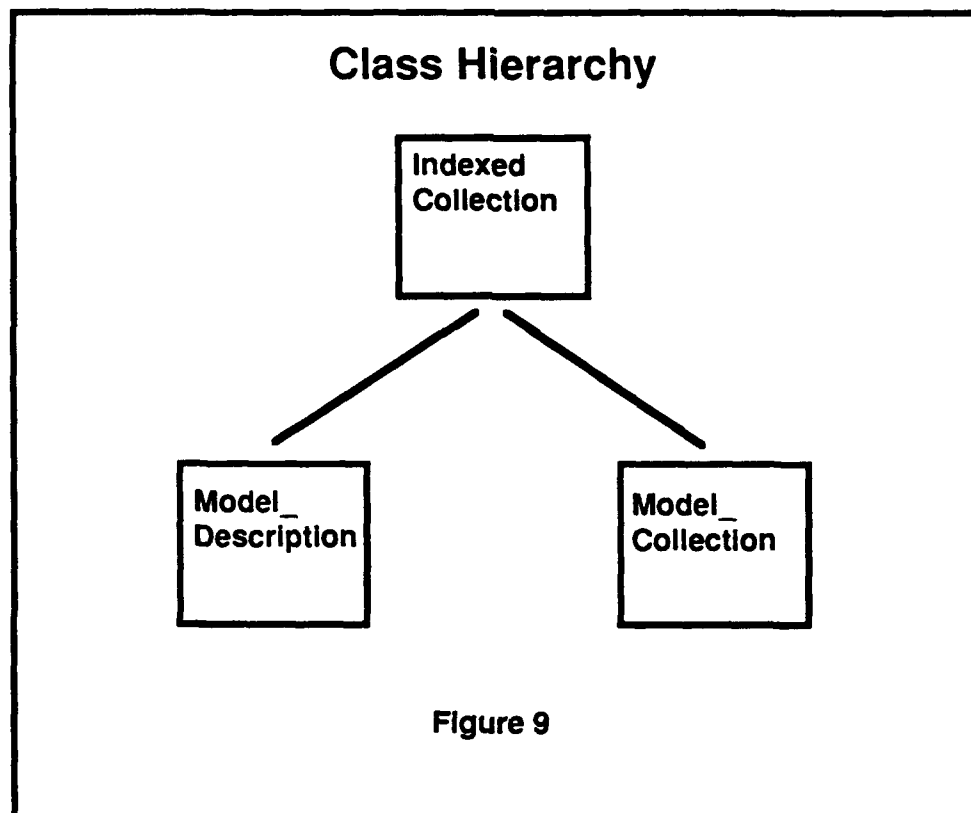
Figure 9

Figure 9 illustrates a class hierarchy in which the class Indexed_Collection is a generalization of the two other classes. Observe that Indexed_Collection is a complex object. Complex objects are an integral part of object-oriented programming because they provide a structure for a class hierarchy and therefore make inheritance possible[9].

Both Model_Description and Model_Collection are Indexed_Collections and therefore require Indexed_Collection manipulation messages. Generalization permits inheritance of variables and messages from Indexed_Collection to both Model_Description and Model_Collection. At the same time, Model_Description and Model_Collection require their own respective messages. The Model_Description class requires all of the query processing messages. The Model_Collection class does not require these. The Model_Collection class requires user interface messages. The Model_Description class does not require these.

The final phase of the development methodology consisted of determining what queries would be supported by the database, writing code to answer these queries, and deriving a user interface for user interaction with the Prototype Database. There are two queries to search for keywords. Both of these queries return only the name of the attribute descriptions in which the keyword occurs. If the actual description of any of the named attributes is required, the third query is used.

Prototype Database Queries

Query:	FindKeywordInAttributeDescriptions
Purpose:	Given a model and a keyword search every attribute description for an occurance of the keyword
Input:	Model to search through, keyword to search for
Output:	Name of all attributes that contain keyword

Query:	FindKeywordInNamedAttributeDescription
Purpose:	Given a model, a keyword, and an attribute name, search the named attribute for an occurance of the keyword
Input:	Model, keyword, attribute
Output:	Attribute name

Query:	FindNamedAttributeDescription
Purpose:	*Given a model and an attribute name, return the named attribute's description*
Input:	Model, attribute name
Output:	Attribute description

Figure 10

Because of limited reference documentation, the sophistication of the user interface was limited. The Smalltalk/V Reference Manual documents setting up a standard Apple Menu and a one multipane window[10]. These components were sufficient to support a working Prototype Database user interface.

Database Operations

Enter Model Names
Enter Model Attribute Names
Enter keywords
Return attribute name if keyword occur in attribute - all attributes
Return attribute name if key word occurs in named attribute
Return description of named attribute

Figure 11

User Display Screen

Instructions	User Input
System Output	

Figure 12

Figures 11 and 12 illustrate the user interface. Consider a query to all models for occurrences of the keyword "communications". This example is run by executing the following sequence of actions:

1. Type "all" in user input window pane

2. Select "Enter model names" from DB Operations Menu

3. Type "communications" in user input window pane

4. Select "Enter keywords" from DB Operations Menu

5. Select "Return attribute name if keyword occur in
   attribute - all attributes" fromDB Operations Menu

Figure 13

If "communications" occurs somewhere in a Model or Simulation, the system responds with the model name and the attribute name where "communications" occurs.

In its final form the Prototype Database performed correctly but it seemed slow. In order to examine the speed of the Smalltalk/V language, a program was written in Think C to duplicate the functionality of the Prototype. No object-orientation was used in the C program. An execution comparison showed the C program outperformed the Prototype by a factor of 10. However, once the display of query results to the screen was removed from the Prototype, another execution comparison showed roughly equivalent performance. It appears the Smalltalk/V screen output is a bottleneck.

Even though I have programming experience with the C language, writing a simple program was not a simple task. In C, reading string data from the disk is not straightforward. Moreover, if a C program requires a large amount of data stored in memory, the stack size must be increased and the memory used must be specifically allocated. Although it is not difficult to do this, another argument against C is programmers should be free of the issues of language implementation.

Smalltalk/V's main asset is that it is a truly object-oriented language. Programmers work soly with objects. For me, working with objects is easy and accurate because a program is built out of concrete units -- objects, and messages -- actions to give to objects. The object-oriented paradigm orients the programmer's thinking to task behavior and task completion. Generally, writing a program is

simply achieved by completing successively more complex tasks. It was easier for me to write Smalltalk/V code than to write C code.

I found four deficiencies with the Smalltalk/V language: slow output, lack of multiple inheritance, a cryptic debugger, and an equally cryptic method for producing executable code. In my opinion the latter two deficiencies could be worked though with approximately one month of research. It is difficult to say whether the former two deficiencies can be worked though.

The following resources provided the know-how for my completion of the Prototype Database. The Smalltalk/V Reference Manual provided useful instruction and many examples (unfortunately, no other books found on Smalltalk/V were helpful). Digitalk's customer support answering service was very helpful in directly solving many problems. In addition two other sources of instruction, Professor Spetka of SUNY Tech and a consultant working under Professor Rhody of RIT, made numerous helpful suggestions in the Prototype Database's design.

There are two potential solutions to enhance the Prototype performance (other than improving the output display speed). The first suggestion was made by Digitalk's customer support service and it is to utilize a predefined Smalltalk class called a Dictionary to record keyword location and thereby permit faster keyword queries. A second approach is to define a hierarchical representation of Rome Laboratory Models and Simulations. This too, would permit faster keyword keyword queries.

The next generation of the Prototype Database should have a better performance. First it should be the better performer of a program built with the aforementioned enhancements, and an existing Object-Oriented Database Management System (OODBMS). Secondly, a standard for the user interface should be agreed upon and implemented.

This Autumn I will be exploring the issues of object-oriented programming and OODBMSs through my work on my Master's thesis. I will be working with my advisor Professor Spetka at SUNY Tech. One issue I will work on is the effects of enhancements such as Dictionaries and hierarchical representation on query processing time. Another issue I will research is how to represent multimedia data. Finally, I will compare whether a C based Database, a C++ based Database (with enhancements), or an OODBMS has the best performance both in terms of capability for efficient representation for a variety of data types, and fastest query processing time.

References

[1] Rome Laboratory Models and Simulations.

[2] Rome Laboratory Modeling and Simulation Department (XPX) DMSO Proposal.

[3] Smalltalk/V Reference Manual, Digitalk Corporation. 1991.

[4] Korth and Silberschatz, Database Systems Concepts, 1991, McGraw-Hill, Inc. p427.

[5] Rumbaugh, Blaha, et al, Object-Oriented Modeling and Design, 1991, Prentice Hall, p 97.

[6] Rumbaugh, Blaha, op cit., p 97.

[7] Korth, op cit., p 428.

[8] Rumbaugh, op cit., p 57.

[9] Smalltalk/V Reference Manual, Digitalk Corporation. 1991.

[10] Smalltalk/V Reference Manual, Digitalk Corporation. 1991.

Appendices

Attached in the Appendices is the source code for the Prototype Database. This code has been tested and works correctly.

```
"-----------------------------------------------------------------
Definition of the Class AllModels1 and its methods.   This class is referenced in the
summary paper as the Class ModelCollection
-----------------------------------------------------------------"
OrderedCollection  subclass:  #AllModels1
    instanceVariableNames:
        'inputStream  replyStream  inputPane  replyPane  inputString  modelStream
attributeStream  keywordStream  queryid  textPane '
    classVariableNames:  "
    poolDictionaries:
      'CharacterConstants ' !


!AllModels1 class methods ! !



!AllModels1  methods !


a
"Method  to  enter  model  names"
I line I

modelStream := ReadWriteStream  on:  String  new.
modelStream := (inputPane  contents)  asStream.
self  changed:  #input.
CursorManager  execute  change.
line := modelStream  nextLine.
modelStream := line asStream.
CursorManager  normal  change.!


b
"Method  to  enter  model  attributes"
I line I

attributeStream := ReadWriteStream  on:  String  new.
attributeStream := (inputPane  contents)  asStream.
self  changed:  #input.
CursorManager  execute  change.
line := attributeStream  nextLine.
attributeStream := line asStream.
CursorManager  normal  change.!
```

```
c
"Method to enter keywords"
lline I

keywordStream  :=  ReadWriteStream  on:  String  new.
keywordStream  :=  (inputPane  contents)  asStream.
self  changed:  #input.
CursorManager  execute  change.
line  :=  keywordStream  nextLine.
keywordStream  :=  line  asStream.
CursorManager  normal  change.!


CollectAllModelDescriptions
     "Method to load memory with all model descriptions"
linput aModelDescription index word word1 word2l
input  :=  File  pathName:  ':oneliners'.
index := 1.
[input  atEnd]
     whileFalse: [
          aModelDescription := ModelDescription new.
           aModelDescription FormModelDescription: input.
          word := ('model', (index asString)).
          aModelDescription PutName: word.
          self add: aModelDescription.
        index := index + 1].
^self!


d
"Method  to  return  attribute  names  if  keyword  occurs  in  attribute
- all attributes"
I amodelname modelStream1 startTime jl

startTime  :=  Time  millisecondClockValue.
modelStream  reset.
j := 0.
amodelname  :=  modelStream  nextWord.
(amodelname  =  'all')
     ifTrue: [modelStream1 := self GetModelNames]
   ifFalse: [
        modelStream reset.
        modelStream1 := modelStream].
modelStream1  reset.
[modelStream1  atEnd]
     whileFalse: [
        keywordStream reset.
        amodelname := modelStream1 nextWord.
        j := j + (self ModelDescriptionSupplier: amodelname with: 1)].
inputStream  nextPutAll:  'occurances:  ';
          nextPutAll: (j asString); cr;
          nextPutAll: 'elapsed time: ';
             nextPutAll: ((Time millisecondClockValue) - startTime) asString!
```

e
"Method to return attribute names if keyword occurs in named attribute"
| amodelname modelStream1 |

modelStream reset.
amodelname := modelStream nextWord.
(amodelname = 'all')
    ifTrue: [modelStream1 := self GetModelNames]
   ifFalse: [
        modelStream reset.
        modelStream1 := modelStream].
modelStream1 reset.
[modelStream1 atEnd]
    whileFalse: [
        attributeStream reset.
        amodelname := modelStream1 nextWord.
         self ModelDescriptionSupplier: amodelname with: 3]!

f
"Method to return description of named attribute"
| amodel anattributename amodelname modelStream1 |

modelStream reset.
amodelname := modelStream nextWord.
(amodelname = 'all')
    ifTrue: [modelStream1 := self GetModelNames]
   ifFalse: [
        modelStream reset.
        modelStream1 := modelStream].
modelStream1 reset.
[modelStream1 atEnd]
    whileFalse: [
        attributeStream reset.
        amodel := modelStream1 nextWord.
         self ModelDescriptionSupplier: amodel with: 2]!

GetModelDescription: aname
"--
Purpose: Return the model description corresponding to the given model
       name.
Input: aname
Output: aModelDescription
Receiver: AllModels
---"
|word|
self do: [:amodel |
         word := amodel GetName.
         (amodel GetName) = aname
            ifTrue: [^amodel]].
^nil!

16 - 23

```
GetModelNames
"-----------------------------------------------------------------------
Purpose: Return the names of all the models
Output:  answerStream
Receiver:  AllModels
------------------------------------------------------------------------"
| answerStream word|
answerStream := ReadWriteStream on: String new.
self do: [:aModelDescription |
    word := aModelDescription GetName.
    answerStream nextPutAll: word; space].
^answerStream!

initWindowSize
        "Answer the initial window extent."
    ^(Screen boundingBox insetBy: 95@50) extent!

input
        "Initialize inputPane with inputString."


^String new!

inputMenu
^(Menu
    labels: 'Enter model names\Enter model attribute names\Enter keywords\Return
attribute name if keywords occur in attribute - all attributes\Return description of
named attribute\Return descriptions if keyword occurs in named attribute'
breakLinesAtBackSlashes

    selectors:#(a b c d e f) )
    title:' DB Operations '!

instrut
        "Initialize inputPane with a String."
| aStream |

aStream := ReadWriteStream on: String new.
aStream nextPutAll: 'Enter DB Query in Right Pane'; cr;
        nextPutAll: 'in the following order:'; cr;
        nextPutAll: '1. Query Range'; cr;
        nextPutAll: '2. Attribute Names'; cr;
        nextPutAll: '3. Keywords'.
^(aStream contents)!
```

```
ModelDescriptionSupplier: amodelname  with:  aqueryid
"-----------------------------------------------------------------------
Purpose: Supply  one  or  more  model  descriptions  (specified  by  user)  for
         a user specified query
Input:  amodelname,  aqueryid
Output:  answerString
Receiver:  aModelCollection
-----------------------------------------------------------------------"
I  keyword  attributename  aModelDescription  jl

j := 0.
(aqueryid  =  1)
   ifTrue: [
       keywordStream  reset.
       [keywordStream  atEnd]
        whileFalse: [
              keyword := keywordStream  nextWord.
               aModelDescription := self GetModelDescription: amodelname.
                j := (aModelDescription FindKeywordInAllAttributeDescriptions: amodelname
with:  keyword  with:  replyStream)]].
(aqueryid  =  2)
   ifTrue: [
       keywordStream  reset.
       [keywordStream  atEnd]
        whileFalse: [
              keyword := keywordStream  nextWord.
               aModelDescription := self GetModelDescription: amodelname.
           attributeStream  reset.
           [attributeStream  atEnd]
           whileFalse: [
                   attributename := attributeStream  nextWord.
                    aModelDescription FindKeywordInNamedAttributeDescription:
amodelname  with:  keyword  with:  attributename  with:  replyStream]]].
(aqueryid  =  3)
   ifTrue: [
       attributeStream  reset.
       [attributeStream  atEnd]
        whileFalse: [
              attributename := attributeStream  nextWord.
               aModelDescription := self GetModelDescription: amodelname.
                aModelDescription FindNamedAttributeDescription: amodelname  with:
attributename  with:  replyStream]].
^j!
```

```
openOn
        "Create a  window with aString
        as its initial script."
    | topPane instrutPanel
    (topPane := TopPane new)
    label: 'MODEL  DESCRIPTION  QUERIES';
     model: self.
    topPane addSubpane:
        (instrutPane := TextPane new
           menu: #inputMenu;
         model: self;
          name: #instrut;
            framingRatio: (0@0 extent: (1/2) @ (1/4))).
    topPane addSubpane:
        (inputPane := TextPane new
         model: self;
          name: #input;
            framingRatio: ((1/2)@0 extent: 1 @ (1/4))).
    topPane addSubpane:
        (replyPane := TextPane new
         model: self;
          name: #reply;
            framingRatio: (0 @ (1/4) extent: 1 @ (3/4))).
     inputStream := inputPane dispatcher.
     replyStream := replyPane dispatcher.
     topPane dispatcher open scheduleWindow!

reply
        "Initialize reply pane with an empty string."

                                      ^String new! !
```

```
"-------------------------------------------------------------
Definition of the class attribute description and its methods
----------------------------------------------------------"
Object subclass:  #AttributeDescription
    instanceVariableNames:
     'name description '
  classVariableNames:  ''
  poolDictionaries:  '' !


!AttributeDescription class methods !

name:  aString
    "comment"
    | temporaries |

name := aString! !


!AttributeDescription methods !

description:  aStream
    "comment"
    | temporaries |

description := aStream!

GetDescription
    "comment"
    | temporaries |

^description!

GetName
    "comment"
    | temporaries |

^name!

name:  aString
    "comment"
    | temporaries |

                                name := aString! !
```

Appendix   3

"--
Definition of the class Database and its methods
--"
Object subclass: #Database
    instanceVariableNames:
    'aModelCollection '
  classVariableNames: "
  poolDictionaries:
    'CharacterConstants ' !

!Database class methods ! !

!Database methods !

GetModelCollection

^aModelCollection!

initModelCollection

                aModelCollection := AllModels1 new.! !

Appendix 4

```
"------------------------------------------------------------------
Definition of class ModelDescription and its methods
------------------------------------------------------------------"
OrderedCollection subclass: #ModelDescription
    instanceVariableNames:
    'name '
    classVariableNames: ''
    poolDictionaries:
      'CharacterConstants ' !

!ModelDescription class methods ! !


!ModelDescription methods !

FindKeywordInAllAttributeDescriptions: amodelname with: akeyword with:
replyStream
"Purpose: Return all attribute descriptions for this model where the attribute
          descriptions contain the input keyword"
ItempStream jl

tempStream := ReadWriteStream on: String new.
tempStream := nil.
j := 0.
1 to: (self GetEndPosition) do: [:i |
    tempStream := (self at: i) GetDescription.
    (True = (tempStream FindKeyword: akeyword))
    ifTrue: [
      j := j + 1.
            replyStream nextPutAll: amodelname; tab; tab;
                  nextPutAll: ((self at: i) GetName); tab; tab; tab; tab;
                  nextPutAll: akeyword; cr]].
^j!
```

```
FindKeywordInNamedAttributeDescription: amodelname  with:  akeyword  with:
anattributename  with:  replyStream
"*************************************************************
**********
Purpose: Find  keyword  in  named  attribute  description  of  named  model
        description
Input:  akeyword,  anattribute
Output:  attribute  description  string  ("  if  not  found)
Receiver:  ModelDescription
*************************************************************
**********"
| tempStream |

tempStream := ReadWriteStream  on:  String  new.
tempStream := nil.
self  do:  [:anAttribute |
      (anAttribute  GetName)  =  anattributename
      ifTrue: [
              tempStream := anAttribute  GetDescription.
            anAttribute := self  GetEndPosition]].
tempStream  reset.
[tempStream  atEnd]
    whileFalse: [
        ((tempStream  nextWord)  =  akeyword)
        ifTrue: [
                replyStream  nextPutAll: amodelname;  tab;  tab;
                      nextPutAll: anattributename;  tab;  tab;  tab;  tab;
                  nextPutAll: akeyword;  cr.
            tempStream  setToEnd]].!


FindNamedAttributeDescription:  amodelname  with:  anattributename  with:
replyStream
"-------------------------------------------------------------
Purpose: Find  and  return  the  named  attribute  description.
Input:  anattributename
Output:  attribute  description  stream  (nil  if  not  found)
Receiver:  ModelDescription
-------------------------------------------------------------"
replyStream  nextPutAll: amodelname;  tab;  tab.
self  do:  [:anAttribute |
      (anAttribute  GetName)  =  anattributename
      ifTrue: [
              replyStream  nextPutAll: (anAttribute  GetDescription);  cr.
            anAttribute := self  GetEndPosition]].!
```

```smalltalk
FormModelDescription: input
   "Purpose: Retrieve the contents of one model description"

|line  word  titleposition  anAttribute|
anAttribute  :=  AttributeDescription  new.
titleposition := input position.
line := input nextLine.
word := line asStream nextWord.
self initPositions: 14.
(word isNil)
    ifFalse: [ ('TITLE' = word)
       ifTrue: [
               anAttribute name: word.
                 anAttribute description: (line asStream).
             self add: anAttribute]].
[input atEnd]
whileFalse: [
     titleposition := input position.
     line := input nextLine.
     word := line asStream nextWord.
     (word isNil)
        ifFalse: [('TITLE' = word)
            ifTrue: [input setToEnd]
           ifFalse: [
                   anAttribute := AttributeDescription new.
                 anAttribute name: word.
                   anAttribute description: (line asStream).
               self add: anAttribute]]].
(word = 'TITLE')
      ifTrue: [input position: titleposition]
      ifFalse: [input setToEnd].
^self!

GetEndPosition

^endPosition!

GetName
"------------------------------------------------------------------
-
Purpose: Return name of the ModelDescription
Output: String
Receiver: aModelDescription
------------------------------------------------------------------
-"

^name!
```

PutName: aString
"-- ----------------------------

Purpose: Return input string as name of the Model Description
Input: aString
Receiver: aModelDescription

-----"

name := aString! !

Synthesis

Mark J. Wierman

Ph.D. Student

Department of Systems Science

Binghamton University

Bingnamion, NY 13902-6000

Final Report For:

Graduate Student Research Program

Rome Laboratory

Sponsored By:

August 1993

Synthesis

Mark J. Wierman
Ph.D. Student
Department of Systems Science
Binghamton University

Abstract

Modeling and Simulation is an activity that touches all aspects of science. The modeling and Simulation Department at Rome Laboratory developed a vision after a broad survey of the work being done at the four super laboratories of the Air Force Materiel Command (AFMC). One result of this survey was the creation of the first Rome Laboratory database publicly available over the INTERNET. The vision created by the Modeling and Simulation Summer Research Group aided in the development of a dedicated modeling and simulation laboratory. This paper describes these efforts and concludes with an example of how a dedicated modeling and simulation facility can impact the course of research here at Rome Laboratory.

Synthesis

Mark J. Wierman

INTRODUCTION

This paper gives an overview of my summer research at Rome Laboratory in the office of modeling and simulation.

The tasks accomplished varied and included three main projects. The first was to establish how the office of modeling and simulation can best meet the needs of the DoD. It became clear that one of the primary goals of this office was to promote community awareness of the latest in tools and technology for Modeling and Simulation. This led to the second task, putting the list of models on the INTERNET, a world wide computer service. The third task was the creation of a dedicated Modeling and Simulation learning laboratory-a facility that contains the latest in software and hardware for use by researchers in modeling and simulation projects. These three tasks are described in the following sections.

PART 1: The creation of the Rome Laboratory Modeling and Simulation Department.

Modeling and Simulation is an endemic activity. It occurs whenever and wherever science is practiced. Whenever an engineer writes down an equation, creates a drawing, or programs a computer, she is creating an abstract model of a physical object or process. Often she is unaware of the modeling assumptions she is making when she scripts a differential equation, draws a blueprint, or types a program into a computer, but they are critical to the success of her endeavor and are usually only examined when the model starts to fail.

A model is a physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process. A simulation is a method for implementing a model over time: a technique for testing, analysis, or training in which real-world systems are used, or where real-world and conceptual systems are reproduced by a model. The type of modeling and simulation that is the focus of this paper is computer based modeling and simulation techniques which are primarily aimed at technology development and warfighting simulations.

Recently the Department of Defense (DoD) has decided that the spiraling costs of acquisition, training and wargaming necessitate a greater emphasis on modeling and simulation. The DoD has created the Defense Modeling and Simulation Office (DMSO) and empowered it to select and fund projects that will further the growth of modeling and simulation across the spectrum. The Air Force has established a Modeling Simulation & Analysis Office (AF/XOM) under B/G Frank B. Campbell that will start funding projects in the near future. The recent Air Force 4-Star Review of the impact of modeling and simulation on Command Requirements (AFMC/ST/XR) will also influence modeling and simulation funding.

In response to these and other initiatives Rome Laboratory created a dedicated Modeling and Simulation Department (RL/XPX) with Mr. Richard Simard as Director of Modeling and Simulation. Mr. Simard in turn organized a Modeling and Simulation Summer Research Group consisting of Scott Spetka, Ph.D., a professor at SUNY Technology-Utica/Rome, Leland Hovey, a master's student at SUNY Technology-Utica/Rome, Mark Wierman, a Ph.D. Student at Binghamton University, and Mary Dobransky, also a Ph.D. Student at Binghamton University

The first task of the Modeling and Simulation Summer Research Group was to bring their varied expertise to bear on refining the mission statement of RL/XPX in the AFMC.

Previous to the creation of the Summer Research Group, the Air Force Materiel Command had completed a survey of the models used by the four Super Labs: Armstrong Laboratory, Philips Laboratory, Rome Laboratory, and Wright Laboratory. A report detailing the purpose, classification, method, use, hardware and software requirements for every model in the Air Force Materiel Command's Research Laboratories was published under the title: The FOURMOSST Report.

Rome Laboratory contributed 99 models to the FOURMOSST report. A summary of the hardware usage is presented in Table 1.

DEC	37
SUN	24
IBM PC	9
SILICON GRAPHICS	9
IBM MF	7
SPECIAL PURPOSE	7
DATA GENERAL	4
MAC	2

TABLE 1: Classification of Rome Laboratory model hardware use, from the FOURMOSST Report.

The Summer Research Group examined the Rome Laboratory models and made the following observations about the state of modeling and simulation hardware use

- Over 50% of Rome Laboratory M&S run on processors operating at < 5 MHz

- None run at clock speeds greater than 50 MHz

- Processor currently available from 150-300 MHz

- 1 gHz processors in development.

FORTRAN	53
C	31
LISP	4
C++	2
PASCAL	1
SIMSCRIPT II.5	1
NETWORK II.5	1

TABLE 2: Classification of Rome Laboratory model software use, from the FOURMOSST Report.

The following observations about the software used by Rome Laboratory were also made

17-5

- Predominance of scientific computational models (Fortran 77 & C)

- Written for old generation slow processor architecture's

- Not designed for internetting or distributed processing

- Little use of COTS packages or Object-Oriented languages.

As a result the Modeling and Simulation Department made the following recommendations

- Need efforts to exploit new generation of multi-task & parallel processors

- Need M&S laboratory resource with latest multimedia workstations

- Need COTS M&S development resource libraries

- Need to Emphasize Interoperability & Reusability.

After much research and deliberation it was decided that the purpose of a the Rome Laboratory

Modeling and Simulation Department should be threefold:

- **Community awareness**

- **Information sharing**

- **Creation of a Modeling and Simulation center of excellence.**

The FOURMOSST document was comprehensive and informative. Unfortunately it was also

cumbersome if only specific model description or class of information was needed. It really was not going

to be successful in achieving **information sharing.**

Richard Simard, the director of the Rome Laboratory Modeling and Simulation Department decided

that the information would have to be packaged electronically if it was to be disseminated to the outside

world. Leland Hovey worked at the creation of an object-oriented database of models and Scott Spetka was

tasked with putting the catalog of models on the INTERNET.

PART 2:   The INTERNET

"We will do business electronically"

- Task Force 2000 Working Paper ARCHITECTURAL PRINCIPLES

Recently newspapers from the New York Times to the Chronicle of Higher Education have given bold headlines and major space to the phenomenal growth in size and importance of the electronic communication grid called the INTERNET. The INTERNET is a child of ARPA (Advanced Research Projects Agency, then DARPA) which created standards for inter-computer transmissions so that communication via mail, text files, etc., would be easily available to all end users no matter what type of computer and operating system they use.

The INTERNET has become a vital research resource. Draft versions of papers are distributed on the INTERNET for circulation and feedback. Conference agendas and requests for papers are sent out over the INTERNET. Bibliographies and papers are archived for anonymous-FTP access by interested users world-wide. Electronic mail is becoming more common than written mail. Software, both beta and final versions are distributed on the INTERNET as freeware, shareware, or for sale.

The backbone of the INTERNET is the TCP/IP protocol which allows fast and efficient data transmission between host computers everywhere on "the net". The user can accesses this power through two basic means, the FTP (File Transfer Program), and TELNET (remote terminal emulator).

As an example, suppose you wanted to connect to a network information center (nic) that happens to be located at White Sands. First you would need to know the INTERNET address of the computer you are trying to access. There is always a mnemonic and numerical address, for example

nic.ddn.mil   **192.112.36.5**   **netinfo, RFCs, IEN, IETF**

is a line from a file available on the INTERNET, usually titled "ftp.list" but colloquially called "the list". There are three pieces of information on this line, the verbal address, the numerical address and the start of the list of information available . On LONEX, the Rome Laboratory computer system, the way to connect to the nic is to type

```
ftp<enter>
argument #1  nic.ddn.mil<accpt>
argument #2  <accpt>
```

and the computer will respond with

```
ftp  "nic.ddn.mil"
Connected  to  nic.ddn.mil.
220-*****Welcome  to  the  Network  Information  Center*****
     *****Login  with  username  "anonymous"  and  password  "guest"
     *****You  may  change  directories  to  the  following:
     ddn-news          · DDN  Management  Bulletins
     domain           - Root  Domain  Zone  Files
     iesg            - IETF  Steering  Group
     ietf            - Internet  Engineering  Task  Force
     internet-drafts   - Internet  Drafts
     netinfo          - NIC  Information  Files
     netprog           - Guest  Software  (ex.  whois.c)
     protocols         - TCP-IP  &  OSI  Documents
     rfc             - RFC  Repository
     scc             - DDN  Security  Bulletins
     std             - Internet  Protocol  Standards
220 And  more!
Name  (nic.ddn.mil:wiermanm):
```

here you type the name anonymous and hit <accpt> and the computer responds with

```
331 Guest  login  ok,  send  "guest"  as  password.
Password:
```

type the word guest hit <accept> and you are on (on most systems you would type your email address, i.e., "wiermanm@lonex.rl.af.mil"). You can find out what is on the host computer nic.ddn.mil by entering commands such as **dir** (directory listing) or **cd** (change directory) and transfer files using the **get** and **put** commands. You exit by typing **quit.**

More important, if you ftp to **lonex.rl.af.mil** and change into the **pub** (public) directory you will find another directory called **models.des**. This directory contains a list of publicly release models. Rome laboratory is on the INTERNET.

A complete introduction to the INTERNET is beyond the scope of this paper and there are many excellent books available, but some of the best information is in two documents available on the INTERNET itself, "The Zen Guide to the INTERNET" and "The Hitchhikers Guide to the INTERNET".

In fact the growth of the INTERNET is so rapid that dedicated software aids to finding the files you need have been developed. The two most prominent at this time are called "Archie" and "Gopher". For more information about these programs consult the above mentioned INTERNET guides.

Probably the heaviest use of the INTERNET is Email. By subscribing to UserGroups a researcher will receive Email focused on a particular subject. This way she will always be aware of upcoming conferences, receiving the latest information about them, as well as receiving digests of conferences and requests from researchers who need help in finding information.

For example the following excerpt from the LISTSERV.LIST document describes a UserGroup dedicated to modeling and simulation;

C   *Simulation   simulation@UFL.EDU   simulation-request@UFL.EDU   Simulation software
Paul Fishwick <fishwick@FISH.CIS.UFL.EDU>   All topics connected with simulation are welcome; some sample topics are: Real time simulation methods, Flight simulation, Parallel architectures for simulation analysis and modeling, Simulation and training, Distributed simulation, Artificial intelligence and simulation, Automatic generation and analysis of MODELs, Analog vs. digital methods, hybrids, Continuous, discrete.*

Here is one on military matters;

P   *Military   military@ATT.ATT.COM   military-request@ATT.ATT.COM Military technology   Bill Thacker <military@ATT.ATT.COM> Mailing list for discussions of military technology*

and related matters. The list is gatewayed bi-directionally with the Usenet newsgroup "sci.military", which is moderated; it is distributed in the form of a daily digest.

Some other important nodes for researchers in modeling and simulation are:

lonex.rl.af.mil	Rome Laboratory Models
ftp.cis.ufl.edu	Univ. of Florida M&S node (User Groups)
AJPO.sei.cmu.edu	Ada Joint Program Office
Charm.isi.edu	CHARM
DMSO.dtic.dla.mil	Defense Modeling and Simulation Office
DGIS.dtic.dla.mil	DoD Gateway Information System
IDSS.ida.org	Interoperability Decision Support System

PART 3: Creating the Modeling and Simulation LAB

Wargames are very expensive and so their use is cost-constrained. One problem in developing a computer based Virtual War capability for training is that complete simulation of a battle is beyond the current capabilities of the hardware and software. But what is worse, even if photonic computers and multi-processor interoperable system software were available today, Virtual War would still be impossible. The Air Force does not have complete models for all the systems that it needs to simulate. For example it lacks an up-to-date model for the C17. It lacks a comprehensive model for the war in space. It lacks a complete model for the medical aspects of war.

As a result of the Four-Star review of Modeling and Simulation mentioned above the following recommendations were made

- **Refocus current expenditures at same level**
- **Investment Strategy for Integrated Architecture & Standards**
- **Standard Cost Element to Track Modeling and Simulation**
- **Policy for buying Modeling and Simulation based on Architecture & Standards TBD**
- **AFMC Center of Excellence for Modeling and Simulation Development**
- **National Test Facility Host Models?**

17-10

One of the tasks of the Summer Research Group at Rome Laboratory, in which I was a principal player, was the construction of a prototype Modeling and Simulation Laboratory. It is hoped that this research facility might influence the development of an Air Force Center of Excellence in Modeling and Simulation. As encapsulated in a proposal submitted to the DMSO this laboratory is envisioned as a distributed multimedia laboratory with involvement by a broad spectrum of government, industrial and educational institutions (as shown in Figure 1).
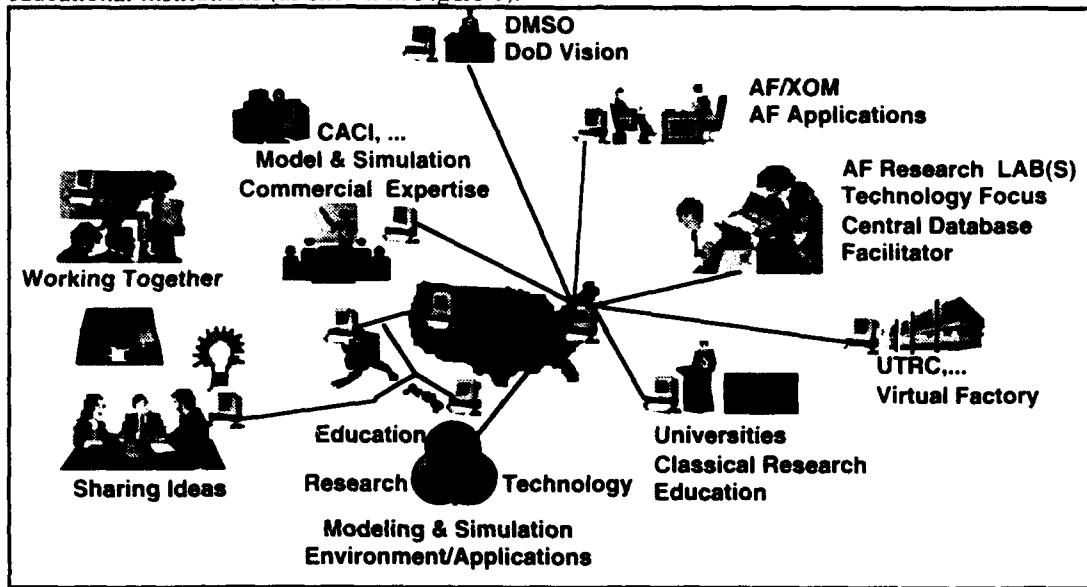

FIGURE 1: A distributed modeling and simulation laboratory.

The physical requirements for the Modeling and Simulation Laboratory were prototyped this summer in Rome Laboratory and include hardware such as a DEC Alpha workstation, a Macintosh Quadra multimedia workstation with direct video I/O, two 1 Gigabyte hardrives, a color scanner and CD ROM drive, as well as black and white and color printers. Software included UNIX operating systems for both workstations as well as a broad spectrum of modeling, simulation, prototyping, and presentation packages. All the machinery is connected on a Ethernet LAN, and is connected to the INTERNET via a T1 line. The ultimate configuration of the Modeling and Simulation Laboratory is indicated in Figure 2.

Internet ISDN

Large Screen Projection

Workstations

MACINTOSH   DEC   SUN   IBM PC

$z = 1.2^y \sin(x)$

Mathematical & Graphics Tools

Software-Modeling & Simulation & Intelligent System Tools

Optical Disks

Color Printer

Scanner

Mic   Camera
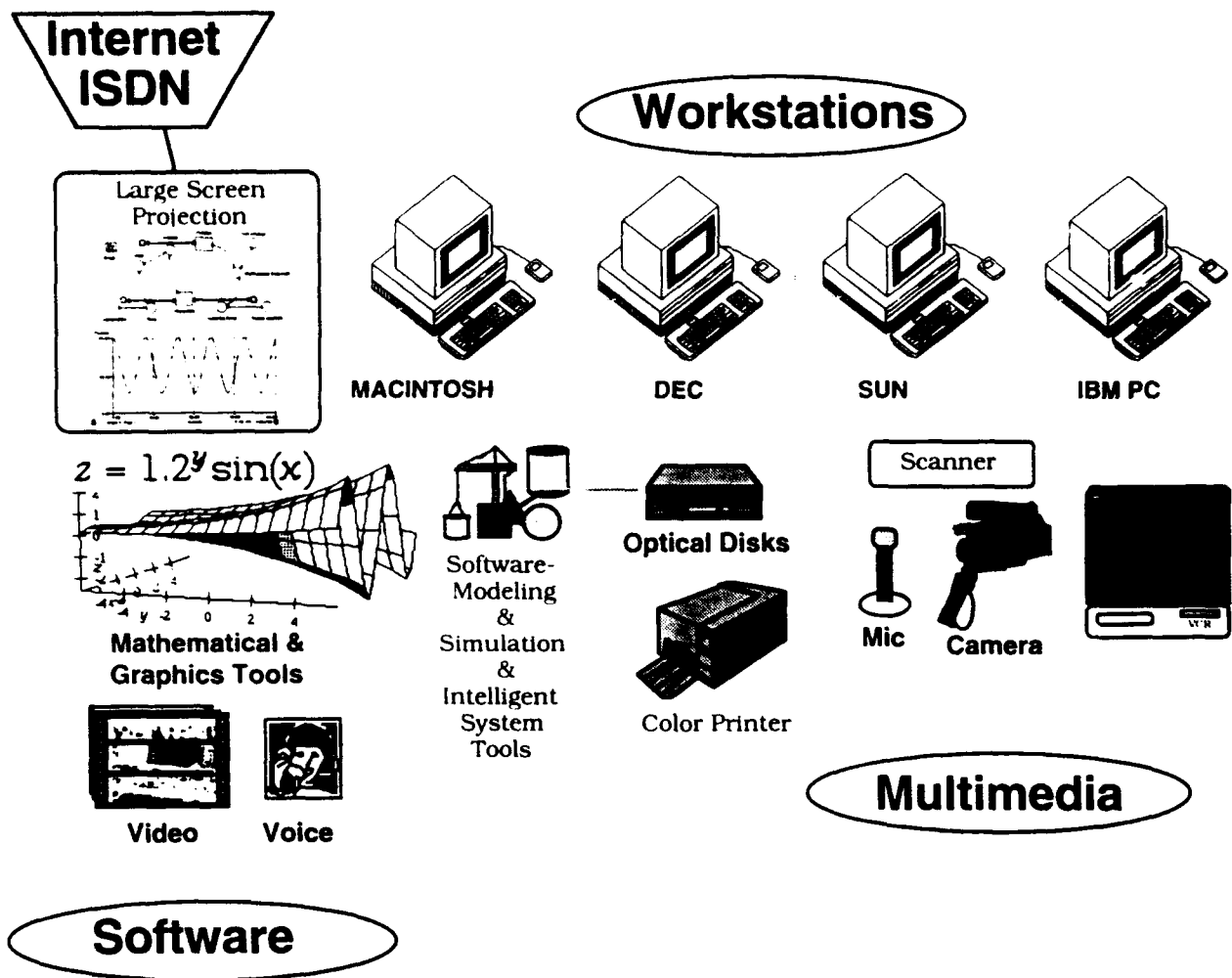
Video   Voice

Multimedia

Software

FIGURE 2: The conceptual plan for a modeling and simulation distributed laboratory node.

The multimedia requirements of the laboratory are critical to Virtual War. The pandemonium of actual combat cannot be simulated in a soundless environment. (Work should also be done on smells since they are the most powerful stimulus for releasing stored memories.) Figure 3 shows an image of the multimedia Macintosh Quadra workstation. It was connected to a Video Camera and a VCR and converted its own image into a computer graphic capable of pasting into any document or presentation.

FIGURE 3: The actual Rome Laboratory Modeling and Simulation Laboratory.

The modeling and simulation laboratory will provide the hardware, software, and connections to outside resources to;

- **Enable C3I technology sharing**

- **Establish a research and educational facility for modeling and simulation**

- **Create video prototypes for RFP**

- **Rapid prototyping (1-3 weeks) in-house environment for selected new initiatives**

- **Development of standards for models, interconnectability and VV&A**

- **System level development for warfighter, global shield, etc.**

- **Inve stigate synthetic environments for acquisition, training & technology**

One example of use of these abilities was demonstrated by the rapid construction of a prototype

modeling and simulation information network by Mary Dobransky. It took her less than three weeks to

integrate the lab hardware and software with the bibliographic, model, and INTERNET databases via SuperCard. This result is described in the paper "PROTOTYPING A MODELING AND SIMULATION INFORMATION NETWORK" also included in this volume.

PART 4: Synthesis

So let me tell you a story.

It took almost two months to get the catalog of Rome Laboratory models on-line but finally it happened. An Email message was sent out stating that the models were available via Anonymous-FTP to one and all.

The very next day the Modeling and Simulation Department received a visit from the Reliability Department asking why their models were not on-line. We told them that if they would write out their models in a standard format we would assuredly put them on-line too. While they were visiting we asked if they would like to go upstairs and see the Modeling and Simulation Laboratory.

Sure they said, and so we did. They were particularly impressed by the ability to capture images off the computer and create videotape demos. We also showed them that we could do the reverse. We could use a video-camera as a source and produce a computer graphic for inclusion in a paper (like the picture of the lab above).

Of course we also demonstrated the other capabilities of the laboratory for rapid prototyping with COTS software like STELLA™, Mathematica™, and SuperCard™ but that is not important to this story.

About a week later one of the members of the Reliability team (Mark Stoklosa) was talking to a friend (James Vaccaro RL/ERSD) whose work is trying to make a computer recognize faces. It seemed that Vaccaro was having trouble acquiring images. Mark said to him "You know, I think I saw the answer to your problems last week. Let's go visit the Modeling and Simulation Laboratory!"

They did, and sure enough James liked what he saw. He wrote down the names of the hardware and software used in the lab and went off to order some for himself. Another technical problem solved.

End of story.

17-14

The point of the story is that making connections is what a Modeling and Simulation Department is all about. Virtual War demands distributed processing, it demands a connected systems of high reliability and varying resolution simulations.

To ensure that the technology developed here at Rome Laboratory is at the forefront means staying on top of the information deluge and using the most current software/hardware. No researcher would ignore a journal in her field. A researcher cannot ignore the INTERNET. She must stay connected with the rapid developments taking place worldwide. This is especially important at a site like Rome Laboratory that is not physically close to academic and industrial leaders.

She needs a place to try out the most current software/hardware where experimentation and creativity is stressed. Where there is no sales pressure. Where the only thing wanted of her is an evaluation of how well it worked for her. A place where she can make connections.

Making connections is where the Modeling and Simulation Department can have the greatest, as well as the most immediate, impact.

The vision of the Rome Laboratory Modeling and Simulation Department is;
- **Community awareness**
- **Information sharing**
- **Creation of a Modeling and Simulation center of excellence.**

The Modeling and Simulation Summer Research Group hopes that we have made a start in fulfilling that vision.

Bibliography

Anderson, Robert H., Steven C. Bankes, Paul K Davis, H. Edward Hall, and Norman Z. Shapiro,(forthcoming),."Toward a Comprehensive RAND Environment for Computer Modeling, Simulation, and Analysis."

Bonder, Seth,1984." Summary of a Verification Study of Vector-2 with the Arab-Israeli War." In Reiner Huber (ed.), Systems Analysis and Modeling in Defense: Developments, Trends, and Issues,Plenum Press, New York",.

Brewer, Gary, and Martin Shubik,1979. "The War Game: A Critique of Military Problem Solving." Harvard University Press.

Davis, Paul K,1990,"An Analyst's Primer for the RAND-ABEL Programming Language",RAND, N-3042-NA

Davis, Paul K.,1993,"An Introduction to Variable-Resolution Modeling and Cross-Resolution Model Connection".RAND, R-4252-DARPA.

Davis, Paul K, and Donald Blumenthal, 1991, "The Base of Sand Problem: A White Paper on the State of Military Combat Modeling", RAND, N-3148-OSD/DARPA",.

Davis, Paul K. Steven C. Bankes, and James P. Kahan, 1986, "A New Methodology for Modeling National Command Level Decision making in War Games and Simulations", RAND, R-3290-NA",.

Defense Modeling and Simulation Office, 1991, "Final Report of the Applications and Methodology Working Group, December 1991",.

Department of Defense/Defense Modeling and Simulation Office, 1992, "Defense Modeling and Simulation Initiate, May 1"

Dupuy, Trevor, 1987, Understanding War. Paragon House, New York

Elzas, Maurice S., Tuncer Oren, and Bernard Zeigler (eds.), 1986, "Modeling and Simulation Methodology in the Artificial Intelligence Era." North-Holland, Amsterdam",.

Gass, Saul, 1983, "Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis", Operations Research, Vol. 21, No. 4, July-August 1983

Helmbold Robert L., 1990, "Point Paper: CAA History Activities." 1980-1990, U.S. Army Concepts Analysis Agency, Bethesda, Maryland",.

Hodges, James S., and James A. Dewar, 1992, "Is It You or Your Model Talking? A Framework for Model Validation. RAND, R-4114-AF/IVOSD

Hughes, Wayne (ed.), 1989, "Military Modeling, Second Edition." Military Operations Research Society, Alexandria, Virginia",.

Klir, G.J., 1989. "Inductive Systems Modeling: An Overview,", in Maurice S. Elzas, Tuncer Oren, and Bernard Zeigler (eds.), Modelling and Simulation methodology in the Artificial Intelligence Era, North-Holland, Amsterdam",.

Martin Marietta, 1990, Confidence Methodology Guide, Third Edition, Final. National Test Bed Technical Report NTB-237-0220G-02, August 15

McQuie, Robert, 1987, "Historical Characteristics of Combat for Wargames (Benchmarks)." Army Concepts Analysis Agency, CAARP-87-2",.

Military Operations Research Society (MORS), 1989, " Human Behavior and Performance as Essential Ingredients in Realistic Modeling of Combat—MORIMOC II, Alexandria, Virginia." Proceedings of a conference held 22-24 February 1989"

Miser, Hugh J., and Edward S. Quade, 1988, Handbook of Systems Analysis: Craft Issues and Procedural Choices. Elsevier Science Publishing Co., New York, New York

Oren, Tuncer, 1989, "Bases for Advanced Simulation: Paradigms for the Future", in Maurice S. Elzas, Tuncer Oren, and Bernard Zeigler (eds.), Modeling and Simulation Methodology in the Artificial Intelligence Era, North-Holland, Amsterdam",.

Pace, Dale K. and Dennis P. Shea, 1992, "Validation of Analysis Which Employs Multiple Computer Simulations", Proceedings of the Summer Simulation Conference, pp. 144-149, Society for Computer Simulation",.

Palmore, Julian, 1992, "Analysis and Verification and Validation of Complex Models", Proceedings of the Summer Simulation Conference, pp. 139-144, Society for Computer Simulation",.

Sargent, Robert, 1986, "An Exploration of Possibilities for Expert Aids in Model Validation,", in Maurice S. Elzas, Tuncer Oren, and Bernard Zeigler (eds.), Modeling and Simulation Methodology in the Artificial Intelligence Era, North-Holland, Amsterdam",.

Sargent, Robert E., 1987, "An Overview of Verification and Validation of Simulation Models", Proceedings of the 1987 Winter Simulation Conference, Society for Computer Simulation",.

Shubik, M., and G. D. Brewer "A Survey of Models, Simulations, and Games", RAND, RM-7821-1-ARPA.

Simard, Richard J., 1993. Personal correspondances.

Stockfisch, John A., 1975, "Models, Data, and War: A Critique of the Study of Conventional Forces, RAND, R-1526-PR",.

Thomas, Clayton, 1983, "Verification Revisited", in Wayne Hughes (ed.) (1989), Military Modeling, Second Edition. Military Operations Research Society, Alexandria, Virginia.

U.S. Army, 1992, Army Regulation 5-11, "Army Model and Simulation Management Program, June".

U.S. General Accounting Office, 1980, "Models, Data, and War: A Critique of the Foundation for Defense Analyses." GAO/PAD-80-21, Washington, D.C.",.

U.S. General Accounting Office, 1987, "DoD Simulations: Improved Assessment Procedures Would Increase the Credibility of Results." GAO/PEMD-88-3, Washington, D.C.

Zeigler, Bernard, 1976, "Theory of Modeling and Simulation, John Wiley and Sons, New York; reissued by Krieger Pub. Co., Malabar, Florida, 1985",.

Zeigler, Bernard, 1984, "Multifaceted Modeling and Discrete Event Simulation", Academic Press, London and Orlando, Florida